SORT 38 (1) January-June 2014, 89-102

Testing extreme value copulas to estimate the quantile

Zuhair Bahraoui, Catalina Bolancé and Ana M. Pérez-Marín*

Abstract

We generalize the test proposed by Kojadinovic, Segers and Yan which is used for testing whether the data belongs to the family of extreme value copulas. We prove that the generalized test can be applied whatever the alternative hypothesis. We also study the effect of using different extreme value copulas in the context of risk estimation. To measure the risk we use a quantile. Our results have been motivated by a bivariate sample of losses from a real database of auto insurance claims.

MSC: MSC62-07 MSC62F05.

Keywords: Extreme value copula, extreme value distributions, quantile.

1. Introduction

Let S be the sum of k dependent random variables $X_1,\dots,X_k$, i.e. $S = X_1 + \cdots + X_k$. The distribution of S depends on the multivariate distribution, i.e. on the relationship between the random variables X_j, $j = 1,\dots,k$ (see Sarabia and Gómez-Déniz, 2008, for a review about the methods of construction of multivariate distributions). Analyzing the distribution of S is essential in finance and insurance for quantifying the risk of loss. In this regard, there are studies that have analyzed the stochastic behaviour of the sum of dependent risks and the way in which the dependency between these marginal risks may affect the total risk of loss (see, Denuit et al., 1999; Kaas et al., 2000; Cossette et al., 2002; Bolancé et al., 2008b). The aim of this paper is to analyze the test proposed by Kojadinovic et al. (2011) that allows to test whether or not our data have been generated

* bahraouizuhair@gmail.com, bolance@ub.edu and amperez@ub.edu
Department of Econometrics, Riskcenter-IREA, University of Barcelona, Av. Diagonal, 690, 08034 Barcelona, Spain.
Received: December 2013
Accepted: February 2014

by an extreme value copula. We conclude that weak convergence of the test statistic is true for any of the alternative hypothesis. Using a real data base, we have analyzed how the error in the selection of the copula can affect the risk estimate. Throughout this paper we simplify the notation to the bivariate case.

As noted by Fisher (2000), copulas are interesting for statisticians for two basic reasons: firstly, because of their application in the study of nonparametric measures of dependence and, secondly, as a starting point for constructing multivariate distributions that capture dependency structures, even when the marginals follow extreme value distributions (EVD). Also, we know that the choice of the marginals may be crucial to model the dependency behaviour of variables. According to Nelsen (2006), when coupling the marginals in the joint distribution, the copula captures the link between the two behaviours. The relationship between the joint distribution and the marginals is established in the fundamental theorem proposed by Sklar (1959). This theorem shows that a bivariate cumulative distribution function (CDF) H of a random vector of variables (X_1, X_2) with marginal cumulative distribution functions (CDFs) F_1 and F_2 includes a copula C according to the following expression:

$$H(x_1,x_2) = C(F_1(x_1),F_2(x_2)) \forall x_1,x_2 \in \mathbb{R}. \quad (1)$$

Due to the fact that the joint distribution (and therefore the dependency structure) is unknown, specific tests for choosing the best copula are necessary. This has been the motivation for developing tests for the adequacy of copulas. It is worth mentioning the paper by Genest and Rivest (1993) on inference for bivariate Archimedean copulas, the test proposed in Scaillet (2005) on the positive quadrant dependence hypothesis and, finally, the test of symmetry in bivariate copulas introduced in Quessy et al. (2012).

Regarding the inference for extreme value copulas, we can mention the test proposed in Genest et al. (2011) based on a Cramér-von Mises statistic and the test analyzed in Ghorbal et al. (2009) based on an U-statistic. However, Kojadinovic et al. (2011) uses the $max - stable$ property to test the adequacy of an extreme value copula that is also based on the Cramér-von Mises statistic. In our study we find a similar result for the bivariate case and we obtain the weak convergence of the statistic proposed in the general case.

In Section 2, first, we present our main result and, second, we describe three examples of copulas which are extreme value copulas: Gumbel, Galambos and Hüsler-Reiss. In Section 3 we describe a real database of auto insurance claims which we use in the empirical application. In Section 4 we report the results of our empirical study, firstly we apply the test described in Section 2 and, secondly, we calculate the quantile using different extreme value copulas and compare these results with those obtained when using a widely known non extreme value copula, such as a Gaussian copula. We use two alternative marginal distributions and we compare them: the log-normal, that is a EVD Type I (Gumbel), and the Champernowne distribution, which converges to a Pareto in the tail and therefore is an EVD Type II (Frechet). We also note that the Champernowe distribution looks more like a log-normal near 0. We conclude in Section 5.

