Journal of Statistics & Applied Science Issue -5
ISSN 2718-0999

Research Article

Received: date:11.21.2021
Accepted: date:15.06.2022
Published: date:30.06.2022

The Incorporation of Generalized Linear Models into Bivariate Gaussian Copula and An Application

Övgücan Karadağ Erdemir [1]* Meral Sucu[2]

[1]Hacettepe University, Department of Actuarial Science, Ankara, Turkey; ovgucan@hacettepe.edu.tr
[2] Hacettepe University, Department of Actuarial Science, Ankara, Turkey; msucu@hacettepe.edu.tr
Orcid: 0000-0002-4725-3588[1] Orcid: 0000-0002-7991-1792[2]
* Corresponding Author, e-mail: ovgucan@hacettepe.edu.tr

Abstract: In non-life insurance mathematics, analyses and premium or reserve calculations are carried out in the presence of dependency between the claim variables in recent years. And, thus over- or underestimation of aggregate loss caused by the assumption of dependency between the claim severity and frequency are prevented. The Gaussian copula function, which is frequently used for dependency modeling, is integrated into the marginal generalized linear models to obtain a mixed copula-based regression model called "copula regression". In this study, a copula regression model is created using a bivariate Gaussian copula, Gamma and Poisson marginal generalized linear models for claim severity and frequency, respectively. An application is performed with a simulated data where there is a dependence between the claim severity and frequency using the R package "CopulaRegression". The importance of the modeling of dependency between claims is investigated by the comparison of the independent and dependent models and the results of application show that the copula regression model in which dependency is considered has lower relative mean square errors compared the independent marginal generalized linear models.

Keywords: Bivariate Gaussian Copula Function; Gamma GLM, Poisson GLM, Dependence, Mixed Copula Approach

1. Introduction

Insurance is a multivariate system created under certain conditions and the complex structures caused by the multivariate situations in life and non-life insurance calculations are generally ignored with some basic assumptions. One of the main assumptions used in calculations is that the random variables are independently and identically distributed. Although the independence is a very basic assumption, it is not a very realistic one and it can cause over-or underestimation [1]. The two main components of the non-life insurance mathematics are the claim severity which represents the monetary losses of claims and the claim frequency expressing the numbers of claims. The calculations such as pricing and reserve are carried out under the assumption of independence for many years, but in recent years, the dependency between variables is included to obtain more accurate and realistic estimation, pricing and reserve calculations. In the presence of the dependence, various dependency structures are encountered. Therefore, in studies where the dependency is taken into consideration, first of all, it is necessary to decide the dependence structure, and to include the dependency in the calculations by modeling.

In ratemaking studies, generally claim severity and frequency are modeled by generalized linear models (GLMs) separately, thereafter, expected values of claim severity and frequency are multiplied to calculate the aggregate loss under independence assumption [2, 3, 4]. To avoid the effects of independence assumption, some approaches were proposed to model the dependence. Copula which is the most used method for modeling dependence in financial and statistical studies is also introduced in actuarial studies. Firstly, copula was used to model dependency in contingent life insurance [5], and over time it was used for dependency modeling studies in non-life insurance mathematics [6, 7].

The dependence between claims can be modeled only with copula functions, or marginal GLMs can be included in the copula. Song [6] defined the mixed copula approach using GLMs and Gaussian copula

Citation: Ö.K Erdemir and M. Sucu , "The Incorporation of Generalized Linear Models into Bivariate Gaussian Copula and An Application", Journal of Statistics and Applied Sciences, no. 5, 1-9, Jul. 2022, doi: 10.52693/jsas.1039360

function and proposed Vector GLM (VGLM) in order to model the dependency between mixed variables. The mixed copula approach lays the groundwork to model dependency between mixed variables such as continuous claim severity and discrete claim frequency. Kastenmeier [8] established a joint regression model for claim severity and frequency using the mixed copula approach. Song et al. [7] modeled dependency between continuous, discreteand mixed variables by using a joint regression analysis. Kolev and Paiva [9] gave some results about regression models based on copulas. Mixed copula model was proposed by Czado et al. [1] to model dependency between claim severity and frequency using Gaussian copula. The usage of copula-based regression models for mixed variables in medical was examined by De Leon and Wu [10].

Generally Gaussian bivariate copula function was used in mentioned studies. However, Krämer et al. [11] used the other parametric copulas such as Clayton, Gumbel and Frank copulas besides Gaussian copula and they referred the models contain GLMs and copula as the copula-based regression models. Krämer et al. [11] also modeled the dependency between claim components only using the copula and claims without GLMs and the approach is entitled as the copula-based models. Copula-based models are useful for modeling the dependence according to only the distribution of claim components without any explanatory variables.

