Athens University of Economics and Business

MSc in Statistics

Statistical Modeling in Actuarial Applications

Georgios Panagou

Department of Statistics

Co-supervised by Prof. Dimitris Karlis and Prof. Luis Bermudez

20 October 2020

Declaration

I hereby certify that the material, which I now submit for assessment on the programme of study leading to the award of Master of Science, is entirely my own work using advice and help from my supervisors and has not been taken from the work of others except to the extent that such work has been cited and acknowledged within the text of my own work. No portion of the work contained in this thesis has been submitted in support of an application for another degree or qualification to this or any other institution. All research was carried out in Risk Center of Universtiat de Barcelona and in Athens University of Economics and Business, Department of Statistics.

_____

Georgios Panagou
20 October 2020

Acknowledgements

I would like to express my special thanks of gratitude to my supervisors, Prof. Dimitris Karlis and Prof. Lluis Bermudez that gave me this golden opportunity to work on such an interesting project in statistical modeling. Using their helpful advice and studying all bibliography proposed, I enhanced my skills and extended my knowledge on different sections of statistical and actuarial science.

List of Figures

List of Tables

Contents

Abstract

In actuarial applications and especially in automobile insurance statistical modeling contributes in a priori ratemaking.
Generalised linear models based on various distributions and techniques can help us distinguish between different types of count claims and examine their charachteristics. Poisson regression is our basic tool,while negative binomial and generalised Poisson regression give us the opportunity to overcome difficult issues like overdipsersion. Zero-inflated models are quite useful since we face an excess of zeros, something really common when it comes to insurance applications. This analysis, goes through modeling two different types of claims through different methods.

Our analysis is separated to different methods of modeling and different distributions. It starts with explaining all techniques of modeling and specificely trivariate reduction method, models based on multiplicative factors and copula-based models. The next step of our analysis focuses on different distributions for count claims, like Poisson, negative binomial and mixed Poisson distributions in general and generalised Poisson too. We examine differences between independence between types of claims and correlation existence. Furthermore, we begin from the most strict modeling technique, which allows only positive dependence between the types of guarantees and as we go through our analysis we relax these assumptions. Complicated models with extra parameters, especially negative binomial regression ones, that explain dependence and overcome overdispersion issues are used through this analysis.
Insurance pricing is carried out, summing premiums for each type of guarantee, after drivers' represantative profiles are constructed. These models are applied in an automobile insurance database belonging to a Spanish insurance company. A subset of 6000 contracts is used for our purposes.

1 Introduction

Technology's rapid evolution in recent years is fully attached to science's evolution. Statistics is a great example of this special relationship. Computers nowadays provide statisticians with lots of tools when it comes to data handling and statistical modeling.

This analysis is based on modeling bivariate count data in actuarial and insurance applications. Designing an efficient tariff structure has been a great challenge for any insurance company. Insurers use data from former claims of company's clients and try to price each individual of the portfolio in the most efficient way.

In the case of one type of claims, modeling is much more simplified and statisticians have proposed a variety of ideas to deal with it, like David, Mihaela & Jemna, Danut. (2015) proposed univariate Poisson and negative binomial models in automobile insurance claim counts. But insurance companies face multiple type of claims from their clients, so models' complexity increases and literature of the bivariate and especially multivariate case is not so wide. A univariate model can use only a single type of claim as a response variable. Therefore, since an automobile insurance claim can arise as a result of an accident or for example windshield damage or theft, it is essential that the response variable of a model has to be of greater dimension. A multivariate model of course can explain better the connection between the response and explanatories. Different types of claims can be explained from different characteristics of the driver. For example, a claim for an accident can be alleged to driver's experience, while a claim for theft can be alleged to the territory that driver lives. Consequently, a univariate response variable cannot distinguish characteristics of different types of claim.

Our purpose is to design ratemaking systems using a variety of models and deal with issues that arise, comparing the efficiency of each one. Through our analysis, we try to examine which characteristics of insured drivers, have an important impact to the number of claims. In other words, we want to construct drivers profile and through the implementation of different models, to identify the best model that fits the data. All these models are applied on automobile insurance data. Claims of insured are divided in two groups (third party liability and the

rest). Many other variables are included also such as gender of insured, type of car, place of residence etc. They will be used as explanatory variables in our models, helping us to handle the heterogeneity of portfolio.

Individuals differ a lot from each other explaining the heterogeneity of the portfolio. In a priori ratemaking there are important variables of drivers' behavior that can not be measured, like driver's aggressiveness or drinking behaviour. In order to deal with this problem, risk measures should be included in the models.

Our first approach relies on modeling through bivariate Poisson and bivariate negative binomial regression through trivariate reduction technique. The trivariate reduction is in many contexts an appealing method for constructing bivariate distributions. Basically, the idea is to create a pair of dependent random variables from three or more random variables. If we assume that numbers of claim counts,which are our variables, between our two categories are independent, then everything is quite simple. However, the most interesting part begins when this assumption does not hold. Measuring dependence between these two different categories of claim counts is a big challenge and modeling becomes immediately harder. Bivariate Poisson models and zero inflated versions based on trivariate reduction on count data were used by Karlis and Ntzoufras (2003) on sports data and by Bermudez (2009) on insurance claim counts. Through trivariate reduction only positive dependence between claim counts is allowed. This limitation encourages us to explore different ways of modeling more flexible correlation structures.

The second approach of our analysis relies on mixture of bivariate Poisson regression combined with different distributions used for random effects. Mixtures of bivariate Poisson models can be implemented in different ways. We focus on bivariate Poisson models using a random variable α for mixing , where α follows a distribution. Ghitany & Karlis & Al-Mutairi & Al-Awadhi (2012) used three multivariate mixed Poisson models (negative binomial, inverse Gaussian and lognormal) for modeling, that allow only positive correlation between claim counts. Some models based on mixture models proposed from Steyn (1976) and Aitchinson and Ho (1989), using bivariate normal and bivariate lognormal respectively as the mixing distribution. Moreover, in all these models,random effects can follow any

type of distributions not only continuous but discrete also.

Another idea was proposed from Bermudez and Karlis (2012), as they used an m-finite mixture of bivariate Poisson regressions extending Karlis and Meligkotsidou (2007) case. The main advantages of this approach are: overdispersion is allowed, interpretation of the model is easier due to mixture model's clustering and different regression lines are used for every component. This finite mixture model allows also negative correlation if the parameters are negative correlated.

Previous models deal mainly with overdispersion issues, but there are plenty more focusing on correlation between variables. Our third approach uses a different way to model dependence between claim clounts. The advantage is that in opposition to modeling with trivariate reduction method, these models allow a flexible correlation structure through a multiplicative factor. A multiplicative factor is a dependency parameter that gives us the opportunity to transform the joint probability function of two variables,in our case our two different claim counts variables, to a product of marginals. If these variables were independent, then naturally the joint probability function could be transformed to this product. Through this multiplicative factor and using appropriate bounded functions, this factorisation is eligible for dependent variables also. Furthermore, anyone can use different distributions for modeling each category of claim counts. Therefore, it is not necessary for both types of claims to follow the same distribution function. It is important to mention that, through this multiplicative factor, any kind of correlation between these two variables is allowed, both positive and negative, zero also. Lakshminarayana (1999) proposed a bivariate Poisson distribution that allows any kind of correlation. To be more specific, the model was a product of Poisson marginal distributions with a multiplicative factor. Moreover, for our analysis we will use bivariate zero inflated Poisson marginal distributions also. Following this particular technique Famoye (2010, 2012) proposed a bivariate negative binomial regression and later Faroughi (2017) different forms of bivariate zero inflated negative binomial regression. Zamani et al. (2016) introduced some nested bivariate generalized Poisson regression models. These models give the opportunity of comparison through likelihood ratio tests and their difference relies on the dispersion parameter. Finally, Hofel and Leitner (2012) presented bivariate count

data models based on Sarmanov distributions. Its formula is less complicated than copula-based model,that we explain in the next part. Moreover, implementation of the model is less time consuming and in parallel it keeps many properties like allowance of both positive and negative correlation between variables.

The last part of our modeling analysis is related to copula functions. Copulas are functions that enable us to separate the marginal distributions from the dependency structure of a given multivariate distribution in a different way than a multiplicative factor. Specificely, copula is a multivariate distribution function with marginally uniform random variables on [0,1]. Copula functions have some appealing properties such as they allow scale-free measures of dependence and are useful in constructing families of joint distributions. Shi and Valdez (2014) proposed multivariate regression models that model correlation through common shock variables and copula functions that allow negative binomial marginals. So et al (2011) used the zero-inflation extension for the negative binomial model based on copulas to deal with correlation issues but in the bivariate case. Faroughi and Ismail (2017) proposed similar models in different forms in order to deal with overdispersion.

In conclusion, after implementation of all models we end up with a priori ratemaking. To construct a tariff structure that reflects the various risk profiles in a portfolio, actuaries usually rely on regression techniques. Such techniques allow for the inclusion of various explanatory variables so that the actuary is able to construct risk classes with more or less similar risk profiles. For insurance pricing, five driver's profiles are constructed based on specific characteristics of insured drivers' profiles and we compare our models mean scores in all of them in order to identify the best models.

Our analysis starts with exploration of our data and explanation of variables. Continuously, we provide every method we use in order to construct our models and how to optimize all of them. The next part, focuses on the analysis of all models based on different probability distributions. In the last part, we present results from our analysis and we go through an a priori ratemaking process in order to show which model fits better our data, based on specific drivers' profiles.

2 Data

2.1 Data acquiring

For our statistical analysis, data from a Spanish automobile insurance company are used. Specifically, every model is implemented on a random subset of 6000 observations and 8 explanatory variables. The first two variables in our dataset are numbers of claims for each category of insureds (N_1: number of claims from third-party liability guarantees and N_2: number of claims from rest of guarantees). Third-party insurance is an insurance policy purchased for protection against the claims of another. Third-party offers coverage against claims of damages and losses incurred by a driver who is not the insured, the principal, and is therefore not covered under the insurance policy. The driver who caused damages is the third-party. The rest variables are used as explanatory variables in our models, defining specific characteristics of insured customers, such as the number and the code of their policy, their gender, age and horsepower of car.

In order to analyze deeper the nature of these variables, one variable describes the gender of an insured person. Another, takes into consideration, whether a person drives in an urban area or not. Essentially, on the one hand, an urban area possibly has lower traffic than a city center, but on the other hand drivers may have different driving behaviour. Another interesting characteristic is whether an insured drives in a deemed high risk region like Northern Spain or in a lower risk. Age also is an important issue, since drivers over thirty years old probably are more experienced in driving than younger drivers. Another variable takes into account vehicle's horsepower. A really strong vehicle with horsepower over 5500 cc might be really dangerous to be driven from an amateur driver. Last but not least, a fact that is really important for an automobile insurance company and for insurance companies in general, is related to the loyalty of the insured. It is reasonable to assume that an insured who has been five or more years with the company is treated in a different way than a new one. Insurance companies can build better someone's profile over the years.
All these characteristics mentioned above, are described from binary variables (Yes/No) in our dataset and response variables from count ones.

List of variables we use in our analysis:

- GEN: Equals 1 if the driver is a woman and 0 if it is a man
- URB: Equals 1 when driving in urban area, 0 otherwise.
- ZON: Equals 1 when zone is deemed high risk (northern Spain), 0 otherwise.
- LOY: Equals 1 if the client has been with the company for more than five years, 0 otherwise.
- AGE: Equals 1 if the insured is 30 years old or younger, 0 otherwise.
- POW: Equals 1 if the vehicle's horsepower is equal to or greater than 5500 cc, 0 otherwise.
- N_1: number of claims from third-party liability guarantee.
- N_2: number of claims from all other guarantees.

All these variables help insurers to have a clear image about profiles of their clients. Some of these variables, characterize the driver, like gender, age and loyalty. Other, characterize the car that the client drives like horsepower and others the area that client drives, urban, high risk or not. Response variables N_1 and N_2 that count how many claims a client reports will be denoted in our models as N_{1i}, N_{2i}, where $i = 1, 2, ..., 6000$ denotes the client that we are referred to.

2.2 Exploratory Analysis

In this section of our analysis we focus on the most important characteristics of our data, in an attempt to highlight them visually also using appropriate plots. Statistical modeling is carried out using variables N_1 and N_2 as a bivariate response variable. As we mentioned above, N_1 refers to the number of claims from third-party liability guarantees and N_2 to the number of claims for the rest of the guarantees. The minimum and the maximum values are 0 and 6 for both of them respectively. Although, the range of these variables equals to six, the mean values are 0.087 and 0.131 with variance 0.139 and 0.200 respectively. Both mean and variance values are close to zero and as Table 1 indicates there is an excess of zero counts. By the term, excess of zeros, we mean that we observe for some reason more often zeros than those expected by our model. It is obvious that zero is the most frequent value for both N_{1i}, N_{2i}, for every $i = 1, 2, 3, ..., 6000$. The difference between zero claim counts and non zero claim counts is big enough, since it is 5611 and 389 respectively for N_1 and 5391 and 609 for N_2. For both response variables N_1, N_2 maximum values equal to six have really low frequencies, 3 and 4 respectively and in general since zero counts are so many for both response variables, it is reasonable for the mean to be close to zero. Moreover, since non-zero counts are a lot less than zero ones, variation is close to the mean, but it exceeds it, as we mentioned above.

Table 1: Frequency table of response variables

Resp Val	N_{1i}	N_{2i}
0	5611	5391
1	290	488
2	74	82
3	19	30
4	4	3
5	1	4
6	1	2

Table 2: Zero and non zero pairs of response variables

Pairs of (N_{1i}, N_{2i})							
N_{1i}	0	1	2	3	4	5	6
N_{2i}							
0	5118	403	66	18	3	3	0
1	208	64	11	7	0	0	0
2	51	16	3	2	0	1	1
3	11	4	2	2	0	0	0
4	3	1	0	0	0	0	0
5	0	0	0	0	0	0	1
6	0	0	0	1	0	0	0

Moreover, a matter of great interest is also to notice the frequency of each pair of claim counts. In other words, how many clients have zero claim counts for both categories at the same time and so on for every possible combination.

It is obvious to notice from Table 2, that for $i = 1, 2, 3, ..., 6000$ zero pairs of response variables N_{1i}, N_{2i} are the most common pair, since 5118 zero pairs exist while the next most common combination equals 403 pairs. Therefore, a great number of the total clients have no claims, neither from third-liability guarantees nor from the rest of the guarantees.

Correlation between N_1 and N_2 is 0.20 indicating that there is positive dependence between different type of claims.
All the other variables are used as covariates in our models except for two variables in our dataset that refer to the number and code of policy and are used from insurance companies in order to discriminate one policy from another. As a matter of fact, these two variables are not important for our purposes, so they will not be taken into consideration for the rest of our analysis. Therefore, six binary explanatory variables are used that take values 0 and 1 (definition in paragraph 2.1). Our dataset does not have any missing value and in addition it has some significant characteristics, that help us understand in a better way policyholders' profiles. Firstly, the difference between young policyholders (younger than 30 years old) and older is important to be mentioned. In our sample there are 5355 policyholders older than 30 years old and 645 younger than them. Maximum values

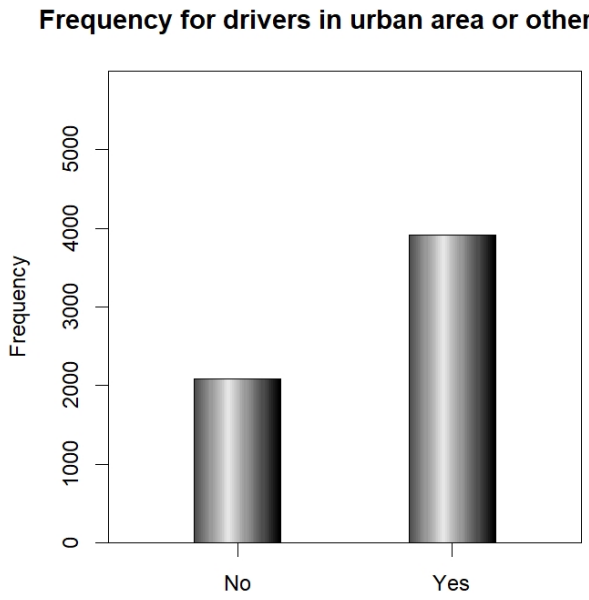


Figure 1: Number of drivers in urban area

for both our response variables occur in older insured.
Moreover, the difference between males and females is an interesting issue, because 83.6% of the population are males and only 16.4% are females. Furthermore, the biggest part of policyholders are loyal to the company, since 4937 out of 6000 have been more than 5 years with the company. It is important for our analysis to mention that more than 80% of the population are males driving a vehicle of 5500 cc or greater. Among policyholders 3918 are driving in an urban area as it is shown in Figure 1, so the area that insured chose to drive reveals a more balanced relationship among the population.