2. Test for extreme value copulas

We know that the class of extreme value copulas corresponds to the class of $max-stable$ copulas (see, for example, Segers, 2012). A copula is $max-stable$ if for every positive real number r and all u_1, u_2 in $[0,1]$, $C(u_1,u_2) = C^r(u_1^{1/r}, u_2^{1/r})$. Then we formulate the null hypothesis and its alternative as:

$$\begin{cases} H_0^r : C(u_1,u_2) = C^r(u_1^{1/r}, u_2^{1/r}), & \forall u_1, u_2 \in [0,1], \forall r > 0 \\ H_1^r : C(u_1,u_2) \neq C^r(u_1^{1/r}, u_2^{1/r}), & \exists u_1, u_2 \in [0,1], \exists r > 0 \end{cases}.$$

Specifically we need to test the $max-stable$ hypothesis,

$$\begin{cases} H_0 : \bigcap_{r>0} H_0^r \\ H_1 : \bigcup_{r>0} H_1^r, \end{cases}$$

in practice we only can test H_0^r for some values of r. From Kojadinovic et al. (2011), it seems that $r < 1$ is not so good, so they consider only values of r greater than 1.

Let (X_{i1}, X_{i2}), $\forall i = 1,\ldots,n$ be a bivariate sample of n independent and identically distributed observations. We consider the functions:

$$\mathbb{D}_n^r(u_1,u_2) = \sqrt{n}\left(C_n(u_1,u_2) - C_n^r(u_1^{1/r}, u_2^{1/r})\right)$$
$$\mathbb{D}^r(u_1,u_2) = \sqrt{n}\left(C(u_1,u_2) - C^r(u_1^{1/r}, u_2^{1/r})\right),$$

where $C_n(u_1,u_2)$ is the empirical copula defined as:

$$C_n(u_1,u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}(\hat{F}_{1_n}(X_{i1}) \leq u_1, \hat{F}_{2_n}(X_{i2}) \leq u_2), \quad u_1,u_2 \in [0,1]^2, \qquad (2)$$

where $I(\cdot)$ is an indicator function that takes value 1 if the condition in brackets is true and 0 otherwise. $\hat{F}_{1_n}$ and $\hat{F}_{2_n}$ are the empirical marginal cumulative distribution functions. To test the $max-stable$ property we need to analyze if we can use $\mathbb{D}_n^r(u_1,u_2)$ as an estimator of $\mathbb{D}^r(u_1,u_2)$. Then we find the convergence to a Gaussian process of the difference $\mathbb{D}_n^r(u_1,u_2) - \mathbb{D}^r(u_1,u_2)$.

We use the result by Fermanian et al. (2004) for the weak convergence of the empirical copula process C_n to a Gaussian process $\mathbb{G}$ in the space of all bounded realvalued functions on $[0,1]^2$, i.e. $l^\infty([0,1]^2)$, which is expressed as follows:

$$\sqrt{n}\,(C_n(u_1,u_2) - C(u_1,u_2)) \rightsquigarrow \mathbb{G}(u_1,u_2) \tag{3}$$
$$= \mathbb{B}(u_1,u_2) - \partial_1 C(u_1,u_2)\mathbb{B}(u_1,1) - \partial_2 C(u_1,u_2)\mathbb{B}(1,u_2), \tag{4}$$

where $\partial_j C(u_1,u_2), j = 1,2$ are the partial derivatives of the function C respect to u_j and $\rightsquigarrow$ indicates weak convergence and $\mathbb{B}$ is a Brownian bridge on $[0,1]^2$ with covariance functions:

$$E[\mathbb{B}(u_1,u_2)\mathbb{B}(u_1',u_2')] = C(u_1 \wedge u_1', u_2 \wedge u_2') - C(u_1,u_2)C(u_1',u_2'),$$

where $\wedge$ is the minimum.