In addition to, copula-based models and copula-based regression models, there are some other approaches for dependency modeling. Gschlöβl and Czado [12] introduced a new approach to model dependency between claim severity and frequency by taking the claim frequency as an explanatory variable in the GLM modeling of total loss and Garrido et al. [13] also used same approach to model dependency in non-life insurance. A copula-based multivariate Tweedie regression model was proposed to model semi-continuous claims. [14]. A copula quantile regression approach was used to estimate the parameters of copula [15].

In this study, it is examined how the estimates will change if the dependency between the claim severity and frequency is considered instead of the independence assumption, which is frequently used in non-life insurance mathematics. Therefore, the dependency between the claim severity and frequency is modeled via copula regression model using the bivariate Gaussian copula function and the marginal gamma and Poisson GLMs. An application is performed with a data simulation where there is a dependence between the claim severity and frequency using the information of a real Turkey comprehensive insurance data with the R package "CopulaRegression". The importance of including dependency in calculations is investigated by the comparison of the independent marginal generalized linear model and copula regression model under the assumptions obtained with the real Turkey data. The study can be a guide for researchers who are interested in pricing studies involving dependency in Turkey.

The remainder of the paper is organized as follows. A general information about the bivariate Gaussian copula function and GLM, which are the components of the copula regression model is briefly given in the Section 2. Mixed copula approach and copula regression model are given in Section 3. In Section 4, using the R package "CopulaRegression" [16], joint cumulative distribution function (c.d.f.) and joint probability density function (p.d.f.) of claim severity and frequency are drawn and a simulation study is carried out. The concluding remarks are given in Section 5.

2. Methods

Copula-based regression models are obtained by the combination of a bivariate copula function and two marginal GLMs [1, 6, 8, 11] In this study, a bivariate Gaussian copula function is integrated with marginal gamma and Poisson GLMs to create a copula regression model. Definitions and properties of the bivariate Gaussian copula function, gamma GLM and Poisson GLM are given briefly as follows for a better understanding of copula regression model.

2.1. The Bivariate Gaussian Copula Function

Copula which is introduced by Sklar [17] is used to model dependence among variables in many disciplines such as economy, finance, econometric, statistics and actuarial science. Copula can be defined as a function which link a multivariate distribution function to their marginal distributions which have standard uniform distributions [18]. Using Sklar's Theorem, where θ is the copula parameter, a $\boldsymbol{C}(.,.|\boldsymbol{\theta})$ bivariate parametric copula can be defined by Equation (1) as follows,

$$F_{XY}(x,y|\theta) = C(F_X(x), F_Y(y)|\theta) \quad (1)$$

where $F_X(x)$ and $F_Y(y)$ are the marginal distribution functions. Copulas can be defined by Kendall's $\tau = 4\int_{[0,1]^2} C(x,y)\, dC(x,y) - 1 \in [-1,1]$ instead θ, due to invariability under monotone transformations of marginal distributions [11].

Gaussian copula function is used in many studies due to the advantages of multivariate normal distribution [6, 19]. The relationship between θ and Kendall's τ for Gaussian copula is $\tau = \frac{2}{\pi} arcsin(\theta)$. Let $\Phi(.)$ and $\Phi_2(.|\Gamma)$ denote the univariate and the bivariate standard normal distribution functions with Γ correlation matrix, respectively. A bivariate Gaussian copula $C: \boldsymbol{I}^2 \to \boldsymbol{I}$ is given as follows,

$$C(u_1, u_2|\boldsymbol{\Gamma}) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2)|\boldsymbol{\Gamma})$$
$$C(u_1, u_2|\boldsymbol{\Gamma}) = \frac{\partial}{\partial u_1}\int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_{i2})} \frac{1}{2\pi\sqrt{1-(\rho_{12})^2}} exp\left\{-\frac{s^2 - 2\rho_{12}st + t^2}{2(1-(\rho_{12})^2)}\right\} dsdt \quad (2)$$

where $(u_1, u_2) \in \boldsymbol{I}^2$, $i = 1, \dots,$ n is the number of observations and $\boldsymbol{\Gamma} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}$ displays the correlation matrix.