The most of the total insured do not drive in a high-risk district of the country as it is shown in Figure 2.
Another really important issue that we will discuss also later in our analysis is that zero counts are the most often value of both our response variables. Apparently, this fact is obvious consulting Figure 3 for both of the response variables. As anyone can notice from both figures, more than 5000 observations of our count

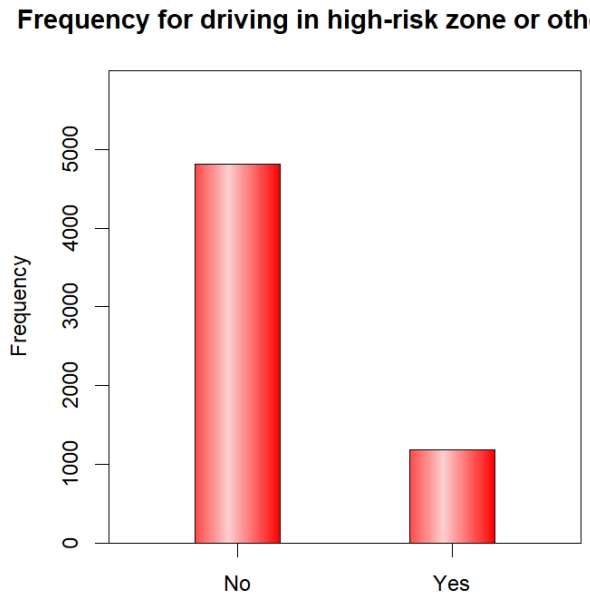
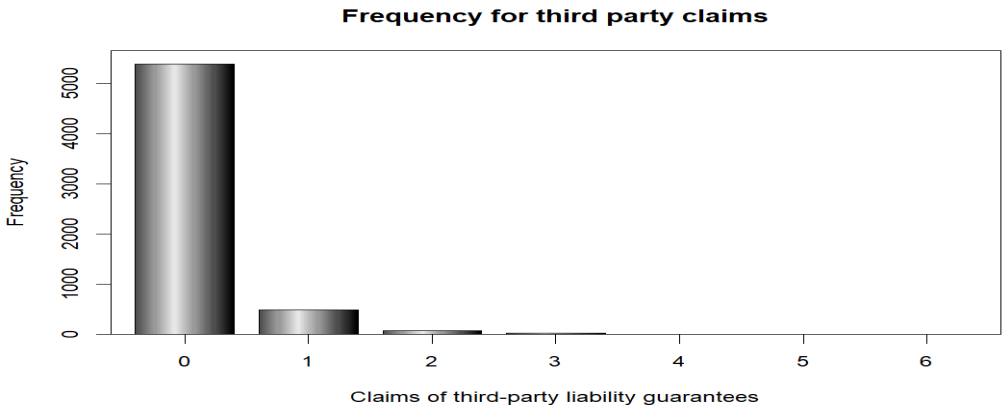


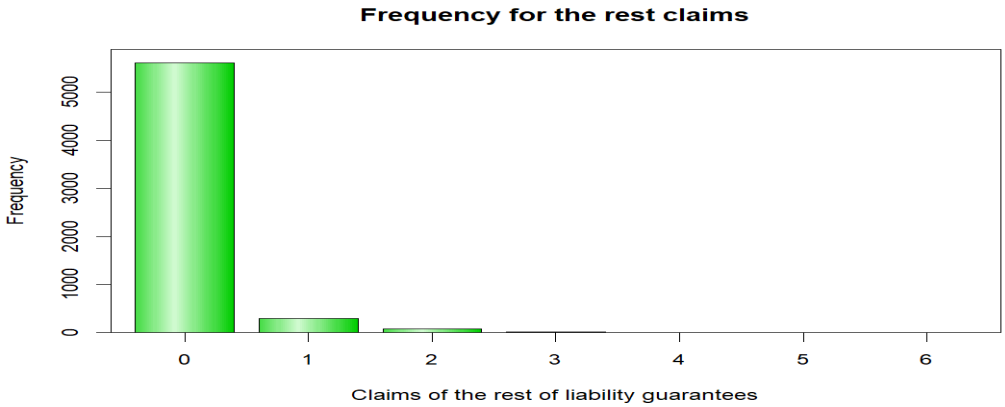
Figure 2: Number of drivers in high-risk zone

data are equal to zero. Phenomena like this, lead to specific problems when it comes to modeling like overdispersion, that we will try to overcome as the analysis goes through.

Modeling excess of zeros in insurance data is a very intriguing situation. It is quite reasonable for anyone to wonder the reason why such a difference between zero and non-zero claims exist. Although, from figures it is not obvious why, this excess exists, modeling comes next to expose this phenomenon. Nevertheless, from these figures it is quite visible, that our observations are left-skewed distributed, since zero claims are the most often observation. These counts lead to skewed distributions, in order to model unobserved heterogeneity that causes overdispersion. In parallel, claim counts equal to five or six are quite a few for both response variables, therefore left-skewness is too sharp. Obviously for that reason, normal distribution does not fit well our data and probably Poisson, negative binomial among others are some distributions that will help us in modeling.



(a) Third party liability count claims (N_1)



(b) Rest of guarantees count claims (N_2)

Figure 3: Both types of guarantees count claims

Many insurance companies use bonus malus systems when it comes to their policies. In insurance, a bonus-malus system (BMS), adjusts the premium paid by a customer according to their individual claim history. Most insurers around the world have introduced some form of merit-rating in automobile third party liability insurance. Such systems penalize at-fault accidents by premium surcharges and reward claim-free years by discounts, commonly known as a "no-claims discount". Therefore, an insured is tempted not to report all accidents in order to claim rewards.As a result, a statistician has to be suspicious about zero counts, so he has to take into consideration these incidents.

3 Method

3.1 Introduction to Statistical Modeling

This chapter aims to introduce the basic probability models for count data and methods of construction that will be applied to automobile insurance data. Using as a response variable, our bivariate count variable of the guarantees and binary variables as explanatories, we try to build appropriate models that can fit our data, expose all of their special characteristics and find these models that fit our data the best. Predictions of the expected frequency of claims can be derived too. Probability models in practice always rely on probability distributions and Poisson distribution is one of them, which has been used widely for insurance data.This distribution has some solid assumptions, like the equality of its mean and variance. When our data do not let us make such an assumption we use specific techniques to introduce several different forms to overcome such difficulties. For example, generalised form of bivariate Poisson and Poisson mixtures are some forms we will go through. Constructing the generalised form of bivariate Poisson we can handle overdispersion. This fact holds, since an extra parameter is introduced that enables the variance to be proportional to the mean. Another problem, is the excess of zeros we described earlier. To be more specific, zero counts are greater than expected from the Poisson distribution. In order to solve such issues, we introduce zero inflated versions of bivariate Poisson or mixtures with other distributions that can fit better our data. It is important to mention another interesting property of bivariate Poisson distribution, the issue of positive or negative dependence.Bivariate Poisson via trivariate reduction, that we will focus in the next section, allows only positive dependence between count variables. Therefore, if our response variables in our data have negative dependence we have to find alternative methods to construct a model. Models based on a multiplicative factor, a factor that allows both positive and negative dependence, also models based on copula functions give the solution to this kind of problems. Techniques for modeling our data vary, but we are going to analyze trivariate reduction based models as well as multiplicative factor and copula based ones.

3.2 Modeling techniques

Literature for building models based on univariate distributions is quite extended since these models are easier to handle than bivariate ones. When someone wants to model a bivariate count reponse variable, like automobile insurance types of guarantees in our case, he has to face many difficulties and challenges like correlation between variables and more. Our first approach for the construction of a bivariate model is through trivariate reduction method, then we go on using a multiplicative factor in order to allow different marginals for our types of guarantees and finally we end up using copulas and common shock variables.
We describe these techniques in this section and we give further details in the next section,while we introduce all models.

3.2.1 Trivariate reduction

It is well known that trivariate reduction, Mardia (1970), a method to generate two dependent random variables from three independent random variables, can be used to generate Poisson random variables with specified marginal distributions and correlation structure. The method,however, works only for positive correlations. Trivariate reduction method is a popular and old technique used for constructing dependent variables. It has been used for both continuous and discrete cases. The method consists of building a pair of dependent random variables starting from three (or more) independent (usually) random variables. The functions that connect initial variables are generally elementary functions, or are given by the structure of the variables that we want to generate. A general approach is the following:
Consider random variables X_1, X_2 and X_3. Then we may define a new pair of variables using $N_1 = g_1(X_1, X_3)$ and $N_2 = g_2(X_2, X_3)$ where $g_i(\cdot, \cdot)$, $i = 1, 2$ are some functions. The central idea is that since N's share the common X_3 they are correlated, the correlation structure is determined by the functions g.
Typical choices are for example $g(X_1, X_2) = X_1 + X_2$ which is used to derive the bivariate Poisson distribution if the X's follow Poisson distributions, and $g(X_1, X_2) = min(X_1, X_2)$ for bivariate exponential distributions if the X's follow exponential distributions. The method can be very flexible, while the choice of the

functions g determine the correlation properties and perhaps put restrictions on them. For example the usage of $g(X_1, X_2) = X_1 + X_2$ leads to necessarily positive correlation. It is important to explain why trivariate reduction leads to positive correlation. As we mentioned earlier, let X_1, X_2, X_3 be three mutually independent Poisson random variables. Then, N_1=X_1+X_3 and N_2=X_2+X_3 are two new random variables that follow Poisson distribution with rate θ_1 and θ_2 respectively. According to Holgate (1984), $X_1 \sim P(\theta_1\text{-}\eta)$, $X_2 \sim P(\theta_2\text{-}\eta)$ and $X_3 \sim P(\eta)$ with $\eta \in [0, min(\theta_1,\theta_2)]$. Then corr($N_1$, N_2) = $\frac{\eta}{\sqrt{\theta_1\theta_2}}$ is non-negative and increasing in η. If $\theta_1 = \theta_2 = \theta$ is fixed, corr(N_1, N_2) takes all values between 0 and 1 as η goes from 0 to θ. Otherwise, if $\theta_1 \neq \theta_2$, then corr(N_1, N_2) $\leq \frac{min(\theta_1,\theta_2)}{\sqrt{(\theta_1\theta_2)}}$.

3.2.2 Multiplicative factor

In this section, we describe an alternative type of modeling, that allows for both positive and negative dependence between random variables. Our models are defined as a product of two marginal distributions with a multiplicative factor. This factor is the reason why dependence can be either positive or negative, overcoming a big limitation that trivariate reduction technique obeys.

The correlation between the two variates can be either positive or negative, depending on the value chosen for the parameter in the above multiplicative factor. We assume that the marginal means of the bivariate model are functions of the explanatory variables. Further, we assume that the relationship between the marginal means and the covariates in our analysis is log-linear. This relationship is usually referred to as the link function in the univariate case. Other link functions can be considered, but we restrict our discussion in this analysis to log-linear relationship. Therefore, let us assume that $F(n_1, n_2)$ is the cumulative function of (n_1, n_2), with marginals $F_1(n_1)$ and $F_2(n_2)$, that follow Poisson distribution with rates θ_1 and θ_2 respectively.

Then the cumulative function,

$$F(n_1, n_2) = F_1(n_1)F_2(n_2)[1 + \alpha[h_1(n_1)h_2(n_2) - E(h_1(n_1))E(h_2(n_2))]] \qquad (1)$$

where $h_1(n_1)$ and $h_2(n_2)$ are bounded functions of $(n_1, n_2) \in \mathrm{R}^2$ and α is the multiplicative factor Panditi et al.(1988).
When α is positive or negative so is the dependence between the two variables respectively.
Let's assume that $h_1(n_1) = \exp(-n_1)$ and $h_2(n_2) = \exp(-n_2)$. Then, $E[N_1] = \theta_1$ and $E[N_2] = \theta_2$. As a result,from the dispersion matrix of (n_1, n_2) the correlation coefficient is:

$$\rho = \alpha\sqrt{\theta_1\theta_2}(1 - e^{-1})^2 e^{(\theta_1+\theta_2)(1-e^{-1})}$$

(2)

Therefore, when α is positive, the dependence between response variables is positive, when it is negative, so is the dependence.
It is important to mention, that multiplicative factor technique allows us to chose different marginals for our models. Therefore, anyone can use for example Poisson distribution to model third-party liability count claims and negative binomial for the rest count claims.

This modeling technique was introduced by Lakshminarayana et al. (1999) using a bivariate Poisson model. The innovation of this model was related to correlation restrictions. Both positive and negative correlation between claim counts is allowed as we mentioned, therefore this model has been used as a primitive idea for more interesting models that can handle problems like overdispersion using for example generalised Poisson or negative binomial marginal distributions. Multiplicative factor α is a measure of dependence that lies in a closed interval ordered by the dispersion matrix of our bivariate response variable. For more information about its structure see Lakshminarayana et al. (1999).

3.2.3 Common Shock and Copulas

Our last construction method is via copulas and common shock models.
Copulas are functions that enable us to separate the marginal distributions from the dependency structure of a given multivariate distribution. They are useful for several reasons. Firstly, as a way of studying scale-free measures of dependence and secondly, as a starting point for constructing families of bivariate distributions. However, they are sometimes used in a "black-box" fashion and understanding the overall joint multivariate distribution can be difficult, when it is constructed by separately specifying the marginals and copula. But, let us define what a copula is.
Let $C(u, v)$ be a bivariate distribution function defined on the unit square with uniform marginals on [0,1]. For all $(u, v) \in (0, 1)^2$, then $C(0, 0) = 0, C(1, 1) = 1, C(u, 1) = u, C(1, v) = v$.
Suppose that $F(n_{1i}, n_{2i})$ is a bivariate distribution function with marginals $G_{1i} = P(N_{1i} \leq n_{1i})$ and $G_{2i} = P(N_{2i} \leq n_{2i})$.
Then through the copula function C we have $F(n_{1i}, n_{2i}) = C(G_{1i}, G_{2i}; \xi)$ where ξ is a dependence parameter, Sklars' Theorem (1959).
Copula's density function, i.e. a PDF, is obtained in the usual manner as:

$$c(u, v) = \frac{dC(u, v)}{dudv} \tag{3}$$

The copula links the marginal distributions together to form the joint distribution. There are plenty of these functions that can link two different marginals with a dependence parameter, Joe (1997), Nelsen (1999). Frank, Gumbel, Gaussian, t-copula are some examples but we will focus on two Archimedean copulas: Frank's and Clayton's copulas, Frank (1979) and Clayton (1978). However, we will present Gumbel's copula also, Gumbel (1960), a copula that gives similar results in modeling with the other two copulas, that we mentioned. Nevertheless, it is an alternative in Archimedean copula modeling. Gaussian and t-copulas belong to elliptical copulas, though we chose Archimedean instead of elliptical, due to the key disadvantage that elliptical copulas do not have closed form expressions and are restricted to have radial symmetry, conditions that do not help with our

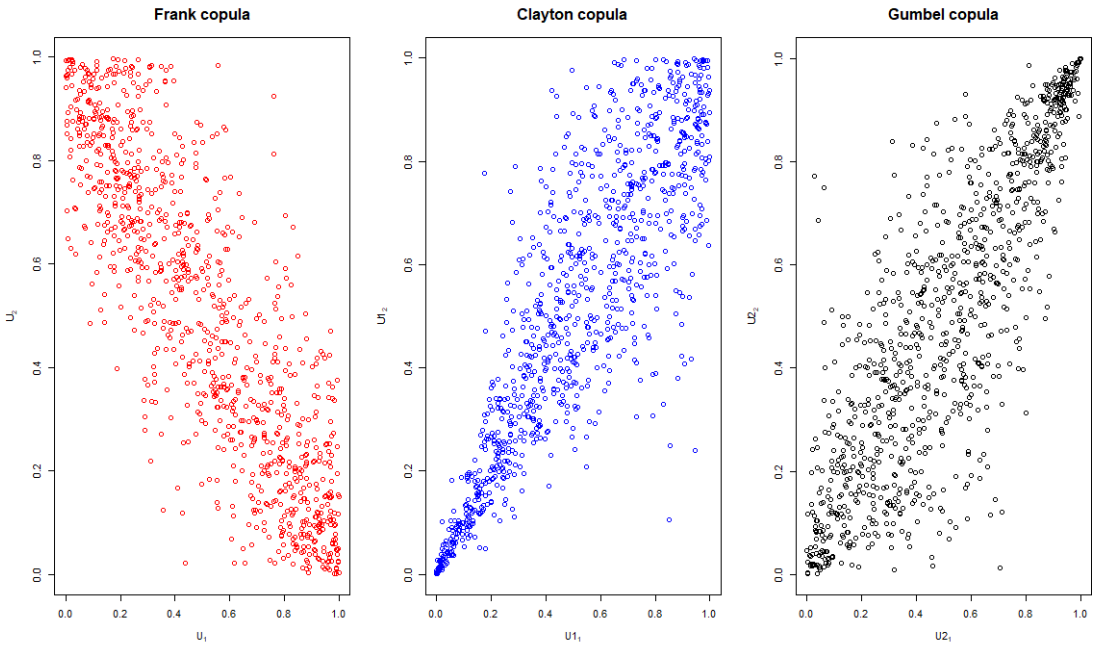
analysis, beacause of copula's density complexity.

While the copula function of the Gaussian distribution does not allow for dependence in the tails, the Student-t copula does not allow for asymmetric tail dependence. However, dependence models contained in the class of Archimedean copulas can indeed capture dependence in the upper and lower tail dependences. To estimate the Archimedean copula, we only need to find functions which will serve as generators and define the corresponding copula. Each has a single parameter that controls the degree of dependence and are defined by:

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)) \tag{4}$$

where ϕ is the generator of the copula.
A function $\phi : R_+ \mapsto I$ is said to be an (outer additive) generator if it is continuous, decreasing and $\phi(0) = 1$, with $\lim_{t\to\infty} \phi(t) = 0$.

Figure 4: Three Different Archimedean Copulas Scatterplots



In Figure 4, we can see scatterplots of our three Archimedean copulas and highlight some of their characteristics. Based on Figure 4, the Gumbel copula and the

Clayton copula can capture the asymmetric dependence between large and small values. For instance, the Gumbel copula exhibits stronger right tail dependence, while the Clayton copula implies the opposite.However, let us firstly define these three copulas and then we will analyze some of their characteristics.
The Frank copula is a symmetric Archimedean copula given by:

$$C_{FR}(u,v) = -\frac{1}{\xi}\log(1 + \frac{(e^{-\xi u}-1)(e^{-\xi v}-1)}{e^{-\xi}-1}) \tag{5}$$

where $\xi \in$R-$\{0\}$.

Frank copulas have some interesting general attributes. Firstly, they have positive slope and from Figure 4 the absence of a tail on either end of the scatter is obvious. Another characteristic is a nebulous but uniform cloud along the full correlation path. Finally, correlation is relatively weak due to the wide and uniform degree of scatter.
The Clayton copula, Clayton (1978), is an asymmetric Archimedean copula,exhibiting greater dependence in the negative tail than in the positive.

$$C_{CL}(u,v) = (u^{-\xi} + v^{-\xi} - 1)^{-1/\xi} \tag{6}$$

Clayton copulas have heavy concentration/density in the left tail and an expanding cloud as we can see in Figure 4.

The Gumbel copula, or Gumbel-Hougard copula is an asymmetric Archimedean copula, exhibiting greater dependence in the positive tail than in the negative. This copula is given by:

$$C_{GB}(u,v) = \exp -[(-\log(u))^{\xi} + (-\log(v))^{\xi}]^{1/\xi} \tag{7}$$

Copulas are a way of modeling and explaining dependence between different claim counts. We introduce another way, to serve our purposes by using common shock variables. A natural approach to modeling this dependence is to assume that all losses can be related to a series of underlying and independent shock processes. When a shock occurs this may cause losses of several different types, the common

shock causes the numbers of losses of each type to be dependent. It is commonly assumed that the different varieties of shocks arrive as independent Poisson processes, in which case the counting processes for the different loss types are also Poisson and can be easily handled analytically.
A simple bivariate model with a common covariance term could be constructed by adding a common parameter to each response variable. This common parameter has the role of a random effect that implies an identical correlation for all pairs of claim type. A good basic reference on such models is Barlow and Proschan (1975) and the ideas go back to Marshall and Olkin (1967). We will define such models in the next chapter of our analysis, based on negative binomial regression models.

