# Quantifying Risk using Copulae and Kernel Estimators

Zuhair Bahraoui

UNIVERSITY OF BARCELONA

FACULTY OF ECONOMICS AND BUSINESS

# Quantifying Risk using Copulae and Kernel Estimators

Zuhair Bahraoui

*Thesis submitted for the degree of*

*Doctor in Statistics*

Director:

Dr. Catalina Bolancé

Universitat de Barcelona

**UNIVERSITY OF BARCELONA**

**FACULTY OF ECONOMICS AND BUSINESS**

# Quantifying Risk using Copulae and Kernel Estimators

Zuhair Bahraoui

Director: Dr. Catalina Bolancé

Tutor: Dr. Montserrat Guillén

Riskcenter

Department of Econometrics, Statistics and Spanish Economy

March 2015

# Acknowledgment

I would like to express my special gratitude to my supervisor for this Thesis -Dr. Catalina Bolancé-, for her guidance, dedication and patience throughout the preparation of this work. Thank you Cati.

I thank Dr. Anna M. Pérez-Marín for her help in the correction of this Thesis.

It is also a pleasure to thank all members of the Riskcenter group for the kindness shown during my stay in the Department of Econometrics, Statistics and Spanish Economy, with special mention for director Dr. Montserrat Guillén and professors Miguel Santolino and Helena Chuliá for their encouragement.

Finally, thanks to my colleagues in the doctoral field which whom I have shared good times and to my family for their unconditional support.

# Contents

# Introduction

The complexity of the dependence structure between the random variables in various fields (see, for example, Frees and Valdez 1998, Genest and Favre 2007) has led statisticians to develop new methods that allow us to model a wide range of dependency structures. The concept was introduced by Sklar (1959) in his pioneer work on copulae. It was not until the 80´s that Genest and MacKay (1986b) shed light on the importance of this type of functions. They show that, by simple calculations, it is possible to derive many of their properties (on what is now called Archimedean copula), giving a geometric interpretation of the Kendall coefficient of concordance.

The outstanding property of copulae is that they can isolate the function of the dependency from the marginals and provide a simple way to generalise the dependence beyond the linear correlation. Since then several methods have been introduced in the estimation, simulation and inference regarding copulae. We can mention, among others, the parametric and nonparametric methods investigated by Genest et al. (1995) and Fermanian et al. (2004).

The essential objective in the insurance and finance field is to analyse the distribution associated with the total loss generated by a multivariate random vector $(X_1, \ldots, X_k)'$ of dependent losses or risk factors; in this work we define the total loss as: $S = X_1 + \ldots + X_k$ and the aim is to estimate the risk of loss. Considering the relationship between the risk factors $X_j$, $j = 1, \ldots, k$, the risk analysis may be faced with two problems: first, what are the copula that best reflect the dependence structure between this factors. Second, how the distribution function of the marginals should be estimated and inserted in the copula.

The most common approach for the adjustment of copulae is to assume that the copula belongs to a particular family, then to estimate the dependence parameters by the maximum pseudo-likelihood method proposed by Genest et al. (1995) and Shih and Louis (1995). To select the copula there are some adequacy tests. This type of inference is relatively recent and began with the work of Fermanian et al. (2004) and Genest et al. (2006), but there are comparatively few to cover the many types of copulae and their multiple properties.

On the other hand, the presence of rare events complicates the study in the sense that a failure in estimating of the model can generate large losses. The press article *The Formula That Killed Wall Street* (see, Salmon 2009) is a clear example of where the misuse of a copula can lead.

We should also bear in mind that the method of estimating the marginals may be crucial to model the dependence behaviour of the variables. According to Nelsen (2006), in the coupling of the joint distribution with marginals, the copula captures the aspect that links them. We test this phenomenon during this work dealing with extreme value data. Although the copula function is independent of the functions of the marginals distributions (except some cases such as the Sarmanov copula as we see below), the selected marginal distribution functions can influence significantly the estimated risk of loss when rare events are detected.

To estimate the parameters of copula, in this work we propose to use smoothing methods based on the kernel estimator instead of the classical pseudo-observations. The use of the range statistics as reference does not affect the dependence structure because all copulae are invariant under monotonous transformation, but a substantial lack of efficiency can occur, caused by the variability of the range statistics when inserted in the copula. This lack of efficiency also occurs when we estimate quantiles using order statistics. It is known that empirical adjustment of the distribution does not always lead to the best estimator.

In this Thesis we investigate how to respond to the previous two questions, i.e. how to select the copula and how to estimate marginal distributions when we have extreme values or rare events.

1. First, we start with the identification of the presence of rare events in the historical data. Once confirmed, by the adequacy test on copula, we proceed to identify the type of dependence structure between the variables or at least rule out the copulae that are not suitable.

2. Estimate the marginals and insert them in the copula.

3. Estimate the risk of loss.

Thus we have three objectives. The first is to develop a test that will be able to detect the existence of extreme values in terms of copulae. In the literature, there is little related research. We can mention the first contribution of Ghoudi et al. (1998) based on the integral probability transformation, later Ben Ghorbal et al. (2009) improve the test by reducing the bias of the statistic used for the contrast. Kojadinovic et al. (2011) analyse the test of extreme value copula using a max-stable hypothesis. Finally, there are other tests of extreme value copula based on the Pickand function of dependence (see, for example, Bucher 2011).

The second point is to estimate the marginal distribution, but unlike the most popular nonparametric estimation with pseudo-observations, we use a modified kernel method to estimate them. This method has never been tested before and its effectiveness has been proved using simulations.

The good fit of the marginal and the rapprochement of the choice of copula lead us to give a best fit to estimate the bivariate function distribution. This leads to the third point which is to estimate the total risk of loss. Here, we also develop new results about the Value-at-Risk, the risk measure most used by analysts.

We can say that this Thesis addresses two fundamental aspects of risk quantification. The first is related to the theory of copula and estimating the risk of loss. The second is related to the use of nonparametric and semiparametrics methods for estimating the cumulative distribution function and the quantiles. All methods presented in this Thesis were programmed with R (the programmes are available from the author).

The work has been divided into six chapters and the final conclusions. The first chapter summarises the required results of the theory of copulae, estimation, inference and simulation. The second chapter contains an introduction to the extreme value theory, where we summarize some results related to some univariate distributions and bivariate extreme value theory and copulae.

Chapters 3, 4, 5 and 6 correspond to four works that have been published and/or presented in different journals and congresses. Concretely, the third chapter corresponds to the paper titled "Testing extreme value copulas to estimate the quantile", published in

*SORT-Statistics and Operations Research Transactions*, Bahraoui et al. (2014c). In this chapter we generalize the test proposed by Kojadinovic et al. (2011).

The fourth chapter contains the work titled "Extreme value copulas and marginal effects", published in *New Perspectives on Stochastic Modeling and Data Analysis* (in preparation), Bahraoui et al. (2014b). We have studied the effect of using different extreme value marginal distribution for estimating the risk. The risk measure used is the Value-at-Risk (VaR). Finally, in order to control the risk, we estimate the bounds of the VaR for the aggregate loss.

The chapter 5 includes the paper titled "Quantifying the risk using copulae with nonparametric marginals" published in *Insurance: Mathematics and Economics*, Bolancé et al. (2014). In this chapter we show that copulae and kernel estimation can be mixed to estimate the risk of an economic loss. We analyse the properties of the Sarmanov copula. We find that the maximum pseudo-likelihood estimation of the dependence parameter associated with the copula, using double transformed kernel estimation to estimate marginal cumulative distribution functions, is a useful method for approximating the risk of extreme dependent losses when we have large data sets.

Moreover, in the chapter 6, we include our latest work titled "Estimating extreme value cumulative distribution functions using kernel approach" that has been presented at two specialised conferences: RISK2013 (*5ª Reunión de Investigación en Seguros y Gestión del Riesgo*) and *R*isk Management in Insurance 2014, Bahraoui et al. (2013; 2014a). This chapter has been divided into two parts; in the first, we analyse the domain of attraction of different mixtures of Extreme Value Distributions (EVDs) and, in the second, we describe a new transformed kernel estimator of the cumulative distribution function based on transformations and bias correction.

For illustrating the applicability of all proposed methods in the four chapters described above we use a bivariate sample of losses from a real database of auto insurance claims. Finally, we include a chapter with the main conclusions and future research. Some definitions and results which are necessary in the calculations are shown in the Appendices.

# Chapter 1

# Dependence

Copulas (derived from the Latin word *copŭlæ*: link, tie) were used by Sklar (1959) in his theorem on multivariate distributions. Fisher (1997) noted in the Encyclopedia of Statistical Sciences that copulae are interesting for statisticians for two basic reasons. Firstly, for their application for studying nonparametric measures of dependence, secondly, as a starting point for constructing multivariate distributions representing dependency structures. Moreover, Genest and Favre (2007) noted that copulae can be used for renewing the parametric estimation methods and constructing goodness of fit tests.

## 1.1 Theory of copulae

In this section, following the definitions of Nelsen (2006), dependence is discussed in terms of copulae. The main contributions of this type of functions are that they are able to explain dependence in a more sophisticated way than the linear correlation does, and that they make it possible the construction of multivariate models with a non Gaussian structure. We present here the most important properties derived from the theory of copulae for the bivariate case, which can be extended to a higher dimension.

**Definition 1.1.1.** A two-dimensional copula is a distribution function defined in $[0, 1]^2$ with uniform marginal distribution functions $U(0, 1)$.

Any two-dimensional copula must fulfill these three conditions:

$$(A_1) \quad C(u_1,u_2) \quad \forall (u_1,u_2) \in [0,1]^2 \quad \text{increasing in each component.}$$

$$(A_2) \quad C(u,1) = C(1,u) = u, \forall u \in [0,1].$$

$$(A_3) \quad C(b_1,b_2) - C(b_1,a_2) - C(a_1,b_2) + C(a_1,a_2) \geq 0,$$

$$\forall (a_1,b_1),(a_2,b_2) \in [0,1]^2 \quad \text{such that} \quad a_1 \leq b_1, a_2 \leq b_2.$$

The first property $(A_1)$ ensures that our copula is a bivariate distribution. The second, $(A_2)$ indicates that the marginals are uniformly distributed $U(0,1)$. The third, $(A_3)$ ensures that the copula is a distribution function.
The inversion method is widely used in order to simulate copulae (and probability distributions in general). Then, it is only necessary to simulate uniformly distributed $U(0,1)$ random values in order to obtain a simulated copula.

> Let $F$ be a continuous distribution function in $\mathbb{R}$ and $F^{-1}$ be the inverse function defined as:
>
> $$F^{-1}(\alpha) = inf\{x \mid F(x) = \alpha\}, \quad \alpha \in ]0,1[.$$
>
> If $U$ is a uniform $U(0,1)$ random variable, then the distribution function of $F^{-1}(U)$ is $F$. On the other hand, if $X$ is a random variable with a distribution function given by $F$, then $F(X)$ is uniformly distributed $U(0,1)$.

Note that if $H$ is the joint distribution of a random variable $(X_1,X_2)$ with marginals $F_1$ and $F_2$ and we consider the function

$$C(u_1,u_2) = H\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right),$$

then we have:

$$C(u_1,1) = H\left(F_1^{-1}(u_1),\infty\right) = P\left(X_1 \leq F_1^{-1}(u_1)\right) = P\left(F_1(X_1) \leq u_1\right) = u_1.$$

The same occurs with $u_2$, therefore $C$ fulfills condition $A_2$. The remaining two conditions result from the properties of a probability distribution. The theory of copulae started with the research carried out by Sklar (1959) and it is almost impossible to accomplish a study on dependence without using this theorem. By using the well-known Sklar´s theorem, the author proves that any multivariate distribution can be written in terms of a copula where the dependency structure is isolated.

> Let $H$ be a two-dimensional distribution function with marginals $F_1$ and $F_2$, then there exist a copula $C : [0,1]^2 \longrightarrow [0,1]$ such that:
>
> $$H(x_1, x_2) = C\big(F_1(x_1), F_2(x_2)\big) \quad \forall x_1, x_2 \in \overline{\mathbb{R}}. \tag{1.1}$$
>
> Moreover, $C$ is unique if $F_1$ y $F_2$ are continuous, otherwise, $C$ is uniquely determined on $Rang(F_1) \times Rang(F_2)$.
> Conversely, if $C$ is a copula and $F_1$ and $F_2$ are distribution functions, then the function $H$ defined in (1.1) is a joint distribution function with marginals $F_1$ and $F_2$.

It is clear that the joint probability distribution of $(X_1, X_2)$ includes the marginals $F_1$ and $F_2$ and the dependence represented by $C$. For modeling a two-dimensional sample we can choose a model for the marginals and, independently, a suitable model for the copula.

Next, we describe the most important parametric copulae, although it is difficult to find a unique classification in the literature. Firstly, the so-called **fundamental copulae** are associated to extreme dependency cases. Secondly, the **implicit copulae**, which are extracted from the Sklar´s theorem and have closed-form expressions. Finally, the **explicit copulae**, which also have a closed-form expression and the structure depends only on a single dependency parameter. Here we describe the Gaussian and t-Student copula in the implicit case, and additionally the Sarmanov copula. After that, we also analyze the properties of Archimedean copulae for the explicit case. Then we examine the properties of the most relevant **extreme value copulae**, which describe the dependency structure of extreme values. In the nonparametric case, we also consider the **empirical copula** which can be useful in the estimation stage.

## 1.2 Fundamental copulae

One of the simplest copulae is the product copula or independence copula, which results from the independency of the two variables:

$$C^{\Pi}(u_1, u_2) = u_1 u_2. \tag{1.2}$$

Other copulae describing the dependence between two random variables are the counter-monotonicity and monotonicity copulae, which coincide with the lower and upper Fréchet bounds, respectively. The lower and upper Fréchet bounds are defined by Fréchet (1957).

$\forall u_1, u_2 \in [0, 1]$, is fulfilled for each copula $C$:

$$\max\big(u_1 + u_2 - 1, 0\big) = C^{\min}(u_1, u_2) \leq C(u_1, u_2) \leq C^{\max}(u_1, u_2) = \min(u_1, u_2). \tag{1.3}$$

## 1.3 Implicit copulae

The implicit copulae are defined as copulae associated to elliptic distributions. Their most important feature is that they represent symmetric dependency relationships, and then becomes irrelevant whether we are analyzing the right or left tail of the distribution. The most popular examples are the Gaussian copula and t-Student copula, which are defined next.

**Definition 1.3.1.** Let $\rho$ be the linear correlation coefficient between two random variables $X_1$ and $X_2$, the Gaussian copula with parameter $\rho$ is:

$$\begin{aligned} C_\rho^{Ga}(u_1,u_2) &= \Phi_\rho\left(\Phi^{-1}(u_1),\Phi^{-1}(u_2)\right) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)}\int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(\frac{2\rho st-s^2-t^2}{2(1-\rho^2)}\right)dsdt, \end{aligned}$$

where $\Phi_\rho$ is the two-dimensional standard Normal distribution with correlation coefficient equal to $\rho$, and $\Phi$ is the standard Normal one-dimensional distribution.

The t-Student copula is defined in the same way. Specifically, let $t_\upsilon$ be the (central) t-Student distribution function, where $\upsilon$ indicates the degree of freedom. Then, it is equivalent to:

$$t_\upsilon(x) = \int_{-\infty}^{x} \frac{\Gamma((\upsilon+1)/2)}{\sqrt{\pi\upsilon}\Gamma(\upsilon/2)}\left(1+\frac{s^2}{\upsilon}\right)^{-\frac{\upsilon+1}{2}} ds,$$

where $\Gamma$ is Euler function.

The corresponding two-dimensional distribution function with correlation parameter $\rho$ is equal to:

$$t_{\upsilon,\rho}(x_1,x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \frac{1}{2\pi\sqrt{1-\rho^2}}\left(1+\frac{s^2+t^2-2\rho st}{\upsilon(1-\rho^2)}\right)^{-\frac{\upsilon+2}{2}} dsdt.$$

**Definition 1.3.2.** The two-dimensional t-Student copula with parameter $\rho$ is defined as:

$$\begin{aligned} C_{\rho,\upsilon}^{t}(u_1,u_2) &= t_{\upsilon,\rho}\left(t_\upsilon^{-1}(u_1),t_\upsilon^{-1}(u_2)\right) \\ &= \int_{-\infty}^{t_\upsilon^{-1}(u_1)}\int_{-\infty}^{t_\upsilon^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}}\left(1+\frac{s^2+t^2-2\rho st}{\upsilon(1-\rho^2)}\right)^{-\frac{\upsilon+2}{2}} dsdt. \end{aligned}$$

When the degree of freedom increases, the t-Student copula becomes more similar to the Normal copula.

In Figures 1.1 and 1.2 the simulated Gaussian and t-Student copulae are represented, respectively. We observe differences in the tails of the two copulae, when the values get close to (0.0) or (1.1), the t-Student copula presents a higher degree of dependence compared to the Normal copula. This is due to the fact that the t-Student distribution decreases slower than the Normal distribution, and therefore the possibility to achieve dependence in the tail is higher.

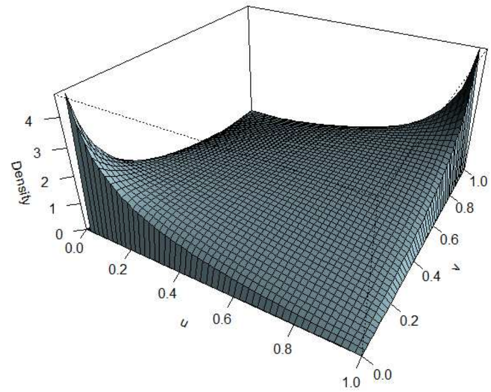


FIGURE 1.1: *Density of a Gaussian copula* $\rho = 0.5$.

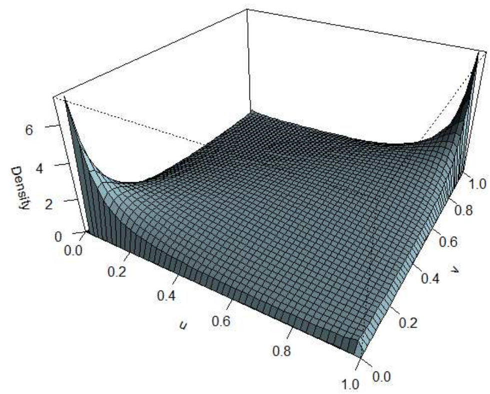


FIGURE 1.2: *Density of a t-Student copula* $\rho = 0.5$, $\upsilon = 4$.

### 1.3.1 Sarmanov copula

Let $(X_1, X_2)$ be a bivariate random vector with marginal probability distribution functions (pdfs) $f_1$ and $f_2$. Also, let $\phi_1$ and $\phi_2$ two bounded non-constant function such that:

$$\int_{-\infty}^{+\infty} f_1(t)\phi_1(t)dt = 0, \qquad \int_{-\infty}^{+\infty} f_2(t)\phi_2(t)dt = 0,$$

then the joint bivariate pdf introduced by Sarmanov (1966), is defined as:

$$h(x_1, x_2) = f_1(x_1)f_2(x_2)\Big(1 + \omega\phi_1(x_1)\phi_2(x_2)\Big).$$

From Sklar's theorem (1.1) the associated copula can be expressed as:

$$C(u_1, u_2) = u_1u_2 + \omega \int_0^{u_1}\int_0^{u_2} \phi_1\left(F_1^{-1}(t)\right)\phi_2\left(F_2^{-1}(s)\right) dtds \tag{1.4}$$

and its density is:

$$c(u_1, u_2) = 1 + \omega\phi_1\Big(F_1^{-1}(u_1)\Big)\phi_2\Big(F_2^{-1}(u_2)\Big), \tag{1.5}$$

where $F_1$ and $F_2$ are the cumulative distribution functions (cdfs) of $X_1$ and $X_2$, respectively.

Parameter $\omega$ is a real number that satisfies the condition $1 + \omega\phi_1(x_1)\phi_2(x_2) \geq 0$ for all $x_1$ and $x_2$.

Note that when $\omega = 0$, $X_1$ and $X_2$ are independent. This parameter is related to the correlation between $X_1$ and $X_2$ (if it exists) (see, Lee 1996) as:

$$corr(X_1, X_2) = \omega\frac{\nu_1\nu_2}{\sigma_1\sigma_2}, \tag{1.6}$$

where $\nu_1 = E(X_1\phi_1(x_1))$, $\nu_2 = E(X_2\phi_2(x_2))$ and $\sigma_1^2 = var(X_1)$, $\sigma_2^2 = var(X_2)$. When we take $\phi_1(x_1) = 1 - 2F_1(x_1)$ and $\phi_2(x_2) = 1 - 2F_2(x_2)$, we have the classical Farlie-Gumbel-Morgenstern (FGM) copula. In this case the dependence parameter has the range $-1/3 \leq \omega \leq 1/3$.

Another special case is when we consider functions of the type:

$$\phi_1(x_1) = x_1 - \mu_{X_1} \quad \text{and} \quad \phi_2(x_2) = x_2 - \mu_{X_2}, \tag{1.7}$$

where $\mu_{X_1} = E(X_1)$ and $\mu_{X_2} = E(X_2)$. In this case, Lee (1996) shows that, if the support of $f_1$ and $f_2$ is contained in $[0,1]$, the range of the dependence parameter is:

$$\max\left(\frac{-1}{\mu_{X_1}\mu_{X_2}}, \frac{-1}{(1-\mu_{X_1})(1-\mu_{X_2})}\right) \leq \omega \leq \min\left(\frac{1}{\mu_{X_1}(1-\mu_{X_2})}, \frac{1}{(1-\mu_{X_1})\mu_{X_2}}\right). \tag{1.8}$$

## 1.4 Archimedean copulae

The term "Archimedean" for this type of copulae was introduced by Ling (1965). This family was developed by Sklar and Schweizer (1983) and became widely known as a result of their applications to finance and other areas. Before that, the Canadian statistician Christian Genest contributed with his research to increase the application of these functions in the field of statistics. There are several reasons that justify their application, among others:

1. There are many parametric formulations, and therefore many dependency structures.
2. They have interesting properties.
3. They are easily constructed and simulated.

An Archimedean copula is obtained by using its generator, as it is explained in the following definition.

**Definition 1.4.1.** Let $\varphi : [0,1] \to [0,\infty]$ be a decreasing function that fulfills the condition $\varphi(1) = 0$. The pseudo-inverse of $\varphi$ is defined as:

$$\varphi^{[-1]}(u_1) = \begin{cases} \varphi^{-1}(u_1) & \text{if } 0 \leq u_1 \leq \varphi(0) \\ 0 & \text{if } \varphi(0) \leq u_1 \leq +\infty. \end{cases}$$

The pseudo-inverse $\varphi^{[-1]}$ is continuous, not increasing in $[0, +\infty]$ and strictly decreasing in $[0, \varphi(0)]$, moreover:

$$\varphi\big(\varphi^{[-1]}(t)\big) = \min(t, \varphi(0)) = \begin{cases} t & \text{if } 0 \leq t \leq \varphi(0) \\ \varphi(0) & \text{if} \varphi(0) \leq t \leq +\infty. \end{cases}$$

Specifically, this type of copula is characterized by (see, Nelsen 2006):

> Let $\varphi$ and $\varphi^{[-1]}$ be the functions defined previously and let $C : [0,1]^2 \longrightarrow [0,1]$ be the function:
>
> $$C(u_1, u_2) = \varphi^{[-1]}\big(\varphi(u_1) + \varphi(u_2)\big). \tag{1.9}$$
>
> Then $C$ is a copula if and only if $\varphi$ is convex.

When $\varphi(0) = +\infty$ we say that the generator is strict. In this case $C(u_1, u_2) = \varphi^{-1}\big(\varphi(u_1) + \varphi(u_2)\big)$ and the copula is an strict Archimedean copula.
The Archimedean family of copulae fulfills the symmetric property, which means that $C(u_1, u_2) = C(u_2, u_1)$, the associative property, that is $\big(C(u_1, u_2), u_3\big) = C\big(u_1, C(u_2, u_3)\big)$ and, finally, its generator is invariant, that is to say, if we multiply it by any positive constant $k$, $k\varphi$ is also a generator of $C$ (see, Nelsen 2006). The most well-known examples of Archimedean copulae are described next.

### 1.4.1 Gumbel copula

Let $\varphi(t) = (-\ln t)^{\theta}$, then according to (1.9) and for $\theta \in [1, +\infty)$ we have:

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

In the limits of $\theta$ we have $C_1 = C^{\Pi}$, which represents the product copula, and $C_{+\infty} = C^{\max}$ is the upper bound Fréchet copula.
In Figure 1.3 the density of a Gumbel copula with parameter $\theta = 2$ is represented. We observe that it takes large values in the extreme $(1, 1)$, whereas in the extreme $(0, 0)$

takes low values. This is due to the fact that the Gumbel copula includes a dependency structure only in the right tail, as we will see later on.

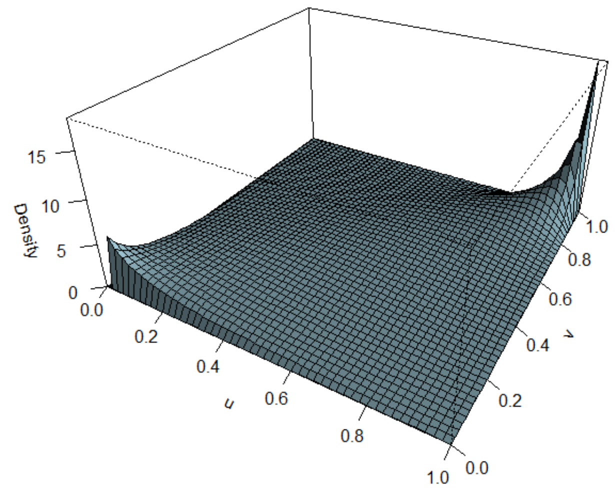


FIGURE 1.3: *Density of a Gumbel copula* $\theta = 2$.

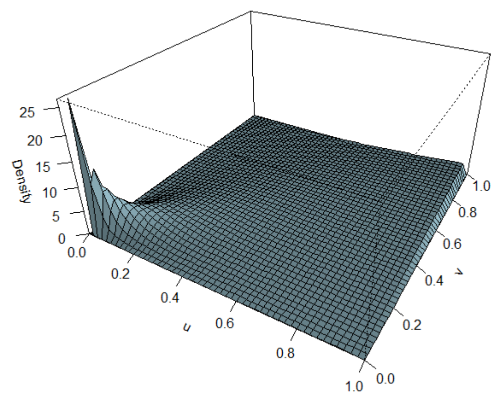


FIGURE 1.4: *Density of a Clayton copula* $\theta = 2$ *(left).*

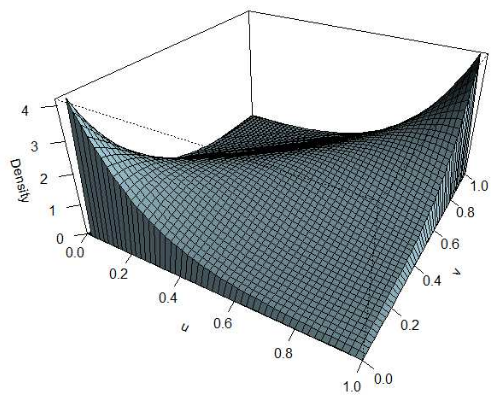


FIGURE 1.5: *Density of a Frank copula with* $\theta = 5$ *(right).*

### 1.4.2 Clayton copula

The generator of a Clayton copula is defined by $\theta \in [-1, 0[\cup]0, +\infty)$,

$$\varphi_\theta(t) = \frac{t^{-\theta} - 1}{\theta}.$$

By using expression (1.9), we obtain the expression of the Clayton copula family given by:

$$C_\theta(u_1, u_2) = \max\left((u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, 0\right).$$

For $\theta > 0$, the Clayton copula is strict and equal to:

$$C_\theta(u_1, u_2) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}.$$

For $\theta > 0$, the Clayton copula is complete, that is to say, $\lim_{\theta \to 0} C_\theta(u_1, u_2) = C^{\Pi}$ is the product copula and $C_{-1} = C^{\min}$ is the lower Fréchet bound. Finally, $C_{+\infty} = C^{\max}$ is equivalent to the upper Fréchet bound. The density of a Clayton copula presents large values in the extreme (0,0), as shown in Figure 1.4, and low values in (1,1). Its structure is similar to the Gumbel copula but in the opposite way. This is due to the fact that this type of copulae are able to identify dependence in the left tail.

### 1.4.3 Frank copula

It is defined by the parameter $\theta \in (-\infty, 0[\cup]0, +\infty)$ and its generator is:

$$\varphi_\theta(t) = \ln\left(\frac{1 - e^{\theta}}{1 - e^{-\theta t}}\right).$$

The Frank copula is defined as:

$$C_\theta(u_1, u_2) = -\frac{1}{\theta}\ln\left(1 - \frac{(1 - e^{\theta u_1})(1 - e^{\theta u_2})}{1 - e^{-\theta}}\right).$$

The main characteristic of the Frank copula is that it does not present dependence in the extremes, but in the center, as it is shown in Figure 1.5.

## 1.5 Extreme value copula

This type of copulae are associated with the extreme value theory. Nelsen (2006) noted that extreme value copulae are obtained by transforming other copulae. To obtain an extreme value copula we use the following result.

Let $(X_{11}, X_{21}), (X_{12}, X_{22}), \ldots, (X_{1n}, X_{2n})$ be a two-dimensional sample of independent and identically distributed observations, with joint distribution denoted by $H$. We assume that it has an associated copula $C$ with marginals $F_1$ and $F_2$. If we define the order statistics $X_{(1n)} = \max\left\{X_{11}, X_{12}, \ldots X_{1n}\right\}$ for $X_1$, we observe that its distribution is given by:

$$P\big(X_{(1n)} \leq x_1\big) = P\big(\cup X_{1i} \leq x_1\big) = \big[P(X_{11} \leq x_1)\big]^n.$$

In a similar way, the distribution of $X_{(2n)} = \max\left\{X_{21}, X_{22}, \ldots X_{2n}\right\}$ is obtained. Therefore, $F_{(1_n)}(x_1) = [F_1(x_1)]^n$ y $F_{(2_n)}(x_2) = [F_2(x_2)]^n$, and we conclude that:

$$\begin{aligned} H_{(n)}(x_1, x_2) &= P\left(X_{(1n)} \leq x_1, X_{(2n)} \leq x_2\right) \\ &= \Big[H(x_1, x_2)\Big]^n = \Big[C(F_1(x_1), F_2(x_2))\Big]^n \\ &= \Big[C([F_{(1_n)}(x_1)]^{1/n}, [F_{(2_n)}(x_2)]^{1/n})\Big]^n. \end{aligned}$$

So, we have that

$$C_{(n)}(u_1, u_2) = C^n\left(u_1^{1/n}, u_2^{1/n}\right), \quad \forall u_1, u_2 \in [0, 1]. \tag{1.10}$$

This proves the next result (see, Nelsen 2006).

> If $C$ is a copula and $n$ is a positive integer number, then the function $C_{(n)}$ defined in (1.10) is a copula. Moreover, if for $i = 1 \ldots n$, $(X_{1i}, X_{2i})$ are independent and identically distributed, with copula C, then $C_{(n)}$ is the copula associated to variables $X_{(1n)} = \max\{X_{1i}\}$ and $X_{(2n)} = \max\{X_{2i}\}$.

**Definition 1.5.1.** A copula is $max - stable$, if for any positive real number $r$ and for all $u_1$, $u_2$ in the interval $[0, 1]$, we have that:

$$C(u_1, u_2) = C^r \left( u_1^{1/r}, u_2^{1/r} \right).$$

**Example 1.5.1.** The Gumbel copula $C_\theta$:

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

belongs to the $max - stable$ class of copulae. It can be easily proved that $C_\theta(u_1, u_2) = C_\theta^r \left( u_1^{1/r}, u_2^{1/r} \right)$.

Following a similar approach as in extreme value theory, which will be presented in the next chapter, an extreme value copula is the limit of a copula in the maximum when it exists. Therefore, as a definition we have:

**Definition 1.5.2.** A copula $C_*$ is an extreme value copula, if a copula $C$ exists such that:

$$C_* = \lim_{n \to \infty} C^n(u_1^{1/n}, u_2^{1/n}),$$

for $u_1, u_2 \in [0, 1]$. Moreover, $C$ is the so-called domain of attraction of $C_*$.

The connection between a $max - stable$ copula and an extreme value copula is clear and coincide when the limit exists (see, Nelsen 2006).

## 1.6 Empirical copula

The empirical copula, introduced by Deheuvels (1979), is based on an approximation of the multivariate distribution by using ranges. The empirical distribution is simply the distributions of the ranges along the sample. A natural approximation would be to use the Sklar's theorem:

$$C_n(u_1, u_2) = H_n \left( F_1^{-1}(u_1), F_2^{-1}(u_2) \right),$$

where $H_n$ is the empirical distribution of $(X_1, X_2)$:

$$H_n(x_1, x_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\Big(X_{1i} \leq x_1, X_{2i} \leq x_2\Big),$$

therefore, if $F_{1_n}$ and $F_{2_n}$ are the empirical distributions of $X_1$ and $X_2$, being $R_i$ and $S_i$ their corresponding ranges (the order of each element in the sample), then the empirical copula is:

$$C_n(u_1, u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\Big(X_{1i} \leq F_{1_n}^{-1}(u_1), X_{2i} \leq F_{2_n}^{-1}(u_2)\Big), \tag{1.11}$$

$$C_n(u_1, u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\left(\frac{R_i}{n} \leq u_1, \frac{S_i}{n} \leq u_2\right). \tag{1.12}$$

## 1.7 Measures of association

In general, we say that two variables are associated if there is not independence between them. However, there are several concepts of association, for example: linear correlation (measured by the linear correlation coefficient of Pearson), concordance (which is different from dependence, and it is measured by the Spearman's $\rho$ and the Kendall's $\tau$) and, finally, the dependence in the tail (see, Cherubini et al. 2004).

### 1.7.1 Pearson's $\rho$

Let $X_1$ and $X_2$ be two continuous random variables, the correlation between $X_1$ and $X_2$ is:

$$\rho_{X_1,X_2} = \frac{E\Big((X_1 - \mu_{X_1})(X_2 - \mu_{X_2})\Big)}{\sigma_{X_1}\sigma_{X_2}}, \tag{1.13}$$

where $\mu_{X_1} = E(X_1)$, $\mu_{X_2} = E(X_2)$, $\sigma^2_{X_1} = Var(X_1)$ and $\sigma^2_{X_2} = Var(X_2)$. This measure is widely used due to its nice properties. When $X_1$ and $X_2$ are independent then $\rho_{X_1,X_2} = 0$. Nevertheless, the opposite is in general false, except for the case when the

distribution is Normal. That is to say, it is possible that the coefficient is equal to zero but the variables $X_1$ and $X_2$ are not independent. Due to this property it is difficult to capture dependence in the no-linear case.
The measures incorporated by using copulae have the advantage of being invariant under monotone transformations of the marginals and, therefore, they avoid the difficulties of using the Pearson's $\rho$.

### 1.7.2 Range correlation

Range correlation coefficients are the Spearman's $\rho$ and the Kendall's $\tau$. They are also called **association measures** and they consist on measuring the degree of intensity of a monotone relationship between two variables.

**Definition 1.7.1.** Let $(X_{1i}, X_{2i}), (X_{1j}, X_{2j})$ two observations of a continuous random variable vector $(X_1, X_2)$, then we say that $(X_{1i}, X_{2i})$ y $(X_{1j}, X_{2j})$ are concordant if:

$$(X_{1i} - X_{1j})(X_{2i} - X_{2j}) > 0.$$

On the contrary, we say that they are discordant if:

$$(X_{1i} - X_{1j})(X_{2i} - X_{2j}) < 0.$$

**Definition 1.7.2.** Copula $C_1$ is less concordant than copula $C_2$ if:

$$C_1(u_1, u_2) \leq C_2(u_1, u_2) \quad \forall u_1, u_2 \in [0, 1]$$

and it is denoted as $C_1 \prec C_2$.

