Fundação Getulio Vargas
Escola de Matemática Aplicada

Marcelo Orgler

Multivariate loss reserving using factor copulas

Rio de Janeiro

2020

Marcelo Orgler

Multivariate loss reserving using factor copulas

Dissertação submetida à Escola de Matemática Aplicada como requisito parcial para a obtenção do grau de Mestre em Modelagem Matemática da Informação.

Orientador: Rodrigo dos Santos Targino
Coorientador: Eduardo Fonseca Mendes

Rio de Janeiro
2020

Dados Internacionais de Catalogação na Publicação (CIP)
Ficha catalográfica elaborada pelo Sistema de Bibliotecas/FGV

Orgler, Marcelo
Multivariate loss reserving using fator copulas / Marcelo Orgler. – 2020.
41 f.

Dissertação (mestrado) -Fundação Getulio Vargas, Escola de Matemática Aplicada.
Orientador: Rodrigo dos Santos Targino.
Coorientador: Eduardo Fonseca Mendes.
Inclui bibliografia.

1. Seguros – Reserva – Modelos matemáticos. 2. Copulas (Estatística matemática). 3. Análise multivariada. 4. Administração de risco. I. Targino, Rodrigo dos Santos. II. Mendes, Eduardo Fonseca. II. Fundação Getulio Vargas. Escola de Matemática Aplicada. III. Título.

CDD – 519.535

Elaborada por Márcia Nunes Bacha – CRB-7/4403

MARCELO ORGLER

"MULTIVARIATE LOSS RESERVING USING FACTOR COPULAS".

Dissertação apresentado(a) ao Curso de MESTRADO EM MODELAGEM MATEMÁTICA do(a) ESCOLA DE MATEMÁTICA APLICADA para obtenção do grau de Mestre(a) em MODELAGEM MATEMÁTICA.

Data da defesa: 07/05/2020

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA

Presidente da Comissão Examinadora: Profº/ª **Rodrigo dos Santos Targino**

Rodrigo dos Santos Targino

Rodrigo dos Santos Targino
Orientador

p/ Rodrigo dos Santos Targino

Eduardo Fonseca Mendes
Co-Orientador

p/ Rodrigo dos Santos Targino

Yuri Fahham Saporito
Membro

p/ Rodrigo dos Santos Targino

Reinaldo Antonio Gomes Marques
Membro

Em cumprimento ao DECRETO nº 46.970 de 13/03/20 - Poder Executivo do Estado do Rio de Janeiro, DOE nº 047-A em 13/03/20, Art 4ª e Portaria MEC nº 343 de 17/03/20, DOU nº 53 de 18/03/20, que dispõe sobre a suspensão temporária das atividades acadêmicas presenciais e a utilização de recursos tecnológicos (em conformidade à legislação vigente), face ao COVID-19, as apresentações das defesas de Tese e Dissertação, de forma excepcional, serão realizadas de forma remota, inclui-se nessa modalidade membros da banca e discente.

César Camacho

César Leopoldo Camacho Manco
Diretor

Antonio de Araujo Freitas Junior
Pró-Reitor de Ensino, Pesquisa e Pós-Graduação FGV
Antonio Freitas, PhD
Pró-Reitor de Ensino, Pesquisa e Pós-Graduação
Fundação Getúlio Vargas

Instrução Normativa nº 01/19, de 09/07/19 - Pró-Reitoria FGV
Em caso de participação de Membro(s) da Banca Examinadora de forma não-presencial*, o Presidente da Comissão Examinadora assinará o documento como representante legal, delegado por esta I.N.

Acknowledgment

I dedicate this thesis to my family.

CONTENTS

ABSTRACT

This work proposes using the copula implied by a latent factor structure to model the dependence among multiple run-off triangles. The marginal distributions are modeled as GLM and estimated using maximum likelihood, whereas the parameters of the factor copula are estimated using simulated method of moments. We apply this model to a data set of two run-off triangles from large insurance companies. When compared to standard Gaussian copula models, we observe our approach better captures tail dependence between triangles, which has impact on the corresponding risk of the portfolio.

Keywords: Claims reserving, Factor copula, Tail dependence, Simulated Method of Moments.

1 INTRODUCTION

Insurers have an important social role in our society. They provide financial protection to their clients against random events. Basically, they profit by receiving more premiums than by paying claims. For many reasons, such claims are not usually settled immediately. The insurers must predict cash flows and set reserves accordingly. Those losses are referred to as outstanding losses and they are an important part of the insurers balance sheets. The process of predicting such losses and setting reserves is referred to as claims reserving.

Two well-known claims reserving methods are the Chain-Ladder (CL) and the Bornhuetter-Ferguson (BF). As Wuthrich (2013) pointed out, they are deterministic algorithms and their use was popular before the development of stochastic methods. They provide only a point estimate for the reserves, so there was a need to formulate stochastic justifications, taking into account the uncertainty in the claims reserving. Nevertheless, many of those stochastic methods follow the idea behind the deterministic CL and BL. For instance, according to Wuthrich (2013), the following stochastic models provide the CL reserves as predictors: the distribution-free CL model of Mack (1993), the over-dispersed poison of Renshaw & Verrall (1998), the bayesian CL model of Gisler & Wuthrich (2008) and the Gamma-Gamma CL model of Merz & Wuthrich (2014).

Lately there has been a growing research interest for multiple claims reserving models. According to Abdallah, Boucher & Cossette (2015) due to new regulatory standards, the insurers need to better quantify the risk of their portfolio. Many property/casualty insurers have more than one line of business (LoBS), which can be dependent on each other. For this reason, it may be interesting to take into account such dependency in order to correctly estimate the aggregate risk of their portfolio. In this study, we are interested in a particular set of dependent loss reserving methods: the ones that make use of explicit copulas.

Shi & Frees (2011) propose a model that uses copulas to model a cell-wise dependence structure among multiple run-off triangles.They also propose regression models for the marginal distributions. They perform an empirical study using the data from a major US insurer and estimate the reserves for a portfolio consisting of a personal auto and commercial auto LoBs. The authors use a lognormal regression with the identity link and a gamma regression with the inverse link for the marginal distributions. They also use a parametric bootstrap in order to introduce uncertainty in the parameters estimates. They conclude that dependencies are important when determining reserves ranges. Their approach is really flexible and allows the user to propose different regression models for

each run-off triangle and choose the dependence structure (the copula).

Copula is a popular method not only in the dependent claims reserving context, but also in other related areas such as Finance and Economics. In Finance, Oh & Patton (2017) used copulas to model dependence in a high-dimension context using copulas implied by a latent factor structure. They observe that many assets that had behaved independently prior the crisis of 2007-2008, suddenly started to move together. They also argue that some models which were used to capture dependence between financial assets were shown inadequate during the crisis because they were built under the assumption of multivariate normality, which may neglect the dependency in the tails. The authors perform an empirical study using returns of stocks in the S&P 100 index and show that factor copulas can be used in high dimensional context.

In this study, we closely follow the ideas of Shi & Frees (2011). We use their cell-wise dependence structure and their formulation for the marginal distributions. We also use Shi & Frees (2011) as the baseline for our Gaussian copula analysis. However, instead of focusing only on a Gaussian Copula, we explore the usage of factor copulas, a very flexible class of copulas analysed in Oh & Patton (2017).

We divide this thesis in 5 chapters : Background, Methodology, Results, Conclusion and Appendix. In the first chapter, we review some concepts that are important for our methodology. In the second chapter we describe our data and models. In the third chapter, we present the result of our empirical study. Finally, we present the conclusions of our study and an analysis of the sensitivity of the dependence measures used in our model estimation with respect to the sample size.

2 BACKGROUND

2.1 Claims Reserving

The aim of this section is to present basic concepts of the claim reserving problem and we use chapter 9 of Wuthrich (2013) as reference. As we have mentioned previously, the claims regarding casualty/property insurance are not usually settled in the accident date.

According to Wuthrich (2013), there is a gap between the accident date and the reporting date. The author argues that the reporting delay can be large when the accident is not immediately noticed. He presents as an example the case of asbestos-induced cancer, which can take several years to be diagnosed and reported. Thus, there might be a significant gap between the accident time and the actual report. The author also states that claims are not usually settled in the moment they are reported. He argues that the insurer sometimes has to wait for external data or judicial decisions.

From the point of view of the insurer at a certain time t, there will be claims that have already been reported but not yet settled and claims that have already been incurred but not yet reported. The author refers to the first as Reported But Not Yet settled (RBNS) and to the latter as Incurred But Not yet Reported (IBNyR). Nevertheless, such claims generate a cash flow which is usually aggregated by their accident year. The claims are usually represent in a triangle, referred to as run-off triangles.

Those triangles can be represented in the form of cumulative payments or incremental payments. Here we consider X_{ij} as the incremental payments regarding claims that occurred at accident year i and were payed at year $i + j$, where j represents the development year. As Wuthrich (2013) points out, it is usually assumed that all payments regarding accident year i are settled no later than year $i + J$. Considering the current year as I, we have $i \in \{0, 1, \ldots, \mathrm{I}\}$ and $j \in \{0, 1, \ldots, J\}$. Table 1 illustrates an incremental run-off triangle at time I.

accident year	development year							
	0	1	$\dots$	$I-i$	$\dots$			J
0	$X_{0,0}$		$\dots$		$\dots$			$X_{0,J}$
$\vdots$	$\vdots$		$\dots$		$\dots$			$\vdots$
$I-J$	$X_{I-J,0}$		$\dots$		$\dots$			$X_{I-J,J}$
$I-J+1$	$X_{I-J+1,0}$		$\dots$		$\dots$		$X_{I-J+1,J-1}$	
$I-J+2$	$X_{I-J+2,0}$		$\dots$		$\dots$	$X_{I-J+2,J-2}$		
$\vdots$	$\vdots$		$\dots$		$\iddots$			
i	$X_{i,0}$		$\dots$	$X_{i,I-i}$				
$\vdots$	$\vdots$		$\iddots$					
$I-1$	$X_{I-1,0}$	$X_{I-1,1}$						
I	$X_{I,0}$							

Table 1 – Example of incremental run-off triangle. Adapted from Wuthrich (2013).