3.3 Optimization

Since we explained all modeling techniques we use to construct bivariate models for our automobile insurance claims data, it is essential to mention the way we will compare these models. Trivariate reduction, multiplicative factor based and copula-based models lie to different assumptions while they can handle different modeling problems as we explained earlier. Therefore, it is crucial to find appropriate methods to compare these models efficiency. As a result, we use optimization techniques in order to maximize their likelihood function. The safest criterion for this comparison in our analysis is based to maximum likelihood estimation, but quantities like Akaike is also used.

3.3.1 Maximum likelihood estimation

In statistics, maximum likelihood estimation (MLE) is a method of estimating the parameters of a probability distribution by maximizing a likelihood function, so that under the assumed statistical model the observed data is most probable. The point in the parameter space that maximizes the likelihood function is called the maximum likelihood estimate. The logic of maximum likelihood is both intuitive and flexible and as such the method has become a dominant mean of statistical inference.
The likelihood of a sample of observations is defined as the joint density of the data, with the parameters taken as variable and the data as fixed (multiplied by any

arbitrary constant or function of the data but not of the parameters). Specifically, let N_{1i}, N_{2i} for $i = 1, ..., 6000$ be a set of independent and identically distributed outcomes with probability mass function $P(\cdot|e)$ where e is a vector of parameters. The likelihood function is the probability of observing the data $N_{ji} = n_{ji}$ for $i = 1, ..., 6000$ and $j = 1, 2$ that is,

$$L(e) = \prod_{i=1}^{6000} P(n_{1i}, n_{2i}|e) \tag{8}$$

The key idea for estimation in likelihood problems is that the most reasonable estimate is the value of the parameter vector that would make the observed data most likely to occur. For our purposes we use the logarithm of this quantity called log-likelihood.

To sum up, for both types of count claims we maximize the log-likelihood of each bivariate model by optimizing all explanatory variables. These variables, are induced in our models, through a log-linear relationship between the mean of our response variable and the explanatory variables. In these models, the response variable (N_{1i}, N_{2i}) is assumed to follow an exponential family distribution with mean θ_{ji}, which is assumed to be some function of $x_{ji}^T\beta_j$. Some would call these "nonlinear" because θ_{ji} is often a nonlinear function of the covariates, but McCullagh and Nelder (1982) consider them to be linear, because the covariates affect the distribution of the response variable only through the linear combination $x_{ji}^T\beta_j$. Furthermore, it is essential for someone to wonder about methods that can be used to achieve our goal of optimization.

Optimization of likelihood function carried out with two different algorithms.

For models via trivariate reduction and mixed Poisson models we implement an EM algorithm proposed by Karlis and Ntzoufras (2003) and Ghitany Karlis et al. (2012) for our models respectively. Standard errors of all variables obtained through bootstrap method, Efron (1979).

More information about the construction of the algorithm exists in the previous mentioned papers and in the next sections of our analysis.

When it comes to the other two categories, we implement a quasi-Newton algorithm because Hessian matrix was too expensive to be computed due to com-

plicated forms of functions. Standard errors in this case were derived from the diagonal elements of square root inverse Hessian matrix, a matrix that was developed in the 19th century by the German mathematician Ludwig Otto Hesse and later named after him.

AIC, Akaike (1971), value was obtained through maximum value of likelihood, since it is a function of this quantity. We define AIC value in the next section of our analysis.

3.3.2 Optimization methods

Our analysis goes through two optimization methods, quasi Newton-Raphson and expectation-maximization (EM) algorithms. The first method (NR) is a time consuming technique that needs derivatives of likelihood function. The second one (EM) is a more elegant method that does not need derivatives, it is less complicated but not eligible in every model. Of course, these two methods can be combined. For example, an EM algorithm with inner NR steps, though we will not go through such a combination.

3.3.2.1 Quasi Newton-Raphson algorithm To begin with, quasi Newton-Raphson is an iterative routine that is used to either find zeroes or local maxima and minima of functions, as an alternative to Newton's method. It can be used if the Jacobian or Hessian is unavailable or is too expensive to compute at every iteration.
We want to maximize $\log(L(N_{1i}, N_{2i}|\nu))$ over every coefficient in parameter ν. So, for $\log(L(N_{1i}, N_{2i}|\nu)) = 0$ we need every partial derivative on ν in order to form the Jacobian matrix J.

$$J = \begin{bmatrix} \frac{\partial^2 \log L}{\partial {\nu_1}^2} & \frac{\partial^2 \log L}{\partial \nu_1 \nu_2} & \cdots & \frac{\partial^2 \log L}{\partial \nu_1 \nu_k} \\ \frac{\partial^2 \log L}{\partial \nu_2 \nu_1} & \frac{\partial^2 \log L}{\partial {\nu_2}^2} & \cdots & \frac{\partial^2 \log L}{\partial \nu_2 \nu_k} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 \log L}{\partial \nu_k \nu_1} & \cdots & \cdots & \frac{\partial^2 \log L}{\partial {\nu_k}^2} \end{bmatrix} \tag{9}$$

for k equals the number of coefficients of the parameter vector ν.
Quasi-Newton methods are a generalization of the secant method (finite difference

method) to find the root of the first derivative for multidimensional problems. In multiple dimensions the secant equation is under-determined, and quasi-Newton methods differ in how they constrain the solution, typically by adding a simple low-rank update to the current estimate of the Jacobian matrix.

In our analysis we optimize log-likelihood function by a simple finite difference method. In numerical analysis, the secant method is a root-finding algorithm that uses a succession of roots of secant lines to better approximate a root of a function f. The secant method can be thought of as a finite-difference approximation of Newton's method. To conclude, by giving initial values to every parameter of our models, we extract the parameter vector that maximizes our function. As starting values we use a random guess and then, after implementing iteratively the algorithm, we end up with the convergence point. It is important to mention that the implementation of quasi Newton-Raphson method is done in R programming with nlm function.

3.3.2.2 EM algorithm The second optimization algorithm we use is called expectation maximization algorithm. One of the main disadvantages for using bivariate regression models is that the log-likelihood is complicated, thus its maximization needs a special effort. For some models we use an EM algorithm for maximum likelihood estimation (ML) as proposed by Karlis (2001, 2005).
To specify, this EM type algorithm is easy to use for finding the MLEs of the model's parameters, since we do not use derivative based optimization.
The EM algorithm, Dempster, Laird and Rubin (1977), in general is a powerful algorithm for maximum likelihood estimation for data containing missing values or being considered as containing missing values. The EM algorithm is an iterative approach that cycles between two modes. The first mode attempts to estimate the missing or latent variables, called the expectation-step or E-step. The second mode attempts to optimize the parameters of the model that best explain the data, called the maximization-step or M-step.
E-Step : Estimate the missing variables in the dataset.
M-Step : Maximize the parameters of the model in the missing data.

We choose initial values for EM algorithm, by optimizing parameters for N_1 and

N_2 independently. In other words, we fit two univariate regression models, with the appropriate response every time, based on the related distribution. We optimize parameters with NR algorithm and then we use these optimized values as initials in our EM algorithm. We go in further details in EM algorithm in next sections of our analysis in mixture models, where we explain all steps of this routine that is also implemented in R programming.

3.3.3 Akaike (AIC)

AIC is a penalized-likelihood criterion. It is sometimes used for choosing best predictor subsets in regression and often used for comparing nonnested models, like BIC value, which ordinary statistical tests cannot do. The AIC value for a model is usually written in the form:

$$AIC = -2 \log L + 2p \tag{10}$$

where L is the likelihood function, p is the number of parameters in the model.
AIC is an estimate of a constant plus the relative distance between the unknown true likelihood function of the data and the fitted likelihood function of the model, so that a lower AIC means a model is considered to be closer to the truth. Akaike information criterion (AIC) (Akaike, 1974) is a fined technique based on in-sample fit to estimate the likelihood of a model to predict/estimate the future values. A good model is the one that has minimum AIC among all the other models.
In every model we use, we compute AIC criterion in order to rank them and find the most efficient models.

3.4 Models

Modeling insurance count claims is a very challenging issue and lots of interensting results can be derived. As we mentioned earlier we have two different categories of claims, third-party liability guarantees and the rest. So, for 6000 different policyholders we have a bivariate response variable that counts the number of claims for each type in a period of time. Moreover, we have explanatory variables in our dataset that characterize every policyholder and are used in our models.

Our analysis begins with Poisson distribution as the probability distribution of the response variable, in many versions in order to overcome difficulties that occur. We continue with, mixed Poisson distribution like negative binomial distribution and then we end using generalised Poisson distribution. Methods that are used in every probability distribution may vary between techniques we mentioned above. Poisson models refer to homogeneous populations, a fact that does not exist in insurance applications. Our population given all the characteristics that we use as explanatory variables in our models is discriminated based on several criteria. Every insured is unique, with different driving abilities and habits, a fact that highlights the heterogeneity of our sample.

As a result this big problem that arises in automobile insurance applications is called unobserved heterogeneity due to the reasons we mentioned above.
In count data models, one consequence of this heterogeneity is overdispersion which practically is the case, when the observed variance exceeds that implied by the assumed model.
Eventhough, we know that in Poisson distribution the mean equals the variance and our data are overdispersed, it is quite interesting to examine the fit of Poisson models despite the fact that equidisperision assumption is not valid and propose alternatives to overcome this issue.
Furthermore, one other consequence is the excess of zeros, which practically happens because some policyholders avoid to report some accidents. Therefore, it is interesting to find the appropriate models to overcome these difficulties. Poisson models cannot deal with overdispersion since the variance equals the mean, so negative binomial models are introduced. Zero-inflation occurs when the proportion of

zeros in a sample exceeds what would be expected from the assumed distribution.

All the models implemented are parametric, since probabilities are known functions depending on finite and real valued parameters. Therefore, they are estimated via maximum likelihood method. As we go through our analysis, we attempt to model count claims and their dependence through various techniques.

3.4.1 Poisson distribution

The Poisson distribution, Poisson (1837), is the most important we will go through. It is a probability distribution of the number of occurencies of an event that happens rarely but has many opportunities to happen. It is used for independent events which occur at a constant rate within a given interval of time. Under this definition, it is obvious that it has various applications as a distribution since anyone can model the number of traffic accidents at a particular intersection, the number of house fire claims per month that are received by an insurance company or for example the number of people who are infected with the AIDS virus in a certain neighborhood. In a binomial distribution, if the number of trials, n, gets larger and larger as the probability of success, p, gets smaller and smaller, we obtain a Poisson distribution.
The Poisson probability distribution function is given by:

$$P(N_{ji} = n_i) = \frac{{\theta_j}^{n_i} e^{-\theta}}{n_i!} \tag{11}$$

where $n_i = 0, 1, 2, ...,\ j = 1, 2,\ i = 1, 2, ..., 6000$. A basic property of Poisson is the equality between its mean and variance $E[N_j] = Var[N_j] = \theta_j$ for $j = 1, 2$. Moreover each event is independent from all other events. This function refers to the univariate case and it would be useful if we were trying to model separately N_1 (number of third-liability claims) and N_2 (number of the rest claims), using as θ_j the expected number of count claims.
In order to model these response variables and using what we mentioned above we are lead to Poisson regression.
Poisson regression is a generalized linear model form of regression analysis used to model count data and contingency tables. It assumes the response variable N_j

has a Poisson distribution, and assumes the logarithm of its expected value can be modeled by a linear combination of explanatory variables. In our case, these explanatory variables were explained earlier and describe all the characteristics of a policyholder.
In our problem the univariate Poisson regression model is given by:

$$\log(\theta_{ji}) = x_{ji}^T \beta_j \tag{12}$$

where θ_{ji}=$E(N_{ji}|x_{ji})$, β_j denotes the regression coefficient , $j = 1, 2$ and $i = 1, 2, ..., 6000$.

The regression coefficient β_j is a vector of coefficients for all explanatory variables we use, for every type of guarantee. The x_{ji} is a vector of values from explanatory variables, for each observation i and each type of guarantee j.

Constructing a univariate generalised Poisson linear model seems simple. But, in our case, bivariate Poisson models are not a task as easy as in the univariate case. In order to serve our purposes we will introduce the bivariate version of the Poisson distribution and regression using different techniques because count variables are correlated and we want them to be estimated jointly.

The bivariate Poisson is the most widely used model for bivariate counts. It was proposed by Holgate (1964) and presented by Johnson and Kotz (1969). The definition of the bivariate Poisson distribution is not unique. Several approaches have been discussed by Kocherlakota and Kocherlakota (1992). Here, the first method we adopt to construct the distribution is the trivariate reduction method, Johnson et al., (1997).

3.4.1.1 Bivariate Poisson regression using trivariate reduction method

In order to construct bivariate Poisson, we use trivariate reduction method, we explained earlier. The bivariate Poisson distribution is a typical example given from Kocherlakota and Kocherlakota (1992, Ch.4) and it is very useful when it comes to modeling pairs of count data that correlation exists.

Our first model is a bivariate Poisson model proposed by Bermudez (2009).
Let N_1 and N_2 be the claims for third-party liability and rest of the guarantees respectively.
We consider $X_i, i = 1, 2, 3$ independent random variables that follow Poisson distribution with rate θ_i respectively.
Then, through trivariate reduction $N_1 = X_1 + X_3$ and $N_2 = X_2 + X_3$ follow jointly a bivariate Poisson distribution denoted as:
$(N_1, N_2) \sim BP(\theta_1, \theta_2, \theta_3)$ with joint probability mass function:

$$P(N_1 = n_1, N_2 = n_2) = e^{-(\theta_1+\theta_2+\theta_3)} \frac{\theta_1^{n_1}}{n_1!} \frac{\theta_2^{n_2}}{n_2!} \times \sum_{i=0}^{min(n_1,n_2)} \binom{n_1}{i} \binom{n_2}{i} i! (\frac{\theta_3}{\theta_1\theta_2})^i \quad (13)$$

Bivariate Poisson distribution allows positive dependence between N_1 and N_2 with $Cov(N_1, N_2) = \theta_3 > 0$, which is always positive in our case.

As a matter of fact, if $\theta_3 = 0$ our random variables are independent and the model reduces to a product of two Poisson marginals with rate θ_1 and θ_2 respectively known as Double Poisson, Kocherlakota and Kocherlakota (1992).
Furthermore, since $E[N_j] = \theta_j + \theta_3$ and $V[N_j] = \theta_j + 2\theta_3$, $j = 1, 2$. If $\theta_3 > 0$ then overdispersion exists because variance exceeds the mean.

The most interesting part lies on modeling θ_j with covariates.
Let's assume that N_{1i} and N_{2i} are the random of variables of each type of guarantee for the i-th policyholder. Then:

$$(N_{1i}, N_{2i}) \sim BP(\theta_{1i}, \theta_{2i}, \theta_{3i}) \tag{14}$$

$$\log(\theta_{1i}) = {x_{1i}}^T \beta_1$$
$$\log(\theta_{2i}) = {x_{2i}}^T \beta_2$$
$$\log(\theta_{3i}) = {x_{3i}}^T \beta_3$$

where $i = 1, 2, ..., n$ denotes the policyholder, x_{ji} is a vector of explanatory variables and β_j is the regression coefficient for $j = 1, 2, 3$.

As we noticed earlier, from Table 2, the proportion of zero pairs of (N_{1i}, N_{2i}) is greater than any other combination.
Therefore, the simple bivariate Poisson regression model we just described may does not fit well the data, since zero pairs are more than expected from Poisson distribution. Over 80% of the total number of pairs is a zero pair (0,0) so it is necessary to introduce zero-inflation as an alternative model.

3.4.1.2 Zero-inflation We briefly had a look at the structure of a regular bivariate Poisson model under trivariate reduction. If we modify appropriately this model, we construct a new one that can handle the excess of zero counts.
Normally, we assume that there is some underlying process that is producing the observed pairs of counts as per the bivariate Poisson probability mass function: $P(N_{1i} = n_{1i}, N_{2i} = n_{2i})$. The intuition behind zero-inflation is that there is a second underlying process that is determining whether a pair of counts is zero or non-zero. This process allows for frequent zero-valued observations. It is important to mention, that we use inflation only at (0,0) point.
In our situation, we will use a bivariate zero inflated Poisson model (ZIBP) with probability function:

$$f_{ZIBP}(n_{1i}, n_{2i}) = \begin{cases} p + (1-p) f_{BP}(0, 0 | \theta_{1i}, \theta_{2i}, \theta_{3i}), n_{1i} = n_{2i} = 0 \\ (1-p) f_{BP}(n_{1i}, n_{2i} | \theta_{1i}, \theta_{2i}, \theta_{3i}), \text{ otherwise} \end{cases} \tag{15}$$

where $f_{BP}(n_{1i}, n_{2i} | \theta_{1i}, \theta_{2i}, \theta_{3i})$ is the bivariate Poisson probability mass function

as mentioned in (13), with p determined in [0,1] and θ_{ji} linked with covariates under a log-linear relationship.

A great advantage of zero inflated model is that it allows for overdispersion since marginals are not Poisson distributed.
Mean and variance of N_1:

$$E_{ZIBP}[N_1] = (1-p)(\theta_1 + \theta_3) \tag{16}$$

$$V_{ZIBP}[N_1] = (1-p)((\theta_1 + \theta_3) + p(\theta_1 + \theta_3)^2) \tag{17}$$

In order to implement these models, we use an EM algorithm proposed by Karlis and Ntzoufras (2003) through R programming. This model was proposed by Karlis and Ntzoufras (2003, 2005) in order to allow for overdispersion of the corresponding marginal distributions. It was applied by Wang et al. (2003) to analyze two types of occupational injuries and by Bermúdez (2009) in the automobile insurance context for a bivariate case.

As we mentioned in another section the expectation–maximization (EM) algorithm is an iterative method to find (local) maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables. The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step.

3.4.1.3 Bivariate Poisson Regression (BPR) with multiplicative factor

Our first approach in this type of modeling is based on Lakshminarayana's (1999) bivariate Poisson with multiplicative factor model. The correlation between the two variables can be either positive or negative, depending on the value chosen for the parameter in the above multiplicative factor.

The extension lies on the assumption that the marginal means of the bivariate model are functions of the explanatory variables. Moreover, we assume that the relationship between marginal means and the explanatories is log-linear.

The probability function of this model is given as:

$$P_{BPR}(n_1, n_2) = \frac{\theta_1^{n_1}\theta_2^{n_2}e^{-\theta_1-\theta_2}[1+\alpha(e^{-n_1}-e^{-s\theta_1})(e^{-n_2}-e^{-s\theta_2})]}{n_1!n_2!} \tag{18}$$

where n_1,n_2=0,1,2,.. and $s = 1 - e^{-1}$.

The marginal distribution of N_j (j=1,2) is Poisson with mean θ_j. The covariance between N_1 and N_2 is $\alpha\theta_1\theta_2 s^2 e^{-s(\theta_1+\theta_2)}$, so the correlation coefficient is $\rho = \alpha\sqrt{\theta_1\theta_2}s^2e^{-m(\theta_1+\theta_2)}$.