### 1.7.3 Spearman's $\rho$

The idea is to apply the correlation (1.13) directly to the ranges, and then an empirical version is obtained.
Specifically, if $R_i$, $i = 1, \ldots, n$, is the range of the observations $(X_{11}, \ldots, X_{1n})$ and

$S_i$, $i = 1, \ldots, n$, is the range of the observations $(X_{21}, \ldots, X_{2n})$, the range correlation is equal to the correlation of the sample $(R_1, S_1), \ldots, (R_n, S_n)$ and the Spearman's $\rho$ coefficient is equal to:

$$\rho_S = \rho_{(R,S)}$$

and its estimator is:

$$\widehat{\rho}_S = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (R_i - S_i)^2.$$

This coefficient has the advantage of being invariant under an increasing transformation of the observations, that is to say, if $C_1 \prec C_2$, then we have $\rho_S(C_1) \leq \rho_S(C_2)$.
When $R_i = S_i$, there is perfect positive dependence, (in this case, $\rho_S = 1$), whereas $R_i = n + 1 - S_i$ indicates perfect negative dependence which corresponds to the value $\rho_S = -1$. When copulae are used, the coefficient is equal to (see, Nelsen 2006):

$$\rho_S = 12 \int_0^1 \int_0^1 \Big( C(u_1, u_2) - u_1 u_2 \Big) du_1 du_2 = 12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3.$$

### 1.7.4 *Kendall's* $\tau$

Let $\mathcal{C}_n$ and $\mathcal{D}_n$ be the number of concordant and discordant pairs, respectively, in a sample $(X_{1i}, X_{2i}), i = 1 \ldots n$, the Kendall's $\tau$ is defined as:

$$\tau_n = \frac{\mathcal{C}_n - \mathcal{D}_n}{\binom{n}{2}} = \frac{4}{n(n-1)} \mathcal{C}_n - 1. \tag{1.14}$$

If we introduce the indicator:

$$I_{ij} = \begin{cases} 1 & \text{si } X_{1i} < X_{1j} \quad \text{and} \quad X_{2i} < X_{2j}, \quad \forall i \neq j \\ 0 & \text{otherwise} \end{cases}$$

and $I_{ii} = 1, \quad \forall i = 1 \ldots n,$

we observe that

$$\mathcal{C}_n = \frac{1}{2} \sum_{i=1}^{n} \sum_{i \neq j} (I_{ij} + I_{ji}) = \sum_{i=1}^{n} \sum_{i \neq j} I_{ij} = -n + \sum_{i=1}^{n} \sum_{j=1}^{n} I_{ij}.$$

Let

$$W_i = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{I}_{ij},$$

then formula $\mathcal{C}_n$ reduces to:

$$\mathcal{C}_n = -n + \sum_{i=1}^{n} nW_i = -n + n^2\overline{W}, \tag{1.15}$$

where

$$\overline{W} = \frac{1}{n}\sum_{i=1}^{n} W_i.$$

By doing a simple calculation we obtain the empirical expression of the Kendall's $\tau$. We simply replace $\mathcal{C}_n$ (1.15), in expression (1.14):

$$\tau_n = \frac{4n\overline{W} - n - 3}{n-1}.$$

For calculating $\tau = P(\mathcal{C}) - P(\mathcal{D})$ in its theoretical version, where $P(\mathcal{C})$ and $P(\mathcal{D})$ are the concordance and discordance probabilities, we use $\tau_n$ as an unbiased estimator of $\tau$. In terms of copulae, Genest and Rivest (1993) propose to calculate $\tau$ as:

$$\tau = \tau(C) = 4\int_0^1\int_0^1 C(u_1,u_2)dC(u_1,u_2) - 1. \tag{1.16}$$

The dependency parameter fulfills the same monotonic property as the Sperman's $\rho$, this means that if $C_1 \prec C_2$ we have that:

$$\tau(C_1) \leq \tau(C_2).$$

By using the expression of the Fréchet bounds, regarding the copula $C$, we conclude that:

$$\tau(C^{\min}) = -1 \leq \tau(C) \leq \tau(C^{\max}) = 1,$$

where $C^{\min}$ and $C^{\max}$ are the Fréchet bounds defined in (1.3).

### 1.7.5 *Kendall's* $\tau$ for the Archimedean case

As a result of (1.16), the Kendall's $\tau$ expressed in terms of the expectation is:

$$\tau(C) = 4E\Big(C(u_1, u_2)\Big) - 1.$$

For the Archimedean copula $C_\varphi$, we use the relationships defined in theorem of Genest and MacKay (1986b):

> Let $X_1$ and $X_2$ two random variables of an Archimedean copula, the Kendall's $\tau$ is equal to:
>
> $$\tau(C_\varphi) = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}dt, \qquad (1.17)$$

except for the Frank's copula which cannot be inverted in expression (1.17) (numerical methods are used in this case). By using (1.17) the dependency parameter $\theta$ can be calculated. Specifically we have:

$$\begin{aligned} \tau_{Gumbel}(X_1, X_2) &= 1 - \frac{1}{\theta}, \\ \tau_{Clayton}(X_1, X_2) &= \frac{\theta}{\theta + 2}, \\ \tau_{Frank}(X_1, X_2) &= 1 - 4\frac{\big(1 - D_1(\theta)\big)}{\theta}, \end{aligned}$$

where $D_m(\theta)$ is the Debye function with $m = 1$:

$$D_m(\theta) = \frac{m}{\theta^m}\int_0^\theta \frac{t}{e^t - 1}dt.$$

## 1.8 Dependence in the tail

**Definition 1.8.1.** The dependency coefficient of the right tail of two variables $X_1$ and $X_2$ with distribution functions $F_1$ and $F_2$, respectively, is:

$$\lambda_u := \lambda_{u_1}(X_1, X_2) = \lim_{\alpha \to 1^-} P\big\{X_2 > F_2^{-1}(\alpha) | X_1 > F_1^{-1}(\alpha)\big\} \in [0, 1]. \quad (1.18)$$

Similarly, the dependency coefficient of the left tail is:

$$\lambda_l := \lambda_l(X_1, X_2) = \lim_{\alpha \to 0^+} P\big\{X_2 \leq F_2^{-1}(\alpha) | X_1 \leq F_1^{-1}(\alpha)\big\} \in [0, 1]. \quad (1.19)$$

The coefficient $\lambda_u = 0$ indicates asymptotic independence of the right tail, whereas $\lambda_l = 0$ indicates asymptotic independence of the left tail.
We note that in the case of the Gaussian copula $\lambda_u = \lambda_l = 0$, that is to say, dependence is only identified in the center, whereas there is asymptotic independence in both tails. Nevertheless, the t-Student copula has dependence in the center and also in both tails.

$$\lambda_u = \lambda_l = 2t_{\upsilon+1}\left(\sqrt{\upsilon+1}(-\sqrt{1-\rho_{X_1X_2}})/(\sqrt{1+\rho_{X_1X_2}})\right).$$

In general, for the family of elliptic copulae we have:

$$\lambda_u = \lambda_l.$$

If $\lambda_u > 0$, extreme events tend to occur simultaneously.

In terms of copulae, the dependency coefficient in the right tail can be expressed as:

$$\lambda_u = \lim_{\alpha \to 1^-} \frac{P\big[X_2 > F_2^{-1}(\alpha), X_1 > F_1^{-1}(\alpha)\big]}{P[X_1 > F^{-1}(\alpha)]} = \lim_{\alpha \to 1^-} \frac{\overline{C}(\alpha, \alpha)}{1-\alpha},$$

where $\overline{C}(u_1, u_2) = 1 - 2u_2 + C(u_1, u_2)$ is the survival copula. On the other hand, dependence in the left tail is equivalent to:

$$\lambda_l = \lim_{\alpha \to 0^+} \frac{C(\alpha, \alpha)}{\alpha}.$$

### 1.8.1 Dependency parameters in the tail for the Archimedean families

For the Archimedean copulae we recall the following two results which can be used (see, for example, Joe 1997;for the proff):

- Let $C$ be a strict Archimedean copula with generator $\varphi$. If $(\varphi^{-1})'(0)$ is finite then:

$$C(u_1, u_2) = \varphi^{-1}\big(\varphi(u_1) + \varphi(u_2)\big)$$

does not have dependence in the upper tail. Moreover if $C$ has dependence in the right tail, then $(\varphi^{-1})'(0) = -\infty$ and the coefficient of dependency of the right tail is given by:

$$\lambda_u = 2 - 2 \lim_{\alpha \longrightarrow 0} \left[ \frac{(\varphi^{-1})'(2\alpha)}{(\varphi^{-1})'(\alpha)} \right]. \tag{1.20}$$

- Let $C$ be a strict Archimedean copula. The coefficient of dependency in the left tail for the copula $C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$ is equal to:

$$\lambda_l = 2 \lim_{\alpha \longrightarrow \infty} \left[ \frac{(\varphi^{-1})'(2\alpha)}{(\varphi^{-1})'(\alpha)} \right]. \tag{1.21}$$

For the Gumbel copula the coefficients of dependency in the tail are:

$$\lambda_u = 2 - 2^{1/\theta}, \quad \text{y} \quad \lambda_l = 0,$$

and for the Clayton copula these coefficients are:

$$\lambda_u = 0, \quad \text{y} \quad \lambda_l = \begin{cases} 2^{-1/\theta} & \text{si } \theta > 0 \\ 0 & \text{si } \theta < 0. \end{cases}$$

Finally, the Frank copula does not have dependence neither in the right nor in the left tail:

$$\lambda_u = \lambda_l = 0.$$

## 1.9 Nonparametric estimation of a copula

The dependency structure is frequently estimated by using the maximum likelihood method. A feasible alternative consist of using pseudo-observations. It is worth mentioning the pioneer work carried out by Genest (1987) on the Frank copula. Later, Genest and Rivest (1993) analyzed Archimedean copulae and their inferences. Before presenting these methods, we define the empirical distributions in both the discrete and continuous case, as described by De Matties (2001).

**Definition 1.9.1.** Let $X_{11} \dots X_{1n}$ be an independent random variable. The empirical distribution function is defined as:

$$F_n(x) = \frac{Cardinal\Big\{i|1 \leq i \leq n, X_{1i} \leq x\Big\}}{n}, \quad \forall x \in ]-\infty, +\infty[.$$

In the continuous case the empirical distribution is obtained by using the order statistic.

**Definition 1.9.2.** Let a,b be two real numbers such that:

$$a \leq \min\left(X_{11}, ..., X_{1n}\right) \leq \max\left(X_{11} ... X_{1n}\right) \leq b,$$

and let $X_{(11)}, ... X_{(1n)}$ be the values of a sample $X_{11} ... X_{1n}$ sorted in increasing order. We define $X_{(10)} = a$ and $X_{(1(n+1))} = b$. The continuous empirical distribution function is:

$$P_n\left(x; a, X_{(11)} ..., X_{(1n)}, b\right),$$

which is equal to 0 if $x \leq a$ and equal to 1 if $x \geq b$. For values of $x$ such that $a < x < b$, it takes the corresponding value given by the linear segment which connects the mean points of the bars of the intervals $[X_{(1i)}, X_{(1(i+1))}]$. The mean point of the bar in the left side $[X_{(11)}, X_{(12)}]$ is connected with the point $(a, 0)$ and the one of the bar $[X_{(1(n-1))}, X_{(1n)}]$ in the right side is connected with the point $(b, 1)$.

### 1.9.1 Procedure by Genest and Rivest

Genest and Rivest (1993) propose an estimator based on using pseudo-observations, which are defined as follows:

**Definition 1.9.3.** Let $C$ be a copula associated to a two-dimensional random variable $(X_1, X_2)$, such that $H$ is the joint distribution function and $F_1$, $F_2$ are the marginals. For $i, j = 1 \dots n$, the pseudo-observations are equal to:

$$Z_i = \hat{H}(X_{1i}, X_{2i}) = \frac{Cardinal\Big\{(X_{1j}, X_{2j})|X_{1j} < X_{1i}, X_{2j} < X_{2i}\Big\}}{n-1}. \tag{1.22}$$

The objective is to estimate the univariate distribution function

$$K(t) = P\left(C(u_1, u_2) \leq t\right) = P\left(H(X_1, X_2) \leq t\right), \tag{1.23}$$

taking into account that $u_1 = F_1(X_1)$ y $u_2 = F_2(X_2)$ are uniforms $U(0, 1)$.

**Definition 1.9.4.** A nonparametric estimator of $K(t)$ can be written as:

$$K_n(t) = \frac{1}{n}\sum_{i=1}^{n}\big[i : Z_i \leq t\big].$$

Genest and Rivest (1993) proved that the distribution of $Z_i$ defined in (1.22) converges to $K(t) = P\big[H(X_1, X_2) \leq t\big]$. Moreover, the empirical distribution function $K_n(t)$ of the variable $Z_i$ is $\sqrt{n}$-consistent estimator of $K(t)$ (it converges in probability to $K(t)$).

In general, the function $K(t)$ is not distributed as an $U(0, 1)$ $(K(t) \geq t, \forall t[0, 1])$. This function is known as Kendall distribution function, given that it has a close relationship with the Kendall's $\tau$ defined in (1.16):

$$\tau = 3 - 4\int_0^1 K(t)dt. \tag{1.24}$$

Only for Archimedean copulae the function $K$ guarantees the uniqueness of the copula. In some cases, this function is used to derive inference for the copulae, as the goodness-of-fit tests (see, for example, Genest and Rivest 2001, Genest et al. 2009, Nelson et al. 2001; 2003;among others).

### 1.9.2 Inference regarding the copula parameter

Genest and Rivest (1993) proved that the distribution of the estimator $\tau_n$ of the Kendall's $\tau$ defined in (1.14) is:

$$\frac{\tau_n - \tau}{4S} \sim N(0,1). \tag{1.25}$$

Where

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Z_i + \widetilde{Z} - 2\overline{Z}\right)^2,$$

and $\widetilde{Z}$ , $\overline{Z}$ are two terms equivalent to:

$$\begin{aligned} \widetilde{Z} &= \frac{1}{n-1}\sum_{i=1}^{n} I_{ij} = \frac{1}{n-1}\sum_{i=1}^{n}\{j : X_{1i} \leq X_{1j}, X_{2i} \leq X_{2j}\}, \\ \overline{Z} &= \frac{1}{n}\sum_{i=1}^{n} Z_i. \end{aligned}$$

By applying the Delta method to the transformation of $\theta = g(\tau)$ which results in (1.25) (where $g$ is a differentiable function) the assymptotic distribution of $\widehat{\theta}_n = g(\tau_n)$ is equal to:

$$\widehat{\theta}_n - \theta \sim N\left(0, \frac{1}{n}(4Sg'(\tau_n))^2\right).$$

Finally, if $\xi_{\alpha/2}$ is the quantile of the standard Normal distribution $N(0,1)$, a confidence interval for $\theta$ at a confidence level $100(1-\alpha)\%$ is equal to:

$$\widehat{\theta}_n \pm \xi_{\alpha/2}\frac{1}{\sqrt{n}}4S|g'(\tau_n)|.$$

## 1.10 Maximum likelihood method

The maximum likelihood method (MLE) consist of finding the set of parameters which come from the copula and marginals simultaneously. According to Cherubini et al. (2004) the procedure consist on the following steps:

1. Identification of the marginals.
2. Definition of a suitable copula.

If we consider a two-dimensional sample $(X_{11}, X_{21}), \ldots, (X_{1n}, X_{2n})$ which has a parametric representation, so the marginals associated to the random variables $X_1$ and $X_2$ are of the form $\{F_{1_\alpha}; \alpha \in A\}$ y $\{F_{2_\beta}; \beta \in B\}$, ($A$ and $B$ parameter classes) respectively, the two-dimensional parametric law, according to the Sklar's theorem, is equal to:

$$H(x_1, x_2) = C_\theta \left( F_{1_\alpha}(x_1), F_{2_\beta}(x_2) \right),$$

and the density (when it exists) is given by:

$$h(x_1, x_2) = c_\theta \left( F_{1_\alpha}(x_1), F_{2_\beta}(x_2) \right) f_{1_\alpha}(x_1) f_{2_\beta}(x_2),$$

where $c_\theta = \partial^2 C_\theta(u_1, u_2)/\partial u_1 \partial u_2$, $f_{1_\alpha}$ and $f_{2_\beta}$ are the densities of $F_{1_\alpha}$ and $F_{\beta_2}$, respectively. The log-likelihood function is equal to:

$$L(\theta, \alpha, \beta) = \sum_{i=1}^{n} \ln \left[ c_\theta(F_{1_\alpha}(X_{1i}), F_{2_\beta}(X_{2i})) \right] + \sum_{i=1}^{n} \ln(f_{1_\alpha}(X_{1i})) + \sum_{i=1}^{n} \ln(f_{2_\beta}(X_{2i})). \quad (1.26)$$

The maximum likelihood estimator is obtained by maximizing the function (1.26),

$$(\widehat{\theta}, \widehat{\alpha}, \widehat{\beta})_{MLE} = \max L(\theta, \alpha, \beta).$$

By using this method, the optimization process might be quite time-consuming. Moreover, it can also happen that no solution is found to the maximization problem. In order to solve these problems, Joe (1997) introduced the so-called **inference from marginals** procedure, which consists on:

1. Firstly, the parameters $\alpha$ and $\beta$ are estimated separately by maximizing the log-likelihood function of the marginals

$$\sum_{i=1}^{n} \ln\left(f_{1_\alpha}(X_{1i})\right), \quad and \quad \sum_{i=1}^{n} \ln\left(f_{2_\beta}(X_{2i})\right).$$

2. Secondly, for $i = 1 \dots n$ is established $\widehat{u_{1i}} = F_{1_{\hat{\alpha}}}(X_{1i})$ and $\widehat{u_{2i}} = F_{2_{\hat{\beta}}}(X_{2i})$.

3. Finally, the estimator $\widehat{\theta}$ is chosen as the one which maximizes the function:

$$L(\theta) = \sum_{i=1}^{n} \ln\left(C_\theta(\widehat{u_{1i}}, \widehat{u_{2i}})\right).$$

## 1.11 Pseudo-likelihood estimation

It is a nonparametric method and consists of approximating the marginals $F_1$ y $F_2$ by using the empiricals:

$$F_{1_n}(x) = \frac{1}{n}\sum_{i=1}^{n} I\big(X_{1i} \leq x_1\big), \quad and \quad F_{2_n}(x_2) = \frac{1}{n}\sum_{i=1}^{n} I\big(X_{2i} \leq x_2\big).$$

These estimators have nice properties. Based on the convergence established by the Glivenko-Cantelli theorem:

$$\sup_{x\in R} |F_{1_n}(x) - F_1(x)| \to 0 \quad and \quad \sup_{x\in R} |F_{2_n}(x) - F_2(x)| \to 0.$$

The log-likelihood function in this case is equal to:

$$L(\theta) = \sum_{i=1}^{n} \ln\left[c_\theta\big(F_{1_n}(X_{1i}), F_{2_n}(X_{2i})\big)\right].$$

An estimator of the parameter $\theta$ will have the following expression:

$$\arg\max_{\hat{\theta}\in\Theta} L(\hat{\theta}).$$

Another alternative consists on replacing the pseudo-observations with their ranges. The change can be justified by arguing that it has no impact on the dependency structure, as the copula does not depend on the marginals, the log-likelihood function in this case can be expressed as:

$$L(\theta) = \sum_{i=1}^{n} \ln \left[ c_\theta \left( \frac{R_i}{n+1}, \frac{S_i}{n+1} \right) \right].$$

Note that $n+1$ appears in the denominator in order to avoid numerical problems in $(1, u_2)$ and $(u_1, 1)$ when differentiating the function $\ln[c_\theta(.,.)]$.

## 1.12 K-Plot

The K-Plot is a two-dimensional graphical tool, which shows the dependency structure. It is proposed by Genest and Boies (2003) and it is inspired on the QQ-plot. When data are independent the dots in the graph are close to the diagonal $x_1 = x_2$ and any departure from this diagonal line indicates dependence in the data. Perfect positive dependence, when it occurs, will be depicted with a curve over the diagonal, and perfect negative dependence will result in a line below the diagonal. The technique consist on plotting the pairs $(W_{i,n}, H_{(i)})$ such that:

$$H_{(1)} < \ldots < H_{(n)}$$

is the order statistic associated to the quantities of $H_1 \ldots H_n$ which represent the modified pseudo-observations:

$$H_i = \frac{Cardinal\{(X_{1j}, X_{2j}) | X_{1j} \leq X_{1i}, X_{2j} \leq X_{2i}\}}{n-1} = \frac{nZ_i - 1}{n-1}. \tag{1.27}$$

On the other hand, $W_{in}$ is the expected value of the i-th order statistic under the hypothesis that the copula $W = C(u_1, u_2)$ is an independence copula:

$$W_{in} = n \binom{n-1}{i-1} \int_0^1 t k_0(t) [K_0(t)]^{i-1} [K_0(t)]^{n-1} dt,$$

where

$$K_0(t) = t - t\log(t), \quad t \in [0, 1].$$

Finally, $k_0(w)$ is the density of $K_0(t)$.

**Example 1.12.1.** In Figure 1.6 the perfect positive dependence is shown with the K-Plot by using a sample of 100 generated values of a variable $X \sim exp(1)$ and with its cubic relationship $Y = X^3$. On the other hand, in Figure 1.7 the perfect negative dependence is shown by using a sample of 100 generated values of a variable $X \sim exp(1)$ and with the relationship $Y = -X^2$.

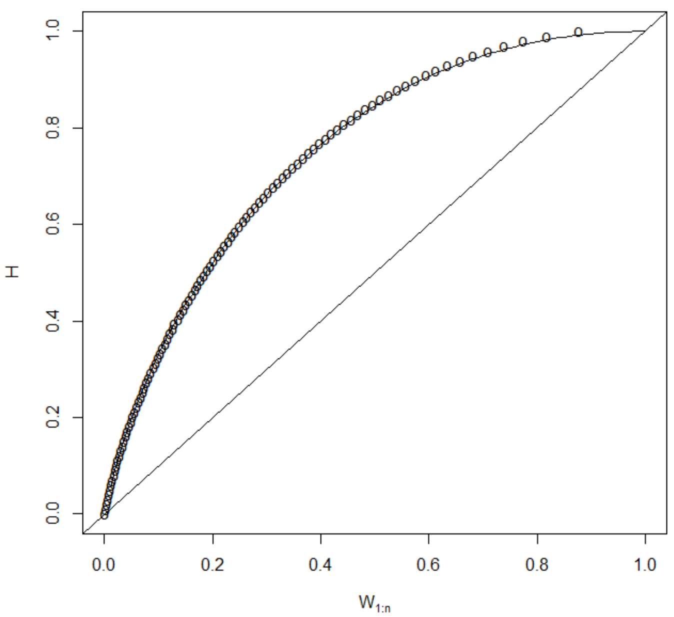


FIGURE 1.6: *K-Plot of a perfect positive dependence.*

## 1.13 Simulating from copula

In general, to generate a two-dimensional random variable from a copula we use a procedure based on the conditional distribution of the random vector $(U_1, U_2)$, (see,

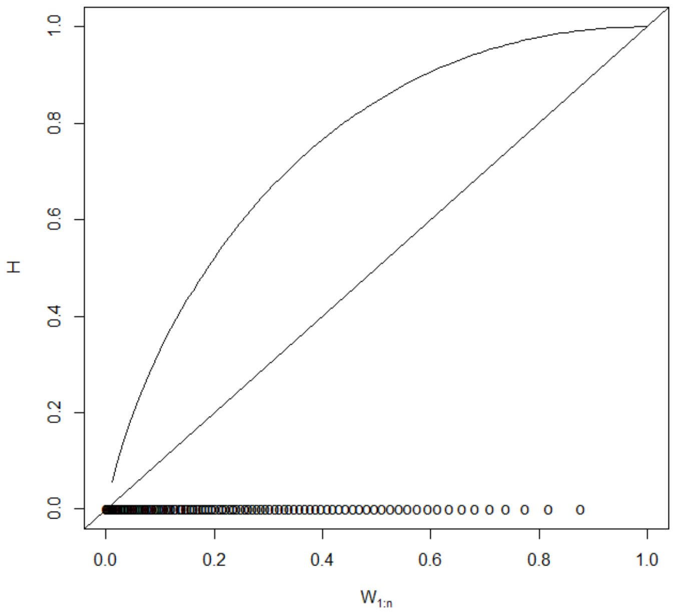


FIGURE 1.7: *K-Plot of a perfect negative dependence.*

Nelsen 2006):

$$\begin{aligned} P(U_2 \leq u_2 | U_1 = u_1) &= C_{u_1}(u_2), \\ \text{where} \quad C_{u_1}(u_2) &= \lim_{\delta u_1 \to 0^+} \frac{C(u_1 + \delta u_1, u_2) - C(u_1, u_2)}{\delta u_1} = \frac{\partial C(u_1, u_2)}{\partial u_1}. \end{aligned}$$

The algorithm is implemented as follows:

1. Firstly, two independent random variables $u_1$ and $t$ are generated from a Uniform distribution $U(0, 1)$.

2. Set $u_2 = C_{u_1}^{[-1]}(t)$ where $C_{u_1}^{[-1]}$ denotes a quasi-inverse of $C_{u_1}$. The quasi-inverse is:

$$C_{u_1}^{[-1]}(t) = \begin{cases} \inf \{x | C_{u_1}(x) \leq t\} & \text{if } t = 0 \\ C_{u_1}^{-1}(t) & \text{if } t \in (0, 1) \\ \inf \{x | C_{u_1}(x) \leq t\} & \text{if } t = 1 \end{cases}$$

3. The desired pair is $(u_1, u_2)$.

However, the conditional distribution for the Gumbel copula is not invertible. So, to generate the random variable from this copula we can use the following two algorithms: Firstly, the algorithm proposed by Genest and MacKay (1986a) (which works only in the bivariate case) consisting on:

1. Generating two independent random variables $\widetilde{u}_1$ and $t$ from the Uniform distribution $U(0,1)$.
2. Setting $z = K^{-1}(t)$, where $K(s) = s - \frac{\varphi(s)}{\varphi'(s)}$, where $\varphi(\cdot)$ is the generator function for the Gumbel Archimedean copula .
3. Setting $u_1 = \varphi^{-1}(\widetilde{u}_1\varphi(z))$ and $u_2 = \varphi^{-1}((1-\widetilde{u}_1)\varphi(z))$.

Secondly, the algorithm proposed by Chambers et al. (1976) (which works for any dimension) consisting on:

1. Generating two independent random variables $d_1$ and $d_2$ from an Exponential distribution $Exp(1)$.
2. Generating a stable distribution S with the parameters $(1/\alpha, 1, 1, 0)$ where $\alpha$ is the estimated parameter of the copula.
3. Setting $u_1 = \varphi^{-1}(\frac{d_1}{S})$ and $u_2 = \varphi^{-1}(\frac{d_2}{S})$

Finally, for the Gaussian and t-Student copulae we propose to use a classical simulation method (see, Devroye 1986). This method ensures that the simulated values for the t-Student have heavier tails than those of the Gaussian. It consists on:

1. Generating two independent random variables $u_1$ and $t$ from the Uniform distribution $U(0,1)$.
2. Setting $X_1 = F^{-1}(u_1)$ and $Z_2 = F^{-1}(t)$, where $F$ is the marginal cdf used to construct the copula.

3. Transforming these variables by rotation about the origin:

$$\begin{aligned} X_1 &= X_1 \\ X_2 &= \rho X_1 + \sqrt{1-\rho^2} Z_2. \end{aligned}$$

4. Transforming back into uniform variables applying cdf:

$$\begin{aligned} u_1 &= F(X_1) \\ u_2 &= F(X_2). \end{aligned}$$

# Chapter 2

# Extreme value distributions

Extreme Value Theory (EVT) deals with the measurement and fitting of events which have a very low probability to occur. According to Coles (2001), extreme values are scarce and very often larger sample sizes are required in order to carry out the estimation. Then, an extrapolation from observable to unobservable levels should be done and EVT provides the procedure in order to do it. EVT is also related to the study of the minimum and maximum of a random variable.

From a historical perspective, EVT was intensively developed at the beginning of 1920, firstly with the pioneer work by Bortkiewicz (1922), who worked with the distribution of the range of random samples from the normal distribution, and later with the contributions of Von Mises (1923) and Fréchet (1927), who calculate the expectation and studied the limiting distribution. Finally, Fisher and Tippet (1928) provided the three possible limiting distributions of an extreme value distribution. Later, Von Mises (1936) provides the conditions for the weak convergence to each of these limit distributions.

In the forties, Gnedenko (1943) unified these results and presented a necessary and sufficient condition for the convergence of the maximum of a sequence of random variables. It was in 1958 when Gumbel (1958) drew the attention of engineers and statisticians interested in a formal application of EVT, but according to Kotz and Nadarajah (2000), Fuller (1914) used EVT for investigating foods, whereas Griffith (1921) applied the same theory to discuss the phenomena of rupture and flow in solids.

Nowadays, EVT is used in different areas of risk management (see, for example, the book by Embrechts et al. 1997;on finance and insurance). Additionally, in the book by Beirlant et al. (2004) there are several examples of the application of EVT in the fields of hydrology, meteorology and natural phenomena. There are also several books on EVT for the two-dimensional case, among others Tiago de Oliveira (1984), Resnick (1987), Galambos (1978) and Coles (2001). Finally, it is worth mentioning the articles by Abdous et al. (1999) and Caperaá et al. (2000) dealing with extremes in terms of the dependency function.

Let $X_1, \ldots, X_n$ be $n$ independent and identically distributed variables, the maximum is:

$$M_n = max\{X_1 \ldots, X_n\},$$

where $X_1, \ldots, X_n$ are independent and identically distributed variables, with distribution function $F$. The exact distribution of $M_n$ is obtained by taking into account that:

$$\begin{aligned} P(M_n \leq x) &= P\big(X_1 \leq x, \ldots, X_n \leq x\big) \\ &= P\big(X_1 \leq x\big) \ldots P\big(X_n \leq x\big) \\ &= F^n(x). \end{aligned} \tag{2.1}$$

As noted by Coles (2001), this is not very useful in practice as the distribution $F$ is unknown. A possibility is to estimate $F$ by using standard statistical methods and then replace it in (2.1), but then it is possible that small errors in the estimation of $F$ could result in bigger errors in the power $F^n$.

An alternative approximation consist of assuming that the distribution is unknown and then finding the distribution family to which $F^n$ belongs to. However, $F^n$ is asymptotically degenerate, i.e. if the the right end point of $F$ is $r(F) = \sup\{t|F(t) < 1\}$, we have:

$$\lim_{n\to\infty} F^n(x) = \lim_{n\to\infty} F_{M_n}(x) = \begin{cases} 0 & \text{si} \quad x < r(F) \\ 1 & \text{si} \quad x \geq r(F). \end{cases} \tag{2.2}$$

This is proved by taking into account that $\forall\epsilon > 0$:

$$
\begin{aligned}
P\big(|M_n - r(F)| \geq \epsilon\big) &= P\Big(M_n \geq r(F) + \epsilon) + P(M_n \leq r(F) - \epsilon\Big) \\
&= P\big(M_n \leq X - \epsilon\big) \\
&= F^n\big(r(F) - \epsilon\big)
\end{aligned}
$$

and as $F(r(F) - \epsilon) < 1$ it can be proved that $M_n$ converges in probability to $r(F)$, (and this implies its convergence in distribution to (2.2). Therefore, the distribution of $M_n$ is asymptotically degenerate.

The problem is solved if the variable $M_n$ is linearly renormalized with

$$
M_n^* = \frac{M_n - a_n}{b_n}, \tag{2.3}
$$

where $a_n > 0$ and $b_n \in \mathbb{R}$ are two sequences. Appropriate values of these constants stabilize the scale and location of $M_n^*$. Fisher and Tippet (1928) prove that:

If two real series $a_n > 0$ and $b_n \in \mathbb{R}$ exist such that:

$$
P\Big(\frac{M_n - a_n}{b_n} \leq x\Big) \to G(x) \quad \text{when} \quad n \to \infty, \tag{2.4}
$$

where $G$ is a non degenerate distribution, then $G$ belongs to the following families:

**Maximal Fréchet**

$$
G(x) = \begin{cases} 0 & \text{if } x \leq \mu \\ \exp(-(\frac{x-\mu}{\sigma})^{-\alpha}) & \text{if} \quad x > \mu. \end{cases} \tag{2.5}
$$

**Maximal Weibull**

$$
G(x) = \begin{cases} \exp(-(\frac{\mu-x}{\sigma})^{\alpha}) & \text{if} \quad x < \mu \\ 1 & \text{si} \quad x \geq \mu. \end{cases} \tag{2.6}
$$

**Maximal Gumbel**

$$G(x) = \exp(-e^{(x-\mu)/\sigma}), \quad x \in \mathbb{R} \tag{2.7}$$

where $\mu \in \mathbb{R}$,$\sigma > 0$ and $\alpha > 0$.

According to Jenkinson (1955), these three types of parametric distributions can be unified in a single extreme value distribution family:

$$\begin{aligned} G_{\mu,\sigma,\xi}(x) &= \exp\left\{-\left(1+\xi\left(\tfrac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right\} & \text{if } \xi \neq 0 \\ G_{\mu,\sigma,\xi}(x) &= \exp\left\{-\exp\left(-\tfrac{x-\mu}{\sigma}\right)\right\} & \text{if } \xi = 0 \end{aligned} \tag{2.8}$$

and it is defined in the set $\{x : 1+\xi(x-\mu)/\sigma > 0\}$, where $\mu$ is the location parameter, $\sigma > 0$ is the scale parameter and $\xi \in \mathbb{R}$ is the shape parameter.

It is also called tail index because it indicates the tail thickness, that is to say, the larger, the heavier the tail of the distribution is. In this chapter we describe the extreme value distribution (EVD) that we will use in posterior analysis.

## 2.1 Domain of attraction

When the limit $G_{\mu,\sigma,\xi}$ exists and it is not degenerate, then we say that the variable $X$ or its distribution function $F$ belongs to the maximum domain of attraction of $G_{\mu,\sigma,\xi}$ and it is denoted by MDA ($G_{\mu,\sigma,\xi}$). Specifically, we look for necessary and sufficient conditions on $F$ so that belongs to the MDA ($G_{\mu,\sigma,\xi}$). Regarding extremes, the problem consist on knowing which is the domain of attraction of a distribution and how to choose the series $a_n$ and $b_n$.

In general, a distribution function $F$ belongs to the domain of attraction of distribution $G_{\mu,\sigma,\xi}$, with normalization parameters $a_n > 0$ and $b_n \in \mathbb{R}$ if:

$$\lim_{n\to\infty} n\bar{F}(a_n x + b_n) = -n\log\left(G_{\mu,\sigma,\xi}(x)\right), \quad x \in \mathbb{R}. \tag{2.9}$$

When $G_{\mu,\sigma,\xi}(x) = 0$ the limit is infinite.

The domains of attraction of some important distributions can be found in (see, Castillo et al. 2004, Embrechts et al. 1997), among them:

1. Pareto, Burr, Cauchy and log-gamma distributions belong to the Fréchet MDA.
2. Normal, log-normal, Gamma and Weibull distributions belong to the Gumbel MDA.
3. Uniform and Beta distributions belong to the Weibull MDA.

In this section, we obtain the domain of attraction of some distributions which have been used in several articles which are part of the core of this thesis, such as the Generalized Champernowne, Weibull and some mixtures of log-normal and Pareto distributions.