2.2. Generalized Linear Model

GLM is a generalized form of a linear model which models the relationship between a response variable and explanatory variables. GLM consists of a density function of the response variable from the exponential family, a linear component and a link function [20]. In GLM, response variable y follows an exponential family distribution such as Poisson, binomial, negative binomial, normal, gamma, inverse Gaussian, etc. Linear component is an instrument that represents the relationship between the response variable and explanatory variables. The linear component for ith observation $\eta_i = \boldsymbol{X_i'\beta} = \sum_{j=1}^{p} \beta_j x_{ij}$ where $\boldsymbol{X_i} = \left(x_{i1}, x_{i2}, \dots, x_{ip}\right)'$ and $\boldsymbol{\beta} = \left(\beta_1, \beta_2, \dots, \beta_p\right)'$ for $i = 1,2, \dots, n$, $j = 1,2, \dots, p$. The number of explanatory variables is $p-1$, the first term of $\boldsymbol{X_i}$ is model constant and equal to 1 for all observations. Monotone and differentiable link function links the expected value of the response variable and the linear component. According to distribution of the response variable, link function is determined as identity, logarithmic, power, square root and logit. For $E(y_i) = \mu_i$, the general form of the link function is $g(\mu_i) = \eta_i$. observation $\eta_i = \boldsymbol{X_i'\beta} = \sum_{j=1}^{p} \beta_j x_{ij}$ where $\boldsymbol{X_i} = \left(x_{i1}, x_{i2}, \dots, x_{ip}\right)'$ and $\boldsymbol{\beta} = \left(\beta_1, \beta_2, \dots, \beta_p\right)'$ for $i = 1,2, \dots, n,\ j = 1,2, \dots, p$. The number of explanatory variables is $p-1$, the first term of $\boldsymbol{X_i}$ is model constant and equal to 1 for all observations. Monotone and differentiable link function links the expected value of the response variable and the linear component. According to distribution of the response variable, link function is determined as identity, logarithmic, power, square root and logit. For $E(y_i) = \mu_i$, the general form of the link function is $g(\mu_i) = \eta_i$.

GLMs for gamma and Poisson distributed response variables are called as gamma and Poisson GLMs, respectively. Logarithmic link function is widely used for both GLMs. Let $X \sim Gamma(\mu, \nu^2)$ and $Y \sim Poisson(\lambda)$, the p.d.f.s of X and Y are given in Equations (3) and (4) according to the mean parametrization. Here, μ and λ are the mean parameters, while ν^2 is the dispersion parameter.

$$f_X(x|\mu, \nu^2) = \frac{1}{\Gamma\left(\frac{1}{\nu^2}\right)}\left(\frac{1}{\mu\nu^2}\right)^{\frac{1}{\nu^2}} y^{\left(\frac{1}{\nu^2}\right)-1} exp\left(-\frac{x}{\mu\nu^2}\right), \quad x \geq 0 \quad (3)$$

$$f_Y(y|\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}, \quad y = 0,1,2,3, \dots \quad (4)$$

Gamma and Poisson GLMS with logarithmic link function are given by Equations (5) and (6), respectively as follows. Here, $\boldsymbol{z_1'} \epsilon R^p$ and $\boldsymbol{z_2'} \epsilon R^q$ are the explanatory variable vectors of the claim severity X with p parameters and the claim frequency Y with q parameters. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the parameter vectors of the random variables X and Y. $ln(e)$ is the offset term, where e is the exposure to risk.

$$g(x) = ln(\mu) = \boldsymbol{z_1'\alpha} \quad (5)$$

$$g(y) = ln(\lambda) = ln(e) + \mathbf{z}_2'\boldsymbol{\beta} \tag{6}$$

3. Copula Regression Model

A mixed copula-based regression model called "copula regression" can be obtained by a bivariate copula function and two marginal GLMs. For the purpose of modeling dependency between claim components, claim severity and frequency are modeled via gamma and Poisson GLMs, respectively. Then the marginal GLMs and a bivariate copula function are combined to define "copula regression model". Generally, Gaussian copula function is used in this combination due to the advantages of the Gaussian distribution [1, 6, 8]. However, Krämer et al. [11] used Clayton, Gumbel and Frank copulas besides Gaussian copula. The integration of bivariate Gaussian copula and the marginal GLMs is briefly summarized by Figure 1.