It is obvious that the correlation can be zero, positive or negative depending on the value of α.

We assumed that $\log(\theta_j) = \mu_{ji}$, where $\mu_{ji} = {x_{ji}}^T\beta_j$ for $i = 1, 2..., n$ and $j = 1, 2$ with β_j the regression coefficient vector and x_{ji} the covariates' vector for every observation.

The log-likelihood of BPR, a quantity we use in order to compare all models efficiency is given:

$$l(\theta_1, \theta_2; n_1, n_2) = \sum_{i=1}^{n}\sum_{j=1}^{2}[n_{ji}\log(\theta_{ji}) - \theta_{ij} - \log(n_{ji}!)] + \sum_{i=1}^{n}\log(1+\alpha(e^{-n_{1i}}-e^{-s\theta_{1i}})(e^{-n_{2i}}-e^{-s\theta_{2i}})) \tag{19}$$

where α is the correlation parameter.

Although, we introduced a method that allows every kind of correlation between counts of claims, we still face the problem of overdispersion. Due to this fact we use zero inflation once again in order to improve our regression model.

3.4.1.4 Bivariate Zero-Inflated Poisson Regression (BZIPR) with multiplicative factor

One way to deal with overdispersion is constructing models with distributions which have an additional parameter that explains dispersion like negative binomial and generalised Poisson.

However, we have to face another phenomenon, the occurence of more zero counts than expected under a standard Poisson model. In our case, such a situation is common, since many policyholders, due to conditions in their policies, they are discouraged to report small claims.
Taking all the previous under consideration, we implement a zero-inflated bivariate Poisson model Li et al. (1999), Wang et al. (2003), Faroughi (2017).

For our model we will use bivariate Poisson model(BPR), as defined from Lakshminarayana (1999) and we will extend it to its zero inflated version.
The regression model comes as a result of mixing BPR regression model with a distribution degenerate at zero. The joint probability mass function of BZIPR regression model is:

$$P_{BZIPR}(n_{1i}, n_{2i}) = \begin{cases} p + (1-p)[\prod_{j=1}^{2} e^{-\theta_{ji}}][1 + \alpha \prod_{j=1}^{2}(1 - e^{-s\theta_{ji}})], n_{1i} = n_{2i} = 0 \\ (1-p)P_{BPR}(n_{1i}, n_{2i}), \text{ otherwise} \end{cases} \tag{20}$$

where $0 \leq p < 1$ is the zero-inflation parameter, $s = 1 - e^{-1}$, α is the correlation parameter and $P_{BPR}(n_{1i}, n_{2i})$ is the joint p.m.f of BPR regression model shown in Equation (18).
The covariates can be included using log links, $\log(\theta_{ji}) = x_{ji}^T \beta_j$, j=1,2 but no covariates will be used for the inflation parameter.
The marginal mean is $E(N_{ji}) = (1-p)\theta_{ji}$ and the marginal variance is $Var(N_{ji}) = E(N_{ji})(1 + p\theta_{ji})$.
From marginal variances it is obvious that if $p = 0$, then BZIPR reduces to BPR. When α is positive or negative, we have same correlations (positive or negative).

The log-likelihood function is given once again by:

$$logL = \begin{cases} \sum_{i=1}^{n} \log(p + (1-p)[\prod_{j=1}^{2} e^{-\theta_{ji}}][1 + \alpha \prod_{j=1}^{2}(1 - e^{-s\theta_{ij}})]), n_{1i} = n_{2i} = 0 \\ \sum_{i=1}^{n} \log((1-p)P_{BPR}(n_{1i}, n_{2i})), \text{otherwise} \end{cases} \tag{21}$$

3.4.1.5 Bivariate Poisson Regression using copulas Besides trivariate reduction and multiplicative factors we analyzed a third method of modeling,copulas. The approach has attracted considerable attention, Genest and Rivest (1993), Joe (1997), Nelsen (1999), Caperaa et al. (2000). In this paragraph, we introduce a bivariate Poisson model using Frank's and Clayton's copula which take the form of Equation (5) and (6) respectively.
For our case, suppose $F(n_{1i}, n_{2i})$ is a joint distribution with corresponding marginal distributions $F_1(n_{1i})$ and $F_2(n_{2i})$. Then $F(n_{1i}, n_{2i})$ can be expressed as:

$$F(n_{1i}, n_{2i}) = C(F_1(n_{1i}), F_2(n_{2i}); \xi) \tag{22}$$

where C is a parametric copula function and ξ is a dependence parameter measuring dependence between the two random variables.
Both F_{ji}, for $i = 1, 2, ..., n$ and $j = 1, 2$, will be univariate Poisson distributions. We add covariates to the mean through $\theta_{ji} = exp(x_{ji}^T \beta_j)$ for $i = 1, 2, ..., n$ and $j = 1, 2$.

If our response variables were continuous then, we could derive the joint probability mass function by partially derive the cumulative function with respect to our two response variables. Since we are in the discrete case, we obtain the joint probability mass function of N_{1i} and N_{2i},using finite differences, Cameron et al. (2004).

$$f(n_{1i}, n_{2i}) = F(n_{1i}, n_{2i}) - F(n_{1i}-1, n_{2i}) - F(n_{1i}, n_{2i}-1) + F(n_{1i}-1, n_{2i}-1) \tag{23}$$

where f is the Poisson density function as mentioned in (13) and F the Poisson distribution function as the summation of each density over each observation. Therefore as $F_1(n_{1i})$ and $F_2(n_{2i})$ we use the following form G_{ji}:

$$\mathrm{G}_{ji} = \begin{cases} \sum_{k=0}^{n_{ji}} f_{Pois}(k; \theta_{ji}), \text{for Poisson distribution} \\ p + (1-p) \sum_{k=0}^{n_{ji}} f_{Pois}(k; \theta_{ji}), \text{for zero-inflated Poisson distribution} \end{cases}$$

where p_j is the inflation parameter for $j = 1, 2$ and $G_{1i} = u$ and $G_{2i} = v$.

Therefore,the log-likelihood function is given by:

$$\log L = \sum_{i=1}^{n} \log(C_{FR}(G_{1i}, G_{2i}) - C_{FR}(G_{1i}-1, G_{2i}) - C_{FR}(G_{1i}, G_{2i}-1) + C_{FR}(G_{1i}-1, G_{2i}-1)). \tag{24}$$

where C_{FR} is the Frank's copula (5), but the same holds for Clayton's (6) with the appropriate form. Cameron et al. (2004) and So et al. (2011) proposed Poisson and zero-inflated negative binomial models using Frank's copula function.

3.4.2 Mixed Poisson Distributions

The second part of our analysis is focused on mixed Poisson models that deal with problems mentioned above.
Our interest lies on bivariate negative binomial, inverse Gaussian and Poisson lognormal regression models like Ghitany, Karlis et al. (2012) proposed in the multivariate case though.

In our case we have two Poisson random variables N_1 and N_2 with means $\alpha'\theta_1$ and $\alpha'\theta_2$ respectively. Random variable α', from a mixing distribution, is common for both random variables.
Mixing distribution introduces overdisperision and common α' for N_1,N_2 introduces correlation.
Furthermore, we will use covariates for our modeling through log-link:

$$\log(\theta_{ji}) = {x_{ji}}^T \beta_j \tag{25}$$

for $j = 1, 2$ and $i = 1, 2, ..., n$, where β_j is the regression coefficient vector, x_{ji} is the covariate one and $E[\alpha'] = 1$ in order to achieve identifiability of the model.

Let $N_{ji} \sim P(\alpha'_i \theta_{ji})$ with $i = 1, 2, ..., n$ and $j = 1, 2$.
α' is an independent identically distributed random variable from a mixing distribution $G(\alpha'; \phi)$,where ϕ is a vector of parameters.
We allow for regressors using a log-link, through Equation (25).

The joint probability mass function of the model is:

$$P(n_{1i}, n_{2i}; \phi) = \int_0^\infty \prod_{j=1}^{2} \frac{exp(-\alpha'\theta_j)(\alpha'\theta_j)^{n_{ji}}}{x_{ji}!} g(\alpha'; \phi) d\alpha' \tag{26}$$

$g(\alpha'; \phi)$ is the probability density function of α'.

Some important properties to mention are: both count variables have the same mixed Poisson distribution, the variance of N_{ji} is $Var(N_{ji}) = \theta_{ji}(1 + \theta_{ji}\sigma^2)$ for $j = 1, 2,\ i = 1, 2, ..., n$ and $Cov(N_{1i}, N_{2i}) = \theta_{1i}\theta_{2i}\sigma^2$,where σ^2 is the variance of $g(\alpha'; \phi)$.
We will implement three different mixed Poisson models as mentioned earlier allowing for regressors with $log(\theta_{ji}) = {x_{ji}}^T \beta_j$.

3.4.2.1 EM algorithm for Poisson Mixtures One of the main disadvantages for using bivariate mixed Poisson regression models is that the log-likelihood is complicated, thus its maximization needs a special effort. For all three models we use an EM algorithm for maximum likelihood estimation (ML) as proposed by Karlis (2001, 2005).
To specify, this EM type algorithm is easy to use for finding the MLEs of the model's parameters, since we do not use derivative based optimization. The EM algorithm in general is a powerful algorithm for maximum likelihood estimation for data containing missing values or being considered as containing missing values. This formulation is particularly suitable for distributions arising as mixtures since the mixing operation can be considered as producing missing data. The unobserved quantities are simply the realizations α_i' of the unobserved mixing parameter for the i-th observation. Hence, at the E-step one needs to calculate the conditional expectation of some functions of α_i''s and then to maximize the likelihood of the complete model which reduces to maximizing the likelihood of the mixing density. For more details, see Karlis (2001, 2005).

Let's describe EM algorithm for the bivariate negative binomial model:
For the gamma mixing distribution, the posterior distribution of $\alpha'|x$ is a gamma distribution with parameters $\gamma + \sum_{j=1}^{2} n_{ji}$ and $\gamma + \sum_{j=1}^{2} \theta_{ji}$, for $i = 1, 2, ..., 6000$.

The posterior distribution is proportional to prior distribution multiplies by the likelihood function, Bayes (1763).

$$f(\alpha_i'|n_{ji}) \propto f(n_{ji}|\alpha_i')f(\alpha_i') \tag{27}$$

Therefore, if N_{ji}, for $i = 1, 2, ..., n$ and $j = 1, 2$, which are our counts of claims response variables follow Poisson distribution with parameter θ_{ji} and α_i' follows a $Gamma(\gamma, \gamma)$ then:

$$\begin{aligned} f(\alpha_i'|n_{ji}) &\propto \frac{\gamma^\gamma}{\Gamma(\gamma)}{\theta_{ji}}^{\gamma-1}\exp(-\gamma\theta_{ji})\exp(-\sum_{i=1}^{n}\theta_{ji}){\theta_{ji}}^{\sum_{i=1}^{n}n_{ji}} \\ &\propto {\theta_{ji}}^{\sum_{i=1}^{n}n_{ji}+\gamma-1}\exp(-(\sum_{i=1}^{n}\theta_{ji}+\gamma)) \end{aligned}$$

which is a $G(\gamma + \sum_{j=1}^{2} n_{ji}, \gamma + \sum_{j=1}^{2} \theta_{ji})$ distribution.

For the E-step, we use the posterior mean to compute $E(\alpha_i'|n_{ji})$ and we use the posterior distribution to compute $E(\log \alpha_i'|n_{ji})$. The EM-algorithm,in this case, is as follows:

• E-step: Calculate for $i = 1, 2, ..., 6000$, the estimates S_{ji} in order to form pseudo-values B_{ji}.

$$S_{ji} = E(\alpha_i'|n_{ji}) = \frac{\gamma + \sum_{j=1}^{2} n_{ji}}{\gamma + \sum_{j=1}^{2} \theta_{ji}} \tag{28}$$

$$B_{ji} = E(\log(\alpha_i')|n_{ji}) = \Psi(\gamma + \sum_{j=1}^{2} n_{ji}) - \log(\gamma + \sum_{j=1}^{2} \theta_{ji}) \tag{29}$$

where $\Psi(.)$ is the digamma function and $\theta_{ji} = \exp({x_{ji}}^T\beta_j)$, $i = 1, 2, ..., n$ and $j = 1, 2$, are obtained using the current values of β_j

• M-step: Update the regression parameters β_j, j=1,2,...,p, using the pseudo-values B_{ji} as offset values and by fitting a simple Poisson regression model.

Update γ by

$$\gamma_{new} = \gamma_{old} - \frac{\Psi(\gamma_{old}) - \log(\gamma_{old}) - B + S - 1}{\Psi'(\gamma_{old}) - 1/\gamma_{old}} \tag{30}$$

where Ψ'(.) denotes the trigamma function, S and B are the sample means of $S_{j1}, S_{j2}, ..., S_{jn}$ and $B_{j1}, B_{j2}, ..., B_{jn}$,respectively. This is the one step ahead Newton iteration. For further details see Karlis (2001, 2005).
We consider as complete data $(N_{1i}, N_{2i}, \alpha_i')$, the observed data with the unobserved value of mixing distribution.
Then, the loglikelihood function is:

$$l_C(\Theta) = \sum_{i=1}^{n}\sum_{j=1}^{2}(-\alpha_i'\theta_{ji} + n_{ji}\log(\alpha_i'\theta_{ji}) - \log(n_{ji}!)) + \sum_{i=1}^{n}\log(g(\alpha_i';\phi)) \tag{31}$$

where Θ stands for all parameters of the model.

3.4.2.2 Bivariate Negative Binomial Since we explained the way of mixing distributions in order to from bivariate cases, it is essential to present how this method could be applied.
Our first mixture regression model is the bivariate negative binomial regression. There are many different ways to define negative binomial distribution, but we will present two of them.
Firstly, as the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified (and fixed) number of failures occurs. Secondly, the negative binomial distribution can be viewed as a Poisson distribution where the Poisson parameter is itself a random variable, distributed according to a Gamma distribution. Thus, the negative binomial distribution is known as a Poisson-Gamma mixture. Negative binomial regression can be used for overdispersed count data, that is when the conditional variance exceeds the conditional mean. It can be considered as a generalization of Poisson regression since it has the same mean structure as Poisson regression and it has an extra parameter to model the overdispersion.

For our analysis, we need the bivariate negative binomial regression, since we want to model two different count variables.
The bivariate negative binomial can be easily derived if we choose gamma distribution as the mixing one with density function. We assume that $N_{ji} \sim Poisson(\alpha'\theta_j)$ for $j = 1, 2$, while θ_j is the realization of our random variables. Since α_i' follows $G(\phi, \phi)$:

$$g(\alpha_i'; \phi) = \frac{\phi^\phi}{\Gamma(\phi)} \alpha_i'^{\phi-1} exp(-\phi\alpha_i') \tag{32}$$

where ϕ>0. Then the joint probability mass function is, for more information see Johnson(1997):

$$P_G(n_{1i}, n_{2i}; \phi) = \frac{\Gamma(n_{1i} + n_{2i} + \phi)}{\Gamma(\phi) n_{1i}! n_{2i}!} \left(\frac{\phi}{\phi + \theta_1 + \theta_2}\right)^\phi \left(\frac{\theta_1}{\phi + \theta_1 + \theta_2}\right)^{n_{1i}} \left(\frac{\theta_2}{\phi + \theta_1 + \theta_2}\right)^{n_{2i}} \tag{33}$$

In order to introduce the correlation coefficient between two marginals for individual i, we have to include the exposure of each individual E_{ji}, for $i = 1, ..., n$ and $j = 1, 2$.
Then, correlation coefficient for individual i is:

$$corr(N_{1i}, N_{2i}) = \sqrt{\frac{E_{1i}\theta_{1i} E_{2i}\theta_{2i}}{(E_{1i}\theta_{1i} + \alpha_i')(E_{2i}\theta_{2i} + \alpha_i')}} \tag{34}$$

which is a positive quantity for given N_{1i}, N_{2i}.

3.4.2.3 Bivariate Negative Binomial Regression (BNBR) with multiplicative factor Another model we will use in order to deal with overdispersion is the negative binomial one combined with a multiplicative factor. One special characteristic is that it allows for more flexible correlation structure.

Following the Lakshminarayana's technique in the bivariate Poisson model (BPR), Famoye (2010) defined the negative binomial regression with covariates, modeling θ_j through loglinear functions for μ_{ji}, for $i = 1, 2, ..., n$ and $j = 1, 2$.
In our analysis, the multiplicative factor α and the dispersion parameter of the negative binomial distribution m_j will not be functions of covariates. Therefore,

working as earlier, this model is a product of two negative binomial marginals with multiplicative factor.
The joint probability mass function is given by:

$$P(n_{1i}, n_{2i}) = \prod_{j=1}^{2} \binom{m_j^{-1} + n_{ji} - 1}{n_{ji}} \left(\frac{\mu_{ji}}{m_j^{-1} + \mu_{ji}}\right)^{n_{ji}} \left(\frac{m_j^{-1}}{m_j^{-1} + \mu_{ji}}\right)^{m_j^{-1}} [1+\alpha \prod_{j=1}^{2}(e^{-n_{ji}} - c_j)] \tag{35}$$

where $c_j = [(1-\theta_j)/(1-\theta_j e^{-1})]^{m_j^{-1}}$ with $\theta_j = \mu_{ji}/(m_j^{-1} + \mu_{ji})$ for $i = 1, 2, ..., n$ and $j = 1, 2$.

When $m_j \to 0$,then there is no dispersion and the model reduces to bivariate Poisson.
The marginal distributions of N_j is a negative binomial with mean $\mu_j = m_j^{-1} \frac{\theta_j}{1-\theta_j}$ and variance $\sigma_j^2 = m_j^{-1} \frac{\theta_j}{(1-\theta_j)^2}$. The correlation coefficient depends on the value of α, the multiplicative factor parameter and it is given by:

$$\rho = \alpha \frac{E(e^{-N_1})E(e^{-N_2})A_1 A_2}{\sigma_1 \sigma_2} \tag{36}$$

where $A_j = \frac{m_j^{-1}\theta_j e^{-1}}{(1-\theta_j e^{-1})} - \frac{m_j^{-1}\theta_j}{1-\theta_j}$ for $j = 1, 2$. As a result, if α equals to zero, then our variables are independent, if it is positive, they have positive dependence and if it is negative, they have negative dependence. For more information about correlation structure, see Famoye (2009).