### 2.1.1 Champernowne distribution

The Generalized Champernowne distribution was developed by Buch-Larsen et al. (2005). Its probability distribution function (pdf) and cumulative distribution function (cdf) are:

$$f_{\alpha,M,c}(x) = \frac{\alpha(x+c)^{\alpha-1}\big((M+c)^\alpha - c^\alpha\big)}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha}$$

and

$$F_{\alpha,M,c}(x) = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha},$$

with parameters $\alpha > 0$, $M > 0$ and $C \geq 0$. This distribution converges to a Pareto distribution in the tail:

$$f_{\alpha,M,c}(x) \longrightarrow \frac{\alpha[(M+c)^\alpha - c^\alpha]}{x^\alpha},$$

with a restricted domain $x \in [((M+c)^\alpha - c^\alpha)^{1/\alpha}, \infty)$.
Buch-Larsen et al. (2005) analyzed the role of the three parameters associated with the Generalized Champernowne distribution and conclude that $\alpha$ is a parameter that controls the shape of the tail; $M$ is a scale parameter and, finally, $c$ controls the shape

of the distribution near zero, although, it also plays the role of a scale parameter. The Generalized Champernowne distribution has a very flexible shape. It is similar to a Lognormal distribution for low values and tends to a Generalized Pareto for extreme values of the variable (see, Buch-Larsen et al. 2005).

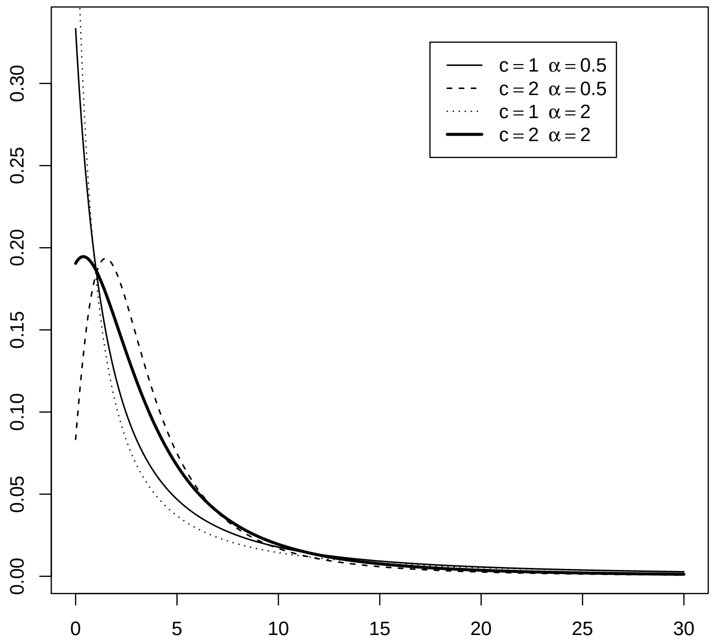


FIGURE 2.1: *Density of the Champernowne distribution with different parameters.*

In Figure 2.1, we visualize the effect of the parameters $c$ and $\alpha$ in the shape of the pdf with $M = 3$. We observe that as $c$ increases the mode of the shape moves to the right and as $\alpha$ decrease the tail is heavier.

In the next proposition we establish what is the MDA of the Champernowne distribution, later we prove this result.

**Proposition 2.1.1.** *The Generalized Champernowne distribution belongs to the MDA of Fréchet.*

*Proof.* (Castillo et al. 2004;p.203) defines the necessary and sufficient condition on $F(x)$ for belonging to the MDA of $G_{\mu,\sigma,\xi}$ defined in expression (2.8):

$$\lim_{x\to 0} \frac{F^{-1}(1-x) - F^{-1}(1-2x)}{F^{-1}(1-2x) - F^{-1}(1-4x)} = 2^{-\xi} \tag{2.10}$$

where $\xi$ is the shape parameter of $G_{\mu,\sigma,\xi}$. The quantile of the Generalized Champernowne distribution with $\alpha > 0, M > 0$ and $c \geq 0$ is:

$$F^{-1}(x) = \left( \frac{x[(M+c)^\alpha - 2c^\alpha] + c^\alpha}{1-x} \right)^{\frac{1}{\alpha}} - c \quad \forall u \in [0,1[. \tag{2.11}$$

Replacing this term in expression (2.10) we obtain:

$$\begin{aligned}
\lim_{x\to 0} \quad & \frac{F^{-1}(1-x) - F^{-1}(1-2x)}{F^{-1}(1-2x) - F^{-1}(1-4x)} = \\
\lim_{x\to 0} \quad & \frac{\left(\frac{(1-x)[(M+c)^\alpha - 2c^\alpha]+c^\alpha}{x}\right)^{\frac{1}{\alpha}} - \left(\frac{(1-2x)[(M+c)^\alpha - 2c^\alpha]+c^\alpha}{2x}\right)^{\frac{1}{\alpha}}}{\left(\frac{(1-2x)[(M+c)^\alpha - 2c^\alpha]+c^\alpha}{2x}\right)^{\frac{1}{\alpha}} - \left(\frac{(1-4x)[(M+c)^\alpha - 2c^\alpha]+c^\alpha}{4x}\right)^{\frac{1}{\alpha}}} \\
= \quad & \frac{1-\frac{1}{2^\alpha}}{\frac{1}{2^\alpha}(1-\frac{1}{2^\alpha})} \\
= 2^\alpha &
\end{aligned}$$

with $\xi = -\alpha < 0$, then the MDA of the Generalized Champernowne distribution is the same as the type Fréchet distribution. Thus a possible choice of normalized parameters is: $a_n = 0$ and

$$b_n = F^{-1}(1 - \frac{1}{n}) = \left( \frac{(1-\frac{1}{n})[(M+c)^\alpha - 2c^\alpha] + c^\alpha}{\frac{1}{n}} \right)^{\frac{1}{\alpha}} - c.$$

The same result can be obtained by using the following condition (see, for example, Embrechts et al. 1997):

$$\begin{aligned}
\lim_{x\to\infty} \frac{1-F(tx)}{1-F(x)} &= \lim_{x\to\infty} \frac{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha}{(tx+c)^\alpha + (M+c)^\alpha - 2c^\alpha} \\
&= t^{-\alpha}.
\end{aligned}$$

□

### 2.1.2 Weibull distribution

This distribution is used in the field of reliability and in the fit of failure times in physical systems. The pdf of the Weibull is:

$$f_{k,\lambda}(x) = \begin{cases} \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}e^{-(\frac{k}{\lambda})^k} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0, \end{cases}$$

and its cdf is:

$$F_{k,\lambda}(x) = 1 - e^{-(\frac{x}{\lambda})^k}.$$

The mean and variance are, respectively:

$$E(X) = \lambda\Gamma(1 + \frac{1}{k}) \quad and \quad Var(X) = \lambda^2\Big[\Gamma(1 + \frac{2}{k}) - \Gamma^2(1 + \frac{1}{k})\Big],$$

such that, $\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt$ is the Euler's Gamma function.

The Weibull distribution has a shape or location parameter $k$ and a scale parameter $\lambda$. A value $k < 1$ indicates that failures decrease with time, when $k = 1$ the failure rate is constant in time and, finally, $k > 1$ indicates that failures increase with time.

In Figure 2.2 we show the pdf of the Weibull. For $k < 1$ we observe that the density tends to $+\infty$ close to $x = 0$ and for $k = 1$ the density tends to a finite value in $x = 0$. When $1 < k < 2$ the density equals zero in the origin and finally when $k > 2$ the distribution is bell-shaped. In general, as $k$ increases the distribution converges to a Dirac distribution with support in $x = \lambda$.

Finally, according to Castillo et al. (2004), this distribution belongs to the MDA of Gumbel.

### 2.1.3 Log-normal-Pareto mixtures

The mixture of these two distributions is interesting because they are flexible and easily fit extreme value data. The approach consist of using the log-normal for the main body

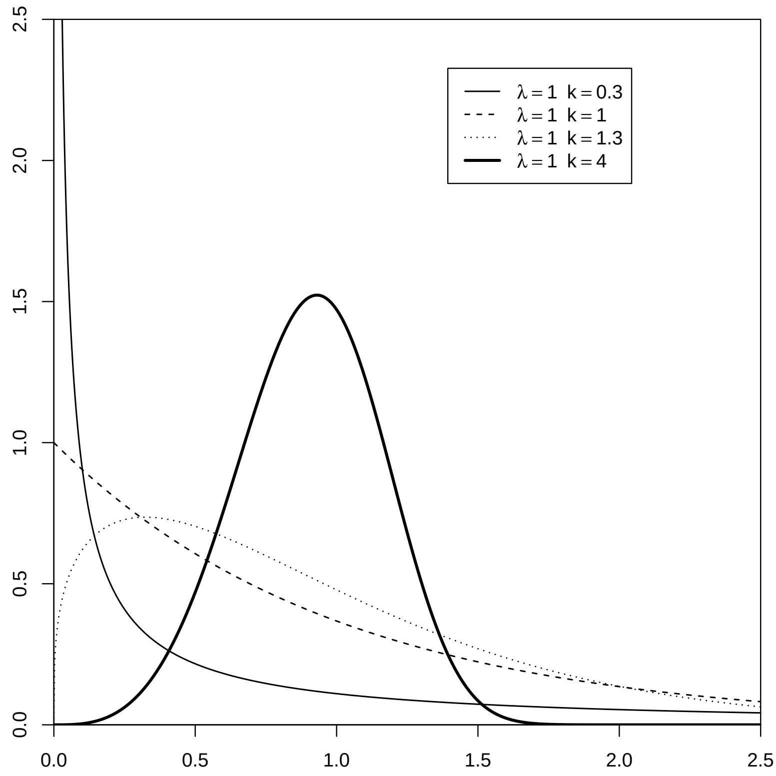


FIGURE 2.2: *Weibull density with different parameters.*

of the distribution and the Pareto for the tail. Both distributions are widely used in insurance and finance (see, for example Klugman et al. 2004). Due to its analytical properties, the log-normal has been used as the basic distribution in the Black-Scholes model in finance. The application of the Pareto has not stopped to increase since firstly used by Pareto (1896) for modeling the income distribution.

The density of the mixture is:

$$\begin{aligned} f(x) &= pf_1(x) + (1-p)f_2(x) \\ f(x) &= p\frac{1}{x\sigma\sqrt{2\pi}}e^{\frac{-1}{2}(\frac{log(x)-\mu}{\sigma})^2} + (1-p)\frac{\rho\lambda^\rho}{(x+\lambda)^{\rho+1}} \quad \forall x > 0, \quad 0 < p < 1 \end{aligned}$$

and its cumulative distribution is:

$$\begin{aligned} F(x) &= pF_1(x) + (1-p)F_2(x) \\ F(x) &= p\int_{-\infty}^{log(x)} \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}(\frac{t-\mu}{\sigma})^2}dt + (1-p)\big(1-(\frac{\lambda}{x+\lambda})^\rho\big). \end{aligned}$$

In the next figure (2.3), we observe that the density of the mixture is flexible and depends on the weight $p$ given to the main body of the variable. As we prove in Chapter 3, the

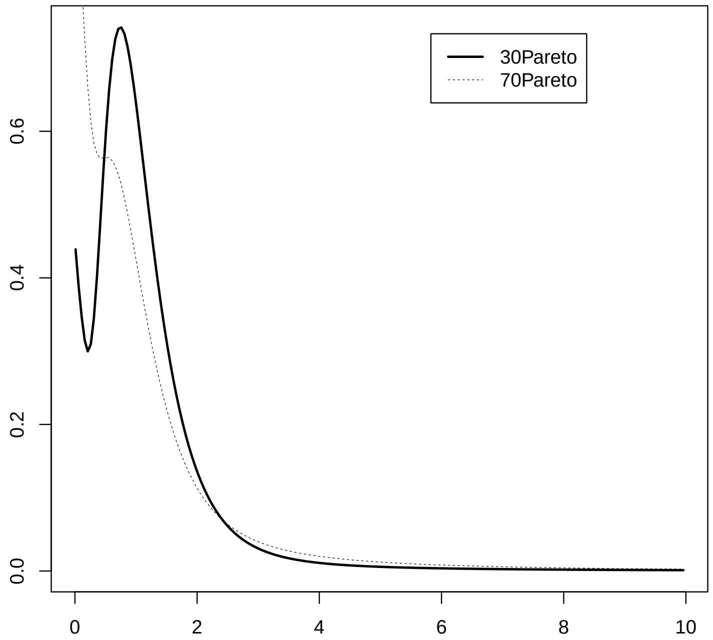


FIGURE 2.3: *Density of the mixture*

mixture for $0 < p < 1$ belongs to MDA of Fréchet.

## 2.2 Bivariate extreme value

The first references about bivariate extreme value distributions are Tiago de Oliveira (1958), Sibuya (1960) and Geffroy (1959). In Beirlant et al. (2004) a review of these approaches can be found. Nevertheless, the expression introduced by Pickands (1981) eclipsed the others for being a general compact formulation in terms of copulae. Nowadays, it is almost impossible to talk about extreme value distributions and copulae without quoting it. Similarly to the univariate case, if we denote $(X_{11}, X_{21}), \ldots, (X_{1n}, X_{2n})$ a two-dimensional sample, the behavior in the extremes is given by the convergence in distribution of the vector $M_n = (M_{X_1,n}, M_{X_2,n})$ where:

$$M_{X_1,n} = \max\left\{X_{11}, \ldots, X_{1n}\right\} \quad \text{and} \quad M_{X_2,n} = \max\left\{X_{21}, \ldots, X_{2n}\right\}.$$

That is, we look for the series $a_{X_1,n}, a_{X_2,n} > 0$ and $b_{X_1,n}, b_{X_2,n} \in \mathbb{R}$ (if exist) such that:

$$\lim_{n\to\infty} P\Big(\frac{M_{X_1,n} - a_{X_1,n}}{b_{X_1,n}} \leq x_1, \frac{M_{X_2,n} - a_{X_2,n}}{b_{X_2,n}} \leq x_2\Big) = M(x_1, x_2) \tag{2.12}$$

As opposed to the univariate case, we observe that $M_n$ does not necessarily belong to the original sample, (see, Coles 2001), and this makes it difficult the generalization to the multivariate case.

In case that the variables $X_1$ and $X_2$ are independent and belong to the domain of attraction of one of the three possible distributions unified in (2.8), the renormalized limit of $M_n$ also belongs to the product of these distributions, which is additionally its MDA. Specifically, if we consider:

$$\begin{aligned} M_n &= (M_{X_1,n}, M_{X_2,n}) \\ A_n &= (a_{X_1,n}, a_{X_2,n}) \quad \text{y} \quad B_n = (b_{X_1,n}, b_{X_2,n}), \end{aligned}$$

we have:

$$\begin{aligned} \lim_{n\to\infty} P\big(\frac{M_n - A_n}{B_n} \leq (x_1, x_2)\big) &= \lim_{n\to\infty} P\Big(\frac{M_{X_1,n} - a_{X_1,n}}{b_{X_1,n}} \leq x_1, \frac{M_{X_2,n} - a_{X_2,n}}{b_{X_2,n}} \leq x_2\Big) \\ &= \lim_{n\to\infty} P\Big(\frac{M_{X_1,n} - a_{X_2,n}}{b_{X_1,n}} \leq x_1\Big) \lim_{n\to\infty} P\Big(\frac{M_{X_2,n} - a_{X_2,n}}{b_{X_2,n}} \leq x_2\Big) \\ &= G_{\mu_{X_1},\sigma_{X_1},\xi_{X_1}}(x_1) G_{\mu_{X_2},\sigma_{X_2},\xi_{X_2}}(x_2). \end{aligned}$$

If the variables are not independent, the two following results (see, Galambos 1978) lead to a limit expression for $M_n$. The first one says, according the theorem of Sklar (1959) that the associated copula to the distribution (2.12) is unique and the second one proves the max-stable property.

1. Any limit that satisfies (2.12) is continuous, additionally the marginals of $M_n$ belong to the three extreme value families for the univariate case.

2. Let $(X_{11}, X_{21}), \ldots, (X_{1n}, X_{2n})$ be a two-dimensional sample, then, if $a_{X_1,n}, a_{X_2,n} > 0$ exists and $b_{X_1,n}, b_{X_2,n} \in \mathbb{R}$ and there exists $M(x_1, x_2)$ defined in 2.12, the distribution function $M$ is not degenerate if only if the marginals belong to the univariate extreme value families and the associated copula satisfies:

$$C(u_1, u_2) = C^k(u_1^{\frac{1}{k}}, u_2^{\frac{1}{k}}),$$

$\forall k \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}$.

The last result does not specify the general expression of the extreme value distribution, although it provides the necessary and sufficient condition in terms of copulae which characterize this types of distributions. Nevertheless, the expression for $M_n$ can be obtained just by knowing the resulting copula and the theorem of Pickands (1981) provides us with this copula when we have standard Fréchet marginals, that is: $F_{X_1}(x_1) = e^{-\frac{1}{x_1}}, x_1 > 0$, $F_{X_2}(x_2) = e^{-\frac{1}{x_2}}, x_2 > 0$.

Before discussing the result of Pickands (1981) it is necessary to standardize the marginal to isolate the dependency structure. According to (Resnick 1987;p253), the transformation of the marginals has no impact on the resulting two-dimensional distribution. Specifically, if we choose the series $a_{X_1,n} = a_{X_2,n} = 0$ and $b_{X_1,n} = b_{X_2,n} = n$ :

$$\begin{aligned} P\Big(\frac{M_{X_1,n}}{n} \leq x\Big) &= P\Big(\frac{M_{X_2,n}}{n} \leq x\Big) = (e^{-\frac{1}{nx}})^n \\ &= e^{-\frac{1}{x}} \end{aligned}$$

and the distribution of the maximum of each marginal does not depend on $n$, therefore if $M_n$ is re-scaled with:

$$M_n^* = \Big(\frac{M_{X_1,n}}{n}, \frac{M_{X_2,n}}{n}\Big)$$

Pickands (1981) proves the following:

If $\big(X_{1i}, X_{i2}\big)$ are independent vectors with standard Fréchet marginal distributions, then if:

$$\lim_{n\to\infty} P\Big(\frac{M_{X_1,n}}{n} \leq x_1, \frac{M_{X_2,n}}{n} \leq x_2\Big) = G(x_1, x_2) \tag{2.13}$$

exists and it is not degenerate, $G$ tends to the following expression:

$$G(x_1, x_2) = e^{-V(x_1,x_2)} \quad x_1 > 0, x_2 > 0,$$

where

$$V(x_1, x_2) = 2\int_0^1 \max\Big(\frac{w}{x_1}, \frac{1-w}{x_2}\Big) dH(w),$$

and $H$ is a distribution function defined in $[0, 1]$ which satisfies:

$$\int_0^1 w dH(w) = \frac{1}{2}.$$

Pickands (1981) also proves that the distribution $G$ can be written as:

$$\log G(x_1, x_2) = -(\frac{1}{x_1} + \frac{1}{x_2})A\Big(\frac{x_2}{x_1 + x_2}\Big) \quad x_1 > 0, x_2 > 0, \tag{2.14}$$

where $A$ is a function defined in $[0, 1]$ which satisfies the conditions: $\max(t, (1-t)) \leq A(t) \leq 1$. It is also related to the function $H$ by the following expression:

$$A(t) = \int_0^1 \max\Big(wt, (1-w)(1-t)\Big) dH(w),$$

and the extreme value copulae is deduced from (2.14)

$$C(u_1, u_2) = \exp\Big(\log(u_1u_2)A(\frac{\log(u_1)}{\log(u_1) + \log(u_2)})\Big) \quad 0 \leq u_1, u_2 \leq 1. \tag{2.15}$$

This copula does not depend on the marginals, therefore the distribution $G$ defined in (2.13) represents the class of two-dimensional extreme values.

**Example 2.2.1.** Here we present some copula functions defined from Pickands dependence function:

1. Independence copula: $A(t) = 1$.
2. Gumbel copula: $A(t) = \left(t^{\theta} + (1-t)^{\theta}\right)^{\frac{1}{\theta}}$, where $\theta \geq 1$.
3. Galambos copula: $A(t) = 1 - \left(t^{-\theta} + (1-\theta)^{-\theta}\right)^{\frac{-1}{\theta}}$, where $\theta > 0$.
4. Hüsler-Reiss copula: $A(t) = (1-t)\Phi(z_{1-t}) + t\Phi(z_t)$, where $z_t = \left(\frac{1}{\theta} + \frac{\theta}{2}\log(\frac{t}{1-t})\right)$, for $\theta \geq 0$, and $\Phi$ being the standard normal cdf.

# Chapter 3

# Testing extreme value copulae to estimate the quantile

We generalize the test proposed by Kojadinovic et al. (2011) 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. Methods are implemented in R.

## 3.1 Introduction

Let $S$ be the sum of $k$ dependent random variables $(X_1, ..., X_k)'$, i.e. $S = X_1 + ... + X_k$. The distribution of $S$ depends on the multivariate distribution, i.e. on the relationship between the random variables $X_j$, $j = 1, ...k$ (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 behavior 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, Bolancé et al. 2008b, Cossette et al. 2002, Denuit et al. 1999, Kaas et al. 2000). 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 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 (1997), copulae are interesting for statisticians due to 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 behavior of variables. According to Nelsen (2006), when coupling the marginals in the joint distribution, the copula captures the link between the two behaviors. 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:

$$H(x_1, x_2) = C\big(F_1(x_1), F_2(x_2)\big) \qquad \forall x_1, x_2 \in \mathbb{R}. \tag{3.1}$$

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 copulae. It is worth mentioning the paper by Genest and Rivest (1993) on inference for bivariate Archimedean copulae, the test proposed in Scaillet (2005) on the positive quadrant dependence hypothesis and, finally, the test of symmetry in bivariate copulae introduced in Genest et al. (2012). Regarding the inference for extreme value copulae, 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 3.2, first, we present our main result and, second, we describe three examples of copulae which are extreme value copulae: Gumbel, Galambos and

Hustler-Reiss. In section 3.3 we describe a real database of auto insurance claims which we use in the empirical application. In section 3.4 we report the results of our empirical study, firstly we apply the test described in section 3.2 and, secondly, we calculate the quantile using different extreme value copulae 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 to log-normal near $0$.

## 3.2 Test for extreme value copulae

We know that the class of extreme value copulae corresponds to the class of $max-stable$ copulae (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 $(X_{i1},X_{i2})$, $\forall i = 1,...n$ be a bivariate sample of $n$ independent and identically distributed observations. We consider the functions:

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

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

$$C_n(u_1, u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\Big(\hat{F}_{1_n}(X_{i1}) \leq u_1, \hat{F}_{2_n}(X_{i2}) \leq u_2\Big), \quad u_1, u_2 \in [0,1]^2, \tag{3.2}$$

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 real-valued functions on $[0,1]^2$ equipped with the uniform metric, i.e. $l^\infty([0,1]^2)$, which is expressed as follows:

$$\begin{aligned} \sqrt{n}\Big(C_n(u_1, u_2) - C(u_1, u_2)\Big) &\rightsquigarrow \mathbb{G}(u_1, u_2) \\ = \mathbb{B}(u_1, u_2) - \partial_1 C(u_1, u_2)\mathbb{B}(u_1, 1) &- \partial_2 C(u_1, u_2)\mathbb{B}(1, u_2), \end{aligned} \tag{3.3}$$

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 (see, Appendix ) and $\mathbb{B}$ is a Brownian bridge (see, Definition in Appendix ) on $[0,1]^2$ with covariance functions:

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

where $\wedge$ is the minimum.

**Proposition 3.2.1.** *If the partial derivatives of a copula $C(u_1, u_2)$ are continuous then for any $r > 0$ we have:*

$$\mathbb{D}_n^r(u_1, u_2) - \mathbb{D}^r(u_1, u_2) \rightsquigarrow \mathbb{C}^r(u_1, u_2) = \mathbb{G}(u_1, u_2) - rC^{r-1}(u_1^{1/r}, u_2^{1/r})\mathbb{G}(u_1^{1/r}, u_2^{1/r}), \tag{3.4}$$

*in $l^\infty([0,1]^2)$. The result in (3.4) 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 3.2.1. 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.* In order to prove the result in Proposition 3.2.1 we consider the function:

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

To find the Hadamard derivative (see, Van der Vaart and Wellner 2000; p. 372) of $\Gamma$ that is denoted by $\Gamma'$, we consider the function:

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

where $t\Delta$ is a function representing a difference, namely, $t$ is a real value and $\Delta$ is a fixed perturbation. Then we calculate $\Gamma'$ 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 3.2.1 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. (2004) and the continuous mapping theorem (see, Van der Vaart and Wellner 2000; Theorem 1.3.6) we obtain:

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

and, finally, applying the delta functional method (see, for example, Van der Vaart and Wellner 2000; Chapter 3.9) we obtain:

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

□

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 (3.4). For hypothesis testing given a fixed $r$, we use a Cramér-von Mises statistic:

$$S_n^r = \int_0^1 \int_0^1 \big(\mathbb{D}_n^r(u_1,u_2)\big)^2 du_1 du_2. \tag{3.5}$$

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

$$S_n^{r_1,\dots,r_p} = \sum_{i=1}^{p} S_n^{r_i}. \tag{3.6}$$

To calculate the critical values we use the method proposed by Van der Vaart and Wellner (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 $\mathbb{D}_n^r$, $\mathbb{D}_n^{r,(1)}, \ldots, \mathbb{D}_n^{r,(N)}$ can be generated, such that

$$\left(\mathbb{D}_n^r, \mathbb{D}_n^{r,(1)}, \ldots \mathbb{D}_n^{r,(N)}\right) \rightsquigarrow \left(\mathbb{D}^r, \mathbb{D}^{r,(1)}, \ldots \mathbb{D}^{r,(N)}\right),$$

where $\mathbb{D}^{r,(1)}, \ldots, \mathbb{D}^{r,(N)}$ 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, $(S_n^{r,(1)}, S_n^{r,(2)} \ldots, S_n^{r,(N)})$ can be calculated by using a numerical approximation of formula (4.3) (see, Kojadinovic et al. 2011), then:

$$\left(S_n^r, S_n^{r,(1)}, S_n^{r,(2)} \ldots, S_n^{r,(N)}\right) \rightsquigarrow \left(S^r, S^{r,(1)}, S^{r,(2)} \ldots, S^{r,(N)}\right),$$

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

3. Obtain the p-value as:

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

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).

### 3.2.1 Three examples of extreme value copulae

In the application presented in next section, we compare three examples of extreme value copulae: Gumbel, Galambos and Hüsler-Reiss, which are described in this section. The functional form of Gumbel copula (see, Gumbel 1960b) 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 $\theta$ is $[0, \infty)$ and the upper tail dependence is $\lambda_U = 2 - 2^{\frac{1}{\theta}}$.

Another example of extreme value copulae 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 $\theta$ is $[0, \infty)$ and $\Phi$ is cdf of the standard Gaussian, $u_1 = -\ln(\hat{u}_1)$ and $u_2 = -\ln(\hat{u}_2)$.

## 3.3 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 3.1 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 3.1 have right skewness. In Figure 3.1 we show the histograms representing the shape of the distributions associated with the variables Cost1 and Cost2.

| 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 |

TABLE 3.1: Descriptive statistics.

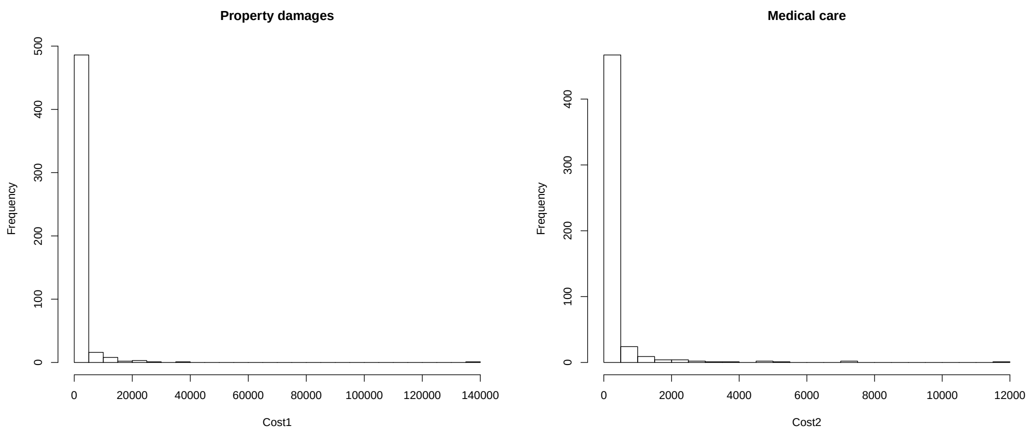


FIGURE 3.1: Histograms.

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 3.2 we show the K-Plot, that compare the order 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.

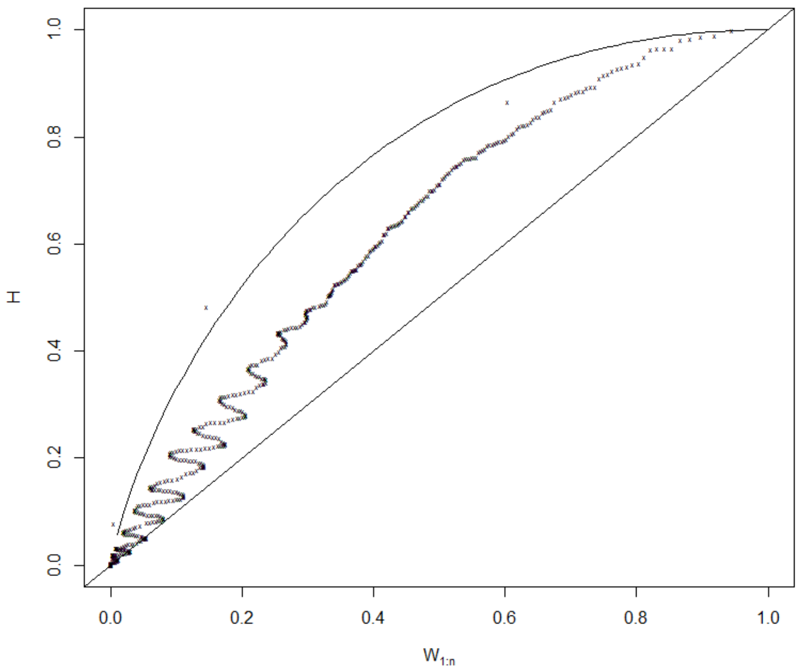


FIGURE 3.2: K-Plot associated to copula of (Cost1, Cost2).

## 3.4 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 (3.6), firstly with $r = 3, 4, 5$. We have estimated the significance level of the test statistic using the method proposed by Van der Vaart and Wellner (2000). In total, we generated $1000$ independent copies of $S_n^{3,4,5}$. The results are shown in Table 3.2 and allow us to conclude that the analyzed bivariate data are generated by an extreme value copula. We estimate the parameters of the three extreme value

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

TABLE 3.2: Cramér-Von Mises statistic.

copulae described in section 3.2.1: Gumbel, Galambos and Hüsler-Reiss. In Table 3.3

we show the estimated parameters for these three copulae together with those obtained for the Gaussian and the t-Student copulae. To estimate the dependence parameter of Gaussian, Gumbel, Galambos and Hüsler-Reiss copulae 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 Grønnrberg and Hjort (2014). The corresponding results are presented in Table 3.3. 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.

| | Gaussian | t-Student* | Gumbel | Galambos | Husler-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 | -212.369 | -217.000 | -246.383 | -243.3305 | -237.854 |
| BIC | -208.119 | -208.500 | -242.133 | -239.0805 | -233.604 |
| CIC | 105.804 | 108.040 | 123.1476 | 121.827 | 119.829 |
| Kendall Tau=0.4252. *d.f.= 9.6442 | | | | | |

TABLE 3.3: *Copula estimation results.*

Once the dependency structure is estimated, the next step is to estimate the marginal distribution functions. Considering the histograms in Figure 3.1, 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,

[1]The cdf of the modified Champernowne distribution is:

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

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).

Alemany et al. 2013, Bolancé 2010, Bolancé et al. 2008a). In Table 3.4 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. For evaluating the risk of

| | Log-normal | Champernowne |
|---|---|---|
| CDFs | $\int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad x \geq 0$ | $\frac{(x+c)^\delta - c^\delta}{(x+c)^\delta + (H+c)^\delta - 2c^\delta}, \quad x \geq 0$ |
| $X_1$=Cost1 | $\mu = 6.4437, \sigma = 1.3349,$ $AIC = 8448.90$ and $BIC = 8452.72$ | $\delta = 1.3271, H = 677, c = 0$ $AIC = 8448.16$ and $BIC = 8453.90$ |
| $X_2$=Cost2 | $\mu = 4.3755, \sigma = 1.5189,$ $AIC = 9425.13$ and $BIC = 9428.96$ | $\delta = 1.1622, H = 88, c = 0$ $AIC = 6443.72$ and $BIC = 6449.45$ |

TABLE 3.4: Maximum likelihood estimation of marginal distributions.

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

1. We generate the pseudo-random sample $\left(\hat{U}_{1i}, \hat{U}_{2i}\right)$, $\forall i = 1, ..., r$, from the bivariate copulae whose estimated parameters are shown in Table 3.3.

2. Using the inverse of the marginal CDFs we calculate $\left(\hat{X}_{1i} = F_1^{-1}(\hat{U}_{1i}), \hat{X}_{2i} = F_2^{-1}(\hat{U}_{2i})\right)$, $\forall i = 1, ..., l$, where the sample volume $l$ is large.

3. We calculate $\hat{S}_i = \hat{X}_{1i} + \hat{X}_{2i}$, $\forall i = 1, ..., l$ and we estimate $q_\alpha(S)$ empirically from the generated pseudo-sample. We generate $l = 10,000$ samples.

In Table 3.5 we show the results of the estimations of $q_\alpha$ for $\alpha = 0.95, 0.99, 0.995, 0.999$. On the first row of Table 3.5 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 $\alpha$; below we show the same $q_\alpha(S)$ that have been estimated by the Monte Carlo simulation method for the five copulae considered here. We note the importance of using an extreme value copula and extreme value marginal distributions when the data indicate this behavior.

In Table 3.5 we show that by using log-normal marginal distributions, the estimated quantile is below the empirical quantile for the five copulae considered here. Therefore,

| $\alpha$ | 0.95 | 0.99 | 0.995 | 0.999 |
|---|---|---|---|---|
| Empirical | 7905.60 | 24821.14 | 28420.87 | 92112.93 |
| | Log-normal | | | |
| Normal | 6635.43 | 15628.80 | 20762.77 | 39733.89 |
| t-Student | 6547.52 | 16638.18 | 22521.18 | 39547.10 |
| Gumbel | 6432.02 | 15464.97 | 22011.38 | 40001.21 |
| Galambos | 6429.16 | 15471.40 | 22066.00 | 39925.67 |
| Husler-Reiss | 6421.03 | 15465.13 | 22122.11 | 39841.56 |
| | Champernowne | | | |
| Normal | 7237.59 | 25504.18 | 38682.44 | 110082.26 |
| t-Student | 7302.17 | 25740.93 | 42223.50 | 117447.02 |
| Gumbel | 7264.83 | 23944.80 | 41461.74 | 119401.41 |
| Galambos | 7253.17 | 24056.95 | 41409.72 | 118982.01 |
| Husler-Reiss | 7241.50 | 24103.04 | 41107.54 | 118539.74 |

TABLE 3.5: Quantiles of total loss.

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.

# Chapter 4

# Extreme value copulae and marginal effects: the bounds of the Value-at-Risk

We have analyzed two fundamental issues when we use copulae with extreme value marginal distributions for estimating the total risk of loss, that has been generated by a multivariate random vector of dependent losses (or risk factors). Firstly, we describe a statistic that allows us to test if the data belongs of a family of extreme value copulae. Secondly, we have studied the effect of using different extreme value marginal distribution for estimating the risk. As risk measure we use the Value-at-Risk (VaR). Finally, in order to control the risk, we estimate the bounds of the VaR for the aggregate loss using two methods. Results have been obtained by using a bivariate sample of losses from a real database of auto insurance claims.