The region in table 1 in white contains all claims paid at time I and before. We denote these values as $X_U = \{X_{ij} : i + j \leq I\}$. The grey area corresponds to the region which contains the losses we want to predict, $X_L = \{X_{ij} : i + j > I\}$. The best nominal estimate for the reserves at time I is defined as : $R = \sum_{i+j>I} E[X_{ij}|X_U]$. Notice that we do not use that same notation as Wuthrich (2013) and we consider the past claims as all information available.

2.2 Chain-Ladder algorithm

The Chain-Ladder (CL) algorithm is a method for estimating the outstanding loss liabilities. As we mentioned previously, although the CL is a deterministic algorithm, the idea behind it is the baseline for many stochastic claim reserving methods. The purpose of this chapter is to briefly describe this algorithm. For this, we follow the exposition in Wuthrich (2013).

In the last section we mentioned that the run-off triangles can also be expressed by the cumulative payments instead of incremental payments. The cumulative payments for accident year i and calendar year $i + j$ can be defined as $C_{ij} = \sum_{t=0}^{j} X_{it}$, where X_{ij} refers to the incremental payment defined in section 2.1.

According to Wuthrich (2013) the idea behind the CL model is that all accident years have similar behaviours. The development pattern in the run-off triangle are determined by the CL factors, f_j^{CL}. Thus, we have $C_{i,j+1} \approx f_j^{CL}\, C_{i,j}$. The main task in the CL algorithm is to determine f_j^{CL} for $j = 1, \dots, J$.

The CL factors estimated as :

$$\hat{f}_j^{CL} = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}} = \sum_{i=0}^{I-j-1} \frac{C_{i,j}}{\sum_{k=0}^{I-j-1} C_{k,j}} \frac{C_{i,j+1}}{C_{i,j}} \tag{2.1}$$

After computing the CL factors, one can predict $C_{i,j}$ for $i + j > I$ and hence the reserves. According to the author, these predictions are given by :

$$\hat{C}_{i,k}^{CL} = C_{i,I-i} \prod_{j=I-i}^{k-i} \hat{f}_j^{CL} \tag{2.2}$$

In the case where $k = J$, we have the ultimate claims :

$$\hat{C}_{i,J}^{CL} = C_{i,I-i} \prod_{j=I-i}^{J-i} \hat{f}_j^{CL} \tag{2.3}$$

The CL reserves at year I for accident years $i > I - J$ are defined by:

$$\widehat{\mathcal{R}}_i^{CL} = \widehat{C}_{i,J}^{CL} - C_{i,I-i} \tag{2.4}$$

The reserves aggregated over all accident years are defined by :

$$\widehat{\mathcal{R}}^{CL} = \sum_{i=I-J+1}^{I} \widehat{\mathcal{R}}_i^{CL} \tag{2.5}$$

2.3 Copula

As mentioned previously, the use of copulas have become frequent in the literature of dependent claims reserving. Not only in the claims reserving context, but also in many multivariate analysis the use of copulas is an interesting and flexible approach. This happens because the copula can describe all the dependence structure in a random vector. In our empirical study, we make use copulas. For this reason we find interesting to dedicate a section for presenting some principles of the Copula theory. We use McNeil, Frey & Embrechts (2005) and Targino (2017) as references for the first part of this section and Oh & Patton (2017) for the second part.

The definition of Copulas presented by McNeil, Frey & Embrechts (2005) is straightforward:

Definition 2.3.1 *A d-dimensional copula is a multivariate distribution function in* $[0, 1]^d$ *with uniform marginals.*

The use of Copulas gives flexibility to multivariate models. The theorem that guarantee such convenience is the Sklar's theorem. Targino (2017) enunciate the Sklar's theorem as follows:

Theorem 2.3.1 *Let F be a joint distribution function with margins $F_1, ..., F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that, for all $x_1, ..., x_d$ in $\mathbb{R} = \mathbb{R} \cup \{-\infty, \infty\}$,*

$$F(x_1, ..., x_d) = C(F_1(x_1), \ldots, F_d(x_d)) \tag{2.6}$$

If the marginals are continuous then C is unique, and given by

$$C(u_1, ..., u_d) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)) \tag{2.7}$$

Conversely, if C is a copula and $F_1, ..., F_d$ are univariate distributions, then the distribution F defined in (2.6) is a joint distribution function with marginals $F_1, ..., F_d$.

Basically, this theorem states we can build a valid multivariate distribution if we combine a copula function C and univariate continuous marginal distributions $F_1, \ldots, F_d$. Therefore, it allows the researcher to choose a model for the marginal distributions and another just for the dependence structure (Copula).

One of the most important properties from copulas is the invariance property. McNeil, Frey & Embrechts (2005) enunciate such property as :

Proposition 2.3.2 (Invariance property) *Let $(X_1, \ldots, X_d)$ be a random vector with continuous margins and copula C and let $T_1, ..., T_d$ be strictly increasing functions. Then $(T_1(X_1), \ldots, T_d(X_d))$ also has copula C.*

In our empirical analysis, one of the copulas we use to formulate a joint distribution model is the Gaussian Copula. This copula refers to the dependence structure implied by joint Normal distribution. Note that using Sklar's theorem (see 2.3.1) it is possible to derive the density of the Gaussian Copula. Consider Y as a d-dimentional random vector $(Y_1, \ldots, Y_d)$ which have distribution $N(\mu, \Sigma)$. Then, define X as $X = A(Y - \mu)$ where A is a diagonal dxd matrix with diagonal $(1/\sigma_1, \ldots, 1/\sigma_d)$, where $1/\sigma_i$ denotes the inverse of the standard deviation of Y_i, for $i = 1, \ldots, d$. We know that X has distribution $N(0, A\Sigma A')$, where $P = A\Sigma A'$ is the correlation matrix of Y. Considering the invariance property from copulas (see 2.3.2), Y and X imply the same copula. Using Sklar's theorem and equation 2.7 we can define the Gaussian copula as:

$$C_g(u_1, ..., u_d) = \Phi_{0,P}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)) \tag{2.8}$$

It's simple to obtain its density :

$$\begin{aligned} c_g(u_1, ..., u_d) &= \frac{\partial^d C(u_1, ..., u_d)}{\partial u_1 \ldots \partial u_d} \\ &= \frac{f_X(F_{X_1}^{-1}(u_1), \ldots, F_{X_d}^{-1}(u_d))}{f_{X_1}(F_{X_1}^{-1}(u_1)) \cdot \ldots \cdot f_{X_d}(F_{X_d}^{-1}(u_d))} \\ &= \frac{\phi_{0,P}(\phi^{-1}(u_1), \ldots, \phi^{-1}(u_d))}{\phi(\Phi^{-1}(u_1)) \cdot \ldots \cdot \phi(\Phi^{-1}(u_d))} \end{aligned}$$

Define q_i as $\Phi^{-1}(u_i)$ in order to simplify notation. It follows :

$$\begin{aligned} c_g(u_1, ..., u_d) &= \frac{(2\pi)^{-d/2}|P|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{q}^T P^{-1}\mathbf{q}\right\}}{\prod_{i=1}^{d}(2\pi)^{-1/2} \exp\left\{-\frac{1}{2}q_i^2\right\}} \\ &= |P|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{q}^T \left(P^{-1} - \mathbf{I}\right) \mathbf{q}\right\} \end{aligned}$$

So the Gaussian copula parameters are given by the matrix P :

$$\begin{bmatrix} 1 & \rho_{12} & \ldots & \rho_{1d} \\ \rho_{21} & 1 & \ldots & \rho_{2d} \\ \vdots & & \ddots & \vdots \\ \rho_{d1} & \ldots & \rho_{d(d-1)} & 1 \end{bmatrix}$$

Where $\rho_{ij} = \rho_{ji}$ for $i = 1, \ldots, d$ and $j = 1, \ldots, d$. $\Phi_{0,P}$ denotes the Multivariate Normal distribution with mean vector of zeros and covariance matrix Σ, ϕ denotes the standard normal density and Φ denotes the standard normal distribution.

The other type of copula we consider is a copula implied by a latent factor structure. The study that inspired the use of such copulas was Oh & Patton (2017). They use the following latent factor structures:

$$H_n = \sum_{k=1}^{K} b_{nk} Z_k + \delta_n, \quad n = 1, 2, \ldots, N \tag{2.9}$$

Where , Z_k is independent but not identically distributed (inid) with distribution F_{z_k} and δ_n is iid and has distribution F_δ.

Oh & Patton (2017) illustrate the flexibility of the copulas implied by 2.9 by presenting the scatter plot from four different joint distributions constructed using different copulas and $N(0,1)$ margins. For each copula, they consider only one common factor (K=1) and they set the variance of each latent variable equal to 1 ($\sigma_z^2 = \sigma_\delta^2 = 1$). The four copulas are :

(1) $F_z = F_\delta \sim N(0,1)$, implying a Gaussian copula, that has zero tail dependence.

(2) $F_z = F_\delta \sim t(4)$, the factor t-t copula. It presents positive and symmetric tail dependence.
(3) $F_z =$ Skew t $(\infty, -0.25)$ and $F_\delta \sim N(0,1)$, the factor skew norm-norm copula. It has asymmetric dependence but zero tail dependence.
(4) $F_z=$ Skew t$(4, -0.25)$ and $F_\delta \sim t(4)$, the factor skew t-t copula. It has asymmetric dependence and presents tail dependence.

Figure 1 was obtained from Oh & Patton (2017) and represent the scatter plots described previously.

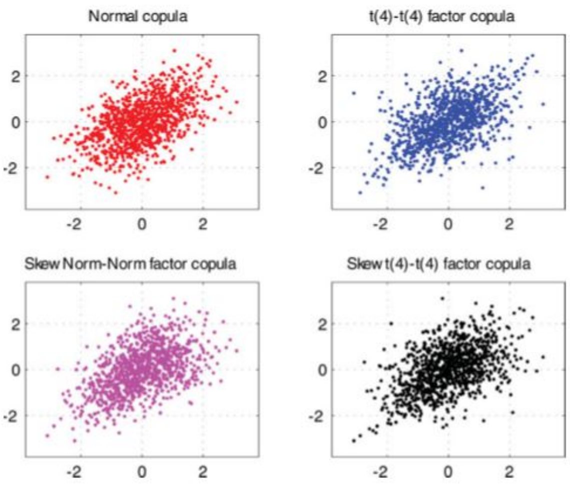

Figure 1 – Scatter plots from the bivariate distributions. It was obtained from Oh & Patton (2017).