The log-likelihood function is given by:

$$logL = \sum_{i=1}^{n} [\sum_{j=1}^{2} [n_{ji} \log \mu_{ji} - m_j^{-1} \log m_j - (n_{ji} + m_j^{-1}) \log(\mu_{ji} + m_j^{-1}) - \log n_{ji}! + \sum_{k=0}^{n_{ji}-1} \log(m_j^{-1} + k)] + \log[1 + \alpha \prod_{j=1}^{2}(e^{-n_{ji}} - c_j)]] \tag{37}$$

where $c_j = (1 + s\mu_{ji}m_j)^{-1/m_j}$, with $s = 1 - e^{-1}$, $i = 1, 2, ..., n$ and $j = 1, 2$, while k is a positive integer counter that take values from zero to $n_{ji} - 1$.

3.4.2.4 Bivariate zero-inflated Negative Binomial Regression (BZINBR) with a multiplicative factor

An alternative to the previous model can be a zero-inflated negative binomial regression model.
This model is obtained by mixing a distribution digenerate at zero and the NB distribution. We will present two types of ZINB models that can be found in Ridout et al. (2004).

The models we will use was proposed by Faroughi (2017) and they are similar to generalised Poisson models proposed by Zamani (2012) that we will explain later. These negative binomial models have different forms with respect to the functional parameter, Shi and Valdez (2012).The advantages of these models are firstly that, they are nested models.Secondly, they allow of additional overdispersion and flexible forms of mean-variance relationship.

The probability mass function of the univariate negative binomial-P regression model is:

$$P_{BN-P}(n_i) = \frac{\Gamma(n_i + m^{-1}\mu_i^{2-P})}{n_i!\Gamma(m^{-1}\mu_i^{2-P})} \left(\frac{m^{-1}\mu_i^{2-P}}{m^{-1}\mu_i^{2-P} + \mu_i}\right)^{m^{-1}\mu_i^{2-P}} \left(\frac{\mu_i}{m^{-1}\mu_i^{2-P} + \mu_i}\right)^{n_i}$$

(38)

where m is the dispersion parameter and P is the functional one.
P is a functional parameter that makes our models nested. This functional parameter, takes values from 1 to P, while P is a positive number greater than 1. Using this parameter, anyone can do statistical tests of these parametric forms. If we replace P with 1 or 2 we have the NB-1 and NB-2 models respectively. Therefore,

from now on model NB-1 will be a negative binomial model with functional parameter equals to one, NB-2 a negative binomial model with functional parameter equals to two and NB-P, a negative binomial model with parameter equals to P, that we can treat later as unknown and as a result, through optimization we can find the value that corresponds to the lowest value of the likelihood function .
Similar to the derivation of the other bivariate models, the joint probability mass function of the BNB-P model, bivariate negative binomial model with functional parameter P, can be derived from the product of two NB-P marginals with a multiplicative factor, Zamani (2012).

$$P_{BNBP}(n_{1i}, n_{2i}) = [\prod_{j=1}^{2} \frac{\Gamma(n_{ji} + m_j^{-1}{\mu_{ji}}^{2-P_j})}{n_{ji}!\Gamma(m_j^{-1}{\mu_{ji}}^{2-P_j})} \left(\frac{m_j^{-1}{\mu_{ji}}^{2-P_j}}{m_j^{-1}{\mu_{ji}}^{2-P_j} + \mu_{ji}}\right)^{m_j^{-1}{\mu_{ji}}^{2-P_j}} \times \left(\frac{\mu_{ji}}{m_j^{-1}{\mu_{ji}}^{2-P_j} + \mu_{ji}}\right)^{n_{ji}}][1 + \alpha \prod_{j=1} 2(e^{-n_{ji}} - c_{ji})] \tag{39}$$

where $c_{ji} = E(e^{-n_{ji}}) = (1 - \theta_{ji}/1 - \theta_{ji}e^{-1})^{m_j^{-1}{\mu_{ij}}^{2-P_j}}$ with $\theta_{ji} = \mu_{ji}/m_j^{-1}{\mu_{ji}}^{2-P_j} + \mu_{ji}$, for $j = 1, 2$ where m_j are the dispersion parameters, P_j are the functional parameters and α the multiplicative factor for the model.

It is easy to observe from Equation (39), when $\alpha = 0$, the response variables N_{1i} and N_{2i} are independent, each follows negative binomial distribution, with functional parameter P. The reason why, they are independent is that the probability function of our bivariate response variable becomes a product of two separate marginals univariate negative binomial distributed (NB-P). When $\alpha < 0$ or $\alpha > 0$ we have negative or positive correlation respectively.

If we mix the BNB-P regression model with a distribution degenerate at zero, we have the zero-inflated extension (BZINB-P), with joint probability mass function:

$$P_{BZINB}(n_{1i}, n_{2i}) = \begin{cases} p + (1-p)[\prod_{j=1}^{2} \left(\frac{m_j^{-1} \mu_{ji}^{2-P_j}}{m_j^{-1} \mu_{ji}^{2-P_j} + \mu_{ji}} \right)^{m_j^{-1} \mu_{ji}^{2-P_j}}] \\ \quad \times [1 + \alpha \prod_{j=1}^{2} (1 - c_{ji})], N_{1i} = N_{2i} = 0 \\ (1-p) P_{BNBP}(n_{1i}, n_{2i}), \text{ otherwise} \end{cases} \tag{40}$$

where $0 \leq p \leq 1$ is the zero-inflation parameter, m_j is the dispersion one, P_j the functional parameters and α the multiplicative factor and $j = 1, 2$.
c_{ji} and θ_{ji} are the same as in BNB-P model, mentioned above.
Covariates are included using log-links, $\log(\theta_{ji}) = x_{ji}^T \beta_j$, $j = 1, 2$, where x_{ji} is the vector of explanatory variables and β_j is the vector of regression coefficients.
The log-likelihood for the BZINB-P regression model is given by:

$$\begin{aligned} \log L = \sum_{i=1}^{n} I_{n_{1i}=0, n_{2i}=0} [\log[p + (1-p)[\prod_{j=1}^{2} \left(\frac{m_j^{-1} \mu_{ji}^{2-P_j}}{m_j^{-1} \mu_{ji}^{2-P_j} + \mu_{ji}} \right)^{m_j^{-1} \mu_{ji}^{2-P_j}}][1 + \alpha \prod_{j=1}^{2} (1 - c_{ji})]]] \\ + \sum_{i=1}^{n} [1 - I_{n_{1i}=0, n_{2i}=0}] [\log(1-p) + \sum_{j=1}^{2} [\log \mu_{ji} + m_j^{-1} \mu_{ji}^{2-P_j} \log(_j^{-1} \mu_{ji}^{2-P_j}) \\ - N_{ji} \log(m_j^{-1} \mu_{ji}^{2-P-j} + \mu_{ji}) - m_j^{-1} \mu_{ji}^{2-P_j} \\ \times \log(m_j^{-1} \mu_{ji}^{2-P_j}) + \sum_{k=0}^{n_{ji}-1} \log(m_j^{-1} \mu_{ji}^{2-P_j+k})] + \log(1 + \alpha(e^{-n_{1i}} - c_{1i})(e^{-n_{2i}} - c_{2i})) \end{aligned} \tag{41}$$

where every quantity is defined in j.p.m.f above.

3.4.2.5 Bivariate Negative binomial using copulas and common shocks

Our last modeling technique for negative binomial distribution is based on common shock variables and copula functions.
A classic way to induce correlation among variables is to employ common shock variables. Using this formulation, we will model dependence through negative

binomial regression.
Copula functions are also a great tool to model bivariate count data and we will use Frank's and Clayton's copulas as previously.

For our purpose, we will use Poisson regression model and two forms of negative binomial regression models (NB-1 and NB-2), with their zero-inflated extensions. The probability function of the Poisson model was Equation (13). When it comes to the negative binomial regression model, we will use two forms that
Shi and Valdez (2012) proposed,the NB-1 and NB-2.

A count variable N_{ji} for $i = 1, 2, ..., 6000$ and $j = 1, 2$ is known to follow a negative binomial distribution if its probability function can be expressed as:

$$P(N_{ji} = n_{ji}) = \frac{\Gamma(\theta_j + n_{ji})}{\Gamma(\theta_j)\Gamma(n_{ji} + 1)} \left(\frac{1}{1+t}\right)^{\theta_j} \left(\frac{t}{1+t}\right)^{n_{ji}} \tag{42}$$

for $n_{ji} = 0, 1, 2, ...$ and λ, $t > 0$.
Including covariates, the NB-1 form is obtained for $\theta_j = \sigma^{-2} exp({x_{ji}}^T \beta_j)$ and its probability function is:

$$f_{NB-1}(n_{ji}|x_{ji}; \beta_j, \sigma^2) = \frac{\Gamma(\sigma^{-2} exp({x_{ji}}^T \beta_j) + n_{ji})}{\Gamma(\sigma^{-2} exp({x_{ji}}^T \beta_j))\Gamma(n_{ji} + 1)} \left(\frac{1}{1+\sigma^2}\right)^{\sigma^{-2} exp({x_{ji}}^T \beta_j)} \left(\frac{\sigma^2}{1+\sigma^2}\right)^{n_{ji}} \tag{43}$$

where σ^2 is the variance of n_{ji}, β_j denotes the vector of regression coefficients and x_{ji} is the vector of explanatory variables.

The NB-2 model is obtained for $\theta_j = \sigma^{-2}$ and takes the form:

$$f_{NB-2}(n_{ji}|x_{ji}; \beta_j, \sigma^2) = \frac{\Gamma(\sigma^{-2} + n_{ji})}{\Gamma(\sigma^{-2})\Gamma(n_{ji} + 1)} \left(\frac{1}{1 + \sigma^2 exp({x_{ji}}^T \beta_j)}\right)^{\sigma^{-2}} \left(\frac{\sigma^2 exp({x_{ji}}^T \beta_j)}{1 + \sigma^2 exp({x_{ji}}^T \beta_j)}\right)^{n_{ji}} \tag{44}$$

Both models assume the same mean structure of the count variable but differ in terms of the dispersion parameter, due to θ_j form.
Their zero-inflated version is formed by mixing these distributions with a digenerate function at zero.

For our first approach in this section we use a common covariance term through

the following construction:
$N_{ji} = U_{ji} + U_{0i}$ for $i = 1, 2, ..., n$ and $j = 1, 2$. $U_{ji} \sim NB(\theta_j, t_{ji})$ and all marginals follow negative binomial distribution with
$N_{ji} \sim NB(\theta, t_{ji} + t_{0i})$.
The key assumption is the common parameter λ for both variables and we add covariates through $\theta_{ji} = exp(x_{ji}^T \beta_j)$ for $j = 1, 2$. Only the NB-1 model is appropriate for this method.

Under these assumptions the vector $N_i = (N_{1i}, N_{2i})$ follows a bivariate negative binomial regression model with j.p.m.f:

$$f_i(n_1, n_2|x_i) = \sum_{s'=0}^{min(n_1,n_2)} f_0(s') f_1(n_1 - s'|x_{1i}) f_2(n_2 - s'|x_{2i}) \tag{45}$$

where $x_i = (x_{1i}, x_{2i})$,$f_0(\cdot) = f_{NB-1}(\cdot|1; \log\theta_0, \sigma^2)$ and $f_j(\cdot|x_{ji}) = f_{NB-1}(\cdot|x_{ji}; \beta_j, \sigma^2)$ for $j = 1, 2$.

The common shock variable U_{0i} has the role of the random effect, implying an identical positive correlation for all the pairs of claim type. The log-likelihood of this model is given by:

$$\log L = \sum_{i=1}^{n} \log(\sum_{s'=0}^{min(n_{1i},n_{2i})} f_0(s') f_1(n_{1i} - s'|x_{1i}) f_2(n_{2i} - s'|x_{ii}))$$

(46)

For our analysis we use two copula functions,Frank's and Clayton's copulas again. As u and v in Equations (5) and (6) we will use univariate negative binomial and zero-inflated negative binomial marginals for both response variables. We add covariates to the mean through $\theta_{ji} = exp(x_{ji}^T \beta_j)$ for $i = 1, 2, ..., n$ and $j = 1, 2$.

Since our marginal distributions are discrete, we obtain the joint probability mass

function of N_{1i} and N_{2i}, using finite differences.

$$f(n_{1i}, n_{2i}) = F(n_{1i}, n_{2i}) - F(n_{1i} - 1, n_{2i}) - F(n_{1i}, n_{2i} - 1) + F(n_{1i} - 1, n_{2i} - 1).$$

(47)

This method lies on the difference of our count variables.
Suppose, for the case of discrete random variables, the variable of interest is the difference $w_i = n_{1i}n_{2i}$, for $i = 1, 2, ..., n$. We present a simple approach using copulas to derive the distribution of w_i.
The joint probability mass function (pmf) is derived by taking finite differences:

$$\begin{aligned} c(F(n_{1i}), F(n_{2i}); \xi) = C(F(n_{1i}), F(n_{2i}); \xi) - C(F(n_{1i}1), F(n_{2i}); \xi) \\ - C(F(n_{1i}), F(n_{2i}1); \xi) + C(F(n_{1i}1), F(n_{2i}1); \xi) \end{aligned}$$

where lower-case c denotes the pmf of the copula function.
With the transformation $w_i = n_{1i}n_{2i}$, the joint pmf can be equivalently expressed in terms of w_i and n_{2i} as:

$$c(F(w_i + n_{2i}), F(n_{2i}); \xi) \tag{48}$$

The pmf of w_i,denoted $q(w_i)$, is obtained by summing over all possible values of $\mathrm{n}_{2i} : q(w_i) = \sum_{n_{2i}=0}^{\infty} c(F(w_i + n_{2i}), F(n_{2i}); \xi)$(49)

For any value of w, (49) gives the corresponding probability mass. The cdf of $q(w_i)$ is calculated by accumulating masses at each point w_i:

$$Q(w_i) = \sum_{t=-\infty}^{w_i} q(t) \tag{50}$$

Both $q(w_i)$ and $Q(w_i)$ characterize the full distribution of w_i so that inference can be made regarding the difference between two count variables. This method can also be applied to any discrete or continuous variables when the marginal

distribution of the components of the differences is parametrically specified.

In conclusion, we will use as marginals of the copula j.p.m.f, function G_{ji} using two different forms like in Poisson section (paragraph 3.4.1.5):

$$\mathrm{G}_{ij} = \begin{cases} \sum_{k=0}^{n_{ji}} f_{NB-2}(k;\theta_{ji}), \text{for negative binomial distribution} \\ p + (1-p)\sum_{k=0}^{n_{ji}} f_{NB-2}(k;\theta_{ji}), \text{for zero-inflated negative binomial distribution} \end{cases}$$

where p_j is the inflation parameter for $i = 1, 2..., n$ and $j = 1, 2$ and probability distribution functions as defined earlier.

Therefore,the log-likelihood function is given by:

$$\log L = \sum_{i=1}^{n} log(C_{FR}(G_{1i}, G_{2i}) - C_{FR}(G_{1i}-1, G_{2i}) - C_{FR}(G_{1i}, G_{2i}-1) + C_{FR}(G_{1i}-1, G_{2i}-1)). \tag{51}$$

where C_{FR} is the Frank's copula, but the same holds for Clayton with the appropriate form.

3.4.2.6 Bivariate Inverse Gaussian Following the same technique as in Poisson-Gamma mixture we want to derive a bivariate inverse Gaussian regression model, Kocherlakota (1986). The inverse Gaussian is a skewed, two-parameter continuous distribution whose density is similar to the Gamma distribution with greater skewness and a sharper peak. It is used as the mixing distribution with density function:

$$g(\alpha';\phi) = \frac{\phi}{\sqrt{2\pi}} exp(\phi^2) \alpha'^{\frac{-3}{2}} exp(-\frac{\phi^2}{2}(\frac{1}{\alpha'} + \alpha')) \tag{52}$$

$\alpha', \phi > 0$. As ϕ tends to infinity, the inverse Gaussian distribution becomes more like a normal (Gaussian) distribution. The inverse Gaussian distribution has several properties analogous to a Gaussian distribution.
The joint probability function is:

$$P_{IG}(n_1, n_2; \phi) = \frac{2\phi exp(\phi^2)}{\sqrt{2\pi}} K_{\sum_{j=1}^{2} n_j - 1/2}(\phi\Delta) \left(\frac{\phi}{\Delta}\right)^{\sum_{j=1}^{2} n_j - 1/2} \prod_{j=1}^{2} \frac{\theta_j^{n_j}}{n_j!} \tag{53}$$

where $\Delta = \sqrt{\phi^2 + 2\sum_{j=1}^{2} \theta_j}$ and $K_r(x)$ is the modified Bessel function of the third kind of order r, Bessel (1824).

Using the same EM algorithm as previously we optimize the complete data log-likelihood function given as:

$$l_C(\Theta) = \sum_{i=1}^{n} \sum_{j=1}^{2} (-\alpha_i' \theta_{ji} + n_{ji} \log(\alpha_i' \theta_{ji}) - \log(n_{ji}!)) + \sum_{i=1}^{n} \log(g(\alpha_i'; \phi)) \tag{54}$$

where Θ stands for all parameters of the model.

3.4.2.7 Bivariate Poisson Lognormal A log-normal (or lognormal) distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable X is log-normally distributed, then $Y = \log(X)$ has a normal distribution.
The Poisson-lognormal model, Stewart (1994), assumes that the intensity parameter of a Poisson process has a lognormal distribution in a sample of observations. This model can yield highly skewed, discrete distributions, but must be estimated by numerical methods.
For the bivariate Poisson lognormal regression model, we use lognormal as the mixing distribution with density function:

$$g(\alpha'; \phi) = \frac{1}{\sqrt{2\pi}\phi\alpha'} exp\left(\frac{-(log(\alpha') + \phi^2/2)^2}{2\phi^2}\right) \tag{55}$$

$\alpha', \phi > 0$. The joint probability function has no closed form so:

$$P_{LN}(n_1, n_2; \phi) = \int_0^\infty \prod_{j=1}^{2} \frac{exp(-\alpha'\theta_j)(\alpha'\theta_j)^{n_j}}{n_j!} \frac{exp\left(\frac{-(log(\alpha') + \phi^2/2)^2}{2\phi^2}\right)}{\sqrt{2\pi}\phi\alpha'} d\alpha' \tag{56}$$

For the bivariate Poisson lognormal model we have to switch to Monte Carlo EM algorithm because numerical integration is needed due to likelihood's lack of closed

form.
The Monte Carlo EM (MCEM) algorithmm Wei and Tanner (1990), is a modification of the EM algorithm where the expectation in the E-step is computed numerically through Monte Carlo simulations. The most flexible and generally applicable approach to obtaining a Monte Carlo sample in each iteration of an MCEM algorithm is through Markov chain Monte Carlo (MCMC) routines such as the Gibbs, Geman (1984), and Metropolis-Hastings, Rosenbluth and Teller (1953), samplers.