Proposition 1 *If the partial derivatives of a copula $C(u_1,u_2)$ are continuous then for any $r > 0$ we have:*

$$\mathbb{D}_n^r(u_1,u_2) - \mathbb{D}^r(u_1,u_2) \rightsquigarrow \mathbb{C}^r(u_1,u_2) = \mathbb{G}(u_1,u_2) - rC^{r-1}(u_1^{1/r},u_2^{1/r})\mathbb{G}(u_1^{1/r},u_2^{1/r}), \tag{5}$$

in $l^\infty([0,1]^2)$. The result in (5) is true under H_0^r and H_1^r.

Kojadinovic et al. (2011) proved the weak convergence under H_0^r of $\mathbb{D}_n^r(u_1,u_2)$ towards the same process defined in Proposition 1. We have proved that the weak convergence of the difference $\mathbb{D}_n^r(u_1,u_2) - \mathbb{D}^r(u_1,u_2)$ is true under H_0^r and H_1^r.

Proof 1 In order to prove the result in Proposition 1 we consider the function:

$$\Gamma : C(u_1,u_2) \longrightarrow \Gamma(C(u_1,u_2)) = C^r(u_1^{1/r},u_2^{1/r}), r > 0.$$

Γ is a differentiable function as defined by Hadamard (see, Ren, 1995). We use the Delta functional method to analyze the weak convergence of $\Gamma(C(u_1,u_2)) = C^r(u_1^{1/r},u_2^{1/r})$. To find the Hadamard derivative of Γ that is denoted by Γ', we consider the function:

$$h(t) = \Gamma((C+t\Delta)(u_1,u_2)) - \Gamma(C(u_1,u_2))$$
$$= (C+t\Delta)^r(u_1^{1/r},u_2^{1/r}) - C^r(u_1^{1/r},u_2^{1/r}),$$

where $t\Delta$ is a function representing a difference, namely, t is a real value and Δ is a fixed perturbation. Then we calculate Γ' as the derivative of function h at $t = 0$. Namely, $\Gamma'(\Delta)$ if the first derivative of function $\Gamma(C(u_1,u_2)) = C^r(u_1^{1/r},u_2^{1/r})$ with respect to t evaluated at $t = 0$.

Using the expression of the Pascal triangle:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k,$$

we obtain that:

$$\begin{aligned} h(t) &= \sum_{k=0}^{r} \binom{r}{k} C^{r-k}(u_1^{1/r}, u_2^{1/r}) t^k \Delta^k(u_1^{1/r}, u_2^{1/r}) - C^r(u_1^{1/r}, u_2^{1/r}) \\ &= \binom{r}{0} C^r(u_1^{1/r}, u_2^{1/r}) + \binom{r}{1} C^{r-1}(u_1^{1/r}, u_2^{1/r}) t \Delta(u_1^{1/r}, u_2^{1/r}) \\ &+ \sum_{k=2}^{r} \binom{r}{k} C^{r-k}(u_1^{1/r}, u_2^{1/r}) t^k \Delta^k(u_1^{1/r}, u_2^{1/r}) - C^r(u_1^{1/r}, u_2^{1/r}). \end{aligned}$$

If we differentiate at $t = 0$, we obtain:

$$\frac{\partial h(t)}{\partial t} |_{t=0} = \Gamma'(\Delta) = rC^{r-1}(u_1^{1/r}, u_2^{1/r}) \Delta(u_1^{1/r}, u_2^{1/r}).$$

The result in Proposition 1 is obtained by observing that:

$$\mathbb{D}_n^r(u,v) - \mathbb{D}^r(u,v) = \sqrt{n}\left((C_n(u_1,u_2) - C(u_1,u_2)) - (C_n^r(u_1^{1/r}, u_2^{1/r}) - C^r(u_1^{1/r}, u_2^{1/r})) \right).$$

Using the convergence of the empirical copula given by Fermanian et al. (see Fermanian et al. (2004)) we obtain:

$$\sqrt{n}\,(C_n(u_1,u_2) - C(u_1,u_2)) \rightsquigarrow \mathbb{G}(u_1,u_2),$$

and, finally, applying the Delta functional method, we obtain:

$$\sqrt{n}\left(C_n^r(u_1^{1/r}, u_2^{1/r}) - C^r(u_1^{1/r}, u_2^{1/r}) \right) \rightsquigarrow \Gamma'(\mathbb{G}(u_1,u_2)).$$

□

Under the hypothesis H_0 we have that $\mathbb{D}^r(u_1,u_2) = 0$ and in this case $\mathbb{D}_n^r(u_1,u_2)$ weakly converges to process (5).