## 4.1 Introduction

In finance and insurance the total loss is usually generated by a multivariate random vector of $k$ dependent losses (or risk factors) $(X_1, ..., X_k)'$, i.e. $S = X_1 + ... + X_k$ is the total risk of loss which depends on the relationship between these risk factors. In this regard, there are studies that have analyzed the stochastic behavior 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, Bolancé et al. 2008b, Cossette et al. 2002, Denuit et al. 2001; 1999, Kaas et al. 2000). The aim of this paper is to estimate the Value at Risk (VaR) of a loss generated by a bivariate random vector with marginal extreme value distributions. We compare the results obtained when using different copulae and we analyze how, given the marginal distributions, the selected copula may affect the risk estimate. Copulae allow us to model a wide range of dependency structures. As noted by Fisher (1997), copulae are interesting to 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 representing dependency structures, even when the marginals follow extreme value distributions (see, Guillén et al. 2007) for nonparametric methods with application in insurance). The aim of this paper is to test if our data have been generated by an extreme value copula. We know that the choice of the marginal may be crucial to model the dependency behavior of variables. According to Nelsen (2006), in the coupling of the joint distribution with marginals, the copula captures the link between them. In recent years, copulae have been widely used by analysts for risk quantification and have many practical applications in the financial and actuarial area. Note that the total cost in insurance results from the sum of dependent costs with marginal extreme value distributions, and modeling the dependency structure is crucial for risk estimation. After adjusting the dependency structure, the relationship between the joint distribution and the marginals is clear. Therein lies the importance of adjustment. This relationship is established in the fundamental theorem proposed by Sklar (1959). This theorem shows that a bivariate distribution function $H$ of a random vector of variables $(X_1, X_2)$ with marginal distribution functions $F_1$ and $F_2$ includes a copula $C$ according to the following expression:

$$H(x_1, x_2) = C\big(F_1(x_1), F_2(x_2)\big) \qquad \forall x, y \in \mathbb{R}. \tag{4.1}$$

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 was the motivation for developing tests for the adequacy of copulae. It is worth mentioning the paper by Fermanian et al. (2004) on the weak convergence of the empirical copula, the contrast of Scaillet (2005) for the positive quadrant dependence hypothesis, and finally, the test of symmetry in bivariate copulae introduced by Genest et al. (2012). Regarding the contrast for extreme value copulae is worth mentioning the test introduced by Ghoudi et al. (1998) derived from the transformation of the bivariate distribution of extreme values, and the one proposed by Kojadinovic et al. (2011), which uses the definition of $max - stable$ as null hypothesis. We use the same statistic of Kojadinovic et al. (2011) to test if our bivariate data come from an extreme value copula. In section 4.2 we describe a test that allows us to know if the data belongs of a family of extreme value copulae and, later, we describe three copulae which are extreme value copulae: Gumbel, Galambos and Hüstler Reiss. In section 4.3 we describe the Champernowne distribution for marginals which converges to a Pareto distribution in the tail, while looking more like a log-normal distribution near $0$. Later, in section 4.4, we analyze the bounds of the estimated VaR, and in section 4.5 we report an application.

## 4.2 Test for extreme value copulae

One way to know if our data has an extreme value copula behavior or not, is to test the $max - stable$ property. A copula is called a $max - stable$ copula 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:

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

Let $(X_{1i}, X_{2i})$, $\forall i = 1, ...n$ be a bivariate sample of $n$ independent and identically distributed observations. We consider the functions:

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

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

$$C_n(u_1,u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\Big(F_{1n}(X_{1i}) \leq u, F_{2n}(X_{2i}) \leq u_2\Big), \quad u_1,u_2 \in [0,1]^2. \tag{4.2}$$

Kojadinovic et al. (2011) prove the weak convergence of $\mathbb{D}_n^r(u_1, u_2)$ to a Gaussian process in[2] $l^\infty([0,1]^2)$ under null hypothesis. Nevertheless Bahraoui et al. (2014c) prove that weak convergence is true under null and alternative hypothesis. To calculate the critical values we use the **Multipier** method proposed by Van der Vaart and Wellner (2000), consisting on generating independent copies of a distribution with the same behavior. 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). In particular, we use the Cramer-von Mises statistic:

$$S_n^r = \int_0^1 \int_1^1 \left(\mathbb{D}_n^r(u_1, u_2)\right)^2 dudv. \tag{4.3}$$

We compare thee examples of extreme value copulae: Gumbel, Galambos and Hustler Reiss. The Gumbel copula is an extreme value copula (see, Genest et al. 2011, Juri and Wüthrich 2002)), and its functional form is given by:

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

where $\theta \in [1, +\infty)$ is the parameter controlling the dependency structure. Finally, the dependence is perfect when $\theta \to \infty$ and we have independence 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. Galambos copula was proposed by Galambos (1975) and has the following form:

$$C(u_1, u_2) = u_1 u_2 \exp\Big(\big[(-\log u_1)^{-\theta} + (-\log u_2)^{-\theta}\big]^{-1/\theta}\Big),$$

where the range of $\theta$ is $[0, \infty)$ and the upper tail dependence is $\lambda_U = 2 - 2^{\frac{1}{\theta}}$. The Hüstler-Reiss copula was developed by Hüsler and Reiss (1989) and its functional form is given by:

$$C(u_1, u_2) = \exp\Big( - \hat{u_1}\Phi\big[\frac{1}{\theta} + \frac{1}{2}\theta \log(\frac{\hat{u_1}}{\hat{u_2}})\big] - \hat{u_2}\Phi\big[\frac{1}{\theta} + \frac{1}{2}\theta \log(\frac{\hat{u_2}}{\hat{u_1}})\big]\Big),$$

[2] The space of all bounded real-valued functions on $[0,1]^2$.

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

## 4.3 Marginals

First of all, we consider the Gumbel and Weibull marginal distribution functions resulting from the generalized theory of extreme values. We also considered the log-normal distribution, which has interesting theoretical properties, and, finally, the modified Champernowne distribution, which has been studied by Buch-Larsen et al. (2005) and Bolancé et al. (2012b) in the context of the transformed kernel density estimation for extremes values distribution. Also, (see, Bolancé et al. 2012a, Nielsen et al. 2012) proposed to use Champernowne distribution to estimate operational risk, where the shape of the distributions of the data are similar to those analyzed in this work.

**Definition 4.3.1.** The modified Champernowne distribution is defined for $x \geq 0$ and its density function is given by:

$$f_{\delta,M,c}(x) = \frac{\delta(x+c)^{\delta-1}\left((M+c)^{\delta} - c^{\delta}\right)}{(x+c)^{\delta} + (M+c)^{\delta} - 2c^{\delta}}$$

with parameters $\delta > 0$, $M > 0$, y $c \geq 0$ and distribution function:

$$F_{\delta,M,c}(x) = \frac{(x+c)^{\delta} - c^{\delta}}{(x+c)^{\delta} + (M+c)^{\delta} - 2c^{\delta}}.$$

The asymptotic behavior of the modified Champernowne distribution is the same as the original, and it is similar to a log-normal distribution for values near 0 and also converges to a Pareto distribution when $x \longrightarrow \infty$.

## 4.4 Bounds of the VaR

One of the risk measures most commonly used in the financial and actuarial field is the VaR. (*Value-at-Risk*). The VaR represents the quantile of a distribution in a given level.

**Definition 4.4.1.** Let $X$ a random variable with cumulative distribution function $F$, then:

$$VaR_\alpha(X) = \inf \{x \in R | F_X(x) \geq \alpha\}.$$

Next, we find the bounds of the VaR for two reasons. On the one hand, because this risk measure violates the subadditivity condition. On the second hand, because in the general case we will not know the joint distribution function or the copula. We use two methods to find the bounds of the VaR. The Bootstrap method and the technique proposed by Mesfioui and Quessy (2005) for the case of unknown marginals (see, Appendix ). Finally, when the dependency structure is unknown and the marginals are known we can use the limits resulting from the Fréchet bounds.

$$W = \max\big(u_1 + u_2 - 1, 0\big) \leq C(u_1, u_2) \leq M = \min(u_1, u_2) \quad \forall u_1, u_2 \in [0, 1].$$

The upper bound $M$ is always a copula in any dimension, however, this is not the case for the lower bound in the multivariate dimension. For the bivariate case, where $F_1$ and $F_2$ are two marginal cumulative distribution functions of $X_1$ and $X_2$, we have the following result by (Embrechts et al. 2003, Embrechts and Puccetti 2006) :

$$\underline{VaR}_\alpha \leq VaR_\alpha(S) \leq \overline{VaR}_\alpha,$$

where

$$\underline{VaR}_\alpha = \sup_{u_1+u_2=\alpha} \big(F_1^{-1}(u) + F_2^{-1}(u_2)\big)$$

$$\overline{VaR}_\alpha = \inf_{u_1+u_2=\alpha+1} \big(F_1^{-1}(u_1) + F_2^{-1}(u_2)\big).$$

In the case of a Champernowne marginal distribution function with parameters $\delta > 0, M > 0$ y $c \geq 0$ we have:

$$F^{-1}(u) = \left(\frac{u[(M+c)^\delta - 2c^\delta] + c^\delta}{1-u}\right)^{\frac{1}{\delta}} - c \quad \forall u \in [0, 1[.$$

By using the fact that the Champernowne density is non increasing after a certain point (it has the same behavior as a Pareto in the tail of the distribution), i.e. there is a point

$x^*$ such that the density is decreasing, $\forall x \geq x^*$ $f^{'}(x) \leq 0$ and

$$(F^{-1})^{''}(u) = \frac{-f^{'}oF^{-1}(u)}{(foF^{-1}(u))^3} \geq 0 \quad \forall u \geq F(x^*).$$

This implies that $F^{-1}$ is convex on $\alpha \leq F(x^*) \leq u \leq 1$. You can find the minimum for the upper bound of VaR seeking $u^*$ such that $h^{'}(u^*) = 0$, and solve it numerically. We get the following result given the convexity of $F^{-1}$:

$$\underline{VaR}_\alpha = \max\left(F_1^{-1}(\alpha) + F_2^{-1}(0), F_2^{-1}(\alpha) + F_1^{-1}(0)\right) \tag{4.4}$$

$$= \max\left(F_1^{-1}(\alpha), F_2^{-1}(\alpha)\right), \tag{4.5}$$

being the upper bound equal to:

$$\overline{VaR}_\alpha = \inf_{\alpha \leq u \leq 1} h(u), \tag{4.6}$$

such that $h(u) = F_1^{-1}(u) + F_2^{-1}(\alpha + 1 - u)$. If $F_1 = F_2$ and $u = \frac{\alpha+1}{2}$ we obtain the same results for $\underline{VaR}_\alpha$ and $\overline{VaR}_\alpha$ as those obtained by Embrechts et al. (2013), who proposed an algorithm for calculating bounds of the $VaR_\alpha$ for high-dimensional portfolios.

## 4.5 Results

In this section we report the results that we have obtained in an empirical application of the methodology that we have presented. 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. 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, 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, 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.

| Cost | Average | Std.Dev. | Kurtosis | Skewness | Min | Max | Median | J-B |
|---|---|---|---|---|---|---|---|---|
| **Cost1** | 182.80 | 686.80 | 297.10 | 15.65 | 13.00 | 137900.00 | 677.00 | 1941868.27 |
| **Cost2** | 283.92 | 863.17 | 82.02 | 8.04 | 1.00 | 11855.00 | 88.00 | 151969.73 |
| **Total Cost** | 211.20 | 752.00 | 286.40 | 15.27 | 32.00 | 149800.00 | 789.00 | 1804742.80 |

TABLE 4.1: *Descriptive statistics and normality test.*

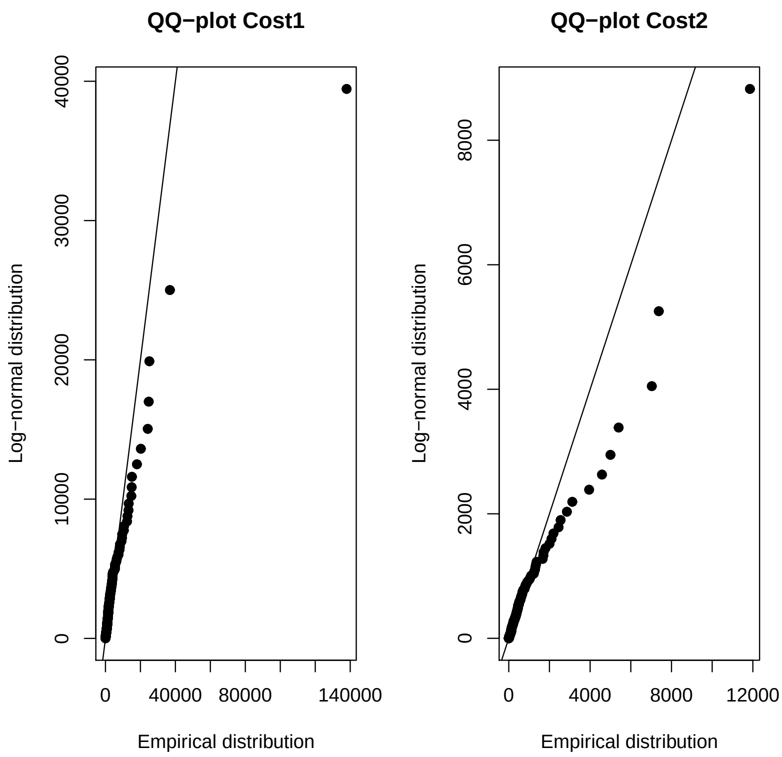


FIGURE 4.1: *Log-normal QQ-Plot.*

In Table 4.1 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. The kurtosis and skewness are very high if we compare them with the corresponding values for the normal distribution. We test normality by using a Jarque-Bera test, and the corresponding p-values are $< 2,21.10^{-16}$ for the three costs, therefore, we reject normality. QQ-plots help us to look at the behavior of the empirical distribution in the tail. Namely, the theoretical quantiles of a given distribution are compared with the quantiles of the empirical

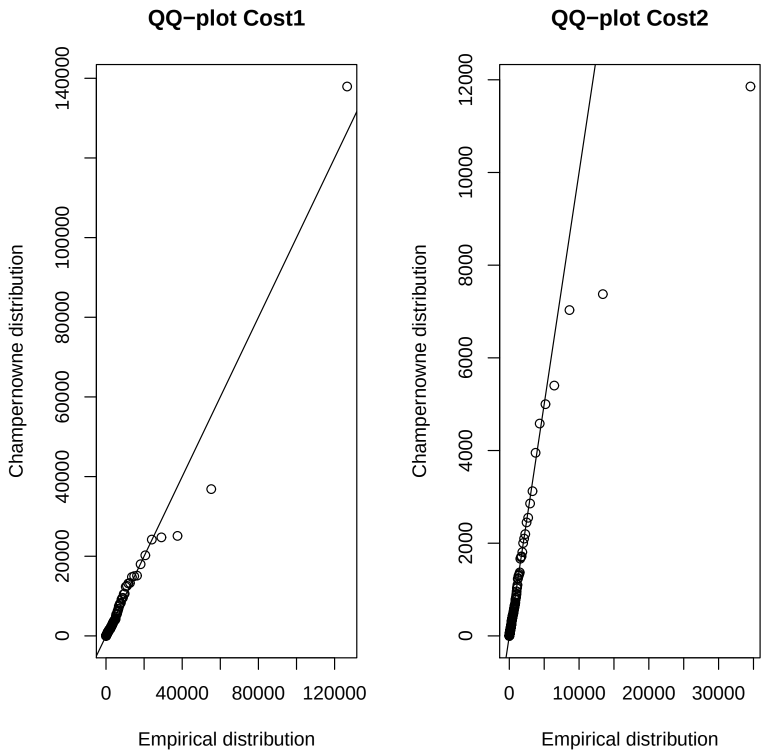


FIGURE 4.2: *Champernowne QQ-Plot.*

distribution. When there is a good fit the empirical quantiles coincide with the theoretical ones. By using a log-normal distribution (Figure 4.1), we note that the empirical quantiles coincide with the theoretical ones for low values of the costs, and then deviate upwards indicating the existence of extreme values. Therefore, in this case heavy tailed distributions may provide a good fit, as shown in Figure 4.2 where the QQ-Plots for a theoretical Champernowne distribution are represented.

### 4.5.1 K-Plot and test of extreme value copula

The K-Plot is a visual method that allows us to analyze if our bivariate data have been generated by an extreme value copula. In Figure 4.3 we show the K-Plot, that compare the order in real data ($H$ pseudo-observations generated by the modified bivariate empirical distribution) with the order supposing that the data have been generated by the independence copula ($W$, expected pseudo-observations, see Genest and Boies (2003)). 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 according to the increasing values of $W$, the data is 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.

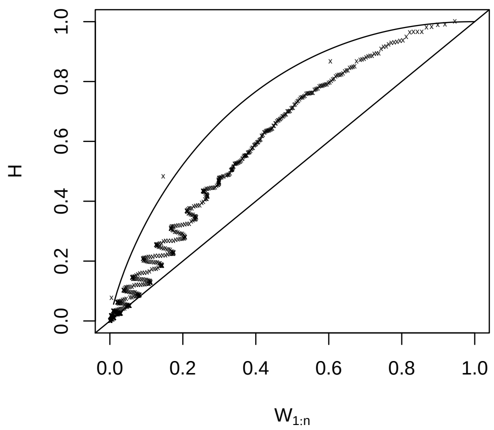


FIGURE 4.3: *K-Plot associated to copula of (Cost1,Cost2).*

The value of the Cramer-Von Mises statistic in (4.3), that is used to test if our data have been generated by an extreme value copula, is equal to 0.26795 with p-value equal to 0.1773227. We do not reject the null hypothesis that the copula associated with this type of data is an extreme value copula at $10\%$ level. For selecting the copula, we can use two known statistical information criteria: the Akaike information criterion that is:

$$AIC = -2 \log L(\theta, u, v) + 2k$$

and the value of the Bayesian information criterion, given by:

$$BIC = -2 \log L(\theta, u, v) + \log(n)k,$$

where $k$ is the number of parameters to be estimated and $L$ the maximum likelihood function. Also, we can use the copula information criteria $CIC$ propose by Grønnrberg and Hjort (2014), the corresponding results are presented in Table 4.2, in practice we observe that AIC and CIC are very similar. Therefore, we conclude that the Gumbel copula reflects the best the dependency structure in our data.

| Copula | AIC | BIC | CIC |
|---|---|---|---|
| Gumbel | -246.3839 | -242.1339 | 123.1476 |
| Galambos | -243.9354 | -238.9922 | 121.827 |
| Hustler-Reiss | -239.6841 | -234.7410 | 119.829 |

TABLE 4.2: *Information Criteria.*

### 4.5.2 Bounding the empirical VaR

In the first row of Table 4.3 we provide the empirical values of the VaR of the aggregate loss $S = X_{Cost1} + Y_{Cost2}$ for different confidence levels $\alpha$. The second and third rows provide the confidence intervals at $95\%$ level of the VaR $_S(\alpha)$ by using Bootstrap. In the last two rows we provide the bounds of the VaR by using the upper and lower limits from the Mesfioui and Quessy (2005), method (MQ).

| $\alpha$ | 0.90 | 0.91 | 0.92 | 0.93 | 0.94 | 0.95 | 0.96 | 0.97 | 0.98 | 0.99 | 0.995 | 0.999 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $VaR_\alpha(S)$ | 3715.90 | 4088.72 | 4519.36 | 5059.16 | 5928.22 | 7905.60 | 8796.40 | 11133.75 | 14975.62 | 24821.14 | 28420.87 | 92112.93 |
| $VaR_\alpha(Cost1) + VaR_\alpha(Cost2)$ | 3597.50 | 3995.83 | 4344.64 | 4672.67 | 6147.62 | 7355.30 | 9351.24 | 11726.40 | 15614.90 | 24329.78 | 30970.58 | 95217.51 |
| $Boostrap.VaR.Inf_\alpha(S)(95\%)$ | 3062.90 | 3316.60 | 3633.04 | 3961.91 | 4261.94 | 5397.50 | 6121.84 | 7597.36 | 9590.88 | 13204.74 | 15857.45 | 24248.93 |
| $Boostrap.VaR.Sup_\alpha(S)(95\%)$ | 4745.60 | 5510.67 | 6383.60 | 7492.32 | 8316.74 | 9805.10 | 12634.32 | 14865.21 | 20270.08 | 25394.25 | 78981.72 | 138863.50 |
| $MQ.VaR.Inf_\alpha(S)(95\%)$ | 2111.92 | 2111.92 | 2111.92 | 2111.92 | 158.67 | 338.27 | 533.80 | 752.29 | 1007.47 | 1334.91 | 1563.87 | 1867.32 |
| $MQ.VaR.Sup_\alpha(S)(95\%)$ | 21119.20 | 23465.78 | 26399.00 | 30170.29 | 32712.77 | 35811.31 | 39986.76 | 46073.21 | 56230.11 | 79036.09 | 111173.50 | 246470.71 |

TABLE 4.3: *Bounding the empirical VaR.*

Figure 4.4 displays the confidence bounds of the empirical VaR. We note that the method by Mesfioui and Quessy provides large values, especially for the upper bound. The Bootstrap technique provides narrower intervals of the empirical VaR. If we plot the VaR of **Cost1** plus the VaR of **Cost2** (thin solid line) we firstly note that their values are within the confidence interval of the aggregate loss at the $95\%$ level and, secondly, we clearly see that the condition of sub-additivity is violated.

### 4.5.3 Simulation of the VaR

In Table 4.4 we calculate the VaR of the total costs by simulation. We use $10,000$ replications and compare the Gumbel copula with other classical implicit copulae: Gaussian

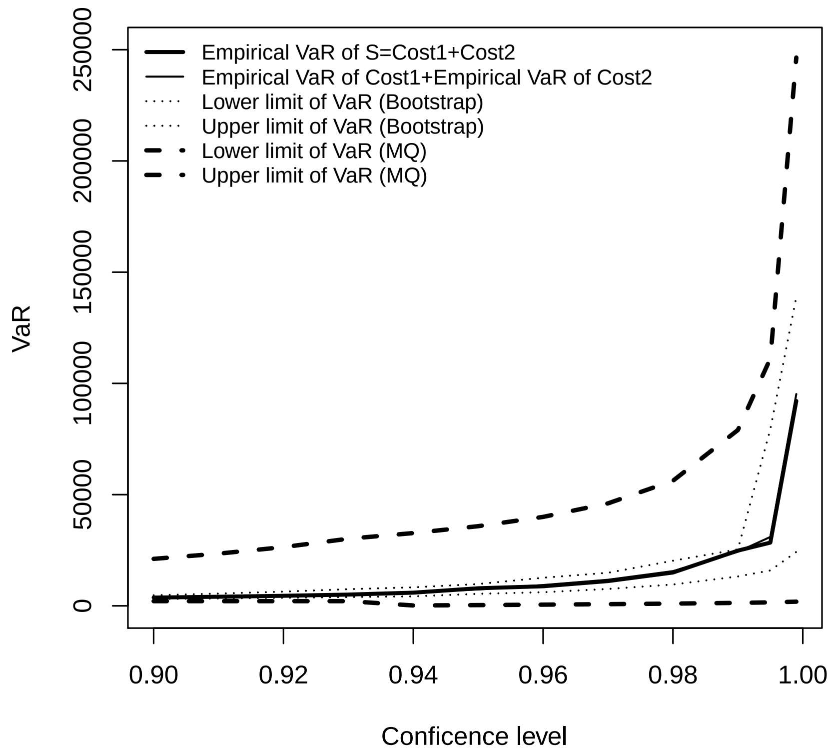


FIGURE 4.4: *Bounds of the empirical VaR.*

and t-Student. Thus, we can analyze the importance of using an extreme value copula when the data indicate this behavior.

The simulation was performed by using the Itau method of R (based on the inversion of the Kendall's $\tau$) to estimate the parameters of alternative copulae and by using the maximum likelihood estimation for parameters of marginal distributions.

We note that by using the Normal copula with Champernowne marginal distribution we obtain a highest value of the VaR at the $95\%$ level. However, if we increase the level of the VaR to $99.5\%$ or to $99.9\%$, as required by Basel III and Solvency II, the highest VaR is obtained for the Gumbel copula with Champernowne marginal distribution, i.e. with the extreme value copula and the heavier tail distribution, the corresponding VaR is near the upper limit from the Mesfioui and Quessy (2005) in Table 4.3 and Figure 4.4. Namely, the only extreme value copula appearing in Table 4.4 is the Gumbel copula, which is also the one previously chosen for the dependency structure of the data. If we consider the VaR as a risk measure for solvency requirements, we can say that the Champernowne distribution ensures a greater capital requirement at the level recommended by Solvency II. Also, QQ plots in Figures 4.1 and 4.2 show that the best fit for both marginals is obtained with the Champernowne distribution.

| Marginals | Log-normal | Weibull | Gumbell | Champernowne |
|---|---|---|---|---|
| Copula | **Normal** | | | |
| VaR (95) | 6635.43 | 11751.43 | 5189.15 | 7237.59 |
| VaR (99.5) | 20762.77 | 26762.99 | 8377.31 | 38682.44 |
| VaR (99.9) | 39733.89 | 23613.46 | 11223.19 | 110082.26 |
| Copula | **t-Student** | | | |
| VaR(95) | 6547.52 | 11889.94 | 5167.74 | 7302.17 |
| VaR(99.5) | 22521.18 | 29476.47 | 8507.31 | 42223.50 |
| VaR (99.9) | 39547.10 | 23269.82 | 11133.09 | 117447.02 |
| Copula | **Gumbel** | | | |
| VaR (95) | 6432.12 | 12065.31 | 5301.55 | 7264.83 |
| VaR (99.5) | 22011.39 | 32434.25 | 7398.00 | 41461.74 |
| VaR (99.9) | 40001.21 | 26504.37 | 12041.83 | 119401.41 |

TABLE 4.4: *Estimated values of the VaR.*

In Figure 4.5 we plot the lower and upper bounds when the associated copula is unknown by calculating the limits numerically by using (4.5) and (4.6). We add in the same graph the VaR resulting from the simulation with a Gumbel copula and Champernowne marginals. We observe that the simulated VaR is identical to the lower bound up to some point, and from this point on it is located within the limits corresponding to the unknown copula. This result shows that the Gumbel copula is suitable for modeling the bivariate behavior of our data.

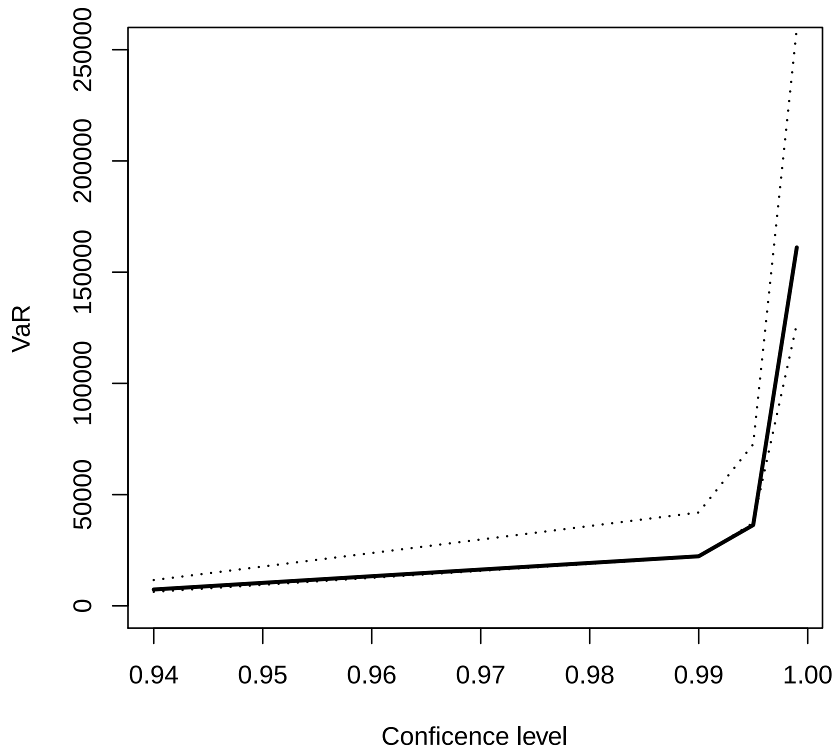


FIGURE 4.5: *Bounds estimated supposing unknown copula (dashed line) and the VaR estimated with Gumbel copula and Champernowne marginals (solid line).*

# Chapter 5

# Quantifying the risk using copulae with nonparametric marginals

We show that copulae and kernel estimation can be mixed to estimate the risk of an economic loss. We analyze the properties of the Sarmanov copula. We find that the maximum pseudo-likelihood estimation of the dependence parameter associated with the copula with double transformed kernel estimation to estimate marginal cumulative distribution functions is a useful method for approximating the risk of extreme dependent losses when we have large data sets. We use a bivariate sample of losses from a real database of auto insurance claims.

## 5.1 Introduction

A major challenge in finance and in insurance is estimating the risk of loss. Contributions by McNeil et al. (2005), Jorion (2007) and Bolancé et al. (2012b) as well as articles of Dhaene et al. (2006), Dowd and Blake (2006) and Alemany et al. (2013), among many others, have focused on just this question. It is known that when the total loss $S$ is generated by a multivariate random vector of dependent losses (or risk factors), i.e. $S = X_1 + ... + X_k$, the total risk of loss depends on the relationship between these risk factors. In this regard, there are studies that have analyzed the stochastic behavior 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, Bolancé et al. 2008b, Cossette et al. 2002, Denuit et al. 2001; 1999, Kaas et al. 2000). The copulae allow us to model a wide range of dependency structures. As noted by Fisher (1997), copulae are interesting to 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 representing dependency structures, even when the marginals follow extreme value distributions. The aim of this paper is to show how the combination of copulae for modeling dependency structures and nonparametric methods to fit marginals is a good tool to estimate the total risk of loss. We know that the choice of the marginal may be crucial to model the dependence behavior of variables. According to Nelsen (2006), in the coupling of the joint distribution with marginals, the copula captures the aspect that links them. A natural nonparametric method for estimating the cumulative distribution function (cdf) is the empirical distribution and its use has been previously analyzed for fitting copulae. Genest et al. (1995) analyzed the consistency of maximum pseudo-likelihood estimation (MPLE) of the parameters of the copula, given empirical marginals. Later, Kim et al. (2007) showed the greater robustness of the MPLE compared to that of the parametric methods: maximum likelihood estimation (MLE) and the so-called inference function for margins (IFM) (see also, Kojadinovic and Yan 2010). However, although obtaining an empirical distribution is very simple, it does not have any smoothing behavior, which can be a problem if the sample is not very large, since we can not represent all values of the distribution if we use the Monte Carlo method to estimate the total risk of loss. Moreover, the fact that the empirical distribution takes the value one in the maximum of the sample does not allow us to calculate the pseudo-likelihood function; therefore a correction needs to be used to avoid this problem. The proposed correction (see, Genest et al. 1995, Kim et al. 2007) increases the efficiency of the estimation of the marginal distributions but incorporates some bias. This is the same as the situation when we compare the kernel smoothing estimation of the cdf with the empirical cdf (without correction)-the kernel smoothing estimation has some bias but, instead, it has lower variance than the empirical. Therefore, we propose using a smoothed estimate of the marginal distributions in the maximum pseudo-likelihood based method to estimate the parameters of the copula. We show that estimating marginals using double transformed kernel estimation (DTKE), as proposed by Alemany et al. (2013), is the method that best fits our purpose especially when we have large data sets. In this study, we use well-known copulae belonging to different families (implicit and explicit copulae, Archimedean and non Archimedean copulae and copulae with and without tail dependence) that represent a wide range of dependence structures. Moreover, we propose using the Sarmanov copula (see, Sarmanov 1966), where the dependence structure is not strictly separated from the marginal distributions, i.e. the marginals are incorporated into the dependence structure. We show how this fact affects the value of the dependence parameter and we analyze tail dependence when the marginal distributions are extreme values. Hernández-Bastida et al. (2009) and Hernández-Bastida and Fernández-Sánchez (2012) used the Sarmanov distribution to obtain a Bayes premium in a collective risk model, in both works parametric marginal distributions were assumed. Yang and Hashorva (2013) analyzed the Sarmanov distribution and proved asymptotic independence when the marginals are extreme value distributions; here we prove that the tail dependence only exists if extreme value marginal distributions are type I (Gumbel). To estimate the total risk of loss, which is obtained as the aggregation of the dependent losses, we use Monte Carlo simulation; Value-at-Risk (VaR) is the selected risk measure. Artzner et al. (1999) discussed other risk measures. They stated that, in practice, the Tail Value-at-Risk (TVaR) is preferred due to its better properties. However, the VaR is used both as an internal risk management tool and as a regulatory measure of risk exposure to calculate capital adequacy requirements in financial and insurance institutions. Moreover, the TVaR can not be calculated if the marginal distributions do not have finite first moment, as occurs with a class of distributions with heavy tail. We apply our proposed method to a real insurance database corresponding to a random bivariate sample of the cost of claims in automobile insurance, which are right skewed and have extreme values. Throughout this study we analyze the bivariate case. The method proposed by Aas et al. (2009) to generalize to multivariate copula is applicable to our case but it is not straightforward. In section 5.2 we define our notation and we describe the VaR estimation using the Monte Carlo simulation method. In section 5.3 we describe the nonparametric estimation of the marginals. Later, in section 5.4, we present the copulae used in our analysis and we describe how we simulate bivariate random samples from each copula. For evaluating the finite sample properties of our proposed Monte Carlo simulation method using nonparametric marginal distributions, in section 5.5 we show the results of a simulation study and in section 5.6 we report an application.

## 5.2 Estimating VaR from bivariate copulae using Monte Carlo simulation

Let $(X_1, X_2)$ be a bivariate vector of random variables representing losses, with marginal cdfs $F_1$ and $F_2$ and bivariate cdf $F$. We are interested in estimating the VaR of the total loss, i.e. $S = X_1 + X_2$, with confidence level $\alpha$. This can be defined as follows,

$$VaR_\alpha(S) = \inf\{s, F_S(s) \geq \alpha\} = F_S^{-1}(\alpha), \tag{5.1}$$

where $S$ is a random variable with probability distribution function (pdf) $f_S$ and cdf $F_S$. Following Sklar's theorem (see, Sklar 1959), if $F_1$ and $F_2$ are continuous distributions, there exists a single copula $C_\theta : [0,1]^2 \longrightarrow [0,1]$, with dependence parameter $\theta$, such that:

$$F(x_1, x_2) = C_\theta(u_1, u_2), \forall x_1, x_2 \in \Re, \tag{5.2}$$

where $u_1 = F_1(x_1)$ and $u_2 = F_2(x_2)$ are, respectively, two values of two random variables $U_1$ and $U_2$ with uniform$(0,1)$ distribution. Let us assume that $(X_{1i}, X_{2i})$, $i = 1, ..., n$, denotes a bivariate sample from the bivariate loss random vector $(X_1, X_2)$, then the Monte Carlo procedure to estimate the $VaR_\alpha(S)$ is described below:

- Estimating with non-parametric method cdfs $\hat{F}_1$ and $\hat{F}_2$.
- Replacing $U_{1i} = \hat{F}_1(X_{1i})$ and $U_{2i} = \hat{F}_2(X_{2i})$, $i = 1, \ldots, n$, in (5.2) in order to estimate the parameter of the copula $\hat{\theta}$ by maximizing the likelihood function associated with copula (pseudo-maximal-likelihood), as we describe later in section 5.3.
- Simulating from the copula the pairs $(\tilde{U}_{1j}, \tilde{U}_{2j})$, $j = 1, \ldots, r$, where $r$ is the number of simulated pairs, as we describe later in section 5.4.
- Calculating simulated losses $\tilde{X}_{1j} = \hat{F}_1^{-1}(\tilde{U}_{1j})$ and $\tilde{X}_{2j} = \hat{F}_2^{-1}(\tilde{U}_{2j})$, $\forall\ j = 1, \ldots, r$.

- Calculating simulated total losses $\tilde{S}_j = \tilde{X}_{1j} + \tilde{X}_{2j}$, $j = 1, \ldots, r$, and estimating $VaR_\alpha(S)$ empirically once a large number of simulated data $r$ are available.