3.4.3 Generalised Poisson distribution

Going further to our analysis, we examined what improvements we have from using negative binomial distribution and mixing distributions in general.
It is well-known that, by construction, the mean of a Poisson distribution equals its variance. Equivalently, its variance-to-mean ratio equals one, a measure also known as statistical dispersion. Accordingly, a Poisson distribution can be expected to fit only poorly to an empirical distribution whose dispersion differs considerably from unity. In applied statistics generally, attention has largely focused on the need to account for overdispersion, that is, distributions with dispersion considerably larger than one. The generalized Poisson (GP) distribution is of interest for modeling count data because it includes the Poisson distribution as a special case, and over the range where the second parameter is positive, it is overdispersed relative to Poisson with a variance to mean ratio exceeding 1.
Despite the fact that both negative binomial distribution and generalised Poisson one can be used to deal with overdispersion and provide us with similar results, they differ a lot.
The GP distribution has one extra parameter than standard Poisson, the so-called dispersion parameter λ. A random variable N is said to have a generalised Poisson distribution if its probability density function is given by, Consul (1989):

$$f(n;\theta,\lambda) = \theta(\theta + n\lambda)^{n-1} e^{-\theta - n\lambda}/n! \tag{57}$$

for $n = 0, 1, 2, ...$
The GPD reduces to the Poisson model when λ=0 and possesses the property of

over-dispersion for all values of $\lambda>0$ and the property of under-dispersion for all values of $\lambda<0$.

It is very useful for our purpose to introduce generalised Poisson regression under a mean parametrization.Since the mean μ for the GPD is given by $\mu = \theta(1-\lambda)^{-1}$, one can write corresponding generalized Poisson regression (GPR) model in the form:

$$P(N = n|\mu, \Phi) = \mu(\mu + (\Phi - 1)n)^{n-1}\Phi^{-n}\frac{e^{-(\mu+(\Phi-1)n)/\Phi}}{n!} \tag{58}$$

where $n = 0, 1, 2, ...$ and $\Phi = (1 - \lambda)^{-1}$.

This distribution form holds for $\lambda > 0$. If $\lambda < 0$ and $n > M$, when M is the largest positive integer for which $\theta + M\lambda > 0$ when λ is negative, then $P(N = n|\mu, \Phi)$ equals to zero.
The difference between the fits of the GP and NB distributions is most apparent for count data with long right tails (heavily right-skewed), Joe and Zhu (2005). Besides, the NB distribution seems to have larger mass at zero than the GP distribution (when the first two moments are fixed). This means their zero-inflated variations tend to have larger discrepancy. However, the fits of their zero-inflated variations may differ when there is a large zero fraction.

3.4.3.1 Bivariate Generalised Poisson Regression (BGPR) Bivariate generalised Poisson distribution has an extra parameter, that can handle over or underdispersion. Similar to the previous model, we define the joint mass probability function of the bivariate generalised Poisson regression model (BGPR) as a product of univariate generalised Poisson marginals with a multiplicative factor, Famoye (2010), also we present several forms of this model, Zamani et al (2016).
The advantages of Zamani's model are firstly that these models can be fitted to bivariate count data with zero, positive or negative correlation.
Secondly, the models allow for under or overdispersion and finally they allow flexible mean-variance relationship. Its probability mass function can be given by:

$$P(n_{1i},n_{2i})=[\prod_{j=1}^{2}\frac{\mu_{ji}\left(\mu_{ji}+\lambda_j\mu_{ji}^{P_j-1}n_{ji}\right)^{n_{ji}-1}}{\left(1+\lambda_j\mu_{ji}^{P_j-1}\right)^{n_{ji}}n_{ji}!}e^{-\frac{\mu_{ji}+\lambda_j\mu_{ji}^{P_j-1}n_{ji}}{1+\lambda_j\mu_{ji}^{P_j-1}}}][1+\alpha\prod_{j=1}^{2}(e^{-n_{ji}}-c_{ij})] \tag{59}$$

where λ_j, $j=1,2$, is the dispersion parameters, α is the multiplicative factor, P_j is the functional parameter that provide us these different forms,
$c_{ji}=E(e^{-n_{ji}})=e^{\frac{\mu_{ji}}{1+\lambda_j\mu_{ji}^{P_j-1}}(s_j-1)}$, $\theta_{ji}=\mu_{ji}(1-\lambda_j)$ and $\log s_j-\frac{\lambda_j\mu_{ji}^{P_j-1}}{1+\lambda_j\mu_{ji}^{P_j-1}}(s_j-1)+1=0$, $j=1,2$.
When λ_j becomes 0 then BGPD reduces to simple bivariate Poisson (BP).
If α=0, then N_1 and N_2 are independent response variables. Furthermore, when $\alpha>0$ and $\alpha<0$ then we have positive or negative correlations respectively.
Moreover, if $\lambda_j>0$, then our model handles overdispersion and if $\lambda_j<0$ it allows underdispersion.
Finally, when $P_1=P_2=1$ and $P_1=P_2=2$,then our model reduces from BGPR-P to BGPR-1 and BGPR-2 respectively.
Since we want to add covariates again in our model, we assume that $\mu_{ji}=\beta_j^Tx_{ji}$ for $i=1,2...,n$ and $j=1,2$ with β_j the regression coefficient vector and x_{ji} the covariates' vector for every observation.

The log-likelihood function for BGPR-P is given by:

$$\log L=\sum_{i=1}^{n}\sum_{j=1}^{2}[\log(\frac{\mu_{ji}\left(\mu_{ji}+\lambda_j\mu_{ji}^{P_j-1}n_{ji}\right)^{n_{ji}-1}}{\left(1+\lambda_j\mu_{ji}^{P_j-1}\right)^{n_{ji}}n_{ji}!}e^{-\frac{\mu_{ji}+\lambda_j\mu_{ji}^{P_j-1}n_{ji}}{1+\lambda_j\mu_{ji}^{P_j-1}}})]+\sum_{i=1}^{n}\log[1+\alpha\prod_{j=1}^{2}(e^{-n_{ji}}-c_{ji})] \tag{60}$$

when $P_j=1$ or $P_j=2$ for j=1,2 we can derive BGPR-1 and BGPR-2 log-likelihood functions respectively. As we have seen in bivariate negative binomial section, paragraph 3.4.2.4, we use a functional parameter P_j in order to acquire all forms of BGPR model, through Equation (59), for P_j equals to one or two respectively.

3.4.3.2 Bivariate Sarmanov regression with a multiplicative factor (BSP-BSGP)

The last models we will present it this section are two models based on

Sarmanov distributions, bivariate Poisson and generalised Poisson regression models through a multiplicative factor.
Sarmanov's bivariate distribution was introduced in the literature by Sarmanov (1966). One of the main interesting properties of the Sarmanov is that the bivariate distribution can support a wide range of marginals, such as in this case,the Poisson and generalised Poisson distributions.

We will go through, the bivariate generalised Poisson regression model that Hofer and Leitner (2012) proposed and we will show that the bivariate Poisson is a special case of the first one.

We select the marginal distributions to be:
$N_1 \sim GP(\theta_1, \lambda_1)$ and $N_2 \sim GP(\theta_2, \lambda_2)$.
The expected values of the response variables are $E(N_1) = \theta_1$ and $E(N_2) = \theta_2$ and the variances are $Var(N_1) = \lambda_1^2\theta_1$ and $Var(N_2) = \lambda_2^2\theta_2$.
When $\lambda_1 = \lambda_2 = 1$ we have the bivariate Poisson case.
In order to acquire a closed form for the probability function we need the expected values $E(e^{-N_1})$ and $E(e^{-N_2})$. We can derive them, from the moment generating function of the GP distribution as it is described from Ambagaspitiya and Balakrishnan (1994).
Therefore, the moment generating function $M(t|\theta, \lambda)$ for a $GP(\theta, \lambda)$ with $\lambda > 1$ takes the form:
$M(t|\theta, \lambda) = exp(-\frac{\theta}{\lambda-1}[W(-(1-1/\lambda)exp(-1+1/\lambda+t)) + 1 - 1/\lambda])$
where W is the Lambert W function, Lambert (1758), defined as $W(x)e^{W(x)} = x$.

As a result the probability function of BSGP is obtained as:

$$P_{BSGP}(n_{1i}, n_{2i}) = \frac{\theta_{1i}(\theta_{1i} + (\lambda_1 - 1)n_{1i})^{n_{1i}-1}}{n_{1i}!\lambda_1^{n_{1i}} exp((\theta_{1i} + (\lambda_1 - 1)n_{1i})\lambda_1^{-1})} \frac{\theta_{2i}(\theta_{2i} + (\lambda_2 - 1)n_{2i})^{n_{2i}-1}}{n_{2i}!\lambda_2^{n_{2i}} exp((\theta_{2i} + (\lambda_2 - 1)n_{2i})\lambda_2^{-1})}\Delta \tag{61}$$

where $\Delta = (1+\alpha(e^{-n_{1i}} - c_{1i})(e^{-n_{2i}} - c_{2i}))$ and $c_{ji} = M(-1|\theta_{ji}, \lambda_j)$ for $i = 1, 2, ..., n$ and $j = 1, 2$.

We can include covariates exactly like the previous models through log-links and the bivariate Poisson case is obtained when the dispersion parameters λ_1, λ_2 are equal to 1.

The log-likelihood of the model is given by:

$$\log L = \sum_{i=1}^{n} \log\left(\frac{\theta_{1i}(\theta_{1i} + (\lambda_1 - 1)n_{1i})^{n_{1i}-1}}{n_{1i}!\lambda_1^{n_{1i}} exp((\theta_{1i} + (\lambda_1 - 1)n_{1i})\lambda_1^{-1})} \frac{\theta_{2i}(\theta_{2i} + (\lambda_2 - 1)n_{2i})^{n_{2i}-1}}{n_{2i}!\lambda_2^{n_{2i}} exp((\theta_{2i} + (\lambda_2 - 1)n_{2i})\lambda_2^{-1})}\Delta\right) \tag{62}$$

4 Results

4.1 Main findings

All models used in our analysis were presented in the previous section and our purpose is to compare them and find which of them has a better fit for our data. In order to find the best model, we will use maximum likelihood estimation plus AIC quantity. The models with the smallest price in these quantities have a better fit for our dataset. We split our models to categories. First category contains models based on Poisson distribution, second one contains mixed Poisson distribution models, like negative binomial models, inverse Gaussian and Poisson lognormal. The third one contains generalised Poisson models. Even though, we presented our models based on the distribution function, it is more interesting to compare them based on modeling technique and distribution function at the same time.

For every model, we present all coefficients for each explanatory variable plus the intercept. Next to estimated quantities of trivariate reduction models there is a standard error between brackets derived from standard bootstrap methods using boot in R programming language. In order to reach our goal we chose a sample size for bootstrap method that equals to 150. Then for each bootstrap sample, we fit our models and we estimate all quantities of interest. Finally, we calculate the standatd deviation of the sample of our model's quantities estimates. For all the other models, mixed Poisson, multiplicative factor-based or copula-based models we obtain standard errors with nlm function in R programming language. The same holds for the commmon shock model. To be more specific, we obtain standard errors from the Hessian matrix. After implementing the optimization algorithm, Newton-Raphson algorithm, we compute the Hessian matrix. Continuously, we compute the square root of diagonal elements of inverse Hessian matrix and these values are our standard errors.
Extra parameters for zero-inflated versions are mentioned and at the bottom anyone can see log-likelihood quantities with the appropriate AIC. It is reasonable that between regression models based on different distribution functions there are differences, but under the same modeling method, results are quite similar.

Therefore, we present Tables for our models based on trivariate reduction, mixed Poisson models, multiplicative factor based, copula-based models and models using common shock variables.

Table 3: Models summary

Model	Name	Method	Citation
Double Poisson	DP	Trivariate reduction	Kocherlakota and Kocherlakota (1992)
Bivariate Poisson	BP-1	Trivariate reduction	Bermudez (2009)
Bivariate Poisson	BP-2	Trivariate reduction	Bermudez (2009)
Bivariate zero inflated Poisson	BZIP-1	Trivariate reduction	Bermudez (2009)
Bivariate zero-inflated Poisson	BZIP-2	Trivariate reduction	Bermudez (2009)
Bivariate negative binomial	BNB	Mixed Poisson	Ghitany, Karlis et al. (2012)
Bivariate inverse Gaussian	BIG	Mixed Poisson	Ghitany, Karlis et al. (2012)
Bivariate Poisson lognormal	BLN	Mixed Poisson	Ghitany, Karlis et al. (2012)
Bivariate Poisson	BPR	Multiplicative factor	Lakshminarayana (1999)
Bivariate negative binomial	BNBR	Multiplicative factor	Famoye (2012)
Bivariate Sarmanov Poisson	BSPR	Multiplicative factor	Hofer and Leitner (2012)
Bivariate Sarmanov Generalised Poisson	BSGPR	Multiplicative factor	Hofer and Leitner (2012)
Bivariate Generalised Poisson	BGPR-1	Multiplicative factor	Zamani et al. (2016)
Bivariate Generalised Poisson	BGPR-2	Multiplicative factor	Zamani et al. (2016)
Bivariate Generalised Poisson	BGPR-P	Multiplicative factor	Zamani et al. (2016)
Bivariate zero inflated Poisson	BZIPR	Multiplicative factor	Faroughi (2017)
Bivariate zero inflated negative binomial	BZINBR-1	Multiplicative factor	Faroughi (2017)
Bivariate zero inflated negative binomial	BZINBR-2	Multiplicative factor	Faroughi (2017)
Bivariate zero inflated negative binomial	BZINBR-P	Multiplicative factor	Faroughi (2017)
Bivariate Poisson Frank copula	BPFR	Copula	Cameron et al. (2004)
Bivariate zero inflated Poisson Frank copula	BZIPFR	Copula	Cameron et al. (2004)
Bivariate negative binomial Frank copula	BNBFR	Copula	So et al. (2011)
Bivariate zero inflated binomial Frank copula	BZINBFR	Copula	So et al. (2011)
Bivariate negative binomial Common Shocks	CS	Common shocks	Shi and Valdez (2012)

Firstly we present in Table 3, all models implemented with the approriate names and citations, in order to clarify to which model we are reffered to every time. As anyone can observe, the first column describes the model, the second one denotes the appropriate shortcut name, then we present a column that explains modeling technique and last column shows the author of the paper that this model has been used. Secondly, we present in Table 4 all models combined with statistics of interest, AIC in descending order and maximum loglikelihood. Based on these two criteria, we compare all models and modeling techniques in order to find the model that fit our data the best.

Table 4: Models statistics - Comparison of loglik and AIC

Model	Name	loglik	AIC
Double Poisson	DP	-4423.78	8875.56
Bivariate Sarmanov Poisson	BSPR	-4356.92	8743.85
Bivariate Poisson	BPR	-4354.85	8739.70
Bivariate Poisson Frank copula	BPFR	-4352.95	8735.91
Bivariate Poisson	BP-2	-4334.63	8711.27
Bivariate Poisson	BP-1	-4338.53	8707.06
Bivariate zero-inflated Poisson	BZIP-2	-4098.60	8247.21
Bivariate zero inflated Poisson	BZIP-1	-4101.66	8241.32
Bivariate zero inflated Poisson	BZIPR	-4098.16	8228.32
Bivariate negative binomial	BNBR	-4092.78	8203.56
Bivariate zero inflated negative binomial	BZINBR-1	-4092.78	8203.56
Bivariate zero inflated Poisson Frank copula	BZIPFR	-4085.11	8202.22
Bivariate Poisson lognormal	BLN	-4074.18	8178.36
Bivariate inverse Gaussian	BIG	-4067.57	8165.14
Bivariate Generalised Poisson	BGPR-2	-4063.92	8161.84
Bivariate negative binomial	BNB	-4064.73	8159.46
Bivariate Generalised Poisson	BGPR-1	-4061.63	8156.68
Bivariate zero inflated negative binomial	BZINBR-2	-4061.59	8141.18
Bivariate negative binomial Common Shocks	CS	-4053.91	8139.83
Bivariate zero inflated negative binomial	BZINBR-P	-4046.59	8133.46
Bivariate zero inflated binomial Frank copula	BZINBFR	-4048.13	8132.26
Bivariate negative binomial Frank copula	BNBFR	-4045.39	8125.70
Bivariate Sarmanov Generalised Poisson	BSGPR	-4040.78	8115.56
Bivariate Generalised Poisson	BGPR-P	-4038.45	8114.90

It is a matter of interest, that the improvement in terms of AIC from the first model to the last is impressive. From the Double Poisson model that assumes independence between response variables, with an AIC equals to 8875.56, to the bivariate generalised Poisson model BGPR-1 with an AIC equals to 8114.90, that fits way better our data.

As we explained earlier, it is more efficient to present results from our models based on modeling technique. Although, explaining based on distribution is important in order to describe how the model assumptions change, when the modeling method changes, presenting models in the following way, gives us the opportunity to ex-

amine every difference between different distributions based on the same modeling method. Table 5, shows all models based on trivariate reduction, bivariate Poisson, double Poisson and zero-inflated Poisson. The double Poisson model, has the highest AIC and the lowest likelihood value, so it is the worst model for our data and as an explanation we can give, that it assumes that response variables N_1 and N_2 are independent. The difference between BP-1 and BP-2 and between BZIP-1 and BZIP-2 respectively, is that the first model does not allow for regressors in λ_3, while the second one allows this condition. In terms of AIC BP-1 and BZIP-1 are better models than BP-1 and BZIP-2 accordingly and the reason is that AIC, is a quantity based on the number of variables, therefore since λ_3 has regressors in the second case, AIC value is smaller. Nevertheless, in terms of likelihood these models, fit worse our data, since these values are lower for BP-1 and BZIP-1. Finally, zero-inflated models, BZIP-1 and BZIP-2 are way better that the other models, a fact that is explained from the high frequency of zeros in both response variables. AIC values are much lower in zero-inflated models, equal to 8241.32 and 8247.21 for BZIP-1 and BZIP-2. while for BP-1 and BP-2, these values are equal to 8707.06 and 8711.27 respectively. We can easily see that variables ZON, LOY and AGE have a significant role when it comes to third party guarantees. In addition, all variables have a big effect when it comes to the rest of guarantees.