For hypothesis testing given a fixed r, we use a Cramér-von Mises statistic:

$$S_n^r = \int_0^1 \int_0^1 (\mathbb{D}_n^r(u_1,u_2))^2 \, du_1 du_2. \tag{6}$$

As proposed by Kojadinovic et al. (2011) for a range of values of r, $r_1,\dots,r_p$, the following statistic can be considered:

$$S_n^{r_1,\dots,r_p} = \sum_{i=1}^{p} S_n^{r_i}. \tag{7}$$

To calculate the critical values we use the method proposed by Van der Vaart (2000), consisting on generating independent copies of S_n^r. The procedure is as follows:

1. If $\partial_j C(u_1,u_2), j=1,2$ are continuous on $[0,1]^2$ then N independent copies of $\mathbb{D}_n^r$, $\mathbb{D}_n^{r,(1)},\dots,\mathbb{D}_n^{r,(N)}$ can be generated, such that

$$(\mathbb{D}_n^r, \mathbb{D}_n^{r,(1)}, \dots \mathbb{D}_n^{r,(N)}) \rightsquigarrow (\mathbb{D}^r, \mathbb{D}^{r,(1)}, \dots \mathbb{D}^{r,(N)}),$$

   where $\mathbb{D}^{r,(1)},\dots,\mathbb{D}^{r,(N)}$ are independent copies of $\mathbb{D}^r$.

2. If $\partial_j C(u_1,u_2), j=1,2$ are continuous on $[0,1]^2$ then, $(S_n^{r,(1)}, S_n^{r,(2)},\dots,S_n^{r,(N)})$ can be calculated by using a numerical approximation of formula (6) (see, Kojadinovic et al., 2011), then:

$$(S_n^r, S_n^{r,(1)}, S_n^{r,(2)}, \dots, S_n^{r,(N)}) \rightsquigarrow (S^r, S^{r,(1)}, S^{r,(2)}, \dots, S^{r,(N)}),$$

   where $(S^{r,(1)}, S^{r,(2)},\dots,S^{r,(N)})$ are independent copies of S^r.

3. Obtain the p-value as:

$$\frac{1}{N}\sum_{k=1}^{N} \boldsymbol{I}(S_n^{r,(k)} \geq S_n^r).$$

The Van der Vaart method is implemented in the software R with the function `evTestC()` included in the package `copula` (see, Hofert et al., 2013).

2.1. Three examples of extreme value copulas

In the application presented in next section, we compare three examples of extreme value copulas: Gumbel, Galambos and Hüsler-Reiss, which are described in this section.

The functional form of Gumbel copula (see, Gumbel, 1958) is given by:

$$C_\theta(u_1,u_2) = \exp\left(-\left[(-\ln(u_1))^\theta + (-\ln(u_2))^\theta\right]^{1/\theta}\right),$$

where $\theta \in [1, +\infty)$ is the parameter controlling the dependency structure. Note that, the dependence is perfect when $\theta \to \infty$, while independence corresponds to the case when $\theta = 1$. For the Gumbel copula, it is well known that lower tail dependence is $\lambda_L = 0$ and upper tail dependence is $\lambda_U = 2 - 2^{\frac{1}{\theta}}$, i.e. the Gumbel copula has upper tail dependence.

The Galambos copula was proposed by Galambos (1975) and has the following form:

$$C(u_1,u_2) = u_1 u_2 \exp\left(\left[(-\ln(u_1))^{-\theta} + (-\ln(u_2))^{-\theta}\right]^{-1/\theta}\right),$$

where the range of θ is $[0,\infty)$ and the upper tail dependence is $\lambda_U = 2 - 2^{\frac{1}{\theta}}$.

Another example of extreme value copulas is the Hüsler-Reiss copula that was developed by Hüsler and Reiss (1989). Its functional form is given by:

$$C(u_1,u_2) = \exp\left(-\hat{u_1}\Phi\left[\frac{1}{\theta} + \frac{1}{2}\theta\ln\left(\frac{\hat{u_1}}{\hat{u_2}}\right)\right] - \hat{u_2}\Phi\left[\frac{1}{\theta} + \frac{1}{2}\theta\ln\left(\frac{\hat{u_2}}{\hat{u_1}}\right)\right]\right),$$

where the range of θ is $[0,\infty)$ and Φ is cdf of the standard Gaussian, $u_1 = -\ln(\hat{u_1})$ and $u_2 = -\ln(\hat{u_2})$.