2.4 Dependence Measures

In the previous section, we presented the definition of copula, some of its properties and some implied copulas. The objective of the current section is to present some dependence measures that are important for the estimation of our copula parameters in our empirical analysis. We use McNeil, Frey & Embrechts (2005) and Oh & Patton (2013) as references. The dependence measures we consider are: rank correlation, tail dependence and quantile dependence. Those measures are functions of the copula and do not depend on the marginal distributions.

The two main rank correlation measures are the Kendal's tau and the Spearman's rho. They are statistics that depend of the order of the data. Before defining Kendal's tau, McNeil, Frey & Embrechts (2005) present the definition of concordance between two points in $\mathbb{R}^2$. According to the authors, the points (x_1, x_2) and $(\tilde{x}_1, \tilde{x}_2)$ are concordant if $(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) > 0$ and discordant if $(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) < 0$.

Consider a random vector (X_1, X_2) and $(\tilde{X}_1, \tilde{X}_2)$ an independent copy of the for-

mer. As the authors points out, if X_1 increases along with X_2, the probability of concordance of (X_1, X_2) and $(\tilde{X}_1, \tilde{X}_2)$ should be high when compared to the probability of discordance. Analogously, if X_1 decreases when X_2 increases, the opposite is expected. The idea behind Kendal's tau is to measure the concordance of a random vector. The author states the definition of this measure as:

Definition 2.4.1 *Kendal's tau of bivariate random vector* (X_1, X_2) *:*

$$\rho_\tau(X_1, X_2) = E\left(\text{sign}\left(\left(X_1 - \tilde{X}_1\right)\left(X_2 - \tilde{X}_2\right)\right)\right)$$

Where $(\tilde{X}_1, \tilde{X}_2)$ *is an independent copy of* (X_1, X_2).

Assuming that X_1 and X_2 are continuous random variables with Copula $C(\ )$ and marginal distributions F_{X_1} and F_{X_2}, Kendal's tau can be expressed as :

$$\begin{aligned}\rho_\tau(X_1, X_2) &= 4E\left[C\left(F_{X_1}(X_1), F_{X_2}(X_2)\right)\right] - 1 \\ &= 4\int_0^1\int_0^1 C(u,v)\mathrm{d}C(u,v)) - 1\end{aligned}$$

According to Oh & Patton (2013) this measure has the following sample counterpart:

$$\hat{\rho}_\tau = \frac{4}{N}\sum_{n=1}^{N}\hat{C}\left(\hat{F}_{X_1}(x_{1n}), \hat{F}_{X_2}(x_{2n})\right) - 1 \tag{2.10}$$

In this section N stands for the sample size, $\hat{F}_{X_i}(x)$ stands for the empirical marginal distribution of X_i and $\hat{C}(u,v)$ stands for the empirical copula.

The Spearman's rho is also an important rank correlation measure. McNeil, Frey & Embrechts (2005) define this measure as:

Definition 2.4.2 *Spearman's Rho of a bivariate random vector :*

$$\rho_S(X_1, X_2) = \rho\left(F_{X_1}(X_1), F_{X_2}(X_2)\right).$$

Where, (X_1, X_2) *is a bivariate random vector and* F_{X_1}*,* F_{X_2} *are their marginal distributions and* ρ *is the Pearson correlation.*

According to definition 2.4.2, this measure is the linear correlation of the transformed variables. Assuming that X_1 and X_2 are continuous random variables, this measure can be represented as :

$$\begin{aligned}\rho_S(X_1, X_2) &= 12E\left[F_{X_1}(X_1)\,F_{X_2}(X_2)\right] - 3 \\ &= 12\int_0^1\int_0^1 uv\ dC(u,v) - 3\end{aligned} \tag{2.11}$$

According to Oh & Patton (2013) this measure has the following sample counterpart:

$$\hat{\rho}_S = \frac{12}{N}\sum_{n=1}^{N}\hat{F}_{X_1}\left(x_{1n}\right)\hat{F}_{X_2}\left(x_{2n}\right) - 3 \tag{2.12}$$

As mentioned previously, the tail dependence is also a measure that depends only on the copula. As observed by McNeil, Frey & Embrechts (2005), the main motivation for analysing the tail dependence coefficients is that they measure the strength of the dependence in the tail of the distribution. The authors define the upper and lower tail dependence as :

Definition 2.4.3 *The upper tail dependence coefficient of a bivariate random vector* (X_1, X_2) *is :*

$$\lambda_{\mathrm{u}}\left(X_1, X_2\right) = \lim_{q\to 1^-} P\left(X_2 > F_{X_2}^{-1}(q) | X_1 > F_{X_1}^{-1}(q)\right)$$

Where, F_{X_1}, F_{X_2} *are their marginal distributions.*

Definition 2.4.4 *The lower tail dependence coefficient of a bivariate random vector* (X_1, X_2) *is :*

$$\lambda_{\mathrm{l}}\left(X_1, X_2\right) = \lim_{q\to 0^+} P\left(X_2 \leq F_{X_2}^{-1}(q) | X_1 \leq F_{X_1}^{-1}(q)\right)$$

Where, F_{X_1}, F_{X_2} *are their marginal distributions.*

When we do not consider the cases where q tends to 0 or 1, those probabilities are referred to as quantile dependence. According to Oh & Patton (2013) the quantile dependence of a random vector (X_1, X_2) with marginal distributions $F_{X_1}(X_1)$, $F_{X_2}(X_2)$ can be expressed as :

$$\lambda_q(X_1, X_2) \equiv \begin{cases} P\left(F_{X_1}\left(X_1\right) \leq q | F_{X_2}\left(X_2\right) \leq q\right) = \frac{C(q,q)}{q}, \text{ for } q \in (0, 0.5] \\ P\left(F_{X_1}\left(X_1\right) > q | F_{X_2}\left(X_2\right) > q\right) = \frac{1-2q+C(q,q)}{1-q}, \text{ for } q \in (0.5, 1) \end{cases} \tag{2.13}$$

Its sample counterpart is :

$$\hat{\lambda}_q \equiv \begin{cases} \frac{1}{Nq}\sum_{n=1}^{N} 1\left\{\hat{F}_{X_1}\left(x_{1n}\right) \leq q, \hat{F}_{X_2}\left(x_{2n}\right) \leq q\right\}, \text{ for } q \in (0, 0.5] \\ \frac{1}{N(1-q)}\sum_{n=1}^{N} 1\left\{\hat{F}_{X_1}\left(x_{1n}\right) > q, \hat{F}_{X_2}\left(x_{2n}\right) > q\right\}, \text{ for } q \in (0.5, 1) \end{cases} \tag{2.14}$$

3 METHODOLOGY

3.1 Data

The data set we use in our empirical analysis has been compiled by Glenn Meyers and Peng Shi [1]. They use the claims data from National Association of Insurance Commissioners (NAIC) Schedule P in order to prepare the data set of loss triangles.

This data set contains claims for personal and commercial lines for casualty/property insurers in the US. The run-off triangles contains data from claims of accident years 1988 to 1997 and 10 development years. They obtained upper triangles data from Schedule P of the year 1997 and the lower triangles from schedule P of year 2006. It contains information from six lines of business: private passenger auto liability/medical,commercial auto/truck liability/medical, workers compensation, medical malpractice, other liability and product liability.

In the loss reserving context, it is natural for an insurer with multiple LoBs to use multivariate claims reserving methods to model dependence between their line of business and include the diversification effect in their risk analysis. Also, one can be interested in using dependent claims reserving for estimating the total reserves needed for two different companies withing the same business line. For instance, imagine a insurer is considering merging with another: they would have to estimate the risk of their joint portfolio. Nevertheless, the methods for dependent claims reserving are general and can be used in both external data or different LoBs from the same insurer. In this study, we used data from two different major personal auto insurers.

We obtain the data for the cumulative paid loss and net premiums in the 'PP Auto Data set' . Then , we transform the cumulative paid loss triangle to incremental paid loss. The incremental paid loss triangles used in our empirical analysis are presented in table 3 . The net premiums earned by insurer 1 and 2 are presented in table 2. The group code for insurer 1 and 2 are 2003 and 1767, respectively.

Insurer/year	1988	1989	1990	1991	1992
Insurer 1	1,106,097	1,195,676	1,331,317	1,517,404	1,749,378
Insurer 2	7,809,394	8,764,863	9,796,463	10,594,952	11,457,922

Insurer/year	1993	1994	1995	1996	1997
Insurer 1	1,964,229	2,104,556	2,156,649	2,170,004	2,187,056
Insurer 2	12,240,633	13,277,675	14,125,898	14,664,665	14,923,375

Table 2 – Net premiums earned in each accident year in thousands USD

[1] Available for download in $https://www.casact.org/research/index.cfm?fa=loss\_reserves\_data$

Table 3 – Incremental paid losses for insurers 1 and 2 (Thousands USD)

	Development Lag									
AY	0	1	2	3	4	5	6	7	8	9
1988	271,778	258,282	170,611	82,938	61,371	25,626	9,337	3,927	1,757	707
1989	296,828	304,995	178,348	108,590	56,209	23,148	8,047	3,056	2,927	
1990	334,331	341,697	202,556	112,094	48,326	22,054	9,804	4,675		
1991	354,530	387,048	208,552	113,348	48,730	19,007	7,160			
1992	423,048	432,495	205,477	96,655	46,835	22,140				
1993	484,849	470,218	216,864	98,809	53,992					
1994	541,361	474,967	198,627	105,175						
1995	547,102	438,923	199,275							
1996	561,687	404,475								
1997	542,021									

	Development Lag									
AY	0	1	2	3	4	5	6	7	8	9
1988	2,439,272	2,283,630	982,744	532,643	281,202	157,935	73,005	37,013	21,365	6,837
1989	2,828,267	2,539,759	1,126,571	601,780	321,492	157,945	79,403	38,023	18,837	
1990	3,186,948	2,726,542	1,227,123	634,002	321,759	154,712	74,098	39,771		
1991	3,192,619	2,685,381	1,197,254	623,205	297,945	143,661	75,745			
1992	3,561,950	2,912,341	1,289,678	640,744	312,213	159,887				
1993	3,895,076	3,129,791	1,318,550	659,703	333,979					
1994	4,323,103	3,267,841	1,333,118	716,036						
1995	4,491,070	3,173,120	1,341,923							
1996	4,444,088	3,042,025								
1997	4,344,144									

3.2 Model

We use the model proposed by Shi & Frees (2011), including their notation. Their approach is not based on the triangle of cumulative payments as in the Chain Ladder. Instead, it is focused on the triangle of incremental payments. They propose using a copula function to model the dependence between N run-off triangles, with the advantage that the copula contains all the dependence structure. Using Sklar's theorem (see 2.3.1), we know that by combining continuous marginal distributions and a copula function we achieve a valid multivariate distribution. Therefore, the use of such framework allows us to propose models for the marginal distributions and the dependence structure separately.