The empirical estimation of the VaR is:

$$\widehat{VaR}_\alpha (S)_n = \inf \left\{ s, \widehat{F}_{S_n} (s) \geq \alpha \right\}, \tag{5.3}$$

where $\widehat{F}_{S_n}$ is the empirical estimation of $F_S$.

## 5.3 Fitting copulae with nonparametric approximation of marginal cdfs

To fit the copula to a bivariate sample $(X_{1i}, X_{2i})$, $i = 1, ..., n$, we use a maximum pseudo-likelihood estimation (MPLE) of the dependence parameter, i.e. first we generate the named pseudo-data $(U_{1i}, U_{2i})$ (see, Genest et al. 1995; 2011) from a nonparametric estimator of the cdf and, second, we estimate the dependence parameter $\hat{\theta}$, maximizing the logarithm of the pseudo-likelihood function given $(U_{1i}, U_{2i})$:

$$ln(L)_\theta = \sum_{i=1}^{n} ln\{c_\theta(U_{1i}, U_{2i})\}, \tag{5.4}$$

where $c_\theta$ is the density associated with the cdf of the copula $C_\theta$. We denote as $\hat{\theta}^*$ the parameter that maximizes the pseudo log-likelihood function. A natural nonparametric method for estimating the cdf is the empirical distribution,

$$\widehat{F}_{l_n}(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_{li} \leq x), \ \ l = 1, 2. \tag{5.5}$$

The empirical distribution is very simple, but it cannot extrapolate beyond the maximum observed data point, where the value of the empirical distribution is $1$, causing numerical problems in calculating the values of $c_\theta(1, U_{2i})$ or $c_\theta(U_{1i}, 1)$. A way of addressing these difficulties (see, Genest et al. 1995) is to correct the empirical distribution in the following way:

$$\tilde{F}_{l_n}(x) = \frac{n}{n+1}\widehat{F}_{l_n}(x), l = 1, 2 \tag{5.6}$$

and generate pseudo-data as $(U_{1i}, U_{2i}) = (\tilde{F}_{1_n}(X_{1i}), \tilde{F}_{2_n}(X_{2i})), i = 1, \ldots, n$. The inefficiency of $\widehat{F}_{l_n}$ is known (see, Azzalini 1981); however, after including the correction, the properties of the empirical distribution change, reducing the variance and adding some bias to the estimation in finite samples. It is easy to deduce that the mean square error (MSE) of $\tilde{F}_{l_n}$ (that we denote as Emp) is:

$$\begin{aligned} E\left\{\tilde{F}_{l_n}(x) - F_l(x)\right\}^2 \sim & \tfrac{F_l(x)[1-F_l(x)]}{n} - \tfrac{2n+1}{(n+1)^2}\tfrac{F_l(x)[1-F_l(x)]}{n} \\ & + \left(\tfrac{1}{(n+1)^2}\left[F_l(x)\right]^2\right), \; l = 1, 2. \end{aligned} \tag{5.7}$$

In expression (5.7) the first term is the MSE of the unbiased estimator $\widehat{F}_{l_n}$, the second term is the reduction in variance after correction and the third term is the square bias. The result of the sum of the second and third terms in (5.7) is negative; therefore, in finite samples the MSE of $\tilde{F}_{l_n}$ is lower than the MSE of $\widehat{F}_{l_n}$ and this difference is $o(n^{-2})$. An alternative to $\tilde{F}_{l_n}$ is the classical kernel estimation (CKE) of cdf. This estimation does not cause numerical problems in calculating the likelihood in expression (5.4) because does not take value 1. The CKE of cdf $F_l$ is:

$$\widehat{F}_l(x) = \tfrac{1}{n}\sum_{i=1}^{n} K\left(\tfrac{x-X_{li}}{b}\right), \; l = 1, 2, \tag{5.8}$$

where $K(t) = \int_{-\infty}^{t} k(u)du$ and $k(\cdot)$ is a pdf, which is known as the kernel function. Some examples of kernel functions are the Epanechnikov and the Gaussian kernels (see, Silverman 1986). Parameter $b$ is the bandwidth or the smoothing parameter, it controls the smoothness of the cdf estimate. The larger $b$ is, the smoother the resulting cdf. The properties of the kernel estimator of the cdf were analyzed by Azzalini (1981). Asymptotically, the MSE of CKE is:

$$\begin{aligned} E\left\{\widehat{F}_l(x) - F_l(x)\right\}^2 \sim & \tfrac{F_l(x)[1-F_l(x)]}{n} - f_l(x)\tfrac{b}{n}\left(1 - \int_{-1}^{1} K^2(t)\,dt\right) \\ & + b^4\left(\tfrac{1}{2}f_l'(x)\int t^2 k(t)\,dt\right)^2, \; l = 1, 2. \end{aligned} \tag{5.9}$$

As with the corrected empirical distribution, the kernel estimation of the cdf also incorporates some bias in the finite sample, the last term in expression (5.9), and it improves the efficiency of the empirical distribution. Azzalini (1981) shows that the optimal bandwidth is $o(n^{-\frac{1}{3}})$ and the sum of the second and third terms of MSE is negative. Then, in finite samples, MSE of $\widehat{F}_l$ is lower than MSE of $\widehat{F}_{l_n}$, the difference is $o(n^{-\frac{4}{3}})$ and this is larger, in absolute value, than the difference between $\tilde{F}_{l_n}$ and $\widehat{F}_{l_n}$. The difficulty of CKE is that the optimal bandwidth depends on the true distribution, which is unknown. Therefore, we propose to estimate this smoothing parameter using the *rule-of-thumb* (see, Silverman 1986), based on the minimization of a weighted mean integrated squared error, so that more weight is given to the accuracy of the estimate in the part of the domain near the quantile in the right tail. The value of this bandwidth using the Epanechnikov kernel is (see, Alemany et al. 2013):

$$\hat{b} = \sigma^{\frac{5}{3}} \left(\frac{8}{3}\right)^{\frac{1}{3}} n^{-\frac{5}{3}}, \tag{5.10}$$

where $\sigma$ can be replaced by a consistent estimation of the standard deviation. Alemany et al. (2013) showed that the transformed kernel estimation of the cdf is more efficient than the classical kernel estimation, although it incorporates a larger bias that can be minimized by selecting an appropriate transformation. These authors proposed double transformed kernel estimation, a method that requires an initial transformation of the data $T(X_i) = Z_i$, where we obtain a transformed variable distribution that is close to uniform$(0,1)$. Afterwards, the data are transformed again using the inverse of the distribution function of a $Beta\,(3,3)$, $M^{-1}(Z_i) = Y_i$. The resulting variable, once the double transformation has been made, has a distribution that is close to a $Beta\,(3,3)$ (see, Bolancé 2010, Bolancé et al. 2008a). The double transformation kernel estimator (DTKE) is:

$$\begin{aligned} \tilde{F}_l(x) &= \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{M^{-1}(T(x)) - M^{-1}(T(X_{il}))}{b}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{y - Y_{il}}{b}\right), \quad l = 1, 2, \end{aligned} \tag{5.11}$$

where the first transformation $T(\cdot)$ is the cdf of the modified Champernowne distribution:

$$T(x) = \frac{(x+c)^{\delta} - c^{\delta}}{(x+c)^{\delta} + (H+c)^{\delta} - 2c^{\delta}}, \quad x \geq 0, \tag{5.12}$$

with parameters $\delta > 0$, $H > 0$ and $c \geq 0$. If we analyze the properties of this distribution, we can conclude that it has a very flexible shape. It is similar to a Lognormal distribution in the low values and tends to a Generalized Pareto in the extreme values. The estimation of transformation parameters is performed using the maximum likelihood method described in Buch-Larsen et al. (2005). The second transformation $M^{-1}(\cdot)$ is the inverse of the following $Beta\,(3,3)$ cdf:

$$M(x) = \frac{3}{16}x^5 - \frac{5}{8}x^3 + \frac{15}{16}x + \frac{1}{2}, -1 \leq x \leq 1, \tag{5.13}$$

that, as Terrell (1990) showed, minimizes the asymptotic mean integrated squared error of the CKE of cdf, i.e. this distribution is the best fit using CKE. The distribution associated with the double transformed variables has been established, with the cdf defined in (5.13). It is crucial here, that this method provides an accurate way to obtain the smoothing parameter based on the minimization of the MSE at $\tilde{F}_l(x)$, $l = 1, 2$ (see, Alemany et al. 2013):

$$b_x = \left( \frac{m(x) \int K(t)\,[1 - K(t)]\,dt}{\left(m'(x) \int t^2 k(t)\,dt\right)^2} \right)^{\frac{1}{3}} n^{-\frac{1}{3}}, \tag{5.14}$$

where $m$ is the pdf of the $Beta\,(3,3)$. We propose DTKE to generate pseudo-data.

**Remark 5.3.1.** Let $F_l(t), l = 1, 2$, be a continuous distribution function, to compare the fit of the nonparametric estimation of marginal cdfs we estimate the Integrated Squared Error (ISE) (see, Scott 2001, Shirahata and Chu 1992):

$$ISE_l = \int_{-\infty}^{+\infty} \left(F_l^*(t) - F_l(t)\right)^2 dt, \;\; l = 1, 2,$$

as

$$\widehat{ISE}_l = \int_{-\infty}^{+\infty} \left(F_l^*(t)\right)^2 dt - \frac{2}{n}\sum_{i=1}^{n} F_{l_i}^*(X_{li}), \;\; l = 1, 2,$$

where $F_l^*$ is a nonparametric estimation of marginal cdf $l$ (Emp, CKE or DTKE) and $F_{l_i}^*$ denotes the nonparametric estimation obtained from the data without observation $X_{li}$, i.e. the leave-one-out estimation.

## 5.4 Copulae under analysis

The copulae analyzed here belong to different families (implicit and explicit copulae, Archimedean and non Archimedean copulae and copulae with and without tail dependence) and represent a wide range of dependence structures. Below, we define the copulae that are compared in our analysis and we describe how we generate bivariate uniform samples from each $C_\theta(u_1, u_2)$. First, we focus on the Sarmanov copula that was defined above in Bairamov et al. (2011). This copula has special characteristics and in this case we detect that the dependence parameter $\omega$ is not scale independent, i.e. it depends on the values of the variables analyzed. Furthermore, we test the upper and lower tail dependence coefficients. Second, we describe the other well-known copulae that we use in our analysis. Finally, we describe the methods that we use to generate pairs of uniforms random variables from the different copulae analyzed.

### 5.4.1 Sarmanov copula

The Sarmanov copula belongs to the family of implicit copulae. Let $(X_1, X_2)$ be a bivariate vector of random variables representing losses with marginal pdfs $f_1$ and $f_2$. Also, let $\phi_1$ and $\phi_2$ be two bounded non-constant functions such that:

$$\int_{-\infty}^{+\infty} f_1(t)\phi_1(t)dt = 0, \qquad \int_{-\infty}^{+\infty} f_2(t)\phi_2(t)dt = 0,$$

then the bivariate pdf introduced by Sarmanov (1966) is defined as:

$$h(x_1, x_2) = f_1(x_1)f_2(x_2)\big(1 + \omega\phi_1(x_1)\phi_2(x_2)\big).$$

From Sklar's theorem, it can be shown that the associated copula can be expressed as (see, Bairamov et al. 2011):

$$C_\omega(u_1, u_2) = u_1 u_2 + \omega \int_0^{u_1} \int_0^{u_2} \phi_1(F_1^{-1}(t))\phi_2(F_2^{-1}(s))dtds \tag{5.15}$$

and its density is:

$$c_\omega(u_1, u_2) = 1 + \omega\phi_1(F_1^{-1}(u_1))\phi_2(F_2^{-1}(u_2)), \tag{5.16}$$

where dependence parameter $\theta = \omega$ satisfies the condition $1 + \omega\phi_1(x_1)\phi_2(x_2) \geq 0$ for all $x_1$ and $x_2$. This parameter is related to the correlation between $X_1$ and $X_2$ (if it exists). As Lee (1996) shows, the correlation between $X_1$ and $X_2$ is:

$$\rho(X_1, X_2) = \omega \frac{\nu_1 \nu_2}{\sigma_1 \sigma_2},$$

where $\nu_1 = E(X_1\phi_1(X_1))$ and $\nu_2 = E(X_2\phi_2(X_2))$ and $\sigma_1$, $\sigma_2$ are standard deviation of $X_1$ and $X_2$, respectively. When we take $\phi_1(x_1) = 1 - 2F_1(x_1)$ and $\phi_2(x_1) = 1 - 2F_2(x_2)$, we have the classical Farlie-Gumbel-Morgenstern (FGM) copula described below, where the dependence parameter is in the range $-1/3 \leq \omega \leq 1/3$ and, in this case, it is scale independent. Another special case is when we consider functions of the type:

$$\phi_1(x_1) = x_1 - \mu_1 \quad \text{and} \quad \phi_2(x_2) = x_2 - \mu_2, \tag{5.17}$$

where $\mu_1 = E(X_1)$ and $\mu_2 = E(X_2)$. Lee (1996) shows that, if the support of $f_1$ and $f_2$ is contained in $[0, 1]$, then the range of the dependence parameter is:

$$\max\left(\frac{-1}{\mu_1\mu_2}, \frac{-1}{(1-\mu_1)(1-\mu_2)}\right) \leq \omega \leq \min\left(\frac{1}{\mu_1(1-\mu_2)}, \frac{1}{(1-\mu_1)\mu_2}\right). \tag{5.18}$$

We extend the result in (5.18) in Proposition 5.4.1.

**Proposition 5.4.1.** *Let $X_1$ and $X_2$ be two random variables with pdfs $f_1$ and $f_2$, respectively. If the support of $f_1$ is contained in $[a, b]$ and that of $f_2$ is contained in $[c, d]$,* *where a,b,c and d are finite real numbers, then:*

$$\max\left(\frac{-(b-a)(d-c)}{(\mu_1-a)(\mu_2-c)},\frac{-(b-a)(d-c)}{(b-\mu_1)(d-\mu_2)}\right)\leq(b-a)(d-c)\omega\leq\min\left(\frac{(b-a)(d-c)}{(\mu_1-a)(d-\mu_2)},\frac{(b-a)(d-c)}{(b-\mu_1)(\mu_2-c)}\right). \tag{5.19}$$

*Proof.* If we consider the variables $Y_1=\frac{X_1-a}{b-a}$, and $Y_2=\frac{X_2-c}{d-c}$, we observe that $Y_l\in[0,1]$ for $l=1,2$ and $\rho(Y_1,Y_2)=\rho(X_1,X_2)$. The results in (5.19) can be deduced immediately if we replace $\mu_{Y_1}=\frac{\mu_X-a}{b-a}$, $\mu_{Y_2}=\frac{\mu_Y-c}{d-c}$ in (5.18), and finally substituting $\omega_{Y_1,Y_2}$ by $(b-a)(d-c)\omega$. □

**Remark 5.4.1.** Using Emp and CKE we have $a,c=0$ and $b\approx max(X_{i1})$, $d\approx max(X_{i2})$, $i=1,...,n$. Also, when $b,d\rightarrow\infty$ the dependence parameter is $\omega\approx0$.

Tail dependence is important when the marginals are extreme value distributions. We discuss the dependence in both tails of the Sarmanov copula and, in particular, we analyze the dependence in the right (upper) tail when the marginals are extreme value distributions.

The lower tail dependence coefficient $\lambda_L$ is:

$$\lambda_L=\lim_{u\rightarrow0^+}\frac{C_\theta(u,u)}{u},$$

where $u=u_1=u_2$. If $\lambda_L$ is in $(0,1]$ $C_\theta$ has lower tail dependence.

The upper tail dependence coefficient is:

$$\lambda_U=\lim_{u\rightarrow1^-}\frac{1-2u+C_\theta(u,u)}{u}.$$

If $\lambda_U$ is in $(0,1]$ $C_\theta$ has upper tail dependence.

**Proposition 5.4.2.** *If $C_\omega$ is a bivariate Sarmanov copula then $\lambda_L=0$.*

*Proof.*
$$\frac{C_\omega(u,u)}{u}=u+\omega u^{-1}\int_0^u\int_0^u\phi_1(F_1^{-1}(t))\phi_2(F_2^{-1}(s))dtds,$$

where $\phi_l$, $l = 1, 2$, are bounded, then $|\phi_l| \leq M_l$, for some finite positive number $M_l$, $l = 1, 2$. The result is deduced taking the limit $u \to 0^+$ in the inequality:

$$\left|\frac{C(u,u)}{u}\right| \leq u + |\omega| u^{-1} u^2 M_1 M_2.$$

□

**Theorem 5.4.1.** *If the marginals $F_l$, $l = 1, 2$, follow a generalized extreme value (GEV) distribution type II or type III then upper tail dependence of the Sarmanov copula does not exist. If the marginals $F_l$, $l = 1, 2$, have a GEV distribution type I then $\lambda_U = 0$.*

*Proof.* Proof of Theorem 5.4.1: The general expression of the cdf of a GEV distribution is:

$$F(x, \mu, \sigma, \xi) = \exp\left(-1 + \xi\left(\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right) \tag{5.20}$$

and its mean is:

$$E(X) = \begin{cases} \mu + \sigma\frac{\Gamma(1-\xi)-1}{\xi} & \text{if} \quad \xi \neq 0, \xi < 1 \\ \mu + \sigma\gamma & \text{if} \quad \xi = 0 \\ \infty & \text{if} \quad \xi \geq 1 \end{cases}, \tag{5.21}$$

where $\Gamma(\Delta)$ is Euler's gamma function and $\gamma$ is Euler's constant. The inverse of cdf $F$ in (5.20) is expressed as:

$$F^{-1}(t) = \begin{cases} \mu + \frac{\sigma}{-\xi}\big(1 - (-\log(t))^{-\xi}\big) & \text{if} \quad \xi \neq 0 \\ \mu - \sigma\big(\log(-\log(t)\big) & \text{if} \quad \xi = 0 \end{cases}.$$

If we assume that $E(X)$ in (5.21) is finite, i.e. $\xi \leq 1$, then for $F_l$, $l = 1, 2$ with $\xi_l = 0$:

$$\begin{aligned}
\frac{1 - 2u + C(u,u)}{u} &= \frac{(1-u)^2 + \omega\int_0^u\int_0^u (F_1^{-1}(t) - \mu_1)(F_2^{-1}(s) - \mu_2)dtds}{u} \\
&= \frac{(1-u)^2 + \omega\sigma_1\sigma_2\int_0^u\int_0^u \big(\log(-(\log(t)) - \gamma\big)\big(\log(-(\log(s) - \gamma\big)dtds}{u} \\
&= \frac{(1-u)^2 + \omega\sigma_1\sigma_2\big(\mu\log(-\log(\mu)) - l(u) + \gamma\mu\big)^2}{u},
\end{aligned}$$

where $l(x) = \int_0^x \frac{1}{\log(t)}dt$.

For $0 < \xi_l < 1$, $l = 1, 2$ we obtain:

$$\begin{aligned}\frac{1-2u+C(u,u)}{u} &= \frac{(1-u)^2 + \omega \int_0^u \int_0^u (F_1^{-1}(t) - \mu_1)(F_2^{-1}(s) - \mu_2)dtds}{u} \\ &= I_1 + I_2,\end{aligned}$$

where $I_1 = \frac{(1-u)^2}{u}$ and

$$\begin{aligned} I_2 &= \frac{\omega \frac{\sigma_1}{\xi_1}\frac{\sigma_2}{\xi_2} \int_0^u \int_0^u \big((-\log(t))^{-\xi_1} - \Gamma(1-\xi_1)\big)\big((-\log(s))^{-\xi_2} - \Gamma(1-\xi_2)\big)dtds}{u} \\ &= \frac{\omega}{\mu}\frac{\sigma_1}{\xi_1}\frac{\sigma_2}{\xi_2}\big(\Gamma(-\xi_1+1, -\log(u)) - \mu\Gamma(1-\xi_1)\big)\big((\Gamma(-\xi_2+1, -\log(u)) - \mu\Gamma(1-\xi_2)\big),\end{aligned}$$

where $\Gamma(a,x) = \int_x^\infty t^{a-1}e^{-t}dt$ the is incomplete Gamma function. Calculating the limit when $u \to 1^-$ the upper tail dependence coefficient is:

$$\lambda_U = \begin{cases} 0 & \text{if} \quad \xi_l = 0 \\ \infty & \text{if} \quad 0 < \xi_l < 1 \end{cases}.$$

When $\xi_l \geq 1$ tail dependence can not be calculated. □

### 5.4.2 Other copulae

The most popular implicit copulae are the Gaussian and the Student t copulae, where $C_\theta$ is, respectively, the standard normal bivariate cumulative distribution function and standard Student t (with $\upsilon$ degree of freedom) bivariate cumulative distribution function, with correlation coefficient $\rho$, i.e. $\theta = \rho$. These copulae represent symmetric dependence structures. The main difference between them is that the Student t copula has heavier tails than those of the Gaussian copula. Additionally, the Gaussian copula does not present tail dependence while the Student t copula has both, lower and upper tail dependence. From the explicit and the Archimedean family of copulae we use the most popular: Gumbel (Gumbel 1960b), Clayton (Clayton 1978) and Frank (Frank 1979) copulae. This class of copulae has a simple closed form and its structure depends only on the dependence parameter. Moreover, they are not derived from a bivariate distribution function. Among these copulae, Gumbel is an extreme value copula (see, Genest et al. 2011, Juri and Wüthrich 2002), and its functional form is given by:

$$C_\theta(u_1, u_2) = \exp\Big(-\big[(-\ln(u_1))^\theta + (-\ln(u_2))^\theta\big]^{1/\theta}\Big), \tag{5.22}$$

where $\theta \in [1, +\infty)$ is the parameter controlling the dependence structure. Finally, the dependence is perfect when $\theta \to \infty$ and we have independence when $\theta = 1$. For the Gumbel copula, it is well known that $\lambda_L = 0$ and $\lambda_U = 2 - 2^{\frac{1}{\theta}}$, i.e. the Gumbel copula has upper tail dependence. The Clayton copula, unlike the Gumbel, it is not an extreme value copula. Its functional form is given by:

$$C_\theta(u, v) = \big(u^{-\theta} + v^{-\theta} - 1\big)^{-1/\theta}, \tag{5.23}$$

where $\theta > 0$. In this case the perfect dependence structure is achieved when $\theta \to \infty$, and independence is achieved when $\theta \to 0$. In contrast to the Gumbel copula, the Clayton copula has lower tail dependence, in this case $\lambda_L = 2^{-\frac{1}{\theta}}$ and $\lambda_U = 0$. The Frank copula is defined by the parameter $\theta \in (-\infty, 0[\cup]0, +\infty)$, and is given by:

$$C_\theta(u_1, u_2) = -\frac{1}{\theta} \ln\left(1 - \frac{(1 - e^{\theta u_1})(1 - e^{\theta u_2})}{1 - e^{-\theta}}\right).$$

The upper and lower tail dependence coefficients of the Frank copula are $0$. An implicit, non-Archimedean copula without tail dependence is the Farlie-Gumbel-Morgenstern (FGM) copula (see, Farlie 1960, Gumbel 1960a, Morgenstern 1956). As noted above, this copula is a particular case of the Sarmanov copula and takes the form:

$$C_\theta(u_1, u_2) = u_1 u_2 \Big(1 - \theta(1 - u_1)(1 - u_2)\Big) \quad \theta \in [-1, 1]. \tag{5.24}$$

As Nelsen (2006) describes, this is the only copula with a quadratic functional form in $u_1$ and $u_2$. When $\theta = 0$, we have a special case of independence. Schucany et al. (1978) analyze the dependence structure of the FGM copula and prove that its correlation parameter obeys the condition $|\theta| \leq 1/3$. In their paper, this authors claim that the upper bound for the correlation coefficient is so small as to restrict the usefulness of the model. Generally, to generate a bivariate random variable from copulae we use a procedure based on the conditional distribution of the random vector $(U_1, U_2)$ (see, Nelsen

2006). However, the conditional distribution for the Gumbel copula is not invertible. So, to generate random variables from this copula we follow the algorithm proposed by Chambers et al. (1976). Finally, for the Gaussian and Student t copulae we opted to use a classical method of simulation for these distributions (see, Devroye 1986). This method ensures that the simulated values for the Student t have heavier tails than those of the Gaussian copula.

## 5.5 Simulation study

In this section we summarize the results of a simulation study. The study has been divided in two parts. Firstly, we analyze the effect of the pseudo-data obtained using the double transformed kernel estimation (DTKE), the empirical estimation (Emp) and the classical kernel estimation (CKE) on the estimated parameter of the analyzed copulae. Secondly, we compare the Mean Square Error (MSE) of estimated $VaR_\alpha$ associated with the random variable $S = X_1 + X_2$ using different copulae and nonparametric marginals: Emp, CKE and DTKE. The principal advantage of CKE and DTKE is that, unlike empirical distribution, they have a smooth shape and are defined at all points of the domain of the variable. We generated $1,000$ bivariate samples of size $n = 500$ and $1,000$ bivariate samples of size $n = 5,000$ from each copula in Table 5.1. The theoretical parameters of the copulae were selected similar to those estimated in the application that we show in section 5.6. We supposed three marginal distributions with positive skewness and different tail shapes: Lognormal and two mixtures of Lognormal-Pareto; we used the same parameters as in Bolancé et al. (2008a) and Alemany et al. (2013).

For each generated sample we estimated the parameter of the copula as we described in section 5.3, then we calculated the mean and the standard deviation of the estimated parameters for each combination of copula, nonparametric marginal and sample size. In Table 5.2 and Table 5.3 we show the means and the standard deviations of the estimated parameters with the simulated $1,000$ bivariate samples. The results reflect that there are

TABLE 5.1: Bivariate copulae and theoretical marginal distributions.

| **Copula** | Gauss | Student t* | Gumbel | Clayton | Frank | FGM | Sarmanov |
|---|---|---|---|---|---|---|---|
| **Parameter** | 0.60 | 0.60 | 1.70 | 0.75 | 4.50 | 0.33 | 9.53E-09 |
| Marginal Distributions | | | | | | | |
| **Ln** | $\int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$ | $(\mu, \sigma = (0, \sqrt{0.25})$ | | | | | |
| **70Ln-30Pa** | $p \int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$ | $(p, \mu, \sigma, \lambda, \rho) = (0.7, 0, 1, 1, 1, -1)$ | | | | | |
| **30Ln-70Pa** | $+(1-p)\left(1-\left(\frac{x-c}{\lambda}\right)^{-\rho}\right)$ | $(p, \mu, \sigma, \lambda, \rho) = (0.3, 0, 1, 1, 1, -1)$ | | | | | |

* Degrees of freedom: 8. 70Ln=70% Lognormal, 30Pa=30% Pareto.

no big differences between parameters estimated with the different pseudo-data. However, when we use pseudo-data generated with CKE or Emp there exist some cases, in bold in Table 5.2, where the estimators are less efficient. We specially want to highlight the value of the standard deviation equal to $3.35$ in Table 5.2, obtained for the Gumbel copula, with marginals 70Ln-30Pa, when pseudo-data was obtained using CKE. In this case, for some of the the $1,000$ generated samples, the Gumbel parameter estimated using the MPLE described in section 5.3 is very large, i.e. this method may tend to estimate perfect dependence when the sample is smaller. To corroborate this result we have repeated the analysis for the Gumbel with $1,000$ new simulated samples and sample size $500$, the results obtained were similar, and we observe that a similar extreme case can occur when the marginals are 30Ln-70Pa.

To analyze the finite sample properties of the $VaR_\alpha$ estimator with the different nonparametric marginal, we estimated the $VaR_\alpha$ with $\alpha = 0.95$, $\alpha = 0.99$ and $\alpha = 0.995$ for each simulated sample. As we described in section 5.2, we used a Monte Carlo method with $10,000$ simulated data. It is important to note that results are influenced by the number of simulated pairs $(U_1, U_2)$ in the Monte Carlo method, especially if we work with extreme value distributions; a greater number of pairs would improve the accuracy of the results, but the computational time would increase considerably. Using the $1,000$ estimated $VaR_\alpha$ for each confidence level $\alpha$, each copula and both nonparametric marginal distributions, we obtained the MSE. To calculate the MSE we used the true values of $VaR_\alpha$ that we show in Appendix in Table 5.10; these theoretical values are also calculated with the Monte Carlo method using the values of the parameters

TABLE 5.2: Means and standard deviations (in italic numbers) of the estimated copula parameter for the $1,000$ simulated samples with $n = 500$.

| | | Ln | | 70Ln-30Pa | | 30Ln-70Pa | |
|---|---|---|---|---|---|---|---|
| Gauss | Emp | 0.60 | *0.03* | 0.60 | *0.03* | 0.60 | *0.03* |
| | CKE | 0.60 | *0.03* | 0.60 | *0.03* | 0.60 | *0.04* |
| | DTK | 0.61 | *0.03* | 0.61 | *0.03* | 0.61 | *0.03* |
| Student t* | Emp | 0.60 | *0.03* | 0.60 | *0.03* | 0.60 | *0.03* |
| | CKE | 0.60 | *0.03* | 0.60 | *0.04* | 0.60 | *0.04* |
| | DTK | 0.61 | *0.03* | 0.61 | *0.03* | 0.61 | *0.03* |
| Gumbel | Emp | 1.71 | *0.07* | 1.71 | *0.07* | 1.71 | *0.07* |
| | CKE | 1.70 | *0.07* | 1.81 | ***3.35*** | 1.72 | ***0.17*** |
| | DTK | 1.73 | *0.07* | 1.73 | *0.07* | 1.73 | *0.07* |
| Clayton | Emp | 0.72 | *0.09* | 0.72 | *0.09* | 0.71 | *0.09* |
| | CKE | 0.70 | *0.09* | 0.70 | *0.10* | 0.71 | ***0.14*** |
| | DTK | 0.75 | *0.09* | 0.73 | *0.10* | 0.73 | *0.10* |
| Frank | Emp | 4.52 | *0.35* | 4.50 | *0.34* | 4.51 | *0.33* |
| | CKE | 4.51 | *0.35* | 4.48 | ***0.40*** | 4.49 | ***0.37*** |
| | DTK | 4.58 | *0.35* | 4.58 | *0.34* | 4.59 | *0.33* |
| FGM | Emp | 0.28 | ***0.13*** | 0.28 | *0.08* | 0.28 | *0.08* |
| | CKE | 0.28 | *0.08* | 0.28 | *0.08* | 0.28 | *0.08* |
| | DTK | 0.28 | *0.08* | 0.28 | *0.08* | 0.28 | *0.08* |
| Sarmanov | Emp | **1.4E-04** | ***2.7E-03*** | 2.4E-07 | *1.6E-11* | 7.6E-08 | *2.4E-08* |
| | CKE | **1.4E-04** | ***2.7E-03*** | 2.4E-07 | *1.6E-11* | 7.6E-08 | *2.4E-08* |
| | DTK | 5.8E-05 | *2.2E-06* | 2.4E-07 | *1.6E-11* | 7.6E-08 | *2.4E-08* |

*Degrees of freedom: 8. 70Ln=70% Lognormal, 30Pa=30% Pareto. Bold values indicate that the corresponding estimator is less efficient than the rest.

of copula and marginal distributions that was showed in Table 5.1. The results that we present in Table 5.4 and Table 5.5 correspond to the ratio between the MSE when the marginal distributions are estimated with CKE or DTKE and when the same marginal distributions are estimated using Emp. In general, for the sample size $n = 500$, the estimated $VaR_\alpha$ associated with a Lognormal marginal distribution results obtained with CKE marginals are better than those obtained with DTKE marginals. When marginal distributions have heavier tails, as in both analyzed mixtures, the results with CKE and DTKE are more similar than before. A special case is the extreme value copulae, in our analysis the Student t and the Gumbel copulae, in both cases, especially for the Gumbel copula, DTKE considerably overestimates the $VaR_\alpha$. When analyzing the bias, we deduced that the larger MSE of DTKE are associated with its larger bias, so the $VaR_\alpha$ obtained using DTKE overestimate the risk (see, Alemany et al. 2013). Analyzing the

TABLE 5.3: Means and standard deviations (in italic numbers) of the estimated copula parameter for the $1,000$ simulated samples with $n = 5,000$.

| | | **Ln** | | **70Ln-30Pa** | | **30Ln-70Pa** | |
|---|---|---|---|---|---|---|---|
| Gauss | Emp | 0.60 | *0.01* | 0.60 | *0.01* | 0.60 | *0.01* |
| | CKE | 0.60 | *0.01* | 0.60 | *0.01* | 0.60 | *0.02* |
| | DTK | 0.60 | *0.01* | 0.60 | *0.01* | 0.60 | *0.01* |
| Student t* | Emp | 0.60 | *0.01* | 0.60 | *0.01* | 0.60 | *0.01* |
| | CKE | 0.60 | *0.01* | 0.60 | *0.02* | 0.60 | *0.03* |
| | DTK | 0.60 | *0.01* | 0.60 | *0.01* | 0.60 | *0.01* |
| Gumbel | Emp | 1.70 | *0.02* | 1.70 | *0.02* | 1.70 | *0.02* |
| | CKE | 1.70 | *0.02* | 1.70 | ***0.05*** | 1.71 | ***0.20*** |
| | DTK | 1.71 | *0.02* | 1.71 | *0.02* | 1.71 | *0.02* |
| Clayton | Emp | 0.70 | *0.03* | 0.70 | *0.03* | 0.70 | *0.03* |
| | CKE | 0.70 | *0.03* | 0.70 | *0.04* | 0.70 | *0.04* |
| | DTK | 0.71 | *0.03* | 0.71 | *0.03* | 0.71 | *0.03* |
| Frank | Emp | 4.50 | *0.11* | 4.50 | *0.11* | 4.50 | *0.10* |
| | CKE | 4.50 | *0.11* | 4.50 | ***0.17*** | 4.50 | ***0.13*** |
| | DTK | 4.52 | *0.11* | 4.52 | *0.11* | 4.52 | *0.10* |
| FGM | Emp | 0.33 | *0.04* | 0.33 | *0.04* | 0.33 | *0.04* |
| | CKE | 0.33 | *0.04* | 0.33 | *0.04* | 0.34 | *0.05* |
| | DTK | 0.33 | *0.04* | 0.33 | *0.04* | 0.33 | *0.04* |
| Sarmanov | Emp | 4.7E-05 | 6.6E-07 | 7.2E-08 | *6.9E-09* | 7.4E-12 | *2.6E-12* |
| | CKE | 4.7E-05 | 6.6E-07 | 7.2E-08 | *6.9E-09* | 7.4E-12 | *2.6E-12* |
| | DTK | 4.7E-05 | 6.6E-07 | 7.2E-08 | *6.9E-09* | 7.4E-12 | *2.6E-12* |

* Degrees of freedom: 8. 70Ln=70% Lognormal, 30Pa=30% Pareto. Bold values indicate that the corresponding estimator is less efficient than the rest.

results obtained with sample size $n = 5,000$ in Table 5.5 we observe that MSE using DTKE marginals improve considerably compared to a smaller sample size, especially when we have an extreme value copula with upper tail dependence and heavier tail marginal distributions. When the theoretical model is the Gumbel copula (extreme value copula with upper tail dependence) with mixture marginal distributions the MSE obtained using DTKE marginals are much lower than CKE and Emp.