3. The data

Our example is motivated by a problem in the context of insurance. We assume that when there is an accident, the total cost to be paid to a policyholder is the sum of two components: (1) the material damage and (2) the bodily injury compensation. The insurance company is interested in evaluating the risk of a given claim exceeding a certain amount. So the right-tail quantiles are important to understand the risk that an accident claim is very costly.

We work with a random sample of 518 observations containing two types of costs: Cost1, representing property damages and compensation of the loss, and Cost2, which corresponds to the expenses related to medical care and hospitalization. In general, the cost of bodily injuries is covered by the National Institute of Health, however the insured has to bear the cost of some medical expenses and rehabilitation, technical assistance, drugs, etc., including compensation for pain, suffering and loss of income.

Bodily injury claims typically take years to be settled. Nevertheless, all the claims in our sample were already settled in 2002, according to the company, (see, Bolancé et al., 2008b). Finally, we should mention that the compensation may include payments to third parties that have been damaged in one way or another.

In Table 1 we summarize the descriptive statistics of the sample for Cost1, Cost2 and the Total Cost. The variables Cost1 and Cost2 are always positive, and there is a big difference between the corresponding maximum and minimum values. Furthermore, we observe that the variables described in Table 1 have right skewness. In Figure 1 we show the histograms representing the shape of the distributions associated with the variables Cost1 and Cost2.

The K-Plot (related to Kendall Plot, see, Genest and Boies, 2003) is a visual method that allows us to analyze in a descriptive way if our bivariate data have been generated by an extreme value copula. In Figure 2 we show the K-Plot, that compare the order

Table 1: *Descriptive statistics.*

Cost	Average	Std.Dev.	Skewness	Min	Max	Median
Cost1	182.80	686.80	15.65	13.00	137900.00	677.00
Cost2	283.92	863.17	8.04	1.00	11855.00	88.00
Total Cost	211.20	752.00	15.27	32.00	149800.00	789.00