Table 5: Trivariate reduction - Optimized models results and standard errors

Var	θ_i	Parameter	DP	BP-1	BP-2	BZIP-1	BZIP-2
N_1							
	θ_1	Intercept	-2.512(0.175)	-2.743(0.177)	-2.745(0.211)	-1.236(0.103)	-1.167(0.093)
		GEN	0.024(0.147)	0.028(0.138)	0.030(0.173)	-0.021(0.078)	-0.083(0.081)
		URB	0.024(0.116)	0.028(0.129)	0.025(0.156)	0.005(0.056)	-0.023(0.071)
		ZON	0.082(0.110)	-0.002(0.104)	-0.036(0.127)	0.119(0.066)	0.067(0.068)
		LOY	0.008(0.144)	0.106(0.146)	0.137(0.168)	0.098(0.074)	-0.011(0.071)
		AGE	-0.021(0.093)	-0.018(0.086)	-0.012(0.111)	-0.080(0.046)	-0.024(0.047)
		POW	0.032(0.112)	-0.029(0.112)	-0.049(0.139)	0.009(0.052)	0.023(0.070)
N_2							
	θ_2	Intercept	-2.316(0.163)	-2.480(0.152)	-2.478(0.159)	-1.004(0.081)	-0.955(0.089)
		GEN	0.149(0.099)	0.170(0.102)	0.171(0.109)	0.115(0.057)	0.150(0.053)
		URB	0.219(0.087)	0.251(0.096)	0.249(0.099)	0.186(0.051)	0.115(0.046)
		ZON	-0.211(0.093)	-0.325(0.089)	-0.351(0.097)	-0.174(0.046)	-0.153(0.053)
		LOY	-0.209(0.136)	-0.190(0.122)	-0.174(0.130)	-0.131(0.066)	0.272(0.062)
		AGE	0.233(0.076)	0.268(0.076)	0.272(0.085)	0.152(0.039)	0.163(0.040)
		POW	0.321(0.097)	0.321(0.095)	0.305(0.106)	0.288(0.047)	-0.254(0.058)
	θ_3	Intercept	-	-4.073(3.773)	-4.248(0.886)	-5.873(5.308)	-53.321(16.268)
		GEN	-	-	-0.005(0.407)	-	0.155(5.081)
		URB	-	-	0.022(0.274)	-	15.079(7.947)
		ZON	-	-	0.500(0.293)	-	13.429(7.786)
		LOY	-	-	-	-	5.186(7.965)
		AGE	-	-	-0.431(0.561)	-	0.201(6.297)
		POW	-	-	-0.054(0.851)		16.772(7.684)
p			-	-	-	0.73	0.73
logL			-4423.78	-4338.53	-4334.63	-4101.66	-4098.60
AIC			8875.56	8707.06	8711.27	8241.32	8247.21

Table 6 shows three models based on mixture poisson modeling. The first one is a bivariate negative binomial, the second one a bivariate inverse Gaussian and the last one a Poisson lognormal. The model with the greatest value of loglikelihood and the lowest value of AIC at the same time is the negative binomial one. Obviously, mixture models provide us with similar results. Coefficients for every variable is pretty similar for every model, as anyone can observe from the table. Standard errors from inverse Hessian matrix are also similar. Likelihood values are slightly different, while the Poisson lognormal model seems to be the worst one. It is important to notice the difference in the random variable α that introduces correlation between response variables. This variable in the BLN model is one unit

Table 6: Mixed Poisson Models - Optimized models results and standard errors

Var	Parameter	BNB	BIG	BLN
N_1				
	Intercept	-2.511(0.206)	-2.499(0.209)	-2.459(0.228)
	GEN	0.024(0.136)	0.026(0.138)	0.152(0.138)
	URB	0.023(0.106)	0.040(0.107)	0.059(0.113)
	ZON	0.083(0.124)	0.072(0.126)	0.105(0.130)
	LOY	0.008(0.134)	-0.016(0.135)	0.017(0.149)
	AGE	-0.019(0.166)	0.020(0.167)	-0.071(0.164)
	POW	0.032(0.157)	0.026(0.159)	-0.055(0.157)
N_2				
	Intercept	-2.306(0.185)	-2.299(0.188)	-2.398(0.202)
	GEN	0.157(0.115)	0.160(0.117)	0.227(0.113)
	URB	0.212(0.094)	0.232(0.095)	0.342(0.099)
	ZON	-0.207(0.114)	-0.219(0.116)	-0.206(0.120)
	LOY	-0.214(0.110)	-0.241(0.112)	-0.187(0.119)
	AGE	0.234(0.133)	0.279(0.135)	0.370(0.129)
	POW	0.316(0.147)	0.315(0.149)	0.296(0.144)
α'		0.267(0.020)	0.475(0.023)	1.453(0.048)
logL		-4064.73	-4067.57	-4074.18
AIC		8159.46	8165.14	8178.36

greater than the others, a fact that may occured from difficutlies in optimization of Poisson lognormal distribution, due to its form that it is not closed.

Table 7: Multiplicative factor models - Optimized models results and standard errors

Var	Parameter	BPR	BNBR	BSPR	BSGPR
N_1					
	Intercept	-2.496(0.181)	-2.330(0.513)	-2.512(0.181)	-2.464(0.208)
	GEN	0.028(0.118)	-0.215(0.135)	0.024(0.118)	0.009(0.136)
	URB	0.025(0.092)	0.111(0.106)	0.024(0.092)	0.045(0.107)
	ZON	0.085(0.107)	0.022(0.120)	0.082(0.107)	0.142(0.122)
	LOY	0.009(0.116)	-0.091(0.133)	0.008(0.116)	-0.011(0.134)
	AGE	-0.026(0.144)	-0.826(0.156)	-0.021(0.144)	0.068(0.161)
	POW	0.044(0.136)	0.242(0.149)	0.032(0.137)	-0.039(0.153)
	m_1	-	1.152(0.188)	-	1.295(0.037)
N_2					
	Intercept	-2.272(0.153)	-1.884(0.497)	-2.316(0.155)	-2.205(0.171)
	GEN	0.141(0.092)	-0.113(0.102)	0.149(0.092)	0.114(0.105)
	URB	0.201(0.078)	0.057(0.086)	0.219(0.078)	0.183(0.087)
	ZON	-0.189(0.095)	-0.271(0.104)	-0.211(0.095)	-0.131(0.105)
	LOY	-0.217(0.087)	-0.325(0.093)	-0.209(0.087)	-0.296(0.096)
	AGE	0.216(0.104)	0.311(0.112)	0.233(0.105)	0.260(0.116)
	POW	0.315(0.124)	0.502(0.136)	0.321(0.124)	0.285(0.139)
	m_2	-	2.021(0.239)	-	1.225(0.025)
α		4.692(0.466)	0.815(5.333)	4.861(0.518)	4.849(0.516)
logL		-4354.85	-4092.78	-4356.92	-4040.78
AIC		8739.70	8203.56	8743.85	8115.56

Next table, Table 7, shows four models based on a multiplicative factor. The first and third models are a bivariate Poisson and a bivariate Poisson based on Sarmanov distribution, the second one a bivariate negative binomial and the last is bivariate generalised Poisson based on Sarmanov distribution. Multiplicative factors of Poisson models, after optimization, are quite similar as we can see from the table. Bivariate negative binomial has a small multiplicative factor equals to 0.815, but it has a standard error equal to 5.333. Coefficients for all explanatory variables are close for Poisson models and differ from the negative binomial one. Standard errors are obtained from every model's Hessian matrix. AIC and loglikelihood values are nearly the same for the Poisson models, while the other

two models fit better the data. Probably, the reason is that they assume an extra dispersion parameter for both response variables, m_1 and m_2. The best model out of these four, is the bivariate generalised Poisson Sarmanov model, BSGPR, with AIC equals to 8115.56 and loglikelihood equals to -4040.78.

Table 8: Multiplicative factor models - Optimized models results and standard errors

Var	Parameter	BGPR-1	BGPR-2	BGPR-P
N_1				
	Intercept	-2.884(0.231)	-2.513(0.233)	-2.518(0.210)
	GEN	0.119(0.138)	-0.008(0.156)	0.163(0.123)
	URB	0.034(0.110)	0.096(0.121)	0.014(0.106)
	ZON	-0.028(0.132)	0.045(0.142)	0.131(0.124)
	LOY	0.061(0.141)	0.072(0.152)	-0.051(0.133)
	AGE	-0.080(0.175)	-0.226(0.190)	-0.035(0.150)
	POW	0.281(0.178)	-0.088(0.180)	0.074(0.143)
	λ_1	0.257(0.031)	3.635(0.411)	0.473(0.043)
	P_1	-	-	1.170(0.062)
N_2				
	Intercept	-2.363(0.184)	-2.105(0.199)	-2.190(0.163)
	GEN	0.004(0.113)	0.191(0.123)	0.192(0.099)
	URB	0.165(0.089)	0.096(0.099)	0.231(0.085)
	ZON	-0.261(0.111)	-0.321(0.122)	-0.197(0.102)
	LOY	-0.291(0.099)	-0.524(0.122)	-0.255(0.096)
	AGE	0.142(0.125)	0.051(0.145)	0.402(0.107)
	POW	0.454(0.152)	0.506(0.157)	0.159(0.130)
	λ_2	0.208(0.023)	2.218(0.239)	0.250(0.013)
	P_2	-	-	1.028(0.070)
α		2.304(0.405)	2.612(0.430)	5.062(0.592)
logL		-4061.63	-4063.92	-4038.45
AIC		8156.68	8161.84	8114.90

Table 8 presents three bivariate generalised Poisson regression models based on a multiplicative factor, Zamani et al. (2016). The best model in terms of AIC and loglikelihood is the BGPR-P. AIC and loglikelihood have values equal to 8114.90 and -4038.45 respectively. This model fits better our data, because BGPR-P allow

for an extra parameter to be optimized, the functional one. BGPR-1 and BGPR-2 have P_1 and P_2 equal to 1 or 2 respectively. BGPR-P has functional parameters with values 1.170 and 1.028 that give us the best results. Since these values are closer to 1, it is reasonable, BGPR-1 to fit better our data than BGPR-2.

Table 9: Multiplicative factor models - Optimized models results and standard errors

Var	Parameter	BZIPR	BZINBR-1	BZINBR-2	BZINBR-P
N_1					
	Intercept	-1.064(0.201)	-2.330(0.513)	-2.093(0.360)	-2.297(0.357)
	GEN	0.005(0.129)	-0.215(0.135)	-0.206(0.156)	-0.003(0.148)
	URB	-0.022(0.102)	0.111(0.106)	-0.226(0.118)	0.066(0.115)
	ZON	0.121(0.119)	0.022(0.120)	0.009(0.140)	0.098(0.133)
	LOY	0.064(0.126)	-0.091(0.133)	-0.169(0.147)	0.022(0.145)
	AGE	-0.123(0.155)	-0.826(0.156)	-0.024(0.184)	0.124(0.173)
	POW	0.033(0.151)	0.242(0.149)	0.247(0.181)	0.143(0.172)
	m_1	-	1.152(0.188)	0.197(0.077)	2.150(0.728)
	P_1	-	-	-	1.612(0.258)
N_2					
	Intercept	-0.857(0.179)	-1.884(0.497)	-1.928(0.335)	-1.813(0.323)
	GEN	0.125(0.104)	-0.113(0.102)	0.102(0.115)	0.117(0.114)
	URB	0.162(0.089)	0.057(0.086)	0.229(0.095)	0.174(0.094)
	ZON	-0.163(0.108)	-0.271(0.104)	-0.203(0.114)	-0.298(0.117)
	LOY	-0.135(0.099)	-0.325(0.093)	-0.281(0.110)	-0.267(0.118)
	AGE	0.112(0.118)	0.311(0.112)	0.109(0.134)	0.303(0.134)
	POW	0.292(0.139)	0.502(0.136)	0.269(0.145)	0.242(0.143)
	m_2	-	2.021(0.239)	0.408(0.166)	1.902(0.288)
	P_2	-	-	-	1.887(0.365)
α		-1.087(0.250)	0.815(5.333)	1.495(0.792)	2.926(1.387)
p		0.766(0.011)	0.245(0.851)	0.252(0.209)	0.291(0.202)
logL		-4098.16	-4092.78	-4061.59	-4046.73
AIC		8228.32	8203.56	8141.18	8133.46

Table 9 shows four models based on multiplicative factor also. Four zero-inflated models, the first one is a bivariate Poisson and the other are bivariate negative binomial models with a functional parameter like BGPR. The Poisson inflated

model is worse than the rest in terms of AIC and likelihood. Obviously, negative binomial distribution seems to fit better our data than Poisson, due to the extra parameter that explains dispersion of variables. BZINBR-P is the best model with AIC equals to 8133.46 and loglikelihood equals -4046.73. Coefficients for all quantities differ for every of these models. The functional parameter is optimized at 1.612 and 1.887, closer to 2. Therefore, BZINBR-2 gives better results that BZINBR-1, due to this parameter P.

Table 10: Frank copula models - Optimized models results and standard errors

Var	Parameter	BPFR	BNBFR	BZIPFR	BZINBFR	CS
N_1						
	Intercept	-2.516(0.131)	-2.646(0.227)	-0.967(0.210)	-2.123(0.205)	-2.757(0.236)
	GEN	0.052(0.111)	0.049(0.152)	-0.080(0.139)	0.028(0.151)	-0.103(0.156)
	URB	0.043(0.036)	0.057(0.118)	-0.082(0.106)	-0.014(0.118)	0.087(0.122)
	ZON	0.065(0.039)	0.108(0.138)	0.147(0.123)	0.111(0.138)	0.027(0.143)
	LOY	0.033(0.036)	0.023(0.148)	0.114(0.134)	0.007(0.146)	0.108(0.158)
	AGE	-0.005(0.139)	-0.013(0.184)	-0.111(0.166)	-0.080(0.184)	0.090(0.183)
	POW	0.047(0.130)	0.056(0.175)	0.083(0.158)	-0.072(0.172)	-0.088(0.169)
	θ_1	-	0.124(0.015)	-	0.185(0.026)	-
N_2						
	Intercept	-2.329(0.077)	-2.368(0.184)	-0.802(0.184)	-2.290(0.261)	-2.397(0.189)
	GEN	0.167(0.087)	0.156(0.114)	0.147(0.108)	0.161(0.113)	0.139(0.112)
	URB	0.230(0.040)	0.218(0.093)	0.145(0.091)	0.162(0.093)	0.223(0.095)
	ZON	-0.210(0.048)	-0.201(0.113)	-0.141(0.109)	-0.120(0.111)	-0.254(0.118)
	LOY	-0.176(0.083)	-0.244(0.109)	-0.037(0.103)	-0.088(0.107)	-0.261(0.105)
	AGE	0.240(0.102)	0.204(0.132)	0.112(0.123)	0.306(0.129)	0.285(0.125)
	POW	0.313(0.118)	0.344(0.145)	0.276(0.143)	0.490(0.150)	0.335(0.152)
	θ_2	-	0.258(0.030)	-	0.407(0.021)	-
ξ		3.940(0.331)	3.890(0.404)	3.775(0.436)	3.913(0.408)	-
p		-	-	0.793(0.010)	0.230(0.019)	-
σ^2		-	-	-	-	0.510(0.046)
θ_{12}		-	-	-	-	0.015(0.002)
logL		-4352.95	-4045.39	-4085.11	-4048.13	-4053.91
AIC		8735.91	8125.70	8202.22	8132.26	8139.83

Last table, Table 10, shows bivariate Poisson and negative binomial models and

their zero-inflated versions, using Frank's copula. Also, a bivariate negative binomial model using common shocks is presented also. We present Frank's copula only, because Clayton's copula gives similar but worse results than Frank's. Coefficients for copula models are quite similar, while the most of them are statistically significant, due to their low standard error, compared to the optimal value of the variable. Especially, intercepts have small standard error, as we can see from the table. Bivariate negative binomial model has the lowest AIC, equals to 8125.70 and the greatest loglikelihood value, equals to -4045.39. The bivariate Poisson model, gives the worst results, while its zero-inflated version improves a lot model's efficiency in terms of AIC and loglikelihood. An interesting issue, is that the zero-inflated negative binomial model, gives slightly higher value of AIC and lower for loglikelihood than the standard bivariate negative binomial, in contradiction to other modeling techniques, that holds exactly the opposite. Finally, the common shocks model gives also good results, close to the BNBFR model.

4.2 Summary

After implementation of all models using Poisson, generalised Poisson and negative binomial distributions with different modeling methods, it is obvious that Poisson regression models give the worst results. However, this fact seems reasonable, since the other distributions have an extra parameter that measures dispersion.
Variance exceeds the mean and as a result Poisson distribution cannot explain very well our data. In parallel, generalised Poisson and negative binomial distributions allow us to relax this condition, so this is the reason we observe huge differences between AIC values, for example BP has an AIC value that equals to 8739, while BNB's equals to 8203.
Zero-inflation helps us overcome such difficulties and improves a lot our models, every zero-inflated Poisson model gives far better results than simple Poisson one.

Nevertheless, there are differences between modeling techniques also. Models constructed via trivariate reduction method give worse results than models based on multiplicative factor.
The explanation stands on the fact that some methods like trivariate reduction have some important limits. For example, only positive correlation between re-

sponse variables is allowed, however our response variables have positive correlation.

The best model among all is the bivariate generalised Poisson model with a functional parameter and a multiplicative factor (BGPR-P). Its' AIC value equals 8114 and it has the highest log-likelihood value equals -4038. Although, it assumes two extra parameters, one for dispersion and the functional one, it fits much better our data than Poisson regression models and zero-inflated versions and slightly better than negative binomial regression models. Furthermore, it allows for both positive and negative correlation through the multiplicative factor and it has a simpler form than copula-based models. Finally, through the functional parameter it allows for several tests since BGPR models are nested and as a result it is easy to notice which variables are statistically significant.

4.3 Insurance ratemaking

In this section, we will consider a credibility model superimposed on an a priori risk classification. The main task is to design a tariff structure in order to fairly distribute the burden of claims among policyholders. Rate making or insurance pricing, is the determination of what rates, or premiums, to charge for insurance. A rate is the price per unit of insurance for each exposure unit, which is a unit of liability or property with similar characteristics. Building homogeneous classes of policyholders is quite difficult since heterogeneity exists. This fact lies on drivers' behaviour and this is difficult to measure.

Pricing of individual risks is an actuarial principle. The pure premium approach, Denuit et al. (2007), defines the price of an insurance policy as the ratio of the estimated costs of all future claims against the coverage provided by the insurance policy while it is in effect to the risk exposure, plus expenses. An insurance premium is the monthly or annual payment you make to an insurance company to keep your policy active. As we mentioned, in order to avoid lapses in a competitive market, actuaries have to design a tariff structure that will fairly distribute the burden of claims among policyholders. Premiums for automobile liability coverage often vary by individual characteristics like the use of vehicle or the territory that

it is driven. If the policyholders misrepresent any of these classification variables in their declaration, they are subject to loss of coverage when they are involved in a claim. Our first task is to construct representative profiles from the portfolio. Therefore, we segment policyholders in five categories: Best-Good-Average-Bad-Worst drivers, regarding the lowest mean score of the number of claims. First category has the lowest mean score and so on. In order to give an example, let us assume the bivariate negative binomial model based on a multiplicative factor, BNBR. It's mean and variance are given as:

$$\mu_j = m_j^{-1} \frac{\theta_j}{1-\theta_j} \tag{63}$$

$$\sigma_j^2 = m_j^{-1} \frac{\theta_j}{(1-\theta_j)^2} \tag{64}$$

Using as values, all values that we got from optimization method we obtain mean and variance for the model. Then, we select appropriate combinations of explanatory variables in order to construct profiles with the lowest mean score. For example, if we select variables ZON and LOY that are equal to one and all the other categorical variables equal to zero, then we have the best combination that gives the lowest mean. Following this method, we obtained all mean scores for our models, for different drivers' profiles.