## 5.6 Application

In this section we compare the risks estimated using different copulae with different nonparametric marginals: Emp, CKE and DTKE. We use a data set corresponding to a random sample of $518$ claims obtained from motor insurance accidents. We were kindly

TABLE 5.4: Ratio between the MSE when the marginal distributions are CKE or DTKE and when the same marginal distributions are Emp, for sample of size $n = 500$.

| | CKE | | | DTKE | | |
|---|---|---|---|---|---|---|
| | $VaR_{0.95}$ | $VaR_{0.99}$ | $VaR_{0.995}$ | $VaR_{0.95}$ | $VaR_{0.99}$ | $VaR_{0.995}$ |
| | | | **Ln** | | | |
| Gauss | 1.653 | 0.164 | 0.071 | 0.170 | 1.837 | 0.300 |
| Student t* | 0.626 | 0.921 | 0.306 | 0.865 | 1.650 | 2.210 |
| Gumbel | 0.855 | 1.000 | 1.000 | 1.138 | 7.682 | 69.074 |
| Clayton | 1.939 | 0.172 | 0.127 | 0.044 | 2.653 | 0.515 |
| Frank | 773.819 | 13747.714 | 34.978 | 0.036 | 30.408 | 0.392 |
| FGM | 0.057 | 0.157 | 0.080 | 2.784 | 4.225 | 0.613 |
| Sarmanov | 4.024 | 0.000 | 0.074 | 6.599 | 7.275 | 0.371 |
| | | | **70Ln-30Pa** | | | |
| Gauss | 1.649 | 0.158 | 0.003 | 0.396 | 0.016 | 0.007 |
| Student t* | 0.408 | 0.992 | 0.835 | 0.295 | 1.585 | 35.464 |
| Gumbel | 126.517 | 5362.956 | 6332.125 | 1.297 | 67.885 | 1588.541 |
| Clayton | 1.872 | 0.168 | 0.007 | 0.000 | 0.039 | 0.006 |
| Frank | 1.351 | 37.624 | 0.012 | 0.187 | 5.092 | 0.001 |
| FGM | 0.023 | 0.154 | 0.004 | 1.199 | 0.246 | 0.008 |
| Sarmanov | 1.091 | 4.111 | 0.004 | 0.972 | 2.754 | 0.001 |
| | | | **30Ln-70Pa** | | | |
| Gauss | 1.616 | 0.159 | 0.004 | 0.647 | 0.004 | 0.010 |
| Student t* | 0.410 | 0.992 | 0.834 | 0.357 | 3.414 | 56.445 |
| Gumbel | 0.807 | 1.000 | 1.000 | 1.374 | 35.017 | 402.053 |
| Clayton | 1.825 | 0.168 | 0.007 | 0.046 | 0.072 | 0.013 |
| Frank | 1.374 | 36.118 | 0.013 | 0.367 | 7.555 | 0.000 |
| FGM | 0.060 | 0.154 | 0.004 | 0.411 | 0.312 | 0.019 |
| Sarmanov | 3.184 | 3.169 | 0.009 | 1.112 | 1.157 | 0.001 |

*Degrees of freedom: 8. 70Ln=70% Lognormal, 30Pa=30% Pareto

given access to these data by an insurance company. We have two costs: Cost1 corresponds to the amount paid to the insured party to compensate for damage to their own vehicle and all other losses attributable to third-party damages and Cost2 corresponds to the expenses incurred in paying for medical treatment and hospitalization as a result of the accident (see, Bolancé et al. 2008b;for more information on these data). This data has also been analyzed by Bahraoui et al. (2014c), Bolancé et al. (2009), Guillén et al. (2011; 2013). In Table 5.6, we show the descriptive statistics for Cost1, Cost2 and $S = Cost1 + Cost2$. Clearly Cost1 represents higher values than those presented by Cost2, but if we compare the dispersion measured by the coefficient of variation (CV), it can be seen that both variables are similar. Likewise, both variables are strongly skewed to the right.

TABLE 5.5: Ratio between the MSE when the marginal distributions are CKE or DTKE and when the same marginal distributions are Emp, for sample of size $n = 5,000$.

| | CKE | | | DTKE | | |
|---|---|---|---|---|---|---|
| | $VaR_{0.95}$ | $VaR_{0.99}$ | $VaR_{0.995}$ | $VaR_{0.95}$ | $VaR_{0.99}$ | $VaR_{0.995}$ |
| | **Ln** | | | | | |
| Gauss | 1.276 | 4.438 | 1210.553 | 4.526 | 9.385 | 1949.681 |
| Student t* | 0.981 | 0.660 | 0.915 | 1.039 | 1.089 | 1.488 |
| Gumbel | 0.969 | 0.637 | 1.000 | 0.980 | 0.487 | 1.901 |
| Clayton | 0.901 | 0.975 | 2.636 | 1.535 | 10.070 | 15.705 |
| Frank | 3234.643 | 81688.838 | 35158.148 | 3.704 | 47.577 | 35.386 |
| FGM | 0.759 | 0.182 | 0.011 | 2.648 | 27.939 | 61.817 |
| Sarmanov | 1.019 | 0.545 | 0.082 | 5.177 | 6.329 | 6.961 |
| | **70Ln-30Pa** | | | | | |
| Gauss | 302.648 | 2296.804 | 4843.762 | 0.003 | 0.003 | 0.652 |
| Student t* | 34.252 | 0.594 | 0.812 | 0.909 | 0.768 | 1.125 |
| Gumbel | 5.202 | 0.004 | 0.052 | 0.000 | 0.000 | 0.004 |
| Clayton | 438.987 | 1366.710 | 515.489 | 0.014 | 1.031 | 0.000 |
| Frank | 376.801 | 12004.490 | 108.824 | 0.007 | 2.409 | 0.000 |
| FGM | 728.268 | 1005.916 | 46.400 | 0.010 | 1.888 | 0.670 |
| Sarmanov | 31.694 | 0.010 | 0.648 | 1.159 | 0.922 | 1.175 |
| | **30Ln-70Pa** | | | | | |
| Gauss | 233.914 | 2319.751 | 9787.799 | 0.002 | 0.152 | 3.864 |
| Student t* | 32.388 | 0.594 | 0.812 | 0.901 | 0.802 | 1.487 |
| Gumbel | 100.716 | 0.024 | 1.000 | 0.001 | 0.000 | 0.343 |
| Clayton | 296.132 | 1503.367 | 734.355 | 0.009 | 1.728 | 4.525 |
| Frank | 270.825 | 9473.642 | 123.888 | 0.005 | 4.328 | 2.028 |
| FGM | 409.650 | 899.208 | 54.221 | 0.007 | 3.342 | 5.648 |
| Sarmanov | 1584.494 | 114.186 | 0.464 | 0.344 | 4.014 | 1.053 |

*Degrees of freedom: 8. 70Ln=70% Lognormal, 30Pa=30% Pareto

TABLE 5.6: Descriptive statistics.

| | Mean | Std.Dev. | CV=$\frac{\text{Std.Dev.}}{\text{Mean}}$ | Skewness | Kurtosis | Min | Max |
|---|---|---|---|---|---|---|---|
| Cost1 | 1827.6004 | 6867.8166 | 3.7578 | 15.7430 | 301.2149 | 13 | 137936 |
| Cost2 | 283.9208 | 863.1695 | 3.0402 | 8.0836 | 83.1602 | 1 | 11855 |
| S | 2111.5212 | 7520.1681 | 3.5615 | 15.3554 | 290.3473 | 14 | 149791 |

Covariance between Cost1 and Cost2 is 4312140.2637.

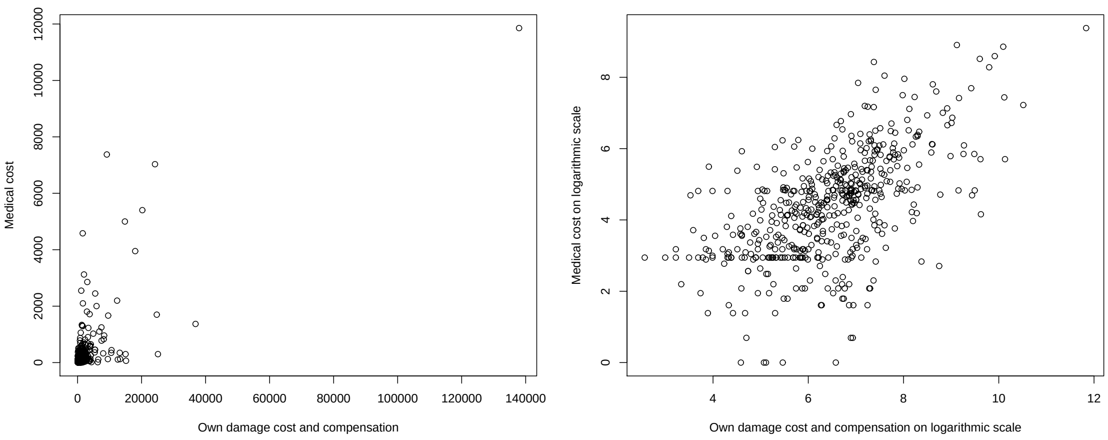


FIGURE 5.1: Cost1 vs Cost2 losses (left) and Cost1 vs Cost2 losses on logarithmic scale (right).

In Figure 5.1, we plot Cost1 versus Cost2 (in the original and logarithmic scale) and observe that both variables show high costs that are generated by an extreme value distribution.

The aim is to calculate the $VaR_\alpha$ of the total cost of claims $S = Cost1 + Cost2$ so as to quantify the risk of loss. To do so we employ different confidence levels: $\alpha = 0.95$, $\alpha = 0.99$, $\alpha = 0.995$ and $\alpha = 0.999$. In our application we decided to include confidence level $\alpha = 0.999$, although we want to note that the number of simulated pairs in the Monte Carlo method is $r = 10,000$. In Table 5.7, we show the estimated dependence parameters for each copula and both nonparametric marginal distributions. Furthermore, we show the values of the Copula Information Criteria (CIC) defined in Grønnrberg and Hjort (2014). We observed that estimated dependence parameters are similar for different estimated marginals. To compare the CKE and the DTKE of marginal cdfs, the results for $\widehat{ISE}$ are shown in Table 5.8. In both cases DTKE has a lower error than that of the CKE.

A comparison of the results in tables 5.7 and 5.8 allows us to conclude that the best fit is obtained when we use the Sarmanov copula with DTKE marginals. Moreover, with the

TABLE 5.7: Results of the fit of the copulae with nonparametric marginals.

| Copula | Marginal | Estimate Dependence Parameter | CIC |
|---|---|---|---|
| Gauss | Emp | 0.5940 | 105.8049 |
| | CKE | 0.6007 | 110.4242 |
| | DTKE | 0.6000 | 108.4305 |
| Student t* | Emp | 0.6007 | 108.0404 |
| | CKE | 0.6077 | 114.0509 |
| | DTKE | 0.6072 | 110.6941 |
| Gumbel | Emp | 1.7007 | 123.1476 |
| | CKE | 1.7026 | 126.0915 |
| | DTKE | 1.7199 | 126.0366 |
| Clayton | Emp | 0.7252 | 50.9524 |
| | CKE | 0.8055 | 60.1762 |
| | DTKE | 0.7347 | 53.2462 |
| Frank | Emp | 4.5291 | 110.0292 |
| | CKE | 4.5308 | 112.7432 |
| | DTKE | 4.6071 | 113.0297 |
| FGM | Emp | 0.3333 | 31.0007 |
| | CKE | 0.3333 | 31.3331 |
| | DTKE | 0.3333 | 52.4245 |
| Sarmanov | Emp | 9.5267E-09 | 89.41598 |
| | CKE | 9.5267E-09 | 124.4873 |
| | DTKE | 9.5267E-09 | 206.5283 |

*Estimated degree of freedom: 9.245659 for Emp, 7.53357 for CKE and 9.147745 for DTKE.

TABLE 5.8: $\widehat{ISE}$ for marginal cdfs of Cost1 and Cost2.

| Marginal | Cost1 | Cost2 |
|---|---|---|
| CKE | 134785.90 | 11349.66 |
| DTKE | 134651.10 | 11320.37 |

Gumbel copula we obtain a good fit. We know that, unlike the Sarmanov copulae, the Gumbel copula is an extreme value copula with a positive upper coefficient of tail dependence. Therefore, this copula is useful when we have a heavy-tailed marginal distribution with dependence at the extreme value. Conversely, the Sarmanov copula is more sensitive to the nonparametric estimated marginal distributions. The $VaR_\alpha$ estimates at different confidence levels are shown in Table 5.9; their standard errors obtained using simple random bootstrap are shown in Appendix, in Table 5.11. We observe that at lower confidence level ($\alpha = 0.95$) the results for the various copulae and marginals are similar. However, at higher confidence levels differences are observed. In these cases, the results are more sensitive to the choice of the marginal estimation method. Focusing

TABLE 5.9: Estimated VaR.

| Copula | Marginal | $\alpha = 0.95$ | $\alpha = 0.99$ | $\alpha = 0.995$ | $\alpha = 0.999$ |
|---|---|---|---|---|---|
| Gaussian | Emp | 8556.35 | 25456.27 | 38112.18 | 141051.50 |
| | CKE | 8190.19 | 24676.76 | 29738.51 | 138402.91 |
| | DTKE | 8347.97 | 39120.32 | 73352.69 | 344582.96 |
| Student t | Emp | 12060.96 | 36587.28 | 48719.14 | 149791.00 |
| | CKE | 10540.10 | 35970.67 | 36978.00 | 149763.44 |
| | DTKE | 11111.43 | 64671.31 | 150355.35 | 867478.98 |
| Gumbel | Emp | 7476.46 | 25526.08 | 41453.65 | 149790.71 |
| | CKE | 6936.31 | 23057.30 | 31157.20 | 145315.16 |
| | DTKE | 7574.88 | 31449.32 | 67146.02 | 454495.55 |
| Clayton | Emp | 8349.68 | 24855.74 | 37034.52 | 138140.44 |
| | CKE | 8030.18 | 24233.34 | 25454.24 | 137984.19 |
| | DTKE | 8746.68 | 39589.80 | 71144.89 | 351897.54 |
| Frank | Emp | 8651.29 | 25143.55 | 37302.93 | 138401.16 |
| | CKE | 8308.93 | 24350.28 | 26437.82 | 138060.45 |
| | DTKE | 8666.48 | 36061.87 | 71192.31 | 264672.15 |
| FGM | Emp | 8311.95 | 24784.13 | 36991.31 | 138078.43 |
| | CKE | 7957.13 | 24197.77 | 25351.70 | 137950.83 |
| | DTKE | 8724.19 | 41415.90 | 75060.06 | 355777.73 |
| Sarmanov | Emp | 8271.39 | 24796.76 | 36991.71 | 138319.93 |
| | CKE | 7531.93 | 24175.01 | 25254.86 | 137962.39 |
| | DTKE | 8414.11 | 33794.75 | 61556.21 | 248524.35 |

on the Gumbel copula, at a confidence level of $\alpha = 0.999$, we observe that the $VaR_\alpha$ estimates with the Emp marginals is equal to the highest value in the sample, while with the CKE marginals the estimation is somewhat lower, i.e. with these methods, quantiles above the sample maximum cannot be extrapolated and, as is shown in Alemany et al. (2013), this can underestimate the risk. However, when we use the DTKE to estimate marginals in the Gumbel copula and calculate the $VaR_\alpha$ at a confidence level of $\alpha = 0.999$ we obtain a result above the maximum of the sample, which we consider to be a reasonable estimation.

To analyze the effect of the copula choice on the estimation of the $VaR_\alpha$, in Figure 5.2 we plot the results of the $VaR_\alpha$ estimate with the four copulae providing the best fit and the DTKE for marginals. We show that the results are similar in the lower quantiles but as the quantile increases the differences also increase.

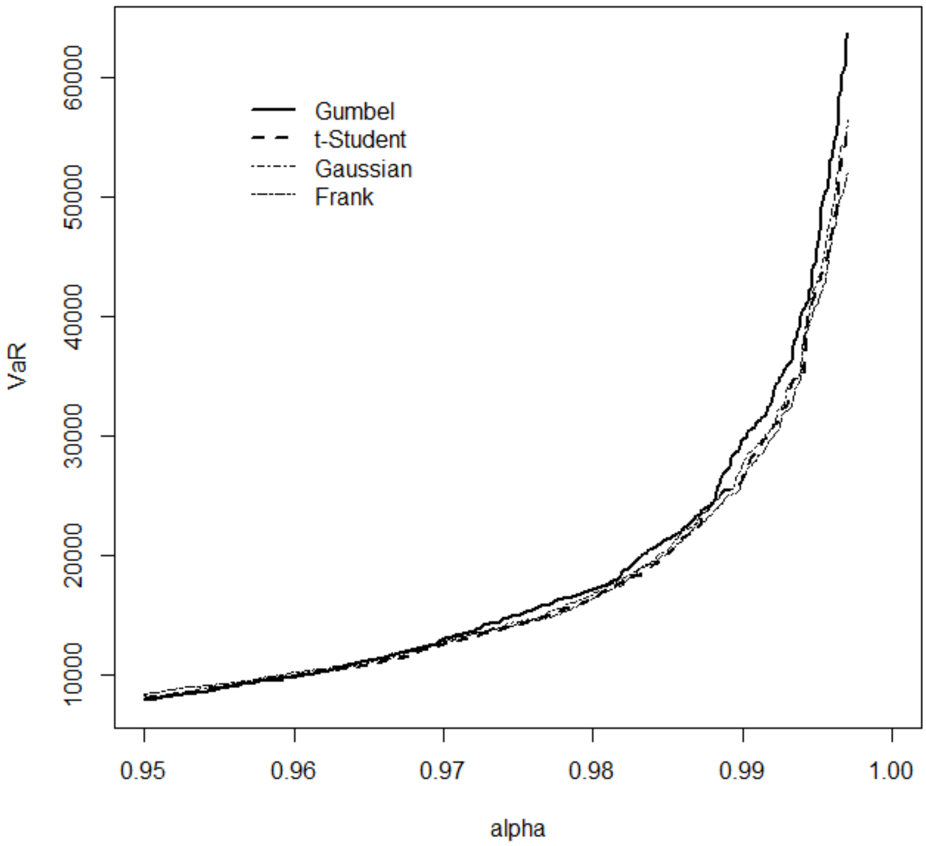


FIGURE 5.2: Estimated VaR.

TABLE 5.10: True VaR obtained with Monte Carlo method.

| Copula | | | |
|---|---|---|---|
| $VaR_{0.95}$ | **Ln** | **70Ln-30Pa** | **30Ln-70Pa** |
| Gauss | 6.7272 | 15.3062 | 28.6315 |
| Student t* | 5.2183 | 27.0457 | 57.0428 |
| Gumbel | 6.7272 | 120.8673 | 281.9592 |
| Clayton | 4.0187 | 15.3549 | 29.8854 |
| Frank | 4.2678 | 15.9630 | 30.7474 |
| FGM | 3.9275 | 14.6989 | 29.3027 |
| Sarmanov | 3.8406 | 17.8347 | 29.0481 |
| $VaR_{0.99}$ | **Ln** | **70Ln-30Pa** | **30Ln-70Pa** |
| Gauss | 10.1906 | 55.2905 | 127.6969 |
| Student t* | 7.2732 | 157.7933 | 368.8149 |
| Gumbel | 10.1906 | 2035.8642 | 4753.0104 |
| Clayton | 5.1154 | 57.9201 | 133.0112 |
| Frank | 5.4090 | 60.4055 | 138.1455 |
| FGM | 5.0114 | 58.1196 | 135.0387 |
| Sarmanov | 4.9725 | 32.7578 | 87.1492 |
| $VaR_{0.995}$ | **Ln** | **70Ln-30Pa** | **30Ln-70Pa** |
| Gauss | 12.6464 | 125.4454 | 289.7663 |
| Student t* | 8.1329 | 321.5879 | 752.3678 |
| Gumbel | 12.6464 | 7396.6610 | 17261.5422 |
| Clayton | 5.6756 | 121.5614 | 281.1708 |
| Frank | 6.0975 | 115.3203 | 267.7573 |
| FGM | 5.5301 | 123.2085 | 286.0225 |
| Sarmanov | 5.4225 | 54.6194 | 126.4926 |

*Degree of freedom: 8. 70Ln=70% Lognormal, 30Pa=30% Pareto.

TABLE 5.11: Standard errors of estimated VaR obtained using simple random bootstrap (based on 200 bootstrap samples).

| Copula | Marginal | $\alpha = 0.95$ | $\alpha = 0.99$ | $\alpha = 0.995$ | $\alpha = 0.999$ |
|---|---|---|---|---|---|
| Gaussian | Emp | 1320.78 | 9443.52 | 47001.20 | 51156.50 |
| | CKE | 1278.62 | 4626.62 | 30035.84 | 51179.84 |
| | DTKE | 1107.33 | 7912.09 | 22356.39 | 244515.05 |
| Student t | Emp | 2033.31 | 13018.11 | 48721.26 | 52231.83 |
| | CKE | 1828.06 | 5327.33 | 30291.67 | 52226.77 |
| | DTKE | 1533.01 | 16785.04 | 54688.48 | 1130082.21 |
| Gumbel | Emp | 1312.53 | 5101.30 | 48422.44 | 51617.14 |
| | CKE | 1244.78 | 4337.21 | 36281.64 | 51452.33 |
| | DTKE | 1145.83 | 6853.63 | 23959.52 | 773653.04 |
| Clayton | Emp | 1354.53 | 14907.98 | 47439.84 | 51201.76 |
| | CKE | 1318.40 | 9686.14 | 29793.08 | 51230.97 |
| | DTKE | 1308.24 | 9128.56 | 24403.25 | 353468.62 |
| Frank | Emp | 1509.56 | 14972.37 | 46926.43 | 51102.38 |
| | CKE | 1395.70 | 5836.76 | 28778.78 | 51163.14 |
| | DTKE | 1291.68 | 9477.98 | 26344.77 | 251070.76 |
| FGM | Emp | 1416.57 | 16815.04 | 47525.89 | 50764.57 |
| | CKE | 1303.96 | 15062.98 | 30732.40 | 50783.16 |
| | DTKE | 1307.85 | 10145.86 | 29242.97 | 449741.88 |
| Sarmanov | Emp | 1455.21 | 18506.48 | 46682.40 | 51386.90 |
| | CKE | 1443.45 | 12493.70 | 32363.81 | 51474.42 |
| | DTKE | 1349.92 | 7944.01 | 19871.09 | 248317.46 |

# Chapter 6

# Estimating extreme value cumulative distribution functions using bias corrected kernel approach

We propose a new kernel estimation of the cumulative distribution function based on transformation and on bias reducing techniques. We derive the optimal bandwidth that minimises the asymptotic integrated mean squared error. The simulation results show that our proposed kernel estimation improves alternative approaches when the variable has an extreme value distribution with heavy tail and the sample size is small.

## 6.1 Introduction

Estimating the cumulative distribution function (cdf) is a fundamental goal in many fields in which analysts are interested in estimating the risk of occurrence of a particular event, for example, the probability of a catastrophic accident or the probability of a major economic loss. Similarly, in risk quantification, risk measurements are usually expressed in terms of the cdf, a good example being the distortion risk measures proposed in Wang (1995; 1996).

Specifically, risk quantification concentrates in the highest values of the domain of the distribution, where sample information is scarce and it is, therefore, necessary to extrapolate the behaviour of the cdf, even above the maximum observed. To extrapolate the distribution we can use parametric models or, alternatively, we can use a nonparametric estimation. In this paper, we propose a nonparametric estimator of the cdf that allows us to extrapolate the behaviour of the cdf with greater accuracy than is possible with existing methods.

A naive nonparametric estimator of the cdf is the empirical distribution. It is known that the empirical distribution is an unbiased estimator of cdf. However, the empirical distribution is inefficient when data are scarce. A nonparametric alternative for estimating the cdf is the kernel estimator. This is more efficient than the empirical distribution but it is, nevertheless, a biased estimator. Furthermore, both the empirical distribution and the kernel estimator of the cdf are inefficient when the shape of the distribution is right skewed and it has a longer right tail, i.e., it belongs to a certain family of extreme value distributions (EVD): the Gumbel or Fréchet types. In these cases, although we have a large sample size, the number of observations in the highest values of the domain of the distribution is small. This kind of distribution is very common in microeconomic, financial and actuarial data, where economic quantities are measured, e.g., costs, losses and wages. Likewise, there are other fields such as demography, geology or meteorology, where the observed phenomena are distributed following an EVD (see, for example, Reiss and Thomas 1997). In this study, we develop a bias-corrected transformed kernel estimator of the cdf that is more accurate than the bias-corrected classical kernel estimator.

With the aim of reducing the bias of the classical kernel estimator (CKE) of the cdf, Kim et al. (2006), based on Choi and Hall (1998), proposed a bias reducing technique, henceforth the bias-corrected classical kernel estimator (BCCKE). Alternatively, Alemany et al. (2013) proved that using the transformed kernel estimator of the cdf, the bias and variance of the CKE could be reduced and they proposed a new estimator based on two transformations, the double transformed kernel estimator (DTKE). However, this estimator has asymptotic properties and needs a large sample size to obtain better results than alternative approaches. In this study, we analyse the properties of the DTKE of the cdf by incorporating the finite sample bias correction proposed by Kim et al. (2006). We refer to this new estimator as the bias-corrected double transformed kernel estimator (BCDTKE).

Estimating the smoothing parameter associated with kernel estimations is also a challenge when the data are generated by an extreme value distribution. When using the two most popular automatic methods. i.e., plug-in and cross-validation, the optimal value frequently degenerates to zero. An alternative for calculating the smoothing parameter is the rule-of-thumb value (Silverman 1986), which is based on a reference distribution. Using the proposed BCDTKE we can estimate the exact rule-of-thumb value based on a known distribution.

The use of nonparametric methods is based on the lack of information about the theoretical distribution associated with the random variable under analysis. This distribution might match one of those belonging to a subfamily of EVDs: Type I (Gumbel) or Type II (Fréchet). Moreover, the distribution might be a mixture of two or more EVDs. An important goal of this study is to analyse the domain of attraction of different mixtures of EVDs. In section 6.2 we carry out this analysis. In section 6.3 we describe the BCCKE of cdf and we propose some new results related to the asymptotically optimal smoothing parameter. These results are then used in section 4, where we describe a new estimator based on transformations and bias correction. In section 5, we show the results of a simulation study. We conclude in section 6.

## 6.2 Maximum domain of attraction of mixtures of extreme value distributions

In this section we prove some results related to the maximum domain of attraction (MDA) of some mixtures of EVDs. The expression of the cdf of a generalised EVD is (see, Jenkinson 1955):

$$\begin{array}{ll} G_\xi(x,\mu,\sigma) = \exp\left\{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right\} & \text{if } \xi \neq 0 \\ G_\xi(x,\mu,\sigma) = \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\} & \text{if } \xi = 0 \end{array} \tag{6.1}$$

and its mean is:

$$E(X) = \begin{cases} \mu + \sigma\frac{\Gamma(1-\xi)-1}{\xi} & \text{if} \quad \xi \neq 0, \xi < 1 \\ \mu + \sigma\gamma & \text{if} \quad \xi = 0 \\ \infty & \text{if} \quad \xi \geq 1 \end{cases}, \tag{6.2}$$

where $\Gamma(\cdot)$ is Euler's gamma function and $\gamma$ is Euler's constant. The MDA of $G_\xi$ ($MDA(G_\xi)$) depends on the shape parameter $\xi$. In the expression (6.1) when $\xi = 0$ a Gumbel type EVD is obtained and when $\xi > 0$ the result is a Fréchet type EVD.

We define the right end point of $G$ as $r(G) = \sup(x|G(x) < 1)$. We know that if two distributions $F$ and $G$ are such that $r(G) = r(F)$ then:

$$\lim_{x \uparrow r(F)} \frac{\bar{F}(x)}{\bar{G}(x)} = c,$$

for some constant $0 < c < \infty$, where $\bar{F}(x) = 1 - F(x)$ and $\bar{G}(x) = 1 - G(x)$. In this case $F$ and $G$ have the same MDA, furthermore, $F$ and $G$ are tail equivalent if (see, for example, Embrechts et al. 1997):

$$\lim_{x \uparrow r(F)} \frac{\bar{F}(x)}{\bar{G}(x)} = 1.$$

**Theorem 6.2.1.** *Let $F$ be a cdf that is expressed as $F(x) = \sum_{i=1}^{m} p_i F_i(x)$, with $\sum_{i=1}^{m} p_i = 1, \forall p_i > 0$, if every $F_i \in MDA(G_{\xi_i})$, with $\xi_i > 0$ (Fréchet), then $F \in MDA(G_{\xi_M})$, where $\xi_M = max(\xi_1, \ldots, \xi_m)$.*

*Proof.* We know that if $F_i \in MDA(G_{\xi_i})$, $\forall i = 1, ..., m$, with $\xi_i > 0$ (Fréchet), then $\bar{F}_i(x) = x^{-\frac{1}{\xi_i}} L_i(x)$, where $L_i$ is a slowly varying function[3] and

$$\begin{aligned}
\bar{F}(x) = 1 - F(x) &= \sum_{i=1}^{m} p_i \bar{F}_i(x) \\
&= \sum_{i=1}^{m} p_i x^{-\frac{1}{\xi_i}} L_i(x) \\
&= x^{-\frac{1}{\xi_M}} \sum_{i=1}^{m} p_i x^{(\frac{1}{\xi_M} - \frac{1}{\xi_i})} L_i(x) \\
&= x^{-\frac{1}{\xi_M}} p_M L_M(x) + x^{-\frac{1}{\xi_M}} \sum_{i \neq M} p_i x^{(\frac{1}{\xi_M} - \frac{1}{\xi_i})} L_i(x) \\
&\sim x^{-\frac{1}{\xi_M}} p_M L_M(x).
\end{aligned}$$

[3] A positive Lebesgue measurable function $L$ on $(0, \infty)$ is slowly varying if

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1, \quad t > 0.$$

The previous result is obtained observing that $x^{(\frac{1}{\xi_M}-\frac{1}{\xi_i})}L_i(x) \to_{x\to\infty} 0$. We conclude that $F$ and $F_M$ are tail equivalents. □

**Theorem 6.2.2.** *(Sufficient condition) If $j \in \{1,\dots,m\}$ is such that $\lim_{x\to\infty} \frac{\bar{F}_i(x)}{\bar{F}_j(x)} = A < \infty$, with $j \neq i$, then if $F(x) = \sum_{i=1}^{m} p_i F_i(x) \in MDA(G_\xi)$, with $\xi > 0$, then $F_j \in MDA(G_\xi)$.*

*Proof.* To prove Theorem 6.2.2 we start with the definition of regular variation. A positive Lebesgue measurable function $L$ on $(0,\infty)$ is a regular variation at infinity with index $\alpha \in \mathbb{R}$ if:

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = t^{\alpha}, \quad t > 0. \tag{6.3}$$

Then, $F \in MDA(G_\xi)$ with $\xi > 0$ (Fréchet), if $\bar{F}$ is a regular variation with index $-\frac{1}{\xi}$, namely:

$$\lim_{x\to\infty} \frac{\bar{F}(tx)}{\bar{F}(x)} = t^{-\frac{1}{\xi}}, \quad t > 0. \tag{6.4}$$

We have:

$$\begin{aligned}
\lim_{x\to\infty} \frac{\bar{F}(tx)}{\bar{F}(x)} &= t^{-\xi} \\
\lim_{x\to\infty} \frac{\sum_{i=1}^{m} p_i \bar{F}_i(tx)}{\sum_{i=1}^{m} p_i \bar{F}_i(x)} &= t^{-\xi} \\
\lim_{x\to\infty} \frac{\left[\sum_{i\neq j}^{m} p_i \frac{\bar{F}_i(tx)}{\bar{F}_j(tx)} + p_j\right] \bar{F}_j(tx)}{\left[\sum_{i\neq j}^{m} p_i \frac{\bar{F}_i(x)}{\bar{F}_j(x)} + p_j\right] \bar{F}_j(x)} &= t^{-\xi},
\end{aligned}$$

taking into account the limit in the interior of the brackets and considering the condition $\lim_{x\to\infty} \frac{\bar{F}_i(x)}{\bar{F}_j(x)} = A < \infty$ we deduce:

$$\lim_{x\to\infty} \frac{\bar{F}_j(tx)}{\bar{F}_j(x)} = t^{-\xi},$$

then $F_j$ is a Fréchet type EVD. □

**Theorem 6.2.3.** *Let $F$ be a cdf that is expressed as $F(x) = \sum_{i=1}^{m} p_i F_i(x)$, with $\sum_{i=1}^{m} p_i = 1$, $\forall p_i > 0$, if $j \in \{1, \ldots, m\}$ is such that $F_j \in MDA(G_{\xi_j})$, with $\xi_j > 0$ (Fréchet), and $F_i$ $\forall i \neq j$ are Lognormal distributions then $F \in MDA(G_{\xi_j})$.*

*Proof.* Firstly we note that:

$$\sup(x|F_j(x) < 1) \subset \sup(x|\sum_{i=1}^{m} p_i F_i(x) < 1)$$

and the right end point of $F$ is $r(F) = \sup(x|F(x) < 1) = \infty$. Besides, we have:

$$\bar{F}(x) = \sum_{i \neq j} p_i \bar{F}_i(x) + p_j \bar{F}_j(x),$$

where:

$$\begin{aligned} \bar{F}_j(x) &= x^{\frac{-1}{\xi_j}} L(x), \text{ where } L(x) \text{ is slowly varying function,} \\ \bar{F}_i(x) &= \bar{\Phi}(\frac{log(x) - \mu}{\sigma}), \text{ where } \Phi \text{ is Normal standard distribution,} \\ \bar{F}_i(x) &\sim \frac{\varphi(\frac{log(x)-\mu}{\sigma})}{(\frac{log(x)-\mu}{\sigma})}, \text{ where } \varphi \text{ is Normal standard density,} \end{aligned} \tag{6.5}$$

the last term in (6.5) is deduced applying l'Hôpital's rule to $\frac{x\bar{\Phi}(t)}{\varphi(t)}$, resulting in $\bar{\Phi}(t) \sim \frac{\varphi(t)}{t}$ when $t$ is high. If $\xi_j > 0$ we can find $\alpha > 0$ such that $\frac{1}{\xi_j} + \alpha > 0$ and

$$\begin{aligned} \frac{\bar{F}_i(x)}{\bar{F}_j(x)} &= \frac{\exp\big(-\frac{1}{2}(\frac{\log(x)-\mu}{\sigma})^2\big)}{\sqrt{2\pi}(\frac{\log(x)-\mu}{\sigma})x^{\frac{-1}{\xi_j}} L(x)} \\ &= \frac{\exp\big(\frac{-1}{2}(\frac{\log(x)-\mu}{\sigma})^2 + (\alpha + \frac{1}{\xi_j})\log(x)\big)}{\sqrt{2\pi}(\frac{\log(x)-\mu}{\sigma})x^{\alpha} L(x)} \to_{x\to\infty} 0 \end{aligned}$$

and

$$\frac{\bar{F}(x)}{\bar{F}_j(x)} = \sum_{i \neq j}^{m} p_i \frac{\bar{F}_i(x)}{\bar{F}_j(x)} + p_j,$$

then we can conclude that $0 < \lim_{x\to\infty} \frac{\bar{F}(x)}{\bar{F}_j(x)} = p_j < \infty$, then $r(F) = r(F_j)$ and both distributions have the same MDA. □

**Theorem 6.2.4.** *Let $F$ be a cdf that is expressed as $F(x) = \sum_{i=1}^{m} p_i F_i(x)$, with $\sum_{i=1}^{m} p_i = 1$, $\forall p_i > 0$, if $j \in \{1, \dots, m\}$ is such that $F_j \in MDA(G_{\xi_j})$, with $\xi_j > 0$ (Fréchet), and $F_i$ $\forall i \neq j$ have $MDA(G_{\xi_i})$, with $\xi_i = 0$ (Gumbel), then $F \in MDA(G_{\xi_j})$.*

*Proof.* **Case 1:** If $r(F_i) = \infty$, $\forall i \neq j$, and we can find $\alpha > 0$ such that $\frac{1}{\xi_j} + \alpha > 0$, we obtain:

$$\frac{\bar{F}_i(x)}{\bar{F}_j(x)} = \frac{\bar{F}_i(x)}{x^{\frac{-1}{\xi_j}} L(x)} = \frac{\bar{F}_i(x)}{x^{-(\alpha+\frac{1}{\xi_j})}} \frac{1}{x^{\alpha} L(x)}.$$

from the properties of the von Mises functions , $\bar{F}_i$ decreases to zero much faster than $x^{-\alpha}$, then when $r(F_i) = \infty$ we have (see, Embrechts et al. 1997;page 139):

$$\lim_{x\to\infty} \frac{\bar{F}_i(x)}{x^{-(\alpha+\frac{1}{\xi_j})}} = 0 \quad \text{and} \quad \lim_{x\to\infty} \frac{1}{x^{\alpha} L(x)} = 0,$$

and we conclude that $F$ and $F_j$ are tail equivalents.