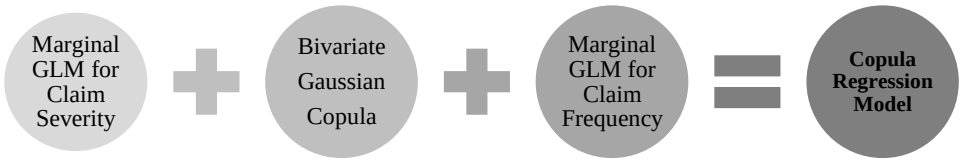

Figure 1. Copula regression model

Copula regression model is comprised of marginal GLMs and a bivariate Gaussian copula function as summarized by Figure 1 using the mixed copula approach proposed by Song [6]. The mixed copula approach is used to model the dependency between mixed variables such as continuous claim severity and discrete claim frequency. For a bivariate Gaussian copula function with the correlation matrix $\boldsymbol{\Gamma}$, copula regression model can be shown as follows by Equation (7).

$$C(Gamma\ GLM, Poisson\ GLM \mid \boldsymbol{\Gamma}) \tag{7}$$

The mixed copula approach, allows the usage of copula functions which, are used with only the continuous random variables, also together with discrete random variables. According to the Sklar's Theorem [16], the joint c.d.f. of mixed variables X and Y is written by Equation (8). The joint p.d.f. of mixed variables X and Y is obtained by mixed copula approach [6] by Equation (9).

$$F_{XY}(x,y) = P(X \le x, Y \le y) = C(F_X(x), F_Y(y)|\boldsymbol{\Gamma}) \tag{8}$$

$$f_{XY}(x,y) = \frac{\partial}{\partial x} P(X \le x, Y = y) \tag{9}$$

For a Gaussian bivariate function, using Radon-Nikodym derivative and the information of $\frac{\partial}{\partial x} C(u_1,u_2) = C_1'(u_1,u_2|\boldsymbol{\Gamma}) = \Phi\left(\frac{\Phi^{-1}(u_2)-\rho_{12}\Phi^{-1}(u_1)}{\sqrt{(1-(\rho_{12})^2)}}\right) := D_{\rho_{12}}(u_1,u_2)$, the p.d.f. of mixed variables X and Y can be expanded by Equation (10) as follows where $F_X(x) = u_1$ and $F_Y(y) = u_2$ [1, 6, 8].

$$f_{XY}(x,y) = \frac{\partial}{\partial x} P(X \le x, Y \le y) - \frac{\partial}{\partial x} P(X \le x, Y \le y-1)$$

$$= \frac{\partial}{\partial x} C(F_X(x), F_Y(y)|\boldsymbol{\Gamma}) - \frac{\partial}{\partial x} C(F_X(x), F_Y(y-1)|\boldsymbol{\Gamma})$$

$$f_{XY}(x,y) = f_X(x)\left[D_{\rho_{12}}\left(F_X(x), F_Y(y)\right) - D_{\rho_{12}}\left(F_X(x), F_Y(y-1)\right)\right] \tag{10}$$

The parameter vector of a copula regression is $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau)$ where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the parameter vectors of gamma and Poisson GLMs and τ is the Gaussian copula parameter. The parameter estimation is carried out by maximum-likelihood techniques.

Krämer et al. [11] were defined the relative MSE as $MSE_{rel} := E\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{\gamma_i - \hat{\gamma}_i}{\gamma_i}\right)^2\right)$ for the parameter vector $\boldsymbol{\gamma} \in R^k$, where $\hat{\gamma}_i$ is the estimate of the parameter. To evaluate the performance of dependent model (copula regression) and independent model (marginal GLMs), the relative MSEs of the estimators

of parameters of GLMs $\boldsymbol{\alpha} \in \boldsymbol{R}^p$ and $\boldsymbol{\beta} \in \boldsymbol{R}^q$, and also the relative MSE of the estimator of the parameter of Gaussian copula τ is calculated.

4. Application

Analyses are performed in R using the R package "CopulaRegression" [16] which presents a bivariate, copula-based model for the joint distribution of a pair continuous and discrete random variables. The R packages "MASS" [21] and "VineCopula" [22] are also used, since they work based on "CopulaRegression" package. The pair of continuous and discrete random variables is composed as a couple of claim severity and frequency. The claim severity and frequency are modeled by marginal gamma and Poisson GLMs, respectively. The marginal GLMS are linked by the bivariate Gaussian copula function to model dependence between claim severity and frequency.

An insurance system is designed using the comprehensive insurance data in Erdemir and Sucu [23] to compare the copula regression and the independent models. The data taken from a Turkish non-life insurance company for year 2017 contains the information about the claim severity, claim frequency and some risk factors related the policyholders and the vehicles; such as age of policyholder (21,…,80), gender (male, female), type of vehicle (automobile, pickup, rent a car, taxi), age of vehicle (0,…,17), usage (private, leasing, commercial), residence (metropolis, little town), engine capacity of vehicle (small, medium, large), type of fuel (benzine, diesel) and status of the policy (new, renewal) of 2820 observations. The descriptive statistics of the comprehensive insurance data is given by Table 1.