In this study, we set $N = 2$, since we only use two run-off triangles as data. Thus, the triangle index $n = 1, ..., N$ is limited to $n = 1, 2$, where $n = 1$ refers to the triangle from insurer 1 and $n = 2$ refers to the triangle from insurer 2. Note that, we have to add the index n to our incremental paid losses, which are denoted as $X_{ij}^{(n)}$ from now on. We denote i as 1988 $-$ the accident year, and j as the development lag. Therefore, we have $i = 0, ..., I$ and $j = 0, ..., J$. Note that I and J are equal to 9 because we have 10 accident years and development lags.

When working with multiple run off triangles, it is possible that there is a disparity in terms of scale between them. For this reason, it is natural to normalize incremental payments by an exposure variable. Thus, we follow the idea of Shi & Frees (2011) and divide the incremental payments by the net premium earned in the corresponding accident year, $Y_{ij}^{(n)} = X_{ij}^{(n)}/w_i^{(n)}$. They assume that $Y_{ij}^{(n)}$ has distribution model $F^{(n)}(y_{ij}^{(n)}; \eta_{ij}^{(n)}; \gamma^{(n)})$.

They use regression models in order to model the marginal distributions of the loss ratios $Y_{ij}^{(n)}$. We follow the idea of Shi & Frees (2011) and use accident years and development years as explanatory variables for the systematic component. Then, it can be expressed as:

$$\eta_{ij}^{(n)} = \xi^{(n)} + \alpha_i^{(n)} + \tau_j^{(n)} \tag{3.1}$$

Where, $\xi^{(n)}$ is a constant effect, $\alpha_i^{(n)}$ is the effect of the accident year i and $\tau_j^{(n)}$ is the effect of the development lag j. If we define $\beta^{(n)}$ as $(\xi^{(n)}, \alpha_0^{(n)}, \dots, \alpha_I^{(n)}, \tau_0^{(n)}, \dots, \tau_J^{(n)})$,

$\eta = (\eta_{00}, \dots, \eta_{I0})$ can be expressed in matrix form:

$$\begin{bmatrix} \eta_{00}^{(n)} \\ \eta_{01}^{(n)} \\ \vdots \\ \eta_{0J}^{(n)} \\ \eta_{10}^{(n)} \\ \vdots \\ \eta_{1(J-1)}^{(n)} \\ \vdots \\ \eta_{I0}^{(n)} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ & & & & \vdots & \vdots & & & \\ 1 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 1 & \dots & 0 & 1 & 0 & \dots & 0 \\ & & & & \vdots & \vdots & & & \\ 1 & 0 & 1 & \dots & 0 & 0 & \dots & 1 & 0 \\ & & & & \vdots & \vdots & & & \\ 1 & 0 & 0 & \dots & 1 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} \xi^{(n)} \\ \alpha_0^{(n)} \\ \vdots \\ \alpha_I^{(n)} \\ \vdots \\ \tau_0^{(n)} \\ \vdots \\ \tau_J^{(n)} \end{bmatrix} \tag{3.2}$$

It is assumed $\alpha_0^{(n)} = 0$ and $\tau_0^{(n)} = 0$ to avoid perfect multicollinearity in the regressions during the estimation of marginal specifications. Within each triangle, the loss ratios $Y_{ij}^{(n)}$ are conditionally independent given $(\eta_{ij}^{(n)}, \gamma^{(n)})$. Considering the dependence between the triangles, we use their pairwise dependence model, which means that each pair $(Y_{ij}^{(1)}, Y_{ij}^{(2)})$ has the same dependence structure given by the Copula function $C(\,.\,;\phi)$, where ϕ is the notation for dependence parameters of the Copula. This means that $Y_{ij}^{(1)}$ depends only of $Y_{ij}^{(2)}$, the corresponding cell of the other triangle. Note that, $Y_{ij}^{(1)}$ and $Y_{i'j'}^{(2)}$ are independent for any i' and j' different from i and j. This means we have an independent but not identically distributed sample $\{(Y_{ij}^{(1)}, Y_{ij}^{(2)})\}$ for i and j satisfying $i + j \leq I$ (the samples are in the upper triangle).

Using Sklar's theorem, the joint distribution model of the loss ratios $(Y_{ij}^{(1)}, Y_{ij}^{(2)})$ can be represented as :

$$F(Y_{ij}^{(1)}, Y_{ij}^{(2)}; \phi) = C(F^{(1)}(Y_{ij}^{(1)}; \eta_{ij}^{(1)}, \gamma^{(1)}), F^{(2)}(Y_{ij}^{(2)}; \eta_{ij}^{(2)}, \gamma^{(2)}), \phi) \tag{3.3}$$

We estimate two different models for the joint distribution of $(Y_{ij}^{(1)}, Y_{ij}^{(2)})$ based on the structure above. Both models have the same marginal specifications, but different Copula functions. We define as ϕ_g as dependence parameter from the Gaussian Copula and ϕ_f the dependence parameters from the factor copula. These models are:

$$F_g(Y_{ij}^{(1)}, Y_{ij}^{(2)}; \phi_g) = C_g(F^{(1)}(Y_{ij}^{(1)}; \eta_{ij}^{(1)}, \gamma^{(1)}), F^{(2)}(Y_{ij}^{(2)}; \eta_{ij}^{(2)}, \gamma^{(2)}); \phi_g) \tag{3.4}$$

and

$$F_f(Y_{ij}^{(1)}, Y_{ij}^{(2)}, \phi_f) = C_f(F^{(1)}(Y_{ij}^{(1)}, \eta_{ij}^{(1)}, \gamma^{(1)}), F^{(2)}(Y_{ij}^{(2)}, \eta_{ij}^{(2)}, \gamma^{(2)}), \phi_f) \tag{3.5}$$

Where $C_g(\,.\,, \phi_g)$ is the bivariate Gaussian Copula function and $C_f(\,.\,, \phi_f)$ denotes the Copula generated by an specific factor structure detailed in equation (3.12).

We describe the estimation procedures for the parameters $(\eta_{ij}^{(1)}, \gamma^{(1)}, \eta_{ij}^{(2)}, \gamma^{(2)}, \phi_g)$ in the next section. The estimation of ϕ_f is detailed in section (3.4). Finally, we explain the simulation procedure of the unpaid losses distributions in section (3.5).

3.3 Marginal and Gaussian Copula parameter estimation

If the Copula has closed form density, the parameters estimations can be obtained by Full Maximum Likelihood Estimation (MLE). The log-likelihood function, given our data, can be represented as :

$$\begin{aligned} L(\beta^{(1)}, \gamma^{(1)}, \beta^{(2)}, \gamma^{(2)}, \phi_g) = &\sum_{i=0}^{I}\sum_{j=0}^{I-i} \log c_g(F^{(1)}(y_{ij}^{(1)}; \eta_{ij}^{(1)}, \gamma^{(1)}), F^{(2)}(y_{ij}^{(2)}; \eta_{ij}^{(2)}, \gamma^{(2)}), \phi_g) + \\ &\sum_{i=0}^{I}\sum_{j=0}^{I-i} (\log f^{(1)}(y_{ij}^{(1)}; \eta_{ij}^{(1)}, \gamma^{(1)}) + \log f^{(2)}(y_{ij}^{(2)}; \eta_{ij}^{(2)}, \gamma^{(2)})) \end{aligned} \tag{3.6}$$

Where $c(\ . \ , \phi_g)$ denotes the gaussian copula density and $f^{(n)}(y_{ij}^{(n)}; \eta_{ij}^{(n)}, \gamma^{(n)})$ denotes the density of $Y_{ij}^{(n)}$.