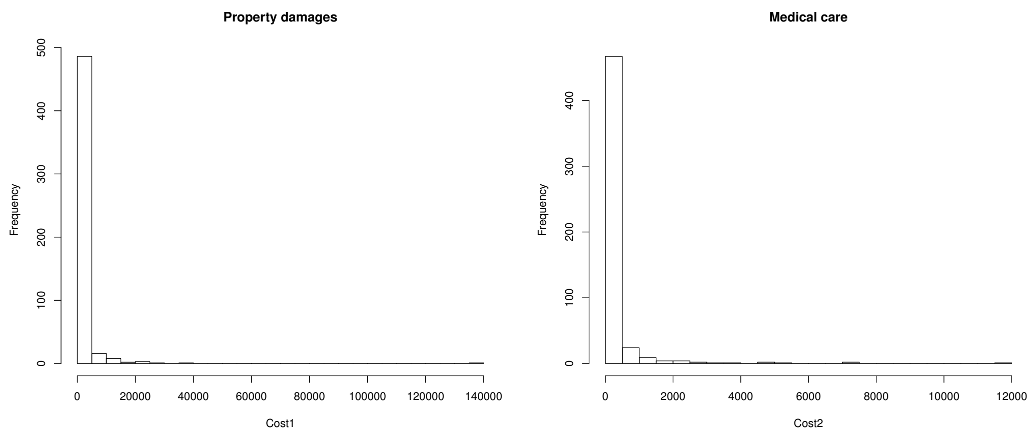

Figure 1: *Histograms.*

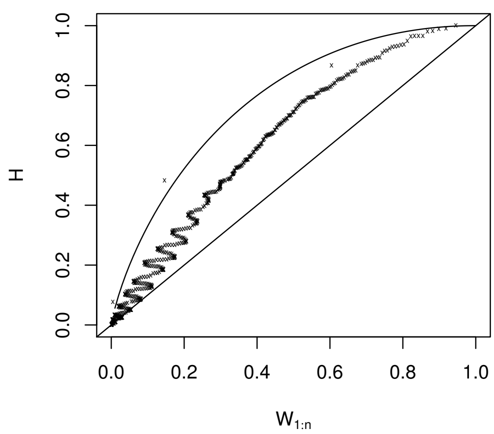

Figure 2: *K-Plot associated with to copula of (Cost1, Cost2).*

in real data (H, pseudo-observations generated by the bivariate empirical distribution) with the order supposing that the data have been generated by the independence copula (W, expected pseudo-observations). We note that costs have a positive association (as shown in the values of the K-plot above the diagonal, which indicates independence). Almost all points are between the straight line and the boundary curve indicating perfect positive dependence. It seems that for larger values of W, the data are closed to the case of a perfect positive dependence. This means that the higher the severity of the claim, the higher is the correlation between the medical costs and compensation.

4. Results

In this section we report the results that we have obtained in an empirical application of the methodology that we have presented. In order to estimate the total risk of loss, our goal is to determine the dependency structure between the data corresponding to a sample of claims provided by a major insurance company which operates in Spain. To test if our data are generated by an extreme value copula we calculate the value of the Cramér-Von Mises statistic in (7), firstly with $r = 3,4,5$. We have estimated the significance level of the test statistic using the method proposed by Van der Vaart (2000). In total, we generated 1000 independent copies of $S_n^{3,4,5}$. The results are shown in Table 2 and allow us to conclude that the analyzed bivariate data are generated by an extreme value copula.

Table 2: *Cramér-Von Mises statistic.*

Statistic	Estimation	p-value
$S_n^{3,4,5}$	0.2680	0.1773

Table 3: *Copula estimation results.*

	Gaussian	t-Student*	Gumbel	Galambos	Hüsler-Reiss
Parameters	0.5905	0.5981	1.7397	1.0208	1.4946
Standard Errors	0.02485	0.02718	0.07538	0.07689	0.09059
AIC	−212.3695	−217.0000	−246.3839	−243.3305	−237.8542
BIC	−208.1195	−208.5000	−242.1339	−239.0805	−233.6042
CIC	−208.1195	−208.5000	−242.1339	−239.0805	−233.6042
Kendall Tau = 0.4252. *d.f. = 9.6442					

We estimate the parameters of the three extreme value copulas described in Section 2.1: Gumbel, Galambos and Hüsler-Reiss.

In Table 3 we show the estimated parameters for these three copulas together with those obtained for the Gaussian and the t-Student copulas. To estimate the dependence parameter of Gaussian, Gumbel, Galambos and Hüsler-Reiss copulas we have used the inversion of Kendall's tau method (Itau). To estimate the dependence parameter and the degree of freedom of the t-Student copula we have used maximum likelihood estimation (MLE). For selecting the copula we have used two known statistical information criterion, the Akaike Information Criterion $AIC = -2\log L(\theta) + 2k$ and the Bayesian Information Criterion $BIC = -2\ln L(\theta) + k\ln(n)k$, where k is the number of parameters to be estimated and L the value of the likelihood function. Also, we have used the copula information criterion *CIC* propose by Gronneberg and Hjort (2014). The corresponding results are presented in Table 3. We observe that BIC and CIC values are very similar and we conclude that the Gumbel copula is the one that best reflects the dependence structure of our data.

Once the dependency structure is estimated, the next step is to estimate the marginal distribution functions. Considering the histograms in Figure 1, we chosed to use two EVD. Namely, we compare the log-normal distribution, that is a EVD Type I (Gumbel), with the modified Champernowne distribution[1], which converges to a Pareto in the tail and therefore it is an EVD Type II (Frechet); besides the Champernowe distribution looks more like a log-normal near 0. Furthermore, the Champernowne distribution have been analyzed in the context of semiparametric estimation of EVD (see, for example, Bolancé, 2010; Bolancé et al., 2008a; Alemany et al., 2013). In Table 4 we show the results for the maximum likelihood estimation of the marginal distributions. We can see that for Cost1, Log-normal and Champernowne have similar AIC and BIC, however for Cost2 Champernowne provides lower values of AIC and BIC.