Table 1. Descriptive statistics of comprehensive insurance data

Variables	Minimum	Maximum	Mean	Median	Variance
Claim Severity (X)	51.24	35477.00	1759.4936	2933.58735	8605934.753
Claim Frequency (Y)	1	4	1.7034	0.4678	0.2190

By the reason of, the package "CopulaRegression" is designed to generate its own dependent data, the information of the data is only used to make some assumptions of simulation study. The dispersion parameter of the gamma distribution is assumed constant as ν^2=0.5 for an easy calculation and the method of moments is used to determine the parameters of the gamma and Poisson distributions. The mean parameters of gamma and Poisson distributions are found as μ=1760 and λ=1.18, respectively [23].

First, the graphs of joint c.d.f. and p.d.f. of claim severity and frequency are plotted considering the dependency between the claim variables using the mixed copula approach and also drawn under the independence assumption to compare the dependent and independent models. "CopulaRegression" package use the Kendall's τ correlation coefficient instead of ρ Spearman correlation coefficient in some analysis. These two coefficients can be easily converted into each other through the copula parameter. The copula parameter is the correlation coefficient for the Gaussian copula function, hence Kendall's τ can be directly associated dependency with $\tau = \frac{2}{\pi} arcsin(\rho)$. Kendall's τ is taken 0 for independent model, since $\frac{2}{\pi} arcsin(0) = 0$. For the copula regression, the values of τ are chosen as 0.3 and 0.9 for the low and high degrees, respectively. The graphs of the joint c.d.f. and p.d.f. are given in Figures 2 and 3. The effect of considering the dependency between claim components and also the effect of τ on dependence modeling are investigated.

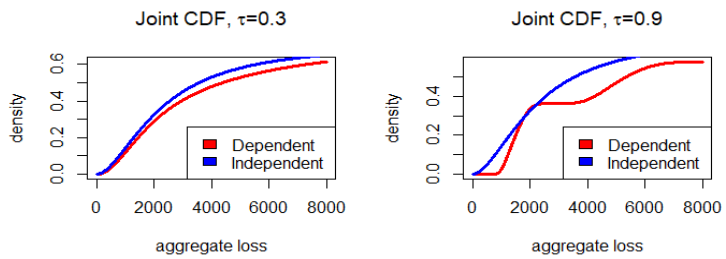

Figure 2. Joint c.d.f. of claim severity and frequency

Since the graphs given in Figure 2 are the graphs of cumulative distribution, it is an expected result to obtain increasing curves for both τ values. The probability values for the dependent model which are displayed by red lines are smaller for both τ values. The effect of dependence cannot be directly analyzed since the probabilities are expressed cumulatively, however, a little fluctuation is observed for highly dependent model when τ=0.9.

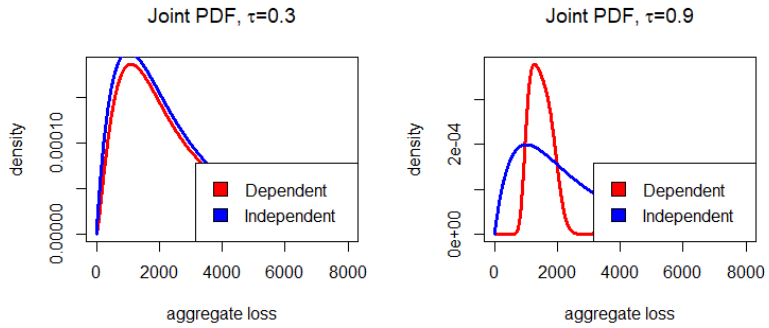

Figure 3. Joint p.d.f. of claim severity and frequency

Right skewness is observed in Figure 3, since gamma distribution assumption is used for claim severity. The effect of considering the dependency between claim components on aggregate loss is observed clearly and it is noticed that the p.d.f. is very sensitive to dependency for higher value of τ. A hillier distribution is obtained with the dependent model. For τ=0.3, that is, when the dependency level is low, both the graphs of c.d.f. and p.d.f. are similar in the two models according to the Figures 2 and 3. However, when the dependency level is high, the graphs differ for the dependent and independent models.