**Case 2:** If $l \neq i \neq j$ is such that $r(F_l) < \infty$

$$\frac{\bar{F}(x)}{\bar{F}_j(x)} = p_l \frac{\bar{F}_l(x)}{\bar{F}_j(x)} + \sum_{i\neq l\neq j}^{m} \frac{\bar{F}_i(x)}{\bar{F}_j(x)} + p_j.$$

Let $X_l$ be a random variable with probability distribution function (pdf) $f_l(\cdot)$ with $E(X_l^k) < \infty$ for every $k > 0$,

$$\frac{\bar{F}_l(x)}{\bar{F}_j(x)} = \frac{\bar{F}_l(x)}{(x - r(F_l)) f_l(x)} \frac{(x - r(F_l)) f_l(x)}{x^{\frac{-1}{\xi_j}} L(x)},$$

the limits of the first term are (see, Embrechts et al. 1997, McNeil et al. 2005):

$$\lim_{x\to\infty} \frac{\bar{F}_l(x)}{(x - r(F_l)) f_l(x)} = \lim_{r(F)\to\infty} \lim_{x\to r(F_l)} \frac{\bar{F}_l(x)}{(x - r(F_l)) f_l(x)} = 0.$$

We obtain:

$$\frac{(x-r(F_l))f_l(x)}{x^{\frac{-1}{\xi_j}}L(x)} = \frac{x^{\frac{1}{\xi_j}+\alpha}(x-r(F_l))f_l(x)}{x^{\alpha}L(x)} \sim \frac{x^{a}f_l(x)}{x^{\alpha}L(x)} \to 0, \text{ with } a>1 \text{ and } \alpha>0$$

and we achieve the same results as in Case 1. □

## 6.3 Classical kernel estimator with bias reducing technique

The BCCKE proposed by Kim et al. (2006) can be expressed as a linear combination of the CKE of the pdf, $f_X$, and the CKE of the cdf, $F_X$. Let us assume that $X_i$, $i = 1, ..., n$, denotes data observations from the random variable $X$; the usual expression for the classical kernel estimator of the pdf is (see, Silverman 1986):

$$\hat{f}_X(x) = \frac{1}{nb}\sum_{i=1}^{n} k\left(\frac{x-X_i}{b}\right)$$

and for the cdf is (see, Azzalini 1981):

$$\widehat{F}_X(x) = \int_{-\infty}^{x} \widehat{f}_X(u)du = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{b}\right),$$

where $K(\cdot)$ is the cdf associated with $k(\cdot)$ which is known as the kernel function (usually a bounded and symmetric pdf). Some examples of very common kernel functions are the Epanechnikov and the Gaussian kernel. The parameter $b$ is the bandwidth or the smoothing parameter and it controls the smoothness of the resulting estimation. The larger the value of $b$, the smoother the resulting estimated function. In practice, the value of $b$ depends on the sample size and satisfies the condition if $n \to \infty$, $b \to 0$ and $nb \to \infty$.

The BCCKE is:

$$\widetilde{F}_X(x) = \frac{\lambda_1 \widehat{F}_1(x) + \widehat{F}_X(x) + \lambda_2 \widehat{F}_2(x)}{\lambda_1 + 1 + \lambda_2}, \tag{6.6}$$

where $\lambda_1, \lambda_2 > 0$ are weights and

$$\widehat{F}_j(x) = \widehat{F}_X(x + l_j b) - l_j b \widehat{f}_X(x + l_j b), \quad j = 1, 2.$$

Kim et al. (2006) proved that if $\lambda_1 = \lambda_2 = \lambda$ then $-l_1 = l_2 = l(\lambda)$, being:

$$l(\lambda) = \left(\frac{(1+2\lambda)\mu_2}{2\lambda}\right)^{1/2},$$

where $\mu_p = \int t^p k(t)dt$. Kim et al. (2006) also proved that the bias of $\widetilde{F}_X(x)$ is $O(b^4)$, while the bias of $\widehat{F}_X(x)$ is $O(b^2)$.

The mean integrated squared error (MISE) can be expressed as the sum of the integrated variance and the integrated square bias:

$$MISE\left(\widetilde{F}_X\right) = E\left(\int (\widetilde{F}_X(x) - F_X(x))^2 dx\right) = \int Var\left(\widetilde{F}_X(x)\right) dx + \int \left[Bias\left(\widetilde{F}_X(x)\right)\right]^2 dx.$$

Based on the asymptotic expression for bias and variance of BCCKE deduced by Kim et al. (2006) we obtain that the asymptotic mean integrated squared error (A-MISE) is:

$$\begin{aligned} A-MISE\left(\widetilde{F}_X(x)\right) &= \frac{1}{n}\frac{2\lambda^2+1}{(2\lambda+1)^2}\int F_X(x)(1-F_X(x))dx + \frac{b}{n}V(\lambda) \\ &+ \frac{b^8}{576}\left(\mu_4 - \frac{3(1+6\lambda)\mu_2^2}{2\lambda}\right)^2 \int \left(f_X'''(x)\right)^2 dx, \end{aligned} \quad (6.7)$$

where, if the kernel $k$ has a compact support $[-1,1]$, it is obtained that:

$$\begin{aligned} V(\lambda) &= \frac{1}{(2\lambda+1)^2}\left[(2\lambda^2+1)\left(\int_{-1}^{1} k^2(t)dt + l\int_{-1}^{1} K^2(t)dt - 1\right)\right. \\ &+ 2\lambda\left(\int_{-1}^{1-l} k(t-l)k(t)dt + \int_{-1+l}^{1} k(t)k(t+l)dt + \int_{1-l}^{1} (k(t)+\lambda k(t+l))dt\right. \\ &- \left.\left.\int_{-1}^{-1+2l} K(t)dt + \lambda\int_{-1+l}^{1-l} (k(t-l)k(t+l) - l^2 K(t-l)K(t+l))dt\right)\right] \end{aligned} \quad (6.8)$$

There exists a value of $\lambda$ that minimises $V(\lambda)$, and this depends on the selected kernel, if the Epanchnikov kernel is used $\lambda = 0.0799$ and $V(0.0799) = -0.1472244$.

**Remark 6.3.1.** Let $F_X$ be a cdf with four bounded and continuous derivatives, the optimal bandwidth that minimises A-MISE is:

$$b^{MISE} = n^{-1/7}\left(\frac{-V(\lambda)}{\frac{\int (f_X'''(t))^2 dt}{72}\left(\mu_4 - \frac{3(1+6\lambda)\mu_2^2}{2\lambda}\right)^2}\right)^{1/7}. \quad (6.9)$$

Kim et al. (2006) did not analyse a method to estimate the optimal bandwidth. Similarly to the CKE, we can use iterative methods such as the plug-in methods or the methods based on cross-validation (see, Jones et al. 1996;for a review). Alternatively, the rule-of-thumb bandwidth is a direct way to estimate the smoothing parameter. Following Silverman (1986), for the BCCKE the rule-of-thumb bandwidth is obtained by replacing in expression (6.9) the functional $\int (f_X'''(x))^2 dx$ with its value assuming that $f_X$ is the density of a normal distribution with scale parameter $\sigma$, then:

$$b^{*} = n^{-1/7}\sigma\left(\frac{-V(\lambda)}{\frac{0.5289277}{72}\left(\mu_4 - \frac{3(1+6\lambda)\mu_2^2}{2\lambda}\right)^2}\right)^{1/7}. \quad (6.10)$$

The smoothing parameter in (6.10) can be estimated by replacing $\sigma$ with a consistent estimation, such as the sample standard deviation $s$. However, Silverman (1986) noted that for long-tailed and right-skewed distributions it is better to use a robust estimation of $\sigma$ based on the interquartile range $R$, that is $\frac{R}{1.34}$. In general, we can use the better estimator of $\sigma$ for each case: $\hat{\sigma} = Min\left(s, \frac{R}{1.34}\right)$.

## 6.4 Transformed kernel estimator with bias reducing technique

In this section we propose a new kernel estimator that combines the greater efficiency of the transformed kernel estimator of the cdf with the bias reduction technique. In general, the transformed kernel estimator involves selecting a transformation function so that the cdf or the pdf associated with the transformed variable can be estimated optimally with the classical kernel estimator or a bias-corrected version. We denote $T(\cdot)$ the transformation function, then the transformed random variable is $Y = T(X)$, and we know that $f_X(x) = f_Y(y)T'(x)$ and $F_X(x) = F_Y(y)$.

Let $T(\cdot)$ be a concave transformation function with at least four continuous derivatives. Assuming equal weights in (6.6), i.e. $\lambda_1 = \lambda_2 = \lambda > 0$, the bias corrected transformed kernel estimator (BCTKE) is:

$$\widetilde{F}_{T(X)}(T(x)) = \frac{\lambda\left[\widehat{F}_1(T(x)) + \widehat{F}_2(T(x))\right] + \widehat{F}_{T(X)}(T(x))}{2\lambda+1} = \widetilde{\widetilde{F}}_X(x). \tag{6.11}$$

We denote $y = T(x)$ and the transformed data are $Y_i = T(X_i)$, $i = 1, ..., n$, then:

$$\widehat{F}_{T(x)}(T(x)) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{T(x)-T(X_i)}{b}\right) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{y-Y_i}{b}\right) = \widehat{F}_Y(y) = \widehat{\widehat{F}}_X(x) \tag{6.12}$$

and

$$\begin{aligned} \widehat{F}_1(T(x)) &= \widehat{F}_{T(X)}(T(x)-lb) + lb\hat{\hat{f}}_X(x-lb) = \widehat{\widehat{F}}_1(x), \\ \widehat{F}_2(T(x)) &= \widehat{F}_{T(X)}(T(x)+lb) - lb\hat{\hat{f}}_X(x+lb) = \widehat{\widehat{F}}_1(x), \end{aligned} \tag{6.13}$$

where $\widehat{\widehat{f}}_X$ is the transformed kernel density estimation (see, for example, Bolancé 2010, Bolancé et al. 2008a, Buch-Larsen et al. 2005, Wand et al. 1991).

$$\hat{\hat{f}}_X(x) = \frac{1}{nb}\sum_{i=1}^{n} k\left(\frac{T(x)-T(X_i))}{b}\right) T'(x). \tag{6.14}$$

**Theorem 6.4.1.** *Let $F_X$ be a cdf with four bounded and continuous derivatives. Let $T(\cdot)$ be a concave transformation function with at least four continuous derivatives.. If the kernel $k$ has a compact support $[-1,1]$, we obtain that the bias and variance of BCTKE are:*

$$E\left(\widetilde{\widetilde{F}}_X(x) - F_X(x)\right) = \frac{b^4}{24}\left(\mu_4 - \frac{3(1+6\lambda)\mu_2^2}{2\lambda}\right)\frac{f_X'''(x)}{T'(x)} D\left(T^{(p)}(x), F_X^{(p)}(x)\right) + o\left(b^4\right), \tag{6.15}$$

$$Var\left(\widetilde{\widetilde{F}}_X(x)\right) = \frac{1}{n}\frac{2\lambda^2+1}{(2\lambda+1)^2}F_X(x)\left(1-F_X(x)\right) + \frac{f_X(x)}{T'\left(x\right)}\frac{b}{n}V(\lambda) + o\left(\frac{b^2}{n}\right). \tag{6.16}$$

*The function* $D\left(T^{(p)}(x), F_X^{(p)}(x)\right)$ *with* $p = 0, ..., 4$*, where the super-index between parentheses refers to the derivative, depends on the transformation* $T$*, the cdf* $F_X$ *and the first four derivatives of these functions, is such that:*

$$D\left(T^{(p)}(x), F_X^{(p)}(x)\right) = 0 \;\; \textit{if} \;\; T(x) = F(x)$$

*and*

$$D\left(T^{(p)}(x), F_X^{(p)}(x)\right) \to 0 \;\; \textit{if} \;\; T^{(p)}(x) \to F_X^{(p)}(x), \;\; \forall p = 0, ..., 4.$$

*Proof.* The bias and the variance of the BCTKE are obtained from the bias and variance of the BCCKE of $\widetilde{F}_Y(y)$, knowing that $F_Y(y) = F_X(x)$ and $f_Y(y) = \frac{f_X(x)}{T'(x)}$ and analysing the derivative $\left(\frac{f_X(x)}{T'(x)}\right)'''$.

$$\begin{aligned} \left(\frac{f_X(x)}{T'(x)}\right)''' &= \frac{f_X'''(x)}{T'(x)} - \frac{3f_X''(x)T''(x)}{T'(x)^2} - \frac{3T'''(x)f_X'(x)}{T'(x)^2} + \frac{6f_X'(x)T''(x)^2}{T'(x)^3} \\ &- \frac{f_X(x)T^{(4)}(x)}{T'(x)^2} - \frac{6f_X(x)T''(x)^3}{T'(x)^4} + \frac{6f_X(x)T'''(x)T''(x)}{T'(x)^3}, \end{aligned}$$

if $T^{(p)}(x) \to F_X^{(p)}(x), \;\; \forall p = 0, ..., 4$ we obtain that $D\left(T^{(p)}(x), F_X^{(p)}(x)\right) \to 0$. □

From the results of Theorem 6.4.1 we prove that if a suitable transformation is found, we can reduce the bias and the variance of the BCCKE.

### 6.4.1 Double transformed kernel estimator with bias correction

The bias-corrected double transformed kernel estimator (BCDTKE) is obtained in a similar manner to that used to obtain the DTKE estimator (see, Alemany et al. 2013).

Let $F$ be a continuous cdf with four bounded and continuous derivatives in a neighbourhood of $x$, we assume that $k$ is the kernel that is a symmetric pdf and with a compact support $[-1, 1]$ and $b$ is the bandwidth. The smoothing parameter $b$ holds that when $n \to \infty$, $b \to 0$ and $nb \to \infty$, then the A-MISE associated with the BCCKE of the transformed random variable $Y$ is:

$$\frac{1}{n}\frac{2\lambda^2+1}{(2\lambda+1)^2}\int F_Y(y)(1-F_Y(y))dx+\frac{b}{n}V(\lambda)$$
$$+\frac{b^8}{576}\left(\mu_4-\frac{3(1+6\lambda)\mu_2^2}{2\lambda}\right)^2\int(f_Y'''(y))^2dx$$

where $V(\lambda)<0$ is the function defined in (6.8).

Given $b$ and the kernel $k$, the A-MISE is minimum when functional $\int\left[f_Y'''(y)\right]^2dy$ is minimum. The proposed method is based on the transformation of the variable in order to achieve a distribution that minimises the A-MISE, i.e. that minimises $\int\left[f_Y'''(y)\right]^2dy$.

Terrell (1990) showed that the density of a $Beta\,(5,5)$ distribution defined on the domain $[-1,1]$ minimises $\int\left[f_Y'''(y)\right]^2dy$, in the set of all densities with known variance. The pdf $h$ and cdf $H$ of the $Beta\,(5,5)$ are:

$$\begin{aligned} h(x) &= \frac{315}{256}(1-x^2)^4 \quad -1\le x\le 1,\\ H(x) &= \frac{1}{256}(35x^4-175x^3+345x^2-325x+128)(x+1)^5. \end{aligned}$$

Then the BCDTKE is:

$$\widetilde{\widetilde{F}}_X(x)=\widetilde{F}_{H^{-1}(T(X))}(H^{-1}(T(x)))=$$
$$\frac{\lambda\left[\widehat{F}_{\{H^{-1}(T(X)),1\}}(H^{-1}(T(x)))+\widehat{F}_{\{H^{-1}(T(x)),2\}}(H^{-1}(T(x)))\right]+\widehat{F}_{H^{-1}(T(x))}(H^{-1}(T(x)))}{2\lambda+1},$$

where $T(.)$ is a first transformation that matches a cdf, so that the transformed sample $T(X_i)$, $i=1,...,n$, takes values from a $Uniform(0,1)$ distribution and, therefore, the double transformed sample $H^{-1}(T(X_i))$, $i=1,...,n$, takes values from a $Beta\,(5,5)$ distribution. Similarly to (6.13), we obtain that

$$\widehat{F}_{\{H^{-1}(T(x)),1\}}(x)=\widehat{F}_{H^{-1}(T(x))}H^{-1}(T(x-lb))+lb\widehat{f}_{H^{-1}(T(x))}H^{-1}(T(x-lb)),$$
$$\widehat{F}_{\{H^{-1}(T(x)),2\}}(x)=\widehat{F}_{H^{-1}(T(x))}H^{-1}(T(x+lb))-lb\widehat{f}_{H^{-1}(T(x))}H^{-1}(T(x+lb)),$$

where $\widehat{f}_{H^{-1}(T(x))}$ is the double transformed kernel density estimation (see, Bolancé 2010, Bolancé et al. 2008a):

$$\widehat{f}_{H^{-1}(T(X))}(H^{-1}(T(x))) =$$
$$\frac{1}{nb}\sum_{i=1}^{n} k\left(\frac{H^{-1}(T(x)) - H^{-1}(T(X_i))}{b}\right) H^{-1'}(T(x))\, T'(x).$$

The smoothing parameter $b$ in BCDTKE can be calculated from expressions (6.9) replacing $f'''$ by $Beta(5,5)$ pdf:

$$b^* = n^{-1/7}\left(\frac{-V(\lambda)}{\frac{1288.6}{72}\left(\mu_4 - \frac{3(1+6\lambda)\mu_2^2}{2\lambda}\right)^2}\right)^{1/7}. \tag{6.17}$$

## 6.5 Simulation study

We compare four kernel estimation methods: CKE, BCCKE, DTKE and BCDTKE. The first transformation $T(\cdot)$ that we use for obtaining DTKE and BCDTKE is the cdf of the modified Champernowne distribution[4] analysed by Buch-Larsen et al. (2005). These authors also proposed a method based on maximising a pseudo-likelihood function to estimate the parameters. We use the rule-of-thumb bandwidth based on minimising A-MISE.

To compare estimated cdfs with theoretical cdfs we use two distances:

$$\begin{aligned} L_1(\check{F}) &= \int \left|\check{F}(t) - F(t)\right| dt \\ L_2(\check{F}) &= \int \left(\check{F}(t) - F(t)\right)^2 dt, \end{aligned} \tag{6.18}$$

where $\check{F}$ represents the different estimators. Distances $L_1$ and $L_2$ evaluate the fit of the cdf differently. Distance $L_2$ attaches greater importance to the major differences

[4]The cdf of the modified Champernowne distribution is:

$$T(x) = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha},\ \ for\, x \geq 0,\ \ \alpha, M, c > 0.$$

between the theoretical cdf and the fitted cdf than is attached by distance $L_1$. When the aim is to fit an extreme value distribution, estimation errors tend to increase as the cdf approaches 1, due to a lack of sample information on the extreme values of the variable. Therefore, distance $L_2$ will be more strongly influenced by the estimation errors at the extreme values of the variable.

We generated 2000 samples for each sample size analysed: $n = 100$, $n = 500$, $n = 1000$ and $n = 5000$ and for each distribution in Table 6.1. We selected four distributions[5] that are positively skewed and which present different tail shapes: Lognormal, Weibull (both Gumbel types) and two mixtures of Lognormal-Pareto (both Fréchet types). The Lognormal and Weibull distributions both have an exponential tail. Specifically, we define the Weibull distribution with a scale parameter equal to 1 and shape parameter $\gamma$, so that the smaller the value of $\gamma$ the slower is the exponential decay in the tail, i.e. the lower the value of $\gamma$, the lighter the tail. For the Lognormal distribution, the shape parameter is $\sigma$. In this case, the higher the value of $\sigma$, the lighter the tail. Furthermore, we analyse two mixtures of Lognormal-Pareto, that is, distributions with "fat" tails or heavy-tailed distributions. As we proved in section 6.2, these mixtures are Fréchet type and have a Pareto tail; thus, in this case the smaller the value of shape parameter $\rho$, the heavier is the tail.

For each simulated sample, we estimated the cdf using the four methods: CKE, BCCKE, DTKE and BCDTKE and we calculated the distances defined in (6.18). Finally, for each sample size, we calculated the mean of the 2000 replicates. To calculate distances $L_1$ and $L_2$ with each simulated sample we used the grid proposed by Buch-Larsen et al. (2005) based on the change of variable defined by Clements et al. (2003), $y = \frac{x-M}{x+M}$, where $M$ is the sample median.

To obtain CKE and BCCKE we used two smoothing parameters: the rule-of-thumb, estimating $\sigma$ from the sample standard deviation $s$ and from $Min\left(s, \frac{R}{1.34}\right)$, where $R$ is the sample interquartile range. The results obtained with $s$ are shown in Tables A-1 and A-2 in the Appendix. Specifically, from the results in Table A-2, we can conclude that both estimators $-$CKE and BCCKE using rule-of-thumb, estimating $\sigma$ from the sample standard deviation $s-$ are not consistent when the distribution is heavy tailed.

[5] We used the same parameters as in Alemany et al. (2012; 2013).

TABLE 6.1: Distributions in the simulation study

| Distribution | $F_X(x)$ | Parameters |
|---|---|---|
| Weibull | $1-e^{-x^{\gamma}}$ | $\gamma = 0.75$ |
| | | $\gamma = 1.5$ |
| | | $\gamma = 3$ |
| Lognormal | $\int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$ | $(\mu, \sigma = (0, 0.25)$ |
| | | $(\mu, \sigma) = (0, 0.5)$ |
| | | $(\mu, \sigma) = (0, 1.0)$ |
| Mixture of Lognormal -Pareto | $p \int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt + (1-p)\left(1-\left(\frac{x-c}{\lambda}\right)^{-\rho}\right)$ | $(p, \mu, \sigma, \lambda, \rho, c) = (0.7, 0, 1, 1, 0.9, -1)$ |
| | | $(p, \mu, \sigma, \lambda, \rho, c) = (0.3, 0, 1, 1, 0.9, -1)$ |
| | | $(p, \mu, \sigma, \lambda, \rho, c) = (0.7, 0, 1, 1, 1.0, -1)$ |
| | | $(p, \mu, \sigma, \lambda, \rho, c) = (0.3, 0, 1, 1, 1.0, -1)$ |
| | | $(p, \mu, \sigma, \lambda, \rho, c) = (0.7, 0, 1, 1, 1.1, -1)$ |
| | | $(p, \mu, \sigma, \lambda, \rho, c) = (0.3, 0, 1, 1, 1.1, -1)$ |

In Tables 6.2 and 6.3 we compare the BCCKE, the DTKE and the BCDTKE with the CKE, i.e., we obtain the ratio between distances $L_1$ and $L_2$ that were obtained with the BCCKE, the DTKE and the BCDTKE and those that were obtained with the CKE. If the ratio is greater than 1, then the CKE is better; if it is lower, then the corrected estimator improves the CKE. The absolute distances are shown in Tables A-3 and A-4 in the Appendix.

The results presented in Tables 6.2 and 6.3 point to differences between distances $L_1$ and $L_2$ and, furthermore, there exist important differences between the results obtained for Gumbel-type and Fréchet-type distributions.

Focusing first on the DTKE, for distance $L_1$ this estimator does not improve the CKE in any case. Furthermore, when the sample is small the $L_1$ obtained for the DTKE is considerably worse than that obtained for the CKE. For distance $L_2$ the DTKE improves the CKE in small and large sample sizes. Focusing on $L_2$, we observe that the largest improvements of the DTKE occur when the distributions are Fréchet-type, although these improvements are not as great as those obtained when bias correction is used.

Focusing now on Gumbel-type distributions, the results in Table 6.2 show that, in general, both boundary correction approaches, the BCCKE and the BCDTKE, make similar improvements to the CKE in distance $L_2$ for all sample sizes. Furthermore, the improvement is greater as the sample size increases. For distance $L_1$ the BCCKE and the BCDTKE do not improve the CKE when the distribution has a lighter tail, i.e., the Weibull distributions with larger shape parameter and the Lognormal distributions with smaller shape parameter.

In Table 6.3 we show the results for the Fréchet-type distributions. We observe that, when the distribution has a heavier tail, the improvement of the BCDTKE with respect to the CKE is more marked than that obtained with BCCKE, for all sample sizes and both distances, except for distance $L_1$ in the case of 70Lognormal-30Pareto ($\rho = 1.1$) and sample size 100. In general, for distance $L_2$ the improvement of the BCDTKE with respect to the BCCKE is around $5\%$. For distance $L_1$ this improvement becomes greater as the sample size increases, exceeding $10\%$ in the case of 70Lognormal-30Pareto ($\rho = 1$).

TABLE 6.2: Comparative ratios obtained with the simulation results for Weibull and Lognormal (Gumbel-type distributions) using rule-of-thumb with scale parameter $Min\left(s, \frac{R}{1.34}\right)$.

| n | 100 | | 500 | | 1000 | | 5000 | |
|---|---|---|---|---|---|---|---|---|
| | Lognormal ($\sigma = 0.25$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| BCCKE | 1.0312 | 0.2002 | 1.0275 | 0.1334 | 1.0238 | 0.1112 | 1.0136 | 0.0742 |
| DTKE | 297.5476 | 0.6184 | 37.9005 | 0.1506 | 13.2495 | 0.1127 | 1.7839 | 0.0734 |
| BCDTKE | 1.0361 | 0.2030 | 1.0307 | 0.1350 | 1.0264 | 0.1124 | 1.0153 | 0.0748 |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| BCCKE | 0.9777 | 0.1882 | 0.9789 | 0.1235 | 0.9830 | 0.1051 | 0.9892 | 0.0703 |
| DTKE | 115.4618 | 0.4155 | 16.0135 | 0.1331 | 6.1701 | 0.1062 | 1.4138 | 0.0701 |
| BCDTKE | 0.9680 | 0.1885 | 0.9693 | 0.1236 | 0.9738 | 0.1052 | 0.9811 | 0.0704 |
| | Lognormal ($\sigma = 1$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| BCCKE | 0.9486 | 0.1625 | 0.9433 | 0.1069 | 0.9416 | 0.0883 | 0.9505 | 0.0588 |
| DTKE | 43.4451 | 0.3195 | 7.4330 | 0.1213 | 3.8088 | 0.0944 | 1.4983 | 0.0604 |
| BCDTKE | 0.9194 | 0.1598 | 0.9137 | 0.1054 | 0.9112 | 0.0871 | 0.9230 | 0.0582 |
| | Weibull ($\gamma = 0.75$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| BCCKE | 0.9626 | 0.1768 | 0.9444 | 0.1169 | 0.9454 | 0.0988 | 0.9383 | 0.0661 |
| DTKE | 15.5507 | 0.2372 | 2.1968 | 0.1254 | 1.4867 | 0.1024 | 1.1714 | 0.0657 |
| BCDTKE | 0.9338 | 0.1740 | 0.9139 | 0.1147 | 0.9140 | 0.0965 | 0.9015 | 0.0630 |
| | Weibull ($\gamma = 1.5$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| BCCKE | 1.0148 | 0.1996 | 0.9874 | 0.1321 | 0.9828 | 0.1114 | 0.9762 | 0.0731 |
| DTKE | 55.5021 | 0.2919 | 5.9624 | 0.1322 | 2.1170 | 0.1094 | 1.0632 | 0.0725 |
| BCDTKE | 1.0121 | 0.2006 | 0.9849 | 0.1327 | 0.9801 | 0.1119 | 0.9733 | 0.0733 |
| | Weibull ($\gamma = 3$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| BCCKE | 1.0644 | 0.2103 | 1.0396 | 0.1384 | 1.0328 | 0.1155 | 1.0168 | 0.0761 |
| DTKE | 53.9094 | 0.2656 | 2.4343 | 0.1357 | 1.3076 | 0.1134 | 1.0787 | 0.0755 |
| BCDTKE | 1.0699 | 0.2126 | 1.0440 | 0.1397 | 1.0365 | 0.1165 | 1.0200 | 0.0766 |

TABLE 6.3: Comparative ratios obtained with the simulation results for Mixtures of Lognormal-Pareto (Fréchet-type distributions) using rule-of-thumb with scale parameter $Min\left(s, \frac{R}{1.34}\right)$.

| n | 100 | | 500 | | 1000 | | 5000 | |
|---|---|---|---|---|---|---|---|---|
| | 70Lognormal-30Pareto ($\rho = 0.9$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| BCCKE | 0.9972 | 0.0767 | 0.9981 | 0.0434 | 0.9983 | 0.0342 | 0.9986 | 0.0260 |
| DTKE | 7.1804 | 0.2000 | 3.6744 | 0.0878 | 2.9466 | 0.0637 | 2.4752 | 0.0441 |
| BCDTKE | 0.9656 | 0.0724 | 0.9377 | 0.0411 | 0.9121 | 0.0322 | 0.9283 | 0.0247 |
| | 70Lognormal-30Pareto ($\rho = 1$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| BCCKE | 0.9950 | 0.0851 | 0.9970 | 0.0463 | 0.9975 | 0.0360 | 0.9982 | 0.0209 |
| DTKE | 10.3436 | 0.2193 | 4.6247 | 0.0895 | 3.3365 | 0.0614 | 2.3266 | 0.0318 |
| BCDTKE | 0.9948 | 0.0814 | 0.9490 | 0.0441 | 0.9259 | 0.0342 | 0.8928 | 0.0199 |
| | 70Lognormal-30Pareto ($\rho = 1.1$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| BCCKE | 0.9930 | 0.0943 | 0.9953 | 0.0519 | 0.9960 | 0.0401 | 0.9975 | 0.0222 |
| DTKE | 14.2912 | 0.2441 | 6.0630 | 0.0954 | 4.4212 | 0.0655 | 2.3962 | 0.0301 |
| BCDTKE | 1.0007 | 0.0908 | 0.9650 | 0.0499 | 0.9512 | 0.0386 | 0.9164 | 0.0214 |
| | 30Lognormal-70Pareto ($\rho = 0.9$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| BCCKE | 0.9975 | 0.0804 | 0.9982 | 0.0464 | 0.9984 | 0.0373 | 0.9986 | 0.0264 |
| DTKE | 4.6250 | 0.1790 | 2.3205 | 0.0757 | 2.0276 | 0.0571 | 1.7288 | 0.0347 |
| BCDTKE | 0.9698 | 0.0759 | 0.9123 | 0.0435 | 0.9084 | 0.0351 | 0.9479 | 0.0252 |
| | 30Lognormal-70Pareto($\rho = 1$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| BCCKE | 0.9972 | 0.0842 | 0.9976 | 0.0476 | 0.9980 | 0.0373 | 0.9983 | 0.0249 |
| DTKE | 6.2399 | 0.1963 | 2.9780 | 0.0794 | 2.3838 | 0.0570 | 1.7963 | 0.0319 |
| BCDTKE | 0.9619 | 0.0794 | 0.9227 | 0.0448 | 0.9182 | 0.0353 | 0.9360 | 0.0237 |
| | 30Lognormal-70Pareto ($\rho = 1.1$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| BCCKE | 0.9958 | 0.0911 | 0.9967 | 0.0508 | 0.9973 | 0.0391 | 0.9982 | 0.0233 |
| DTKE | 8.6851 | 0.2189 | 3.8710 | 0.0860 | 2.9610 | 0.0604 | 1.7962 | 0.0296 |
| BCDTKE | 0.9716 | 0.0867 | 0.9472 | 0.0483 | 0.9288 | 0.0372 | 0.9011 | 0.0223 |

# Conclusion

In this Thesis we addressed two very important themes related to quantitative risk management. On one hand, we provided relevant results about the analysis of extreme value distributions; on the other hand, we also presented different results concerning the dependence modelling between extreme value distributions. These results will be useful in the calculation of the capital requirement in the context of Solvency II in terms of quantifying the risk when the data contain extreme values. This can occur when we analyse operational risk and subscription risk, where the company could have losses that have a very low probability but can reach high values, i.e. "rare cases".

To obtain the results presented, the author programmed in R all the proposed estimators (these programmes are available for readers).

Specifically, two lines of research were examined in this Thesis: the dependence between two random variables from the viewpoint of the copulae and the nonparametric methods to estimate the cumulative distribution function and quantile. In addition some questions related to the theory of extreme values were considered: extreme value copulae and maximum domain of attraction of extreme value mixture distributions.

Inference on copulae was necessary for analysing the structure of dependence between variables. For this, using the definition of max-stable, we generalised the test of extreme value copula to cover a more extensive alternative hypothesis.

In the context of copulae, nonparametric estimation of the cdf was useful for obtaining the pseudo-observations and for estimating the marginals. We proposed the use of new nonparametric methods that improve the accuracy in the risk estimations.

To illustrate the usefulness of the methods analysed in this Thesis, we used data on the costs of accidents in auto insurance. Specifically, we used two databases, the first contains information from a sample of bivariate costs and the second contains information related to a sample of univariate cost for different types of policyholders.

In chapter 3, we introduced a test for the adequacy of extreme value copulae that allows us to determine the most suitable copula, especially when the data include extreme values. In this chapter, we presented an empirical application, which was extended in chapter 4. Among others, the K-Plot identified a positive and increasing dependence between variables related to automobile insurance claims and the test we presented for extreme value copulae confirms that, in our case, we should use an extreme value copula.

In the selection of the marginal distributions, 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. In fact, in order to fulfill the capital requirements driven by Solvency II, we find that using a Gumbel copula with Champernowne marginals yields the highest value of the VaR (Value-at-Risk) at the $99.5\%$ and $99.9\%$ levels, followed by the case where Weibull marginals are considered.

Furthermore, in chapter 4 we analysed the bounds of the VaR. The non-coherency of some risk measures is the motivation for some statisticians to try to estimate the bounds of the VaR of the aggregate loss. The results of Embrechts and Puccetti Embrechts and Puccetti (2006) using the Fréchet bounds when the marginal distributions are known and the copula is unknown are noteworthy. We have used some of these results and we have compared them with the confidence intervals provided by the Bootstrap method. In our study we found that the empirical VaR violates the subadditivity condition at some points, but it is always within the confidence interval of the VaR for the aggregate loss (using Bootstrap) at the $95\%$ level. Finally, we also considered the case of an unknown associated copula with the Champernowne marginal distributions and we concluded that the VaR simulated by using Gumbel copula and Champernowne marginal distributions is within the limits resulting from the bounds developed by Embrecht and Puccetti Embrechts and Puccetti (2006).

In chapter 5, we have shown how mixing copulae with the DTKE (double transformed kernel estimation) for marginals cdfs is useful when estimating the total risk of correlated losses from extreme value distributions. Among the copulae proposed, we analysed the Sarmanov copula, which has special characteristics. Additionally, we provided proof of some results related to the properties of the Sarmanov copula. From our data, we found that the Gumbel copula with DTKE marginals provides a good fit. With this copula we obtained a balanced risk estimation that guarantees that the risk is not underestimated and, where it is relevant, not overestimated in excess.

In general, in chapter 5 we proposed a method for estimating the total risk of loss when we have a multivariate sample of losses with upper tail dependence and heavy-tailed marginal distributions.

In a lot of analyses -in economics, finance, insurance, demography,...- the fit of cdf is very important for evaluating the probability of extreme situations. In these cases, the data are usually generated by a continuous random variable $X$ whose distribution may be the result of the mixture of different EVDs; then both the classical parametric models and the classical nonparametric estimates do not work for the estimation of the cdf. All those problems were addressed in chapter 6. There we presented a method to estimate cdf that is suitable when the loss is a heavy tailed random variable. The proposed double transformation kernel using the bias-corrected technique, in general, provides good fit results for the Gumbel and Fréchet types of extreme value distributions, especially when the sample size is small.

We show, when the sample size is small, that our proposed BCDTKE (bias-corrected double transformed kernel estimator) improves the classical kernel estimator and bias-corrected classical kernel estimator of the cumulative distribution function when the distribution is a right extreme value distribution and the maximum domain of attraction is the one associated with a Fréchet type distribution.