Table 4: *Maximum likelihood estimation of marginal distributions.*

	Log-normal	Champernowne
CDFs	$\int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \ x \geq 0$	$\frac{(x+c)^\delta - c^\delta}{(x+c)^\delta + (H+c)^\delta - 2c^\delta}, \ x \geq 0$
X_1 = Cost1	$\mu = 6.4437, \sigma = 1.3349$, $AIC = 8448.8950$ and $BIC = 8452.7190$	$\delta = 1.3271, H = 677, c = 0$ $AIC = 8448.163$ and $BIC = 8453.899$
X_2 = Cost2	$\mu = 4.3755, \sigma = 1.5189$, $AIC = 9425.1340$ and $BIC = 9428.9590$	$\delta = 1.1622, H = 88, c = 0$ $AIC = 6443.7150$ and $BIC = 6449.4510$

1. The cdf of the modified Champernowne distribution is:

$$F(x) = \frac{(x+c)^\delta - c^\delta}{(x+c)^\delta + (H+c)^\delta - 2c^\delta}, \ x \geq 0,$$

with parameters $\delta > 0$, $H > 0$ and $c \geq 0$. The estimation of transformation parameters is performed using the maximum likelihood method described in Buch-Larsen et al. (2005).

For evaluating the risk of total loss we estimate the quantile of S at confidence level α ($q_\alpha(S)$). We use the Monte Carlo simulation method and the procedure is as follows:

1. We generate the pseudo-random sample $\left(\hat{U}_{1i}, \hat{U}_{2i}\right)$, $\forall i = 1, \ldots, r$, from the bivariate copulas whose estimated parameters are shown in Table 3.
2. Using the inverse of the marginal CDFs we calculate $\left(\hat{X_{1i}} = F_1^{-1}(\hat{U}_{1i}), \hat{X_{2i}} = \right.$ $\left. = F_2^{-1}(\hat{U}_{2i})\right)$, $\forall i = 1, \ldots, l$, where the sample volume l is large.
3. We calculate $\hat{S}_i = \hat{X}_{1i} + \hat{X}_{2i}$, $\forall i = 1, \ldots, l$ and we estimate $q_\alpha(S)$ empirically from the generated pseudo-sample. We generate $l = 10,000$ samples.

In Table 5 we show the results of the estimations of q_α for $\alpha = 0.95, 0.99, 0.995$, 0.999. On the first row of Table 5 we provide the empirical values of the $q_\alpha(S)$ calculated with the 518 observations in the sample of the aggregate loss $S = X_1 + X_2$ for different confidence levels α; below we show the same $q_\alpha(S)$ that have been estimated by the Monte Carlo simulation method for the five copulas considered here. We note the importance of using an extreme value copula and extreme value marginal distributions when the data indicate this behaviour.

Table 5: *Quantiles of total loss.*

α	0.95	0.99	0.995	0.999
Empirical	7905.6000	24821.1400	28420.8700	92112.9300
		Log-normal		
Normal	6635.427	15628.804	20762.765	39733.894
t-Student	6547.524	16638.175	22521.175	39547.101
Gumbel	6432.017	15464.969	22011.382	40001.210
Galambos	6429.160	15471.400	22066.000	39925.670
Hüsler-Reiss	6421.028	15465.126	22122.110	39841.559
		Champernowne		
Normal	7237.591	25504.175	38682.444	110082.261
t-Student	7302.165	25740.933	42223.504	117447.015
Gumbel	7264.831	23944.798	41461.743	119401.409
Galambos	7253.166	24056.946	41409.717	118982.012
Hüsler-Reiss	7241.504	24103.038	41107.537	118539.744

In Table 5 we show that by using log-normal marginal distributions, the estimated quantile is below the empirical quantile for the five copulas considered here. Therefore, the risk is underestimated. We also note that the selected copula does not have much influence on the risk estimation. However, if we use Champernowne marginal distributions, which has a heavier right tail than log-normal distribution, the influence of the selected copula is not significant at lower confidence levels (0.95 and 0.99) but it is significant for extreme confidence levels (0.995 and 0.999). As indicated by the goodness of fit measures for our data, the best selection is the Gumbel copula with Champernowne marginal distributions.

5. Conclusions

The test we have introduced for the adequacy of extreme value copulas allows us to determine the suitable copula, especially when the data have extreme values.

In our empirical application, the K-Plot identified a positive and increasing dependence between variables related to automobile insurance claims, and the new test we presented for extreme value copulas confirms that, in our case, we should use an extreme value copula.