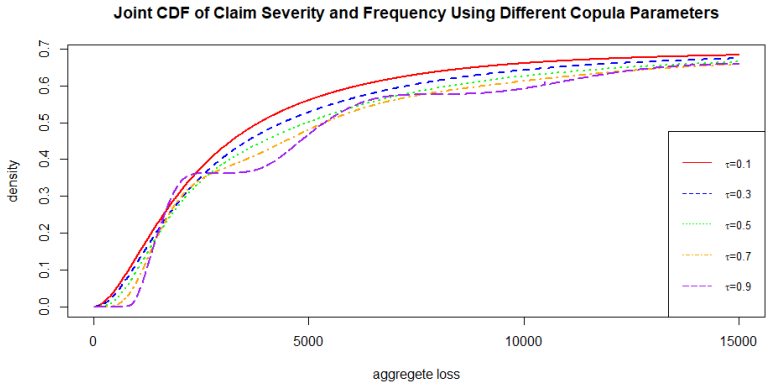

Figure 4. Joint c.d.f. of claim severity and frequency using different Kendall's τ

Five different values of the Kendall's τ (0.1, 0.3, 0.5, 0.7, 0.9) are taken for Gaussian copula function and to analyze the effects of Kendall's τ on joint c.d.f. and p.d.f., the graphs are redrawn by different τ values, the graphs are given by Figure 4 and Figure 5 as follows. The effect of τ cannot be directly observed in Figure 4 as in Figure 2, due to the cumulative structure. However, according to Figure 5, as the τ value increases, the fluctuations in the probability value of the aggregate loss increase. The total loss is more sensitive to higher τ values. It can be deduced that it is important to include dependency in the calculations.

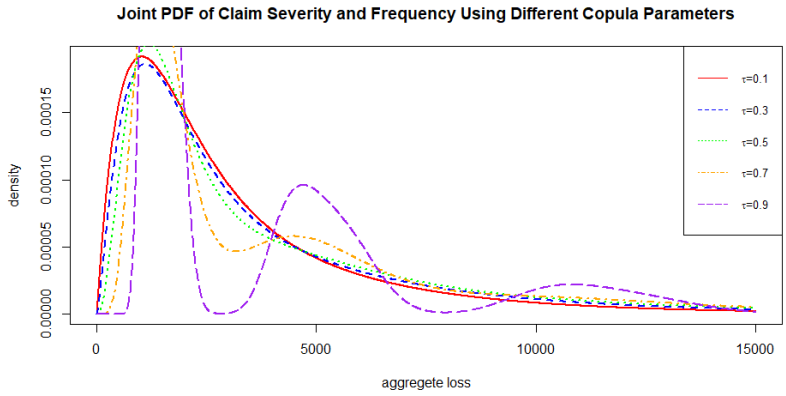

Figure 5. Joint p.d.f. of claim severity and frequency using different Kendall's τ

Joint c.d.f. and p.d.f. are drawn independently of the possible explanatory variables. Explanatory variables can be included by the copula regression model using marginal GLMs. In the R package "CopulaRegression", copreg () function fits a joint, bivariate regression model for a gamma GLM and a (zero-truncated) Poisson GLM. On the purpose of fitting model, the package simulates the joint regression data under some assumption about the distributions and copula function. Then using the simulated data, the copula regression model is created. An insurance system is designed using the comprehensive insurance data in Erdemir and Sucu [23] and a little simulation study with R=50 trials is performed using the R package. The system with n=1000 policy groups with only the automobile type of vehicle is considered and the groups contain insured with gamma-distributed claim severity with the parameters (μ=1760, ν^2=0.5) and Poisson distributed claim frequency with the parameter (λ=1.18). It is assumed that all policy groups contain the same number of policyholders. The claim severity and frequency are modeled by gamma and Poisson GLMs, respectively. Gender, residence and engine power of vehicle are determined as the explanatory variables for GLM modeling and all explanatory variables are assumed as categorical variables. Gender (male-female) and residence (metropolis, little town) are two-category variables, hence they are modeled by only one dummy variables. On the other hand, since the engine capacity of vehicle (low, medium, high) is three-category variable, it is modeled with two dummy variables. Same explanatory variables are used for gamma and Poisson GLMs. Marginal GLMs are designed with an intercept term, hence the first column of the design matrices contains 1's as $Z_1 \coloneqq Z_2 \coloneqq (1, z_{12}, \dots, z_{1n}) \coloneqq (1, z_{22}, \dots, z_{2n}) \epsilon R^{1000x5}$. The second and third columns are dummy variables corresponding to female and metropolis, respectively. The last two columns are the two dummy variables corresponding to low and high the engine capacity of vehicle. All dummy variables are generated randomly such as 0 or 1.

For the copula regression model, three different values are determined for τ as 0.1, 0.5 and 0.9 to represent the low, moderate and high levels of dependency, while τ is taken as 0 for independent model. Under these assumptions, the gamma distributed claim severity and the Poisson distributed claim frequency dependent on each other are generated by the R package. The dependent data changes according to the value of τ in R, hence the values of relative MSE change for independent model for different τ values.