It is possible to estimate the the marginal distributions and copula simultaneously. However, the MLE can be computationally costly when the dimension of the copula is large. For this reason, we chose to estimate the parameters using Inference Functions for Margins (IFM) proposed by Joe & Xu (1996) and Joe (2005). The use of such method is not new in the claims reserving context as it had already been used by Shi (2014).The IFM is a two-step estimation method. The first step consists in finding the parameters that maximize the second term of equation 3.6, the log-likelihood from the marginal distributions. The second step focus on the first term of equation 3.6. Using the estimations of the parameters from the first step in equation 3.6, we maximize the first term of the Likelihood function with respect to the dependence parameters, ϕ, from the Copula. The two steps can be summarized as :

First step : Find $\hat{\beta}^{(n)}, \hat{\gamma}^{(n)}$ for $n = 1, 2$ folllowing the definition :

$$(\hat{\beta}^{(n)}, \hat{\gamma}^{(n)}) = \underset{\beta^{(n)}, \gamma^{(n)}}{\arg\max} \sum_{i=0}^{I}\sum_{j=0}^{I-i} ln f^{(n)}(y_{ij}^{(n)}, \eta_{ij}^{(n)}, \gamma^{(n)}) \tag{3.7}$$

Second step : Let Φ_g be the Gaussian Copula parameter space. Find $\hat{\phi}_g$ by following the definition in equation (2.5).

$$\hat{\phi}_g = \underset{\phi_g \ \in \ \Phi_g}{\arg\max} \sum_{i=0}^{I}\sum_{j=0}^{I-i} \ln c(F^{(1)}(y_{ij}^{(1)}, \hat{\eta}_{ij}^{(1)}, \hat{\gamma}^{(1)}), F^{(2)}(y_{ij}^{(2)}, \hat{\eta}_{ij}^{(2)}, \hat{\gamma}^{(2)}), \phi_g) \tag{3.8}$$

For insurer 1, we assumed that $Y_{ij}^{(1)}$ has distribution $F^{(1)}(y_{ij}^{(1)}, \eta_{ij}^{(1)}, \gamma^{(1)})$, which we now specify as Gamma (k, θ_{ij}), where k is the shape parameter and θ_{ij} is the scale

parameter. We use a log-link function. Therefore, we have :

$$E(Y_{ij}^{(1)}) = k\theta_{ij} = \exp(\xi^{(1)} + \alpha_i^{(1)} + \tau_j^{(1)}) \tag{3.9}$$

$$\eta_{ij}^{(1)} = \log(k\theta_{ij}^{(1)}) \tag{3.10}$$

For insurer 2, we assumed that $Y_{ij}^{(2)}$ has distribution $F^{(2)}(y_{ij}^{(2)}, \eta_{ij}^{(2)}, \gamma^{(2)})$, which we now specify as Lognormal (μ_{ij}, σ). Note that the lognormal distribuion is not in the natural exponential family, but the $\log(Y_{ij}^{(2)})$ follows a normal distribution with mean μ_{ij} and variance σ^2, which belongs to the natural exponential family. The link function considered was the identity link. It follows that :

$$E(\log(Y_{ij}^{(2)})) = \mu_{ij} = \xi^{(2)} + \alpha_i^{(2)} + \tau_j^{(2)} \tag{3.11}$$

$$\eta_{ij}^{(2)} = \mu_{ij} \tag{3.12}$$

We denote the estimated dispersion parameters $\gamma^{(1)}$ and $\gamma^{(2)}$ as $\hat{\gamma}^{(1)} = 1/\hat{k}$ and $\hat{\gamma^{(2)}} = \hat{\sigma}^2$.

The parameters in the first step were obtained using the R package "glm2" and the parameters in the second step are obtained by using the R package "Copula". The parameters estimates and the goodness of fit tests are presented in the results section. Not that the package "glm2' does not give us the maximum likelihood estimate of the dispersion parameter. Thus, it is necessary to pass the correct argument for the dispersion parameter when obtaining the summary of the regressions. We use the function "gamma.dispersion" from the package MASS in order to obtain the MLE of the dispersion parameter in the Gamma regression. In the Lognormal regression, we multiply the estimated dispersion parameter given by "glm2" by $(55-19)/55$ (55 observations and 19 parameters) in order to obtain the MLE of the dispersion.

3.4 Factor Copula

For the factor Copula, we consider a linear, single factor structure.Particularly, our factor t-t structure is:

$$H_n = b_n Z + \delta_n, \quad n = 1, 2 \tag{3.13}$$

Where $Z \sim t(v)$, $\delta_n \sim t(v)$ independently. Also, δ_n and Z are independent for all n. The latent variables vector H has distribution $F_h = C(G_1(\phi_f), G_2(\phi_f))$, where $C(\phi_f)$ represents the implied Copula function and G_n the marginal distribution of the latent variable H_n.

Our target is to estimate $\phi_f = (b_1, b_2, v^{-1})$ and obtain Copula function implied by equation (3.6), $C(\phi_f)$. It is important to point out that we are only interested in $C(\phi_f)$ and not in the marginal distributions, $G_n(\phi_f)$. The objective is to use such Copula to construct our multivariate loss ratio distribution $C(F^{(1)}, F^{(2)}, \phi_f)$, where $F^{(n)}(Y_{ij}, \eta_{ij}^{(n)}, \gamma^{(n)})$ refers to the marginal distributions of Y_{ij} estimated using the specifications in the previous section.

As observed by Oh & Patton (2017), factor Copulas do not usually have closed form densities, making it difficult the use of MLE to estimate the copula parameters. Thus, we use the estimation method proposed by Oh & Patton (2013). Their method is similar to Simulated method of Moments, but instead of using actual distribution moments, it uses "pure" dependence measures. They refer to "pure" dependence measures as measures that are function only of the Copula and are not affected by changes in the marginal distributions. They present as examples of such measures the Kendal's tau, Spearman's rho, quantile dependence and tail dependence.

The estimator $\hat{\phi}_f$ is defined as:

$$\hat{\phi}_f = \underset{\phi_f \, \in \, \Phi_f}{\arg\min}\, g'(\phi_f)\ \hat{W}\ g(\phi_f) \tag{3.14}$$

Where $g(\phi_f) = \hat{m}_T - \tilde{m}_S(\phi_f)$ and $\hat{W}$ is a positive definite matrix. $\tilde{m}_S(\phi_f)$ denotes the vector of dependence measures calculated based on S simulations from F_h, i.e ,$\{H_s\}_{s=1}^{S}$. It is important to point out that our definition of $\hat{m}_T$ is slightly different from the Oh & Patton (2013) because we apply their estimation method in a different context. In this study, $\hat{m}_T$ denotes the vector of dependence measures calculated based on our data, in this case, the sequence of transformed loss ratios $\{(F^{(1)}(y_{ij}, \hat{\eta}_{ij}^{(1)}, \hat{\gamma}^{(1)}), F^{(2)}(y_{ij}, \hat{\eta}_{ij}^{(2)}, \hat{\gamma}^{(2)}))\}$ for all i and j satisfying $i + j \leq I$. In order to simplify our notation, we refer to the latter as $\{(U_1, U_2)_t\}_{t=1}^{T=55}$, where T is our sample size. In this study, we consider $\hat{W}$ as the identity matrix.

3.5 Simulation of Unpaid losses

After estimating all the parameters, we can simulate the predictive loss distributions. We adopt the Monte Carlo simulation procedure used by Shi & Frees (2011):

(i)For all i and j that satisfies $i + j > I$ and $n = 1, 2$, simulate a vector $(u_{ij}^{(1)}, u_{ij}^{(2)})$ from the copula.

(ii) For each cell in the lower triangle, obtain the simulated unpaid loss by:

$$y_{ij}^{(n)} = F^{(n)^{-1}}(u_{ij}^{(n)}, \hat{\eta}_{ij}^{(n)}, \hat{\gamma}^{(n)}) \tag{3.15}$$

(iii) calculate the total unpaid losses at year I by :

$$\sum_{i=0}^{I} \sum_{j=I-i+1}^{I} \sum_{n=1}^{N} \omega_i^{(n)} y_{ij}^{(n)} \tag{3.16}$$

Repeat this process to obtain the distribution of total unpaid losses at year I. Note that in order to simulate from the factor copula, we first need to simulate a vector (h_1, h_2) from $F_h(\hat{\phi}_f)$. Then, one can use the empirical distribution functions of H_n, $\hat{G}_n$, to obtain $u_{ij}^{(n)} = \hat{G}_n(h_n)$ for n= 1,2.

4 RESULTS

4.1 Data analysis

In this section, we follow the ideas of Shi & Frees (2011) and present the loss ratios development pattern. As shown in table 3 , our data is limited to 10 development years and 10 accident years for each triangle. Thus it is assumed the that the claims for accident year i will be settled within ten years. Figures 2 and 3 illustrate the dynamic of the loss ratios across the 10 development years, each line representing one accident year i. It is important to point out that we are only plotting the upper triangle, not including the predicted lower triangle. Note that for each insurer, the loss ratios present similar decreasing behavior, being close to zero as j gets close to J. It is important to point out that in this dataset, the development pattern for each accident year seems pretty similar. However, insurer 2 seems to start with higher loss ratios when compared to insurer 1. The assumption that all claims referring to accident year i will be settled within ten years seems to be reasonable. Note that Shi & Frees (2011) obtained similar development pattern for personal auto LoBs in their study.

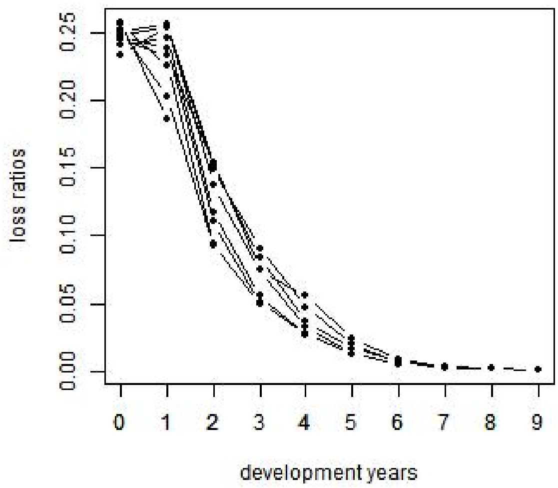

Figure 2 – Loss ratios insurer 1

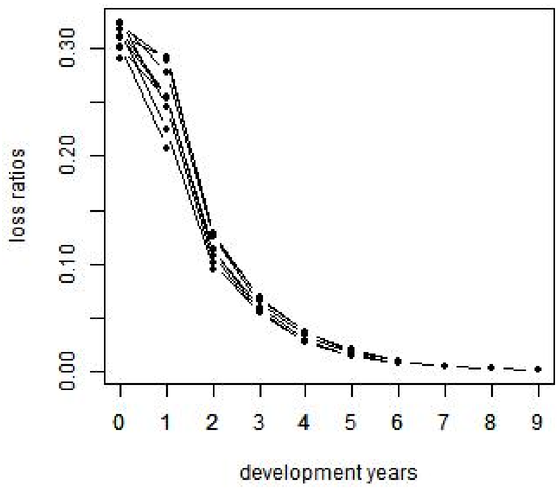

Figure 3 – Loss ratios insurer 2

Figure 4 represents the scatter plot of the loss ratios for the personal auto business line for of insurer 1 and 2. They exhibit a strong positive correlation. If we measure the risk of both insurers separately, it is possible we would underestimate the aggregated risk incurred by these insurers. This preliminary analysis is what motivates our multivariate approach.