In the selection of the marginal distribution we have considered a modified Champernowne distribution. It provides interesting results, due to its similarity to the log-normal distribution for low values of the variable and, additionally, due to its convergence to a Pareto distribution in the right tail.

When the marginal distributions have heavy right tail, as is the case with the Champernowne distribution and if the aim is to estimate extreme quantiles, the results show the importance of testing the adequacy of an extreme value copula to the data.

References

Alemany, R., Bolancé, C. and Guillén, M. (2013). A nonparametric approach to calculating value-at-risk. *Insurance: Mathematics and Economics*, 52, 255–262.

Bolancé, C. (2010). Optimal inverse beta(3,3) transformation in kernel density estimation. *SORT-Statistics and Operations Research Transactions*, 34, 223–237.

Bolancé, C., Guillén, M. and Nielsen, J. (2008a). Inverse beta transformation in kernel density estimation. *Statistics & Probability Letters*, 78, 1757–1764.

Bolancé, C., Guillén, M., Pelican, E. and Vernic, R. (2008b). Skewed bivariate models and nonparametric estimation for cte risk measure. *Insurance:Mathematics and Economics*, 43, 386–393.

Buch-Larsen, T., Guillén, M., Nielsen, J. and Bolancé, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics*, 39, 503–518.

Cossette, H., Gaillardetz, P., Marceau, E. and Rioux, J. (2002). On two dependent individual risk models. *Biometrika*, 30, 153–166.

Denuit, M., Genest, C. and Marceau, E. (1999). Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics*, 25, 85–104.

Fermanian, J., Radulović, D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10, 847–860.

Fisher, N. (2000). *Encyclopedy of Statistical Science*. Kotz, S., Johnson, N. L. and Read, C. B. eds., New York.

Galambos, J. (1975). Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association*, 70, 674–680.

Genest, C. and Boies, J. (2003). Detecting dependence with kendall plots. *Journal of the American Statistical Association*, 57, 275–284.

Genest, C., Kojadinovic, I., Neslehová, J. and Yan, J. (2011). A goodness-of-fit test for bivariate extreme-value copulas. *Bernoulli*, 17, 253–275.

Genest, C. and Rivest, L. P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, 88, 1034–1043.

Ghorbal, B. Genest, C. and Neslehová, J. (2009). arge sample tests of extreme value dependence for multivariate copulas. *The Canadian Journal of Statistics*, 39, 703–720.

Gronneberg, S. and Hjort, N. (2014). The copula information criteria. *Scandinavian Journal of Statistics*, forthcoming.

Gumbel, E. (1958). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55, 698–707.

Hofert, M., Kojadinovic, I., Maechler, M. and Yan, J. (2013). Multivariate dependence with copulas. *R package*.

Hüsler, J. and Reiss, R. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics and Probability Letters*, 7, 283–286.

Kaas, R. Dhaene, J. and Goovaert, M. (2000). Upper and lower bounds for sums of random variables. *Insurance: Mathematics and Economics*, 27, 151–168.

Kojadinovic, I., Segers, J. and Yan, J. (2011). Large sample tests of extreme value dependence for multivariate copulas. *The Canadian Journal of Statistics*, 39, 703–720.

Nelsen, R. (2006). *An Introduction to Copulas*, 2nd Ed. Springer, Portland, OR, USA.

Quessy, J., Genest, C. and Neslehova, J. (2012). Test of symmetry for bivariate copulas. *Annals of Institute of Statistical Mathematics*, 64, 811–834.

Ren, J. (1995). Hadamard differentiability on $d[0,1]^p$. *Journal of Multivariate Analysis*, 55, 14–28.

Sarabia, J. and Gómez-Déniz, E. (2008). Construction of multivariate distributions: a review of some recent results. *SORT-Statistics and Operations Research Transactions*, 32, 3–36.

Scaillet, O. (2005). A kolmogorov smirnov type test for positive quadrant dependence. *The Canadian Journal of Statistics*, 33, 415–427.

Segers, J. (2012). Max-stable models for multivariate extremes. *REVSTAT-Statistical Journal*, 10, 61–82.

Sklar, A. (1959). Fonctions de répartition á n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'université de Paris*, 8, 229–231.

Van der Vaart, A. (2000). *Weak Convergence and Empirical Processes*, 2nd Ed. Springer.