The simulation study is performed for the comparison of the relative MSEs found with copula regression and independent model. A Monte Carlo simulation is carried out by R=50 trials and $\widehat{MSE}_{rel}^{(r)} \coloneqq E\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{\gamma_i - \hat{\gamma}_i^{(r)}}{\gamma_i}\right)^2\right)$ is calculated in the rth step. The mean of all simulations is obtained with $\overline{MSE}_{rel} \coloneqq \frac{\sum_{r=1}^{R}\widehat{MSE}_{rel}^{(r)}}{R}$ and the mean relative MSEs of $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$ and τ are calculated and given in Table 1 as follows. The variance can be calculated with the formula $\frac{\sum_{i=1}^{R}\left(MSE_{rel(i)} - \overline{MSE}_{rel}\right)^2}{R(R-1)}$.

Table 1. Relative MSEs of $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$ and τ with copula regression and independent models (τ=0, 0.1, 0.5, 0.9)

	Copula Regression Model	**Independent Model**
$\overline{MSE}_{rel}(\widehat{\boldsymbol{\alpha}})$ (τ=0.1)	**0.005721400**	0.00600650
$\overline{MSE}_{rel}(\widehat{\boldsymbol{\alpha}})$ (τ=0.5)	**0.004993235**	0.00583436
$\overline{MSE}_{rel}(\widehat{\boldsymbol{\alpha}})$ (τ=0.9)	**0.000778456**	0.00515738
$\overline{MSE}_{rel}(\widehat{\boldsymbol{\beta}})$ (τ=0.1)	**0.018385800**	0.02009030
$\overline{MSE}_{rel}\left(\widehat{\widehat{\boldsymbol{\beta}}}\right)$ (τ=0.5)	0.001899537	**0.00184171**
$\overline{MSE}_{rel}\left(\widehat{\widehat{\boldsymbol{\beta}}}\right)$ (τ=0.9)	**0.004126288**	0.04266449
$\overline{MSE}_{rel}(\hat{\tau})$ (τ=0.1)	**0.505447100**	0.54062490
$\overline{MSE}_{rel}(\hat{\tau})$ (τ=0.5)	**2.14184E-05**	5.7223E-05
$\overline{MSE}_{rel}(\hat{\tau})$ (τ=0.9)	**1.40001E-06**	0.00012083

Smaller values of relative MSE of parameter estimators are displayed bold in Table 1. It is noticed that, in the presence of dependence between of the claim severity and frequency, smaller errors are calculated using the copula regression model.

AIC values of copula regression and independent models are also calculated to compare the models using $AIC = -2l(\widehat{\boldsymbol{\gamma}}) + 2df$ where $l(\boldsymbol{\gamma}) = \sum_{i=1}^{n} lnf_{XY}(x, y)$ and the results of the comparison of AICs are given in Table 2.

Table 2. AIC values of copula regression and independent models

	Copula Regression Model	**Independent Model**
AIC (τ=0.1)	**9946.862**	10021.906
AIC (τ=0.5)	**9251.730**	9942.6440
AIC (τ=0.9)	**6467.568**	10029.176

For all τ values, in that for low, moderate and high degree of dependence, smaller AIC values are calculated for the copula regression models compared to the independent models.

5. Concluding Remarks

In non-life insurance mathematics, the dependence between claim severity and frequency has been modeled and included in the calculations in recent years. Copula-based models are converted into copula-based regression models with GLMs. The effects of the possible explanatory variables are also included in dependency modeling via copula-based models called "copula regression models". Copula regression models can be obtained with a mixed copula approach for continuous and discrete variables. Due to the mixed copula approach, the copula function, which can only be used with continuous variables, can be used with both discrete and continuous variables. It provides a flexible calculation in the branch that includes both discrete and continuous variables such as non-life insurance mathematics.

In this study, the dependency between the claim severity and frequency is modeled via copula regression model using the bivariate Gaussian copula function and the marginal gamma and Poisson GLMs. The effects of considering the dependency between claim variables are investigated by the comparison of the independent model and the copula regression model. The joint c.d.f. and p.d.f of claim severity and frequency are plotted for both models considering different dependency degrees. The importance of modeling dependency claim components is observed clearly with especially the graph of p.d.f. of aggregate loss. In addition, an insurance system is designed under some assumptions using the information of a real Turkish comprehensive insurance data. The relative MSEs are calculated for copula regression and independent models using different τ values. It is noticed that, copula regression models have smaller relative errors. Furthermore, AIC values are calculated for both models and the values support the result found with relative MSE values. In the light of these results, researchers studying on pricing or reserve in the non-life insurance mathematics can make more accurate calculations, including the dependence between the claim frequency and severity. With more accurate pricing policies, companies prevent problems with the ability to meet solvency margin. Since