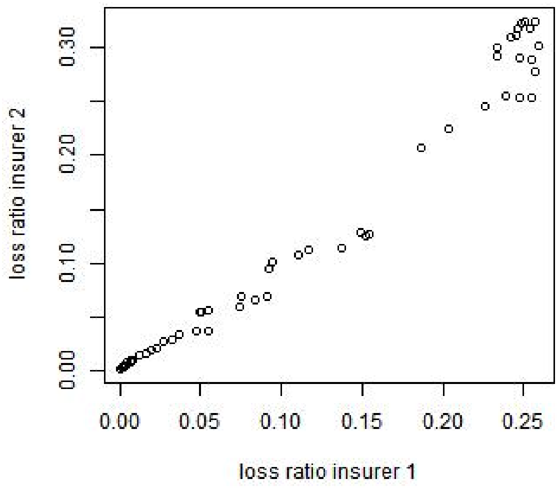

Figure 4 – Scatter plot loss ratios

4.2 Marginal Distributions

The summary of the regressions is presented in table 4. We follow the ideas of Shi & Frees (2011) and use the residuals from the gamma and lognormal regressions to perform goodness-of-fit tests. They refer to $\hat{\epsilon}_{ij}^{(1)}$ and $\hat{\epsilon}_{ij}^{(2)}$ as the residuals from the gamma and lognormal regression, respectively. They define $\hat{\epsilon}_{ij}^{(1)}$ and $\hat{\epsilon}_{ij}^{(2)}$ as :

$$\hat{\epsilon}_{ij}^{(1)} = \frac{y_{ij}^{(1)}}{\hat{\theta}_{ij}} \tag{4.1}$$

$$\hat{\epsilon}_{ij}^{(2)} = \frac{\ln(y_{ij}^{(2)}) - \hat{\mu}_{ij}}{\hat{\sigma}} \tag{4.2}$$

We assumed that $y_{ij}^{(1)}$ has distribution $\mathrm{G}(k, \theta_{ij})$, so $\frac{y_{ij}^{(1)}}{\theta_{ij}}$ has distribution Gamma$(k, 1)$. Analogously, $\frac{\ln(y_{ij}^{(2)})-\mu_{ij}}{\sigma}$ has distribution Nomal$(0, 1)$. Therefore, $\hat{\epsilon}_{ij}^{(1)}$ should have approximate distribution Gamma$(\hat{k}, 1)$ and $\hat{\epsilon}_{ij}^{(1)}$ should have approximate distribution Normal$(0, 1)$. Figures 5 and 6 illustrate the qq-plot for the residuals. They are used for comparing the empirical quantiles with the theoretical quantiles of those distributions. The closer the points are of 45−degree line, the more likely the model is well specified. Their results implies that our marginal models are indeed correctly specified.

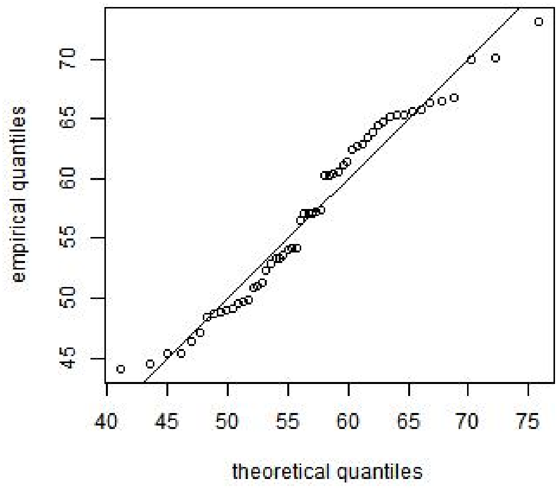

Figure 5 – Gamma QQ-plot (insurer 1)

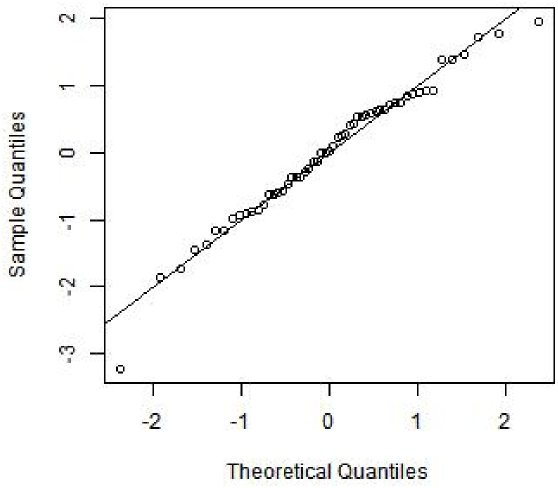

Figure 6 – Normal qq-plot (insurer 2)

Figure 7 is the scatter plot of the transformed residuals. We refer to transformed residuals by $P_1(\hat{\epsilon}_{ij}^{(1)})$ and $P_2(\hat{\epsilon}_{ij}^{(2)})$, where P_1 and P_2 are the Gamma$(\hat{k}, 1)$ and Normal$(0, 1)$ distributions, respectively. Note that these transformed residuals and $\{(U_1, U_2)_t\}_{t=1}^{T=55}$ mentioned in section 3.4 are exactly the same. This is exactly the input we use to estimate the dependence parameters for both copulas. We can interpret Figure 7 as samples from dependent uniforms distributions, where the dependence is given by a copula function.

As observed by Shi & Frees (2011), the transformed residuals purges the effects of accident years and developments lag, removing potential distortions that may affect the claims. After purging such effects, the data still contains a strong positive correlation of 0.62. Furthermore, the dependence structure in figure 4 seems to be symmetric. This means that we should choose Copulas that accommodate positive and symmetric dependence.

Figure 7 – Transformed Residuals

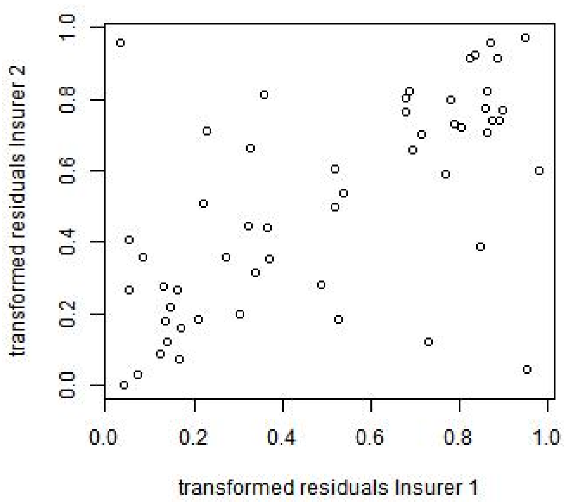

Table 4 – Marginal Estimates

	Insurer 1			Insurer 2		
Parameters	Estimate	Std. Error	t value	Estimate	Std. Error	t value
ξ	-1.16	0.07	-17.52	-0.98	0.03	-35.72
α_1	-0.03	0.06	-0.49	-0.04	0.03	-1.69
α_2	-0.07	0.06	-1.05	-0.11	0.03	-3.88
α_3	-0.25	0.07	-3.62	-0.21	0.03	-7.19
α_4	-0.32	0.07	-4.49	-0.21	0.03	-6.84
α_5	-0.34	0.08	-4.38	-0.21	0.03	-6.43
α_6	-0.35	0.08	-4.21	-0.22	0.04	-6.18
α_7	-0.36	0.09	-3.82	-0.27	0.04	-6.68
α_8	-0.32	0.11	-2.88	-0.30	0.05	-6.33
α_9	-0.24	0.15	-1.62	-0.25	0.06	-4.02
τ_1	-0.07	0.06	-1.11	-0.21	0.03	-8.00
τ_2	-0.72	0.07	-10.90	-1.04	0.03	-37.17
τ_3	-1.35	0.07	-19.51	-1.68	0.03	-57.38
τ_4	-1.99	0.07	-27.16	-2.36	0.03	-76.21
τ_5	-2.81	0.08	-36.11	-3.03	0.03	-91.97
τ_6	-3.76	0.08	-44.52	-3.73	0.04	-104.42
τ_7	-4.55	0.09	-48.41	-4.40	0.04	-110.64
τ_8	-5.03	0.11	-45.58	-5.02	0.05	-107.30
τ_9	-6.20	0.15	-41.91	-6.06	0.06	-96.05
γ	0.017	–	–	0.056	–	–

4.3 Dependence Parameter Estimates

The dependence parameters of the Gaussian copula refers to the elements in the correlation matrix, see matrix 2.3 . In this case, since $N = 2$ and the correlation matrix is symmetric we have only one parameter to estimate. Using the definition in equation 3.8, we estimate $\phi_g = 0.5615$ with standard error of 0.08.

For estimation of ϕ_f we used the following dependence measures : Spearman's rho and quantile dependence for $q = (0.2, 0.3, 0.7, 0.8)$ (see 2.14). Therefore, our vectors $\hat{m}_T$ and $\tilde{m}_S$ have dimensions 5 x 1 . Also, we impose a series of restriction in our factor structure described in equation 3.13. The restrictions we impose are : $b_1 = b_2$ and $v = 5$. The reason we make such strict assumptions is our sample size of $T = 55$. It can be considered small for the use quantile dependence statistics. For instance, with 55 observations there are only 5 observations above 90 th quantile. Futhermore, If we compute the quantile dependence for quantiles too far from the tail, we lose the information we want to capture. As a result, we end up with a restrictive model even though the factor copula structure is flexible. Our inspiration for the choice of dependence measures as moments was Oh & Patton (2017). They used rank correlation and quantile dependence with $q = (0.5, 0.1, 0.9, 0.95)$ as moments in their empirical study. However, their data consists of 696 daily returns for each of the 100 stocks in S&P 100 index. As their sample size is a lot larger than ours, we make an analysis of the sensitivity of our moments with respect to the sample size in the appendix (see 6).