Table 11: Profiles of policyholders

Profile	Kind	GEN	URB	ZON	LOY	AGE	POW
1	Best	0	0	1	1	0	0
2	Good	0	0	0	1	0	1
3	Average	1	0	0	1	0	1
4	Bad	1	1	0	0	1	0
5	Worst	1	1	0	0	0	1

In order to continue to insurance pricing, we will present mean scores of the most representative models considered for the profiles that we constructed above.

Firstly for trivariate reduction modeling, then for mixed Poisson models,models with a multiplicative factor and finally for common shock and copula-based models.

After calculation of the premium for each model, we derive the most efficient models from each category.

Table 12: A priori ratemaking for different profiles

Model	Profiles				
	1st	2nd	3rd	4th	5th
DP	0.153	0.194	0.214	0.263	0.284
BP-1	0.148	0.183	0.184	0.233	0.242
BP-2	0.144	0.179	0.199	0.244	0.251
BZIP-1	0.170	0.203	0.215	0.226	0.256
BNB	0.154	0.194	0.215	0.265	0.285
BIG	0.150	0.191	0.211	0.285	0.293
BLN	0.154	0.183	0.229	0.315	0.330
BP	0.159	0.200	0.209	0.264	0.289
BSP	0.153	0.194	0.210	0.279	0.284
BSGP	0.171	0.189	0.225	0.283	0.288
BGPR-1	0.111	0.189	0.199	0.188	0.197
BZIP	0.175	0.210	0.226	0.223	0.265
BZINB-1	0.130	0.175	0.189	0.220	0.261
BP	0.155	0.199	0.224	0.272	0.291
BNB	0.140	0.180	0.201	0.244	0.275
BZINB	0.167	0.202	0.225	0.232	0.262
CS	0.127	0.162	0.171	0.239	0.242

From this table, the difference between using a bivariate Poisson model and two independent Poisson models for each type of guarantee is visible.
All of the models produce higher means for good risks and lower means for bad risks. In most models average risks have quite similar means.
Moreover, it is important to notice that zero-inflation has a significant impact due to the difference between simple bivariate Poisson and zero-inflated version in all profiles. Especially, when it comes to the first category of best drivers the mean for BZIP-1 is 0.170 and for BP-1 is 0.148.
Our best model in terms of log-likelihood value is the bivariate generalised Poisson model, that produce the lowest mean scores of all models implemented (BGPR-1), but average risks have the highest mean scores.
Futhermore, Poisson model using common shocks gives really interesting results as it allocates low scores of mean for the best drivers and quite high for bad ones.
To conclude, assumption of independence between types of claims clearly does not hold, since correlation affects the premium for each guarantee. Excess of zeros also has a significant role, therefore inflated models give different results than simple bivariate models regardless of distributions. The extra parameter that negative binomial and generalised Poisson distributions have, explains the heterogeneity between our response variables and as a result they produce similar results and quite better than simple bivariate Poisson distribution gives.

5 Discussion

In our analysis, we presented several models based on different methods of modeling and different distribution functions. Firstly, we implemented Poisson models and then, we continued with mixed Poisson models. Modeling methods varied among our analysis. Starting from a simple trivariate reduction technique and its' strict assumptions and continuing with multiplicative factor models and copula-based. We explained advantages and drawbacks for each method and each model. After implementation of a variety of models, we came to the conclusion that the bivariate generalised Poisson model (BGPR-P) is the model that fit our data the best, giving the lowest AIC and the highest loglikelihood value. Implementation of bivariate models is only the top of the iceberg of modeling count data. There is a lot of space for improvements and extensions. Literature for handling univariate models is wide, however, when it comes to bivariate models, like these models we used, everything becomes more difficult and complex. One interesting extension is examining these models' behaviour using multivariate count data. Models complexity becomes greater, but relationship of more types of guarantees in insurance applications is a great goal to discover. Insurance claims may come from several reasons, but only if we add one more category, our models become multivariate and their complexity becomes greater. Probability functions may do not have closed forms, therefore optimization can be really hard to be carried out. Furthermore, it would be interesting to enrich our models with more explanatory variables, in order to characterize with greater precision every client's profile.

Moreover, a matter of great interest is the dependence of our claim counts variables. In our analysis we had to face positive dependence between response variables. Even if we stick to bivariate models, negative correlation betweem claim counts would be a great challenge for modeling. For example, trivariate reduction method is not a feasible method of modeling negative dependence. In this case, probably we could use different modeling methods, like copulas, also different distributions that can handle better this situation. For copulas and multiplicative factor based models, an interesting extension could be a bivariate response variable with marginals that follow different distributions. To be more specific, we can

assume that N_1 is explained better with Poisson distributions as N_2 with lognormal distribution. Since, these modeling methods allow for different marginals, it is interesting to observe how these models fit our data.

The existence of many zero claims is an additional issue for our analysis. It is common in insurance applications, since many policyholders avoid to report small claims due to conditions in their contract. Zero-inflated models improve a lot our models' fit, regardless the distribution we use. Especially, the zero-inflated negative binomial one explains heterogeneity among policyholders and type of claims and deals with overdispersion. In a priori ratemaking, some profiles have to be constructed segmenting policyholders in categories. For this purpose, five different profiles are constructed based on claims mean values, from best drivers to worst. Best drivers' profile is characterized from driving in high risk regions and they are loyal to the insurance company (more than 5 years). Vehicle's horsepower has an important role in our analysis, since if it is greater than 5500 cc, then more claims are reported. All these profiles were made based on mean scores of our models. Changing the method of insurance pricing could be another extension of our analysis. For example, a retrospective rating method.

Finally, since we mentioned that multivariate modeling could be an extension of great interest, new optimizing methods have to be used. If we assume three or more response variables, different categories of count claims, that are correlated, then standard methods we used in this analysis, probably will not work. Correlation matrix of multivariate variables is way more complicated and due to its higher dimension, optimizing tools like Newton-Raphson method seem to be time-consuming and difficult to be implemented methods. Therefore, machine learning techniques for optimization seem to be a great tool for multivariate modeling. Last but not least, a great method for optimizing multivariate count data models parameters of likelihood function is by the inference function of margins, which is divided to two steps. Let assume that we have a multivariate copula model. At the first step of this method the univariate log-likelihoods are maximized independently of the copula parameter and at the second step the joint log-likelihood maximized over the vector of copula parameter with univariate parameters fixed as estimated at the first step of the method. Estimation by this method becomes

more popular as the dimension increases and computational problems arise, for details see Karlis and Nikoloulopoulos (2009).

References

[1] Karlis D., (2001), Exact ML estimation for Mixed Poisson Regression Models Statistical Modelling : An International Journal, 1, 305-319

[2] Karlis D., (2005), An EM algorithm for mixed Poisson distributions ASTIN Bulletin 35, 3-24

[3] Karlis D., and Ntzoufras I. (2003), Analysis of sports data by using bivariate Poisson models. Journal of the Royal Statistical Society: Series D (The Statistician), 52: 381-393. doi:10.1111/1467-9884.00366.

[4] Bermúdez L., (2009), A priori ratemaking using bivariate Poisson regression models, Insurance: Mathematics and Economics, Volume 44, Issue 1, Pages 135-141,ISSN 0167-6687, https://doi.org/10.1016/j.insmatheco.2008.11.005.

[5] Ghitany M., Karlis D., Al-Mutairi D., Al-Awadhi F., (2012), An em algorithm for multivariate poisson regression models and its application. Applied Mathematical Sciences, 6, 6843–6856.

[6] H. S. Steyn, (1976), On the Multivariate Poisson Normal Distribution, Journal of the American Statistical Association, 71:353, 233-236, DOI: 10.1080/01621459.1976.10481521

[7] Aitchison J., Ho C.H, (1989), The multivariate Poisson-log normal distribution, Biometrika, Volume 76, Issue 4, Pages 643–653, https://doi.org/10.1093/biomet/76.4.643

[8] Bermúdez L., Karlis D., (2012), A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking, Computational Statistics Data Analysis, Volume 56, Issue 12, Pages 3988-3999, ISSN 0167-9473, https://doi.org/10.1016/j.csda.2012.05.016.

[9] Karlis D., Meligkotsidou L., (2007), Finite multivariate Poisson mixtures with applications. Journal of Statistical Planning and Inference, 137, 1942–60, ISSN 0378-3758, https://doi.org/10.1016/j.jspi.2006.07.001.

[10] Lakshminarayana J.,Pandit S.N.N.,Srinivasa Rao K., (1999), On a bivariate poisson distribution, Communications in Statistics - Theory and Methods, 28:2, 267-276, DOI: 10.1080/03610929908832297

[11] Famoye F., (2010), On the bivariate negative binomial regression model, Journal of Applied Statistics, 37:6, 969-981, DOI: 10.1080/02664760902984618

[12] Faroughi P., Ismail N., (2017), Bivariate zero-inflated negative binomial regression model with applications, Journal of Statistical Computation and Simulation, 87:3, 457-477, DOI: 10.1080/00949655.2016.1213843.

[13] Zamani H., Faroughi P. Ismail N., (2016), Bivariate generalized Poisson regression model: applications on health care data, Empir Econ 51, 1607–1621, https://doi.org/10.1007/s00181-015-1051-7.

[14] Hofer V. Leitner J., (2012), A bivariate Sarmanov regression model for count data with generalised Poisson marginals, Journal of Applied Statistics, 39:12, 2599-2617, DOI: 10.1080/02664763.2012.724661.

[15] Shi P., Zhang W. and Valdez E.A., (2012), Testing Adverse Selection With Two-Dimensional Information: Evidence From the Singapore Auto Insurance Market, Journal of Risk and Insurance, 79: 1077-1114, doi:10.1111/j.1539-6975.2012.01472.x.

[16] So S., Lee D.H., Jung B.C., (2011), An alternative bivariate zero-inflated negative binomial regression model using a copula, Economics Letters, Volume 113, Issue 2, Pages 183-185, ISSN 0165-1765, https://doi.org/10.1016/j.econlet.2011.07.017.

[17] Faroughi P. & Ismail N., (2017), Bivariate zero-inflated generalized Poisson regression model with flexible covariance, Communications in Statistics - Theory and Methods, 46:15, 7769-7785, DOI: 10.1080/03610926.2016.1165846.

[18] Cameron A.C, Li T., Trivedi P.K., Zimmer M.D, (2004), Modelling the differences in counted outcomes using bivariate copula models with application to

mismeasured counts, The Econometrics Journal, Volume 7, Issue 2, 1 December 2004, Pages 566–584, https://doi.org/10.1111/j.1368-423X.2004.00144.x.

[19] Wang K., Lee A.H, Yau K.K.W, Carrivick P.J.W, (2003), A bivariate zero-inflated Poisson regression model to analyze occupational injuries, Accident Analysis Prevention, Volume 35, Issue 4, 2003, Pages 625-629, ISSN 0001-4575, https://doi.org/10.1016/S0001-4575(02)00036-2.

[20] Ridout M., Demetrio C.G.B, Hinde J., (1998), Models for count data with many zeros, Cape Town, December 1998.

[21] Mardia K.V, (1970), Measures of multivariate skewness and kurtosis with applications, Biometrika, Volume 57, Issue 3, Pages 519–530, https://doi.org/10.1093/biomet/57.3.519.

[22] Pandit S.N.N., Rao C.R, Sarma Y.V.S.S., (1988), On a bivariate binomial distniution, Assam Statistical Review, 2(2), 67-72.

[23] Sklar A., (1959), Fonctions de r´epartition ‘a n dimensions et leurs marges, Publications de l’Institut Statistique de l’Universite de Paris ´ 8 (1959), 229–31.

[24] Joe H., (1997), Multivariate Models and Dependence Concepts, Monographs on Statistics and Applied Probability, 73, Chapmann Hall, London.

[25] Nelsen R.B., (1999), An Introduction to Copulas, volume 139 of Lecture Notes in Statistics, Springer, New York.

[26] Barlow R.E, Proschan F., (1975), Importance of system components and fault tree events, Stochastic Processes and their Applications, Volume 3, Issue 2, Pages 153-173, ISSN 0304-4149, https://doi.org/10.1016/0304-4149(75)90013-7.

[27] Marshall A.W Olkin I., (1967), A Multivariate Exponential Distribution, Journal of the American Statistical Association, 62:317, 30-44, DOI: 10.1080/01621459.1967.10482885.

[28] McCullagh F. and Nelder J.A., (1983), Generalized Linear Models. Chapman and Hall, London.

[29] Efron B., (1979), Computers and the theory of statistics: Thinking the unthinkable, SIAM Review, 21, 460–480, https://doi.org/10.1137/1021092.

[30] Akaike H., (1973), Information theory and an extension of the maximum likelihood principle, in: B.N. Petrov and F. Csaki, eds., 2nd lnternat. Syrup. on Information Theory (Akademia Kiado, Budapest) pp. 267 281.

[31] Dempster, A.P., Laird, N.M. and Rubin, D.B., (1977), Maximum Likelihood from Incomplete Data Via the EM Algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39: 1-22, doi:10.1111/j.2517-6161.1977.tb01600.x.

[32] Schwarz G., (1978), Estimating the dimension of a model, Ann. Statist., 6, 461-464.

[33] Poisson S. D., (1837), Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile, Précédées des Regles Générales du Calcul des Probabilitiés. Bachelier, Imprimeur-Libraire pour les Mathematiques, la Physique, etc.: Paris.

[34] Holgate P., (1964), Estimation for the bivariate Poisson distribution, Biometrika, Volume 51, Issue 1-2, June 1964, Pages 241–287, https://doi.org/10.1093/biomet/51.1-2.241.

[35] Johnson N.L and Kotz S., (1969), Distributions in Statistics: Discrete distributions, Houghton Mifflin, Boston, Massachusetts.

[36] Kocherlakota N.R, (1992), Bubbles and constraints on debt accumulation, Journal of Economic Theory, Volume 57, Issue 1, Pages 245-256, ISSN 0022-0531, https://doi.org/10.1016/S0022-0531(05)80052-3.

[37] LI T.-H., (1999), Multiscale representation and analysis of spherical data by spherical wavelets, SIAM J., Sci. Comput., 21 924– 953, MR1755172.

[38] Genest C. Rivest L.P, (1993), Statistical Inference Procedures for Bivariate Archimedean Copulas, Journal of the American Statistical Association, 88:423, 1034-1043, DOI: 10.1080/01621459.1993.10476372.

[39] Capéraà P., Fougères A.L, Genest C., (2000) Bivariate Distributions with Given Extreme Value Attractor, Journal of Multivariate Analysis, Volume 72, Issue 1, Pages 30-49, ISSN 0047-259X, https://doi.org/10.1006/jmva.1999.1845.

[40] Ridout MS, Besbeas P., (2004), An empirical model for underdispersed count data, Statistical Modelling, 2004;4(1):77-89, doi:10.1191/1471082X04st064oa.

[41] Zamani H. Ismail N., (2012), Functional Form for the Generalized Poisson Regression Model, Communications in Statistics - Theory and Methods, 41:20, 3666-3675, DOI: 10.1080/03610926.2011.564742.

[42] S Kocherlakota K Kocherlakota, (1986), Goodness of fit tests for discrete distributions, Communications in Statistics - Theory and Methods, 15:3, 815-829, DOI: 10.1080/03610928608829153.

[43] Bessel F., (1824), Abh. d. K. Akad. Wiss. Berlin, 1

[44] Stewart J.A, (1994), The poisson-lognormal model for bibliometric/scientometric distributions, Information Processing Management, Volume 30, Issue 2, 1994, Pages 239-251, ISSN 0306-4573, https://doi.org/10.1016/0306-4573(94)90067-1.

[45] Greg C. G. Wei Martin A. Tanner, (1990), A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms, Journal of the American Statistical Association, 85:411, 699-704, DOI: 10.1080/01621459.1990.10474930.

[46] Geman S. and Geman D., (1984), Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no. 6, pp. 721-741, Nov. 1984, doi: 10.1109/TPAMI.1984.4767596.

[47] Metropolis N., Rosenbluth A.W, Rosenbluth M.N, Teller A.H, and Teller E., (1953), Equation of state calculations by fast computing machines, J. Chem. Phys., 21, 1087

[48] Consul P.C, (1989), Generalised Poisson Distribution, Marcel Dekker, New York.

[49] Joe H. and Zhu R., (2005), Generalized Poisson Distribution: the Property of Mixture of Poisson and Comparison with Negative Binomial Distribution, Biom. J., 47: 219-229, doi:10.1002/bimj.200410102.

[50] Sarmanov O.V, Generalized normal correlation and two-dimensional Frechet classes, Dokl. Akad. Nauk SSSR, 168:1 (1966), 32–35.

[51] Balakrishnan N. and Ambagaspitiy R.S., (1994), On skew Laplace distributions, (Tech. Rep.). Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada.

[52] Lambert J.H, (1758), Observationes variae in mathesin puram.

[53] Denuit M., Marechal X., Pitrebois S. and Walhin J.-F., (2007), Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus–Malus Scales (New York: Wiley).

[54] David M., Jemna D., (2015), Modeling the Frequency of Auto Insurance Claims by Means of Poisson and Negative Binomial Models, Annals of the Alexandru Ioan Cuza University - Economics, 62(2), 151-168. doi: https://doi.org/10.1515/aicue-2015-0011.

[55] Gumbel E.J, (1960), Bivariate Exponential Distributions, Journal of the American Statistical Association, 55:292, 698-707, DOI: 10.1080/01621459.1960.10483368.

[56] Clayton D. G., (1978), A model for association in bivariate life-tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65, 141-151.

[57] Frank MJ, (1979), On the simultaneous associativity of F(x,y) and $x + y - F(x, y)$. Aequationes Math., 19:194-226.

[58] Karlis D., Nikoloulopoulos A.K, (2009), Modelling multivariate count data using copulas, Communications in Statistics - Simulation and Computation, Taylor Francis, 2009, 39 (01), pp.172-187, 10.1080/03610910903391262.