the assumptions of application are based on real Turkish comprehensive data, this study can be a good guide for pricing studies for Turkey where the dependency between claim components is considered.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, Ö.K.E. and M.S.; methodology, Ö.K.E.; software, Ö.K.E.; validation, Ö.K.E. and M.S.; formal analysis, Ö.K.E.; investigation, Ö.K.E.; resources, Ö.K.E.; data curation, Ö.K.E.; writing—original draft preparation, Ö.K.E.; writing—review and editing, Ö.K.E. and M.S.; visualization, Ö.K.E. and M.S.; supervision, M.S. All authors have read and agreed to the published version of the manuscript." Please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: No financial resources were provided for this study.

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

References

[1] C. Czado, R. Kastenmeier, E. C. Brechmann and A. Min, "A mixed copula model for insurance claims and claim sizes", Scand. Actuar. J., vol. 4, pp. 278-305, 2012.

[2] Y. K. Tse. "Nonlife actuarial models: theory, methods and evaluation", Cambridge University Press, 2009.

[3] E. Ohlsson, B. Johansson, "Non-life insurance pricing with generalized linear models", Springer, 174, 2010.

[4] M. David, "Automobile insurance pricing with generalized linear models", Proceedings in GV- The 3rd Global Vitual Conference, 6-10 April, 2015.

[5] E. W. Frees and E. A. Valdez, "Understanding relationships using copulas", N. Am. Actuar. J., vol. 2, pp. 1-25, 1998.

[6] P. X. K. Song, "Correlated data analysis: modeling, analytics, and applications", Springer Science & Business Media, 2007.

[7] P. X. K. Song, M. Li and Y. Yuan, "Joint regression analysis of correlated data using Gaussian copulas" , Biometrics, vol. 65(1), pp. 60-68, 2009.

[8] R. Kastenmeirer, "Joint regression analysis of insurance claims and claim sizes", Diploma Thesis, Technische Universitat München, Mathematical Sciences, 2008.

[9] N. Kolev and D. Pavia, "Copula-based regression models: A survey". J Stat Plan Inference, vol. 139(11), pp. 3847-3856, 2009.

[10] A. R. De Leon and B. Wu, "Copula-based regression models for a bivariate mixed discrete and continuous outcome", Stat Med, vol. 30(2), pp. 175-185, 2011.

[11] N. Krämer, E. C. Brechmann, D. Silvestrini and C. Czado, "Total loss estimation using copula-based regression models", Insur Math Econ, vol. 53(3), pp. 829-839, 2013.

[12] S. Gschlößl and C. Czado, "Spatial modelling of claim frequency and claim size in non-life insurance" Scand, vol. 3, pp. 202-225, 2007.

[13] J. Garrido, C. Genest and J. Schulz, "Generalized linear models for dependent frequency and severity of insurance claims", Insur Math Econ, vol. 70, pp. 205-215, 2016.

[14] P. Shi, "Insurance ratemaking using a copula-based multivariate Tweedie model", Scand. Actuar. J., vol. 3, pp. 198-215, 2016.

[15] A. T. Payandeh Najafabadi, M. Qazvini, "A GLM approach to estimating copula models", Comm. Statist. Simulation Comput., vol. 44 (6), pp. 1641-1656, 2015.

[16] N. Krämer, D. Silvestrini and M. N. Krämer, Package 'CopulaRegression', 2013.

[17] A. Sklar, "Fonctions de répartition à n dimensions et leurs marges", Publications de l'Institut de Statistique de L'Université de Paris, vol. 8, pp. 229-231, 1959.

[18] R. B. Nelsen, "An introduction to copulas", Springer Science & Business Media, 2006.

[19] D. Brigo, A. Pallavicini and R. Torresetti, "Credit Models and The Crisis: A Journey Into Cdos, Copulas", Correlations And Dynamic Models, John Wiley & Sons, 2010.

[20] P. McCullagh and J. A. Nelder, "Generalized Linear Models", CRC press, 37, 1989.

[21] B. Ripley, B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, D. Firth and M. B. Ripley, Package 'mass', Cran R, 2013.

[22] U. Schepsmeier, J. Stoeber, E. C. Brechmann, B. Graeler, T. Nagler and T. Erhardt, "VineCopula: Statistical Inference of Vine Copulas", R package version 1, 2012.

[23] Ö. K. Erdemir and M. Sucu, "A comparative study on modeling of dependency between claim severity and frequency", J. Stat.: Stat and Actuar. Sci., vol. 13(1), pp. 18-29, 2020.