We set S as $100 \cdot T$, so $S = 5500$. In order to check how our estimate changes due to different simulations, $\{(h_1, h_2)_s\}_{s=1}^{S}$, we estimated the parameter ϕ_f a 100 times for the same sample, but with different simulations. Figure 8 represents the boxplot of our estimates in this experiment. We chose to use the median of those estimates, $\hat{\phi}_f = \hat{b} = 1.27$.

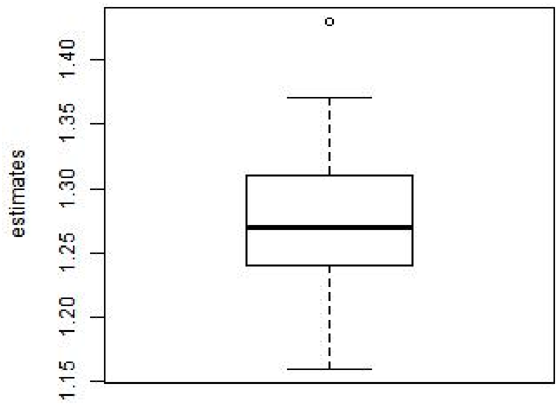

Figure 8 – Boxplot of estimates

4.4 Reserves and Unpaid Losses Distributions

At this point, we have already estimated all parameters we needed. Table 5 and 6 present the best-estimate reserves. We use the mean of the unpaid losses distribution as the best-estimate reserves.

Note that the mean of the unpaid losses distribution of all three models are very similar. The main difference seems to be the standard deviation of the total (combination of insurer 1 and insurer 2) unpaid losses distribution. It is also important to point out that all three reserves are similar the chain-ladder estimates. This result was expected since it was already observed in the study of Shi & Frees (2011).

	Gaussian copula		factor copula	
	reserves	std	reserves	std
Insurer 1	1,953,859	89,392.15	1,955,082	89,124.92
Insurer 2	12,531,968	248,875.48	12,539,480	249,544.40
Total (combined)	14,485,827	307,708.09	14,494,562	312,084.54

Table 5 – Reserves - Gaussian and Factor Copulas

	Independence copula		Chain-Ladder
	reserves	std	reserves
Insurer 1	1,953,754	90,293.59	1,964,892
Insurer 2	12,538,361	254,053.76	12,586,823
Total (combined)	14,492,115	270,501.31	14,551,715

Table 6 – Reserves - Independence Copula and ODP Cl

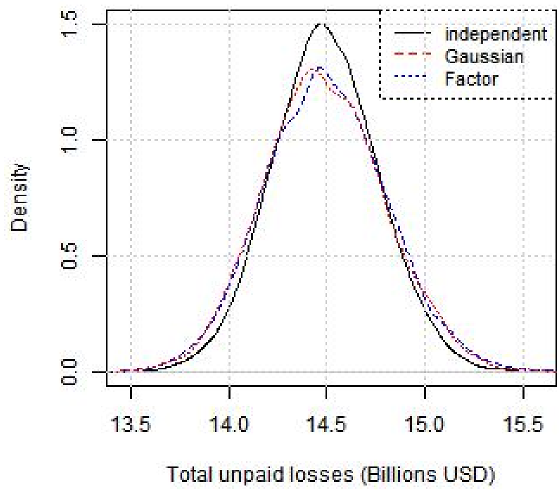

Figure 9 – Total unpaid losses distribution

Figure 9 represent the density of the total unpaid losses. It is clear that the gaussian and factor models generate distribution with fatter tails when compared to the independent case. This happens because they capture the positive dependence between the run-off triangle, leading to a more spread out distribution. This can also be noticed when comparing figures 10, 11 and 12. Each of them illustrate the scatter plot of the predicted loss distribution of insurer 1 and 2.

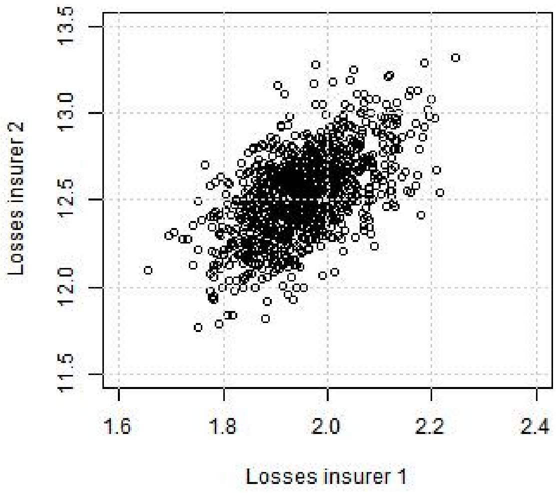

Figure 10 – figure
scatter plot of predicted loss distribution - Gaussian Copula

Figure 11 – figure
scatter plot of predicted loss distribution - factor Copula

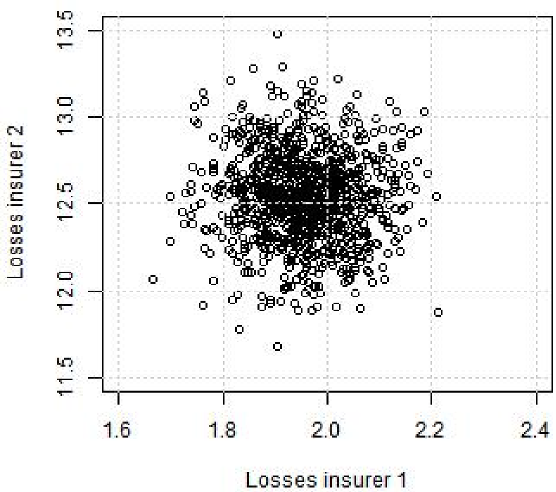

Figure 12 – scatter plot of predicted loss distribution - independence Copula

Our t-t factor copula with the specified parameters has a positive and symmetric tail-dependence. Even though both copulas are similar, the Gaussian copula has tail dependence coefficients equal to zero. Therefore, we should expect that our factor t-t copula leads to unpaid losses distribution with more mass in the tail. Figure 13 represents the qq-plot of the total unpaid losses distributions generated by the Gaussian and Factor copula. As we expected, the factor copula generates a distribution with more mass in the upper tail when compared to the Gaussian copula.

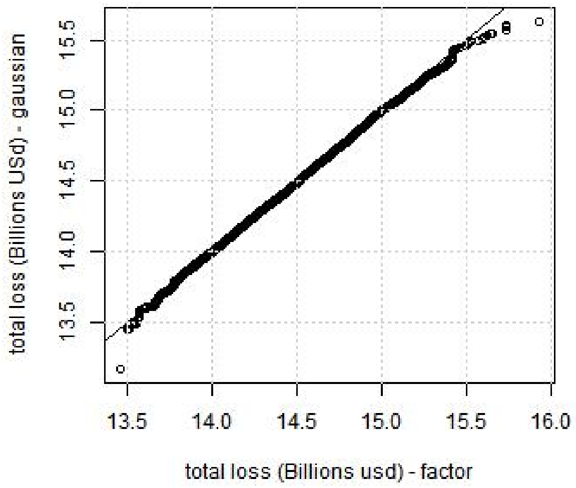

Figure 13 – QQ-plot Total unpaid losses - Gaussian x factor

4.5 Risk Capital

According to Shi & Frees (2011): "The predictive distribution of unpaid losses helps actuaries to determine appropriate reserve ranges, it is also helpful to risk managers in determining the risk capital for an insurance portfolio" . In their study, they use the risk measures Value at risk (VaR) and Expected Shortfall (ES) in order to analyse the impact of the dependency of the run-off triangles in the reserve capital allocation. The risk measures they consider are the Value at Risk and Expected shortfall. In this section, we follow their study and compute the VaR and ES to quantify the risk the total unpaid losses.

Tables 7 and 8 present the VaR and ES, respectively, of the total unpaid losses. It is easy to notice that the Independence Copula model underestimate the risk of the combined run-off triangles, since it leads to smaller VaR and ES values for all confidence levels. This result is expected, since we have already seen in the previous section that both factor and Gaussian copula models leads to a fatter total unpaid losses distribution. Notice that the VaR 1% and ES 1% are significantly higher for the factor copula model when compared to the Gaussian. Thus, the positive tail dependence of the factor copula seems to play an important role here.

	Confidence level		
Model	10%	5%	1%
Gaussian	14,883,852	15,010,253	15,224,994
Factor	14,892,806	15,013,745	15,253,051
Independence	14,840,666	14,941,982	15,136,997

Table 7 – Value at risk - in thousands USD

	Confidence level		
Model	10%	5%	1%
Gaussian	15,041,881	15,140,713	15,340,506
Factor	15,055,997	15,166,126	15,389,370
Independence	14,974,832	15,062,166	15,234,087

Table 8 – Expected Shortfall - in thousands USD

5 CONCLUSION

In our empirical analysis, we noticed that the best nominal total reserves for the Gaussian, factor and independence Copula were similar and that the dependence structure has more impact on the total unpaid losses distribution than on the point estimates of the reserves. These results were expected since Shi & Frees (2011) had similar results in their study. We also noticed that the factor copula generates a loss distribution with more mass in the upper tail when compared to the Gaussian copula. In the section 6.2, we summarize the procedure used in our empirical study.

Also, it is important to point out that according to our simulation study presented in the appendix, we noticed that the use of quantile dependence statistics as moments for the estimation of the factor copula parameters may lead to unreliable parameters estimates if the sample size is small. We observed that the confidence interval for such moments overlaps even when the parameters from the model are significantly different. For this reason, we suggest the use of other methods to estimate the parameters from the factor copula.

BIBLIOGRAPHY

ABDALLAH, A.; BOUCHER, J.-P.; COSSETTE, H. Modeling dependence between loss triangles with hierarchical archimedean copulas. *ASTIN Bulletin*, Cambridge University Press, v. 45, n. 3, p. 577–599, 2015.

GISLER, A.; WUTHRICH, M. V. Credibility for the chain ladder reserving method. *ASTIN Bulletin*, Cambridge University Press, v. 38, n. 2, p. 565–600, 2008.

JOE, H. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, v. 94, n. 2, p. 401–419, 2005.

JOE, H.; XU, J. J. *The Estimation Method of Inference Functions for Margins for Multivariate Models*. 1996. (Faculty Research and Publications).

MACK, T. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, Cambridge University Press, v. 23, n. 2, p. 213–225, 1993.

MCNEIL, A.; FREY, R.; EMBRECHTS, P. Quantitative risk management: Concepts, techniques, and tools. *Journal of The American Statistical Association - J AMER STATIST ASSN*, v. 101, 10 2005.

MERZ, M.; WUTHRICH, M. Claims run-off uncertainty: The full picture. *SSRN Electronic Journal*, 01 2014.

OH, D. H.; PATTON, A. J. Simulated method of moments estimation for copula-based multivariate models. *Journal of the American Statistical Association*, Taylor Francis, v. 108, n. 502, p. 689–700, 2013.

OH, D. H.; PATTON, A. J. Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics*, Taylor Francis, v. 35, n. 1, p. 139–154, 2017.

RENSHAW, A.; VERRALL, R. A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, Cambridge University Press, v. 4, n. 4, p. 903–923, 1998.

SHI, P. A copula regression for modeling multivariate loss triangles and quantifying reserving variability. *Astin Bulletin*, v. 44, 01 2014.

SHI, P.; FREES, E. Dependent loss reserving using copulas. *ASTIN Bulletin*, v. 41, 08 2011.

TARGINO, R. dos S. *Statistical Methods in Financial Risk Management*. Tese (Doutorado) — University College London, 2017.

WUTHRICH, M. Non-life insurance: Mathematics statistics. *SSRN Electronic Journal*, 01 2013.

6 APPENDIX

6.1 Moments Sensitivity Analysis

In order to test the sensitivity of our moments with respect to the sample size, we make the following experiment:

i) Considering the t-t factor copula structure (see 3.13) with $b_1 = b_2 = b$ and $v = 5$, we simulate 500 samples with size $T \in \{100, 1000\}$. Regarding the value of b, we make a grid ranging from -5 to 5 with step equal to 0.1.

ii) For each simulation, compute the rank correlation and quantile dependence.

iii) For each statistic considered in the previous step and for each value of b, obtain the mean, 2.5% -quantile , 97.5%-quantile and the standard deviation.

iv) For each moment plot the mean, the 95% empirical confidence interval and the 95% normal confidence interval.

Figures 14 and 15 represent the plot described above for the Spearman's Rho statistic. Note that the confidence intervals are a lot tighter when considering $T = 1000$. However, even with $T = 100$ the confidence intervals do not overlap when the values of b are not too close.

Figures 16 and 17 represent the plot described above for the quantile dependence with $q = 0.9$. When the sample size is $T = 1000$, this statistic presents tight confidence intervals, which do not overlap when the values of b are significantly different. When the sample size is $T = 100$, we can notice that the confidence intervals are large and they overlap even when the values of b are significantly different. This shows that a small sample size can be a problem for the identification of our model. In order to lessen this issue, we consider quantiles further from the tail and closer to the mean. It is also important to point out that it is not appropriate to consider a quantile too close to the mean because the information we want to capture is the strength and symmetry of the dependence in the tail of the distribution.

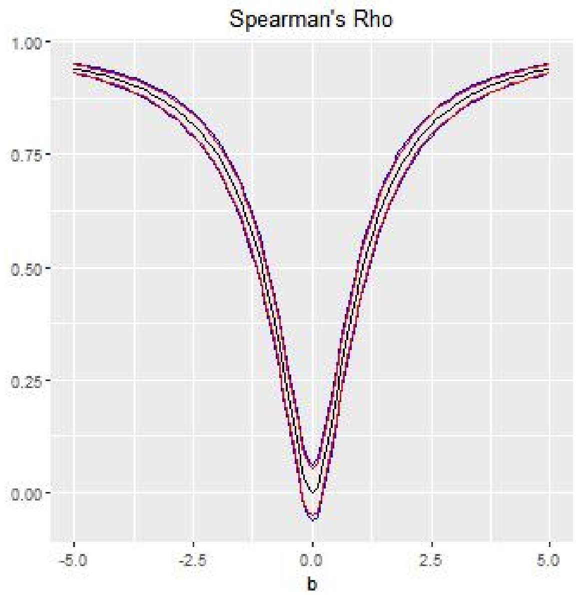

Figure 14 – The plot refering to the Spearman's rho statistic with $T = 1000$.

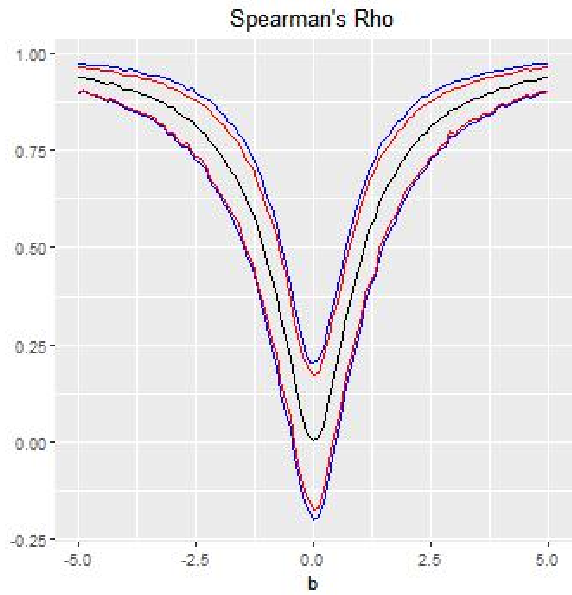

Figure 15 – The plot refering to the Spearman's rho statistic with $T = 100$.

The black line refers to the mean, the red lines refers to the 95% empirical confidence interval and the blue line refers to the 95% normal confidence interval.

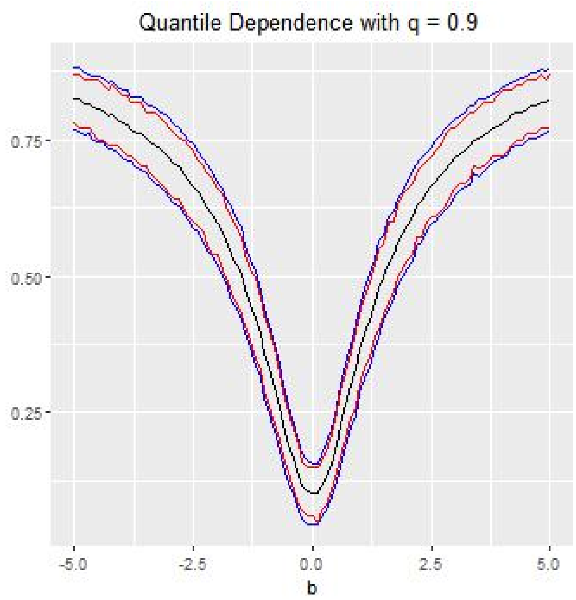

Figure 16 – Quantile dependence with q = 0.9 and T=1000

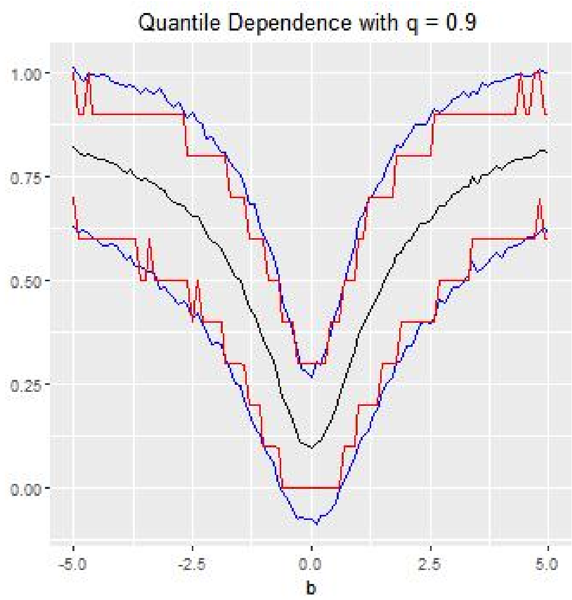

Figure 17 – Quantile dependence with $q = 0.9$ and T = 100

The black line refers to the mean, the red lines refers to the 95% empirical confidence interval and the blue line refers to the 95% normal confidence interval.

6.2 Summary

1. Compute the incremental loss ratios, $Y_{ij}^{(n)} = X_{ij}^{(n)}/w_i^{(n)}$ for $i + j \leq I$ and $n = 1, \ldots, N$.

2. For each triangle, we used a regression model to fit the marginal distribution. We considered the Gamma and Lognormal regression models. As observed by Shi & Frees (2011), both distributions are popular in the claims reserving context. The GLM is usually estimated via Maximum likelihood and is already implemented in most statistical software. It is usually assumed that the dispersion parameter is constant and known. The MLE of $\beta^{(n)}$ is computed and if the dispersion parameter is not known, the software usually gives a moment estimator by default. As mentioned already, we also computed the MLE of the dispersion parameter.

3. After obtaining the parameter for each marginal distribution, we can check the goodness-of-fit using the QQ-plot described in section 4.2.

4. At this point, we already have the estimated marginal distribution for each $Y_{ij}^{(n)}$, $F(\ ;\hat{\eta}_{ij}^{(n)}, \hat{\gamma}^{(n)})$. Therefore, we can compute $F(y_{ij}^{(n)};\hat{\eta}_{ij}^{(n)}, \hat{\gamma}^{(n)})$ for $i+j \leq I$, where $y_{ij}^{(n)}$ denotes the observation of $Y_{ij}^{(n)}$. After that, it is possible to estimate ϕ_g according to equation 3.8.

5. In this step we estimate the parameter of the factor copula model using Simulated Method of Moments. The estimation is described in section 3.4. We use a gridsearch method to solve the minimization problem.

6. At this point we have all parameters we need to simulate from the estimated joint distribution of $(Y_{ij}^{(1)}, \ldots, Y_{ij}^{(N)})$. Note that by keeping each marginal distribution fixed and changing only the copula, we have different joint distributions. We simulate the distribution of unpaid losses using the procedure described in section 3.5.

7. Using the simulations of unpaid losses, we can compute the VaR and ES risk measures.