Furthermore, in chapter 6, we provided some theoretical results about the maximum domain of attraction of extreme value mixture distributions. We concluded that the heavier tail (Fréchet type) prevails over the lighter tails (Gumbel type).

Our future lines of research will focus on further investigation on nonparametric methods for fitting the copula and the marginal cdfs (cumulative distribution functions), in order to improve the accuracy of the nonparametric estimation of the risk. Specifically, we have four principal objectives:

1. To generalise the nonparametric fit of the univariate cdf to the bivariate focusing on the copula. There are two basic works in this field. The first is the paper by Chen and Huang (2007) where the authors analyse the theoretical properties of kernel estimator of the copula and propose the t-Student copula for obtaining the rule-of-thumb optimal bandwidth. The second paper is by Omelka et al. (2009) who present a new kernel estimator based on transformations.

2. To study the theoretical properties of nonparametric estimator of the Kendal function, defined in (1.23), and to compare its asymptotic convergence with the estimator from Genest and Rivest (1993). Our principal aim is to propose a new estimator for Archimedian copulae.

3. To analyse the theoretical properties of the quantile function derived from the bias-corrected double transformed kernel estimator and to compare with alternative estimators, such as those proposed by Charpentier and Ouilidi (2010), Chen (1999), Cheng and Sun (2006), Harrell and Davis (1982), Parzen (1979), Sheather and Marron (1990).

4. To adapt our nonparametric method to the estimation of the expected shortfall.

# Appendix

## Slowly and regularly varying

An L function Lebesgue-measurable on $[0, \infty]$ is said to be regularly varying at infinity with index $\alpha \in \mathbb{R}$ if:

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = t^{\alpha}, \quad t > 0.$$

When $\alpha = 0$, i.e.

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1, \quad t > 0,$$

L is said to be slowly varying at infinity.

## MDA conditions

If we note $F^{-1}(z) = \inf\{x : F(x) \geq z\}$ the generalized inverse function of $F$ (also called quantile when $0 < z < 1$), the possible domains of attraction of a distribution $F$ are summarised, (see, Embrechts et al. 1997):

- $F \in MDA(G_{\mu,\sigma,\xi})$ Fréchet type if, and only if, $1 - F$ is regularly varying with index $\frac{-1}{\xi}$, i.e.:

  $$\lim_{x \to \infty} = \frac{1 - F(tx)}{1 - F(x)} = t^{\frac{-1}{\xi}}, \quad t > 0.$$

  In this case one can choose $a_n = F^{-1}(1 - \frac{1}{n})$ and $b_n = 0$.

- $F \in MDA(G_{\mu,\sigma,\xi})$ Weibull type if, and only if, $r(F) < \infty$ and $1 - F(r(F) - \frac{1}{x})$ is regularly varying with index $\frac{-1}{\xi}$, i.e.:

$$\lim_{x\to\infty} \frac{1 - F(r(F) - \frac{1}{tx})}{1 - F(r(F) - \frac{1}{x})} = t^{\frac{-1}{\xi}}, \quad t > 0.$$

  In this case $a_n = F^{-1}(r(F) - \frac{1}{n})$ and $b_n = r(F)$ can be chosen.

- $F \in MDA(G_{\mu,\sigma,\xi})$ Gumbel type if, and only if, a positive function $\tilde{f}$ exists, such that

$$\lim_{x\uparrow r(F)} \frac{1 - F(x + t\tilde{f}(x))}{1 - F(x)} = e^{-t}.$$

  Here we have $\int_x^{r(F)}(1 - F(s))ds < \infty$, and the condition is true if we choose $\tilde{f}(t) = \frac{\int_x^{r(F)}(1-F(s))ds}{1-F(x)}$. A possible choice of parameters is: $a_n = \tilde{f}(b_n)$ and $b_n = F^{-1}(1 - \frac{1}{n})$.

Another equivalent tool is the following theorem that can be found in Castillo et al. (2004), representing a practical alternative to calculate the limit.

# Brownian bridge

A stochastic process $\{B_t, t \geq 0\}$ is a Brownian motion if the following conditions are fulfilled:

- $B_0 = 0$.
- Given $n$ periods, $0 \leq \cdots \leq t_n$, the increments $B_{t_n} - B_{t_{n-1}}, \ldots B_{t_2} - B_{t_1}$ are independent random variables.
- If $s < t$, the increment $B_t - B_s$ has a Normal law $N(0, t - s)$.
- The steeps are continuous process.

Let $B_t$ Brownian motion, it is said that the process $X_t$ is Brownian bridge if:

$$X_t = B_t - tB_1, \quad t \in [0,1].$$

It follows a Normal process center on with autocovariance function:

$$E(X_tX_s) = (s \wedge t) - st,$$

which verifies $X_0, X_1 = 0 \quad \text{and} \quad s \wedge t = min(s,t)$.

## Weak convergence

Let $(D,d)$ be a metric space. A Borel probability measure on $D$ is a function $P : B(D) \longrightarrow [0,\infty)$, where $B(D)$ is the Borel $\sigma$-algebra on $D$ such that:

- $P(\emptyset) = 0, \quad P(X) = 1.$
- If $B_1, B_2 \dots \in B$ are pairwise disjoint, then:

$$P(\cup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i)$$

Let $(D,d)$ be a metric space and $P, P_1, P_2 \dots P_n$ are Borel finite measures on D. We say that $P_n$ converges weakly to $P$ if:

$$\int_D f dP_n \rightarrow_{n\rightarrow\infty} \int_D f dP$$

$\forall f \in C_B(D)$ where $C_B(D) := \{f : D \rightarrow \mathbb{R}\ f\ bounded\ and\ continuous\}$.

For random variables $X, X_1, X_2 \dots X_n$ that take values in $D$, the weak convergence is equivalent to:

$$E(f(X_n)) \rightarrow E(f(X)) \quad \forall f \in C_B(D).$$

This condition is only valid when $f$ is in set of continuous function that are defined in metric separable spaces as $\mathbb{R}^n$ or $C[0,1]$.

## Bounds of VaR with unknown marginals

In their studies on univariate extreme value distributions, Kaas and Goovaerts (1986) delimit a distribution knowing only its first moments. Specifically, if $X$ is a variable whose distribution $F$ is unknown, $\mu = E(X) > 0$ and $Var(X) = \sigma^2 > 0$, the authors show that:

$$\underline{F}_{\mu_X,\sigma_X}(x) \leq F(x) \leq \overline{F}_{\mu_X,\sigma_X}(x),$$

where

$$\underline{F}_{\mu_X,\sigma_X}(x) = \begin{cases} \frac{\sigma_X^2}{\sigma_X^2+(x-\mu_X)^2} & \text{if } 0 \leq x \leq \mu_X \\ 1 & \text{if } x > \mu_X, \end{cases}$$

and

$$\overline{F}_{\mu_X,\sigma_x}(x) = \begin{cases} 0 & \text{if } 0 \leq x \leq \mu_X \\ \frac{\mu_X - x}{x} & \text{if } \mu_x < x \leq \frac{\sigma_X^2+\mu_X^2}{\mu_X} \\ \frac{(x-\mu_X)^2}{(x-\mu_X)^2+\sigma_X^2} & \text{if } x > \frac{\sigma_X^2+\mu_X^2}{\mu_X}. \end{cases}$$

Exploiting this result, Mesfioui and Quessy (2005) deduce the lower and upper limits of VaR. Let $q(x) = \sqrt{x/(1-x)}$ be an increasing function, then:

$$\begin{aligned} g_{a,b}(x) &= (a - bq(1-x))I(x \geq \frac{b^2}{a^2+b^2}) \\ h_{a,b}(x) &= a + aq^2(x)I\,(x \leq \frac{b^2}{a^2+b^2}) + bq(x)I(x > \frac{b^2}{a^2+b^2}), \end{aligned}$$

and *I* being the standard indicator, then the bounds proposed by Mesfioui and Quessy (2005) are:

$$\underline{VaR}_{\mu_X,\sigma_X}(\alpha) \leq VaR_\alpha(X) \leq \overline{VaR}_{\mu_X,\sigma_X}(\alpha),$$

where

$$\underline{VaR}_{\mu_X,\sigma_X}(\alpha) = \overline{F^{-1}}_{\mu_X,\sigma_x}(\alpha) = g_{\mu_X,\sigma_X}(\alpha) \quad \text{(B-1)}$$

$$\overline{VaR}_{\mu_X,\sigma_X}(\alpha) = \underline{F^{-1}}_{\mu_X,\sigma_x}(\alpha) = h_{\mu_X,\sigma_X}(\alpha). \quad \text{(B-2)}$$

# Tables

TABLE A-1: Simulation results for Weibull and Lognormal using rule-of-thumb with scale parameter $s$.

| n | 100 | | 500 | | 1000 | | 5000 | |
|---|---|---|---|---|---|---|---|---|
| | | | | Lognormal ($\sigma = 0.25$) | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 0.0161 | 0.1239 | 0.0073 | 0.0838 | 0.0051 | 0.0704 | 0.0023 | 0.0476 |
| BCCKE | 0.0163 | 0.0246 | 0.0075 | 0.0111 | 0.0052 | 0.0078 | 0.0024 | 0.0035 |
| | | | | Lognormal ($\sigma = 0.5$) | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 0.0410 | 0.1983 | 0.0180 | 0.1321 | 0.0131 | 0.1126 | 0.0058 | 0.0747 |
| BCCKE | 0.0378 | 0.0361 | 0.0169 | 0.0159 | 0.0124 | 0.0116 | 0.0055 | 0.0052 |
| | | | | Lognormal ($\sigma = 1$) | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 0.1735 | 0.4069 | 0.0848 | 0.2872 | 0.0611 | 0.2442 | 0.0273 | 0.1635 |
| BCCKE | 0.1338 | 0.0592 | 0.0644 | 0.0273 | 0.0462 | 0.0192 | 0.0214 | 0.0087 |
| | | | | Weibull ($\gamma = 0.75$) | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 0.1116 | 0.3265 | 0.0534 | 0.2276 | 0.0387 | 0.1940 | 0.0179 | 0.1321 |
| BCCKE | 0.0966 | 0.0545 | 0.0452 | 0.0253 | 0.0328 | 0.0185 | 0.0151 | 0.0088 |
| | | | | Weibull ($\gamma = 1.5$) | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 0.0360 | 0.1854 | 0.0170 | 0.1278 | 0.0122 | 0.1085 | 0.0054 | 0.0724 |
| BCCKE | 0.0362 | 0.0368 | 0.0167 | 0.0168 | 0.0120 | 0.0121 | 0.0053 | 0.0053 |
| | | | | Weibull ($\gamma = 3$) | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 0.0187 | 0.1330 | 0.0085 | 0.0902 | 0.0061 | 0.0766 | 0.0027 | 0.0515 |
| BCCKE | 0.0198 | 0.0280 | 0.0088 | 0.0125 | 0.0063 | 0.0088 | 0.0028 | 0.0039 |

TABLE A-2: Simulation results for Mixtures of Lognormal-Pareto using rule-of-thumb with scale parameter $s$.

| n | 100 | | 500 | | 1000 | | 5000 | |
|---|---|---|---|---|---|---|---|---|
| | 70Lognormal-30Pareto ($\rho = 0.9$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 8.0598 | 1.9372 | 9.4548 | 2.0452 | 12.6591 | 2.1364 | 13.2365 | 2.3089 |
| CKEbrt | 5.0740 | 0.2238 | 5.0508 | 0.2990 | 6.8743 | 0.3847 | 7.4525 | 0.6335 |
| | 70Lognormal-30Pareto ($\rho = 1$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 2.9700 | 1.4059 | 3.4953 | 1.4343 | 5.0217 | 1.4872 | 5.6733 | 1.4808 |
| CKEbrt | 1.9101 | 0.1472 | 1.9262 | 0.1721 | 2.4518 | 0.2140 | 3.4110 | 0.3214 |
| | 70Lognormal-30Pareto ($\rho = 1.1$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 2.5304 | 1.0949 | 1.7559 | 1.0465 | 2.8612 | 1.0500 | 2.3528 | 1.0210 |
| CKEbrt | 1.2847 | 0.1156 | 1.0331 | 0.1066 | 1.9549 | 0.1304 | 1.3763 | 0.1758 |
| | 30Lognormal-70Pareto ($\rho = 0.9$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 18.5349 | 2.9764 | 23.3374 | 3.2261 | 25.2348 | 3.2729 | 30.4567 | 3.4660 |
| CKEbrt | 11.2088 | 0.3977 | 12.3226 | 0.5722 | 12.5870 | 0.6789 | 19.1204 | 1.0613 |
| | 30Lognormal-70Pareto($\rho = 1$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 10.6396 | 2.1808 | 6.2217 | 2.0786 | 9.6786 | 2.1426 | 9.5012 | 2.0563 |
| CKEbrt | 5.6933 | 0.2687 | 3.5953 | 0.2862 | 5.0031 | 0.3613 | 5.6010 | 0.5123 |
| | 30Lognormal-70Pareto ($\rho = 1.1$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 4.9616 | 1.5990 | 4.8501 | 1.5662 | 4.5138 | 1.5305 | 4.0134 | 1.4511 |
| CKEbrt | 2.5866 | 0.1808 | 3.1662 | 0.2080 | 2.8573 | 0.2260 | 2.1638 | 0.2942 |

TABLE A-3: Simulation results for Weibull and Lognormal using rule-of-thumb with scale parameter $Min\left(s, \frac{R}{1.34}\right)$.

| n | 100 | | 500 | | 1000 | | 5000 | |
|---|---|---|---|---|---|---|---|---|
| | | | | Lognormal ($\sigma = 0.25$) | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 0.0159 | 0.1232 | 0.0073 | 0.0837 | 0.0051 | 0.0703 | 0.0023 | 0.0475 |
| BCCKE | 0.0164 | 0.0247 | 0.0075 | 0.0112 | 0.0052 | 0.0078 | 0.0024 | 0.0035 |
| DTKE | 4.7257 | 0.0762 | 0.2756 | 0.0126 | 0.0678 | 0.0079 | 0.0042 | 0.0035 |
| BCDTKE | 0.0165 | 0.0250 | 0.0075 | 0.0113 | 0.0053 | 0.0079 | 0.0024 | 0.0036 |
| | | | | Lognormal ($\sigma = 0.5$) | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 0.0389 | 0.1931 | 0.0173 | 0.1294 | 0.0126 | 0.1107 | 0.0056 | 0.0739 |
| BCCKE | 0.0380 | 0.0363 | 0.0170 | 0.0160 | 0.0124 | 0.0116 | 0.0056 | 0.0052 |
| DTKE | 4.4878 | 0.0802 | 0.2776 | 0.0172 | 0.0780 | 0.0118 | 0.0079 | 0.0052 |
| BCDTKE | 0.0376 | 0.0364 | 0.0168 | 0.0160 | 0.0123 | 0.0116 | 0.0055 | 0.0052 |
| | | | | Lognormal ($\sigma = 1$) | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 0.1415 | 0.3695 | 0.0680 | 0.2575 | 0.0488 | 0.2182 | 0.0222 | 0.1472 |
| BCCKE | 0.1342 | 0.0600 | 0.0642 | 0.0275 | 0.0460 | 0.0193 | 0.0211 | 0.0087 |
| DTKE | 6.1486 | 0.1180 | 0.5058 | 0.0312 | 0.1859 | 0.0206 | 0.0333 | 0.0089 |
| BCDTKE | 0.1301 | 0.0591 | 0.0622 | 0.0272 | 0.0445 | 0.0190 | 0.0205 | 0.0086 |
| | | | | Weibull ($\gamma = 0.75$) | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 0.1008 | 0.3107 | 0.0476 | 0.2147 | 0.0344 | 0.1829 | 0.0157 | 0.1236 |
| BCCKE | 0.0970 | 0.0549 | 0.0450 | 0.0251 | 0.0326 | 0.0181 | 0.0147 | 0.0082 |
| DTKE | 1.5673 | 0.0737 | 0.1046 | 0.0269 | 0.0512 | 0.0187 | 0.0184 | 0.0081 |
| BCDTKE | 0.0941 | 0.0541 | 0.0435 | 0.0246 | 0.0315 | 0.0176 | 0.0142 | 0.0078 |
| | | | | Weibull ($\gamma = 1.5$) | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 0.0357 | 0.1848 | 0.0169 | 0.1276 | 0.0122 | 0.1083 | 0.0054 | 0.0723 |
| BCCKE | 0.0363 | 0.0369 | 0.0167 | 0.0169 | 0.0120 | 0.0121 | 0.0053 | 0.0053 |
| DTKE | 1.9834 | 0.0540 | 0.1009 | 0.0169 | 0.0258 | 0.0119 | 0.0057 | 0.0052 |
| BCDTKE | 0.0362 | 0.0371 | 0.0167 | 0.0169 | 0.0119 | 0.0121 | 0.0053 | 0.0053 |
| | | | | Weibull ($\gamma = 3$) | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 0.0187 | 0.1330 | 0.0085 | 0.0902 | 0.0061 | 0.0766 | 0.0027 | 0.0515 |
| BCCKE | 0.0199 | 0.0280 | 0.0088 | 0.0125 | 0.0063 | 0.0088 | 0.0028 | 0.0039 |
| DTKE | 1.0059 | 0.0353 | 0.0207 | 0.0122 | 0.0080 | 0.0087 | 0.0030 | 0.0039 |
| BCDTKE | 0.0200 | 0.0283 | 0.0089 | 0.0126 | 0.0063 | 0.0089 | 0.0028 | 0.0039 |

TABLE A-4: Simulation results for Mixtures of Lognormal-Pareto using rule-of-thumb with scale parameter $Min\left(s, \frac{R}{1.34}\right)$.

| n | 100 | | 500 | | 1000 | | 5000 | |
|---|---|---|---|---|---|---|---|---|
| | 70Lognormal-30Pareto ($\rho = 0.9$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 3.0542 | 1.5946 | 2.3282 | 1.4453 | 1.9522 | 1.3488 | 0.9512 | 0.9403 |
| BCCKE | 3.0457 | 0.1223 | 2.3239 | 0.0628 | 1.9490 | 0.0461 | 0.9498 | 0.0210 |
| DTKE | 21.9300 | 0.3189 | 8.5546 | 0.1269 | 5.7525 | 0.0860 | 2.3543 | 0.0356 |
| BCDTKE | 2.9490 | 0.1155 | 2.1832 | 0.0594 | 1.7806 | 0.0434 | 0.8830 | 0.0200 |
| | 70Lognormal-30Pareto ($\rho = 1$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.6856 | 1.1958 | 1.3384 | 1.0788 | 1.1827 | 1.0244 | 0.6770 | 0.8077 |
| BCCKE | 1.6771 | 0.1018 | 1.3343 | 0.0500 | 1.1797 | 0.0369 | 0.6758 | 0.0169 |
| DTKE | 17.4350 | 0.2622 | 6.1896 | 0.0965 | 3.9460 | 0.0629 | 1.5752 | 0.0257 |
| BCDTKE | 1.6768 | 0.0973 | 1.2701 | 0.0476 | 1.0950 | 0.0351 | 0.6045 | 0.0161 |
| | 70Lognormal-30Pareto ($\rho = 1.1$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 1.0170 | 0.9261 | 0.7713 | 0.8170 | 0.6550 | 0.7628 | 0.4306 | 0.6341 |
| BCCKE | 1.0099 | 0.0873 | 0.7677 | 0.0424 | 0.6524 | 0.0306 | 0.4295 | 0.0141 |
| DTKE | 14.5344 | 0.2261 | 4.6766 | 0.0779 | 2.8958 | 0.0500 | 1.0318 | 0.0191 |
| BCDTKE | 1.0177 | 0.0841 | 0.7443 | 0.0408 | 0.6230 | 0.0295 | 0.3946 | 0.0136 |
| | 30Lognormal-70Pareto ($\rho = 0.9$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 6.3599 | 2.3361 | 4.5735 | 2.0597 | 3.4532 | 1.8223 | 1.5239 | 1.1857 |
| BCCKE | 6.3439 | 0.1877 | 4.5654 | 0.0956 | 3.4477 | 0.0679 | 1.5218 | 0.0313 |
| DTKE | 29.4148 | 0.4181 | 10.6127 | 0.1558 | 7.0016 | 0.1041 | 2.6346 | 0.0411 |
| BCDTKE | 6.1676 | 0.1774 | 4.1721 | 0.0897 | 3.1368 | 0.0640 | 1.4445 | 0.0298 |
| | 30Lognormal-70Pareto($\rho = 1$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 3.6494 | 1.7381 | 2.5657 | 1.5209 | 2.0622 | 1.3898 | 0.9779 | 0.9545 |
| BCCKE | 3.6391 | 0.1464 | 2.5596 | 0.0723 | 2.0582 | 0.0519 | 0.9763 | 0.0238 |
| DTKE | 22.7718 | 0.3411 | 7.6408 | 0.1208 | 4.9160 | 0.0792 | 1.7567 | 0.0305 |
| BCDTKE | 3.5104 | 0.1381 | 2.3675 | 0.0682 | 1.8936 | 0.0491 | 0.9153 | 0.0226 |
| | 30Lognormal-70Pareto ($\rho = 1.1$) | | | | | | | |
| | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ | $L_1$ | $L_2$ |
| CKE | 2.0891 | 1.3123 | 1.5079 | 1.1466 | 1.2428 | 1.0595 | 0.7175 | 0.8321 |
| BCCKE | 2.0803 | 0.1196 | 1.5029 | 0.0582 | 1.2394 | 0.0415 | 0.7162 | 0.0194 |
| DTKE | 18.1441 | 0.2872 | 5.8372 | 0.0986 | 3.6800 | 0.0640 | 1.2888 | 0.0246 |
| BCDTKE | 2.0298 | 0.1138 | 1.4284 | 0.0554 | 1.1543 | 0.0395 | 0.6466 | 0.0185 |

# Bibliography


Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44:182–198.

Abdous, B., Ghoudi, K., and Khoudraji, A. (1999). Non-parametric estimation of the limit dependence function of multivariate extremes. *Extremes*, 2:243–268.

Alemany, R., Bolancé, C., and Guillén, M. (2012). Non-parametric estimation of value-at-risk. *XARXA de Referencia en Economia Aplicada*, Working paper XREAP2012-19.

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

Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9:203–228.

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, 68:326–328.

Bahraoui, Z., Bolancé, C., and Alemany, R. (2013). Estimación núcleo transformada corregida: doble transformación versus núcleos asimétricos en la estimación del riesgo de pérdida. In *Risk2013: 5ª Reunión de Investigación en Seguros y Gestión del Riesgo*, Las Palmas de Gran Canaria. ULPGC.

Bahraoui, Z., Bolancé, C., and Alemany, R. (2014a). Using a kernel approach for estimating a cumulative distribution function with gumbel and fréchet type maximum domain of attraction. In *Risk management in Insurance*, Barcelona. UB.

Bahraoui, Z., Bolancé, C., and Pérez-Marín, A. M. (2014b). Extreme value copulas and marginal effects. In Girardin, V., Christos, H., and Skiadas, J. R. B., editors, *New Perspectives on Stochastic Modeling and Data Analysis*.

Bahraoui, Z., Bolancé, C., and Pérez-Marín, A. M. (2014c). Testing extreme value copulas to estimate the quantile. *SORT-Statistics and Operations Research Transaction*, 38:89–102.

Bairamov, I., Altinsoy, B., and Kerns, G. J. (2011). On generalized sarmanov bivariate distributions. *Journal of Applied and Engineering Mathematics*, 1:86–97.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of extremes: theory and applications*. Wiley, England.

Ben Ghorbal, M., Genest, C., and Néslehová, J. (2009). On the test of ghoudi, khoudraji, and rivest for extreme-value dependence. *Canadian Journal of Statistics*, 37:534–552.

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

Bolancé, C., Ayuso, M., and Guillén, M. (2012a). A nonparametric approach to analyzing operational risk with an application to insurance fraud. *The Journal of Operational Risk*, 7:57–75.

Bolancé, C., Bahraoui, Z., and Artís, M. (2014). Quantifying the risk using copulae with nonparametric marginalse. *Insurance: Mathematics and Economics*, 58:46–56.

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

Bolancé, C., Guillén, M., and Nielsen, J. P. (2009). Transformation kernel estimation of insurance claim cost distributions. In Corazza, M. and Pizzi, C., editors, *Mathematical and Statistical Methods for Actuarial Sciences and Finance, Springer*, pages 223–231.

Bolancé, C., Guillén, M., Nielsen, J. P., and Gustafsson, J. (2012b). *Quantitative Operational Risk Models*. Chapman and Hall, CRC Finance Series, New York.

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.

Bortkiewicz, L. (1922). Variationsbreite und mittlerer fehler. *Sitzungsberichte der Berliner Math*, 21:3–11.

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

Bucher, A. (2011). *Statistical Inference for Copulas and Extremes*. PhD thesis, University of Bochum, Germany.

Caperaá, P., Fougéres, A. L., and Genest, C. (2000). Bivariate distribution with given extreme value attractor. *Journal of Multivariate Analysis*, 72:30–49.

Castillo, E., Ali, S. H., Balakrishnan, N., and Sarabia, J. M. (2004). *Extreme Value and Related Models with Applications in Engineering and Science*. Wiley, New Jersey.

Chambers, J. M., Mallows, C. L., and Stuck, B. W. (1976). A method for simulating stable random variables. *Journal of the American Statistical Association*, 71:340–344.

Charpentier, A. and Ouilidi, A. (2010). Beta kernel quantile estimators of heavy-tailed loss distributions. *Statistical Computing*, 20:35–55.

Chen, S. S. (1999). A beta kernel estimator for density functions. *Computational Statistics & Data Analysis*, 31:131–145.

Chen, S. X. and Huang, T. M. (2007). Nonparametric estimation of copula functions for dependence modelling. *Canadian Journal of Statistics*, 35:265–282.

Cheng, M. Y. and Sun, S. (2006). Bandwidth selection for kernel quantile estimation. *Journal of Chines Statistical Association*, 44:271–295.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula methods in finance*. Chichester, wiley edition.

Choi, E. and Hall, P. (1998). On bias reduction in local linear smoothing. *Biometrika*, 85:333–345.

Clayton, D. G. (1978). A model for association in bivariate life tables and its applicationin epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141–151.

Clements, A. E., Hurn, A. S., and Lindsay, K. A. (2003). Möbius-like mappings and their use in kernel density estimation. *Journal of the American Statistical Association*, 98:993–1000.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, London.

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

De Matties, R. (2001). *Fitting copulas to data*. PhD thesis, Institute of Mathematics of the University of Zurich.

Deheuvels, P. (1979). La fonction de dependence empirique et ses propiétés. un test non paramètrique d'independence. *Bulletins de l'Academie Royale des Sciences, des Lettres et des Beaux Arts de Belgique*, 65:274–292.

Denuit, M., Dhaene, J., and Ribas, C. (2001). Does positive dependence between individual risks increase stop-loss premiums? *Insurance: Mathematics and Economics.*, 28:305–308.

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

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer, New York.

Dhaene, J., Denuit, M., Goovarts, M. J., Kaas, R., and Vyncke, D. (2006). Risk measure and comonotonicity. *A Review. Stochastic Models*, 22:573–606.

Dowd, K. and Blake, D. (2006). After VaR: The theory, estimation, and insurance applications of quantile-based risk measures. *The Journal of Risk and Insurance*, 73:193–229.

Embrechts, P., Höing, A., and Juri, A. (2003). Using copulae to bound the value-at-risk for functions of dependent risks. *Finance and Stochastics*, 7:145–167.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Zurich, 4 edition.

Embrechts, P. and Puccetti, G. (2006). Bounds for functions of dependent risks finance and stochastics. *Finance and Stochastics*, 10:341–352.

Embrechts, P., Puccetti, G., and Rüschendorf, L. (2013). Model uncertainty and var aggregation. *Journal of Banking and Finance*, 37:2750–2764.

Farlie, D. G. J. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47:307–323.

Fermanian, J. D., Radulovisc, D., and Wegkamp, M. H. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10:2847–860.

Fisher, N. (1997). *Copulas. In: Encyclopedia ofStatistical Sciences, Update Vol 1. pp 159-163 (Kotz S, Read CB, Banks DL,eds.)*. Wiley, New York.

Fisher, R. A. and Tippet, L. J. C. (1928). Limiting forms of the frequency distribution of largest of smallest member of a sample. *Proc Cambridge Philos Soc*, 24:180–190.

Frank, M. (1979). On the simulaneous associativity of f(x,y) and x+y-f(x,y). *Aequationes mathematicae*, 19:194–226.

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Société Polanaise de Mathematique. Annales*, 6:93–116.

Fréchet, M. (1957). Sur les tableaux de corrélation dont les marges sont données. *Annales University Lyon*, 20:13–31.

Frees, E. W. and Valdez, E. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2:1–25.

Fuller, W. E. (1914). Flood flows. *American Society of Civil Engineers*, 77:564–617.

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

Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statastics*. Wiley, New York.

Geffroy, J. (1959). Contribution a la théorie des valuers extrêmes. *Publications de l' Institut de Statistique de l'Université de Paris 8*, pages 123–184.

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika*, 74:549–555.

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

Genest, C. and Favre, A. C. (2007). Everything you always wanted to know about copulamodeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12:347–368.

Genest, C., Guoudi, K., and Rivest, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82:543–552.

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 MacKay, J. (1986a). Copules archimédiennes et families de lois bidimensionnelles dont les marges sont données. *Canadian Journal of Statistics*, 14:145–159.

Genest, C. and MacKay, J. (1986b). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, 40:280–283.

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

Genest, C., Quessy, J. F., and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33:337–366.

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

Genest, C. and Rivest, L. P. (2001). On the multivariate probability integral transformation. *Statistics and Probability Letters*, 53:391–399.

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44:199–213.

Ghorbal, B., Genest, C., and Neslehová, J. (2009). On the test of ghoudi, khoudraji, and rivest for extreme-value dependence. *Canadian Journal of Statistics*, 37:1–9.

Ghoudi, K., Khoudraji, A., and Rivest, L. P. (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Canadian Journal of Statistics*, 26:187–197.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44:423–453.

Griffith, A. A. (1921). The phenomena of rupture and flow in solids. *Philosophical Transactions of the Royal Society of London*, 221:163–198.

Grønnrberg, S. and Hjort, N. L. (2014). The copula information criteria. *Scandinavian Journal of Statistics*, 41:436–459.

Guillén, M., Nielsen, J. P., and Pérez-Marín, A. (2007). Improving the efciency of the nelson-aalenestimator: the naive local constant estimator. *Annals of the Institute of Statistical Mathematics*, 34:419–431.

Guillén, M., Prieto, F., and Sarabia, J. M. (2011). Modelling losses and locating the tail with the pareto positive stable distribution. *Insurance: Mathematics and Economics*, 49:454–461.

Guillén, M., Sarabia, J. M., and Prieto, F. (2013). Simple risk measure calculations for sums of positive random variables. *Insurance: Mathematicsand Economics*, 53:273–280.

Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press, New York.

Gumbel, E. J. (1960a). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55:698–707.

Gumbel, E. J. (1960b). Distributions des valeurs extrêmes en plusiers dimensions. *Publications de l'Institut de statistique de l'Université de Paris*, 9:171–173.

Harrell, F. E. and Davis, C. E. (1982). A new distribution-free quantile estimator. *Biometrika*, 69:635–640.

Hernández-Bastida, A. and Fernández-Sánchez, M. P. (2012). A sarmanov family with beta and gamma marginal distributions: anapplication to the bayes premium in a collective risk model. *Statistical Methods & Applications*, 21:391–409.

Hernández-Bastida, A., Fernández-Sánchez, M. P., and Gómez-Déniz, E. (2009). The net bayes premium with dependence between the risk profiles. *Insurance: Mathematics and Economics*, 45:247–254.

Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2013). Copula: Multivariate dependence with copulas. *cran.r-project.org*.

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

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological events. *Quarterly Journal of the Royal Meteorological Society*, 81:158–172.

Joe, H. (1997). *Multivariate Models and Dependence Concept*. Chapman and Hall, London.

Jones, M. C., Marron, J. S., and Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91:401–407.

Jorion, P. (2007). *Value at risk. The new Benchmark for managing Financial Risk*. The McGraw-Hill Companies, 2 edition.

Juri, A. and Wüthrich, M. (2002). Copula convergence theorems for tail events. *Insurance: Mathematics and Economics.*, 27:151–168.

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

Kaas, R. and Goovaerts, M. G. (1986). Best bounds for positive distributions with fixed moments. *Insurance: Mathematics and Economics*, 5:287–922.

Kim, C., Kim, S., Park, M., and Lee, H. (2006). A bias reducing technique in kernel distribution function estimation. *Computational Statistics*, 21:589–601.

Kim, G., Silvapulle, M. J., and Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis*, 51:2836–2850.

Klugman, S. A., Panjer, H. H., and Willmot, G. (2004). *Loss Models: From Data to Decisions*. Wiley, New York, 2 edition.

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

Kojadinovic, I. and Yan, J. (2010). Comparison of three semiparametric methods for estimating dependence parameters in copula models. *Insurance: Mathematics and Economics*, 47:52–63.

Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London.

Lee, M. L. (1996). Properties and applications of the sarmanov family of bivariate distributions. *Communications in Statistics-Theory and Methods*, 25:1207–1222.

Ling, C. H. (1965). Representation of associative functions. *Publ. Math. Debrecen*, 12:189–212.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press, Princeton.

Mesfioui, M. and Quessy, J. F. (2005). Bounds on the value at risk for the sum of possibly dependent risks. *Insurance: Mathematics and Economics*, 372:135–151.

Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitteilungsblatt für Mathematische Statistik*, 8:234–235.

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

Nelson, R. B., Quesada-Molina, J. A., and Úbeda Flores, M. (2001). Distribution functions of copulas: a class of bivariate probability integral transforms. *Statistics and Probability Letters*, 54:277–282.

Nelson, R. B., Quesada-Molina, J. A., and Úbeda Flores, M. (2003). Kendall distribution functions. *Statistics and Probability Letters*, 65:263–268.

Nielsen, J. P., Guillén, M., , Bolancé, C., and Gustafsson, J. (2012). Quantitative modeling of operational risk losses when combining internal and external data. *Journal of Financial Transformation*, 35:179–185.

Omelka, M., Gijbels, I., and Veraverbeke, N. (2009). Improved kernel estimation of copulas:weak onvergence and goodness-of-fit testing. *The Annals of Statistics*, 37:3023–3058.

Pareto, V. (1896). *Ecrits sur la courbe de la répartition de la richesse. Reproduit par Busino.G, 1965*. Librairie Droz, Geneve.

Parzen, E. (1979). Nonparametric statistical data modeling (with discussion). *Journal of the American Statistical Association*, 74:105–131.

Pickands, P. (1981). Multivariate extreme value distributions. *In: Bulletin of Proceeding of the 43rd Session of the International Statistical Institute. Buenos Aires*, pages 189–212.

Reiss, R. D. and Thomas, M. (1997). *Statistical Analysis of Extreme Values from Insurance, Finance, Hydrology and Others Fields*. Birkhäuser Verlag, Berlin.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York, 2 edition.

Salmon, F. (2009). Recipe for disaster: The formula that killed wall street function. *Wired Magazine*, 17.03.

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

Sarmanov, O. V. (1966). Generalized normal correlation and two-dimensional fréchet. *Soviet Mathematics. Doklady*, 25:1207–1222.

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

Schucany, W. R., Parr, W. C., and Boyer, J. E. (1978). Correlation structure in farlie-gumbel-morgenstern distribution. *Biometrika*, 25:650–653.

Scott, D. W. (2001). Parametric statistical modeling by minimum integrated square error. *Technometrics*, 43:274–285.

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

Sheather, S. J. and Marron, J. S. (1990). Kernel quantile estimators. *Journal of the American Statistical Association*, 85:410–416.

Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameterin copula models for bivariate survival data. *Biometrics*, 51:1384–1399.

Shirahata, S. and Chu, I. S. (1992). Integrated squared error of kernel-type estimator of distribution function. *Annals of the Institute of Statistical Mathematics*, 579–591:44.

Sibuya, M. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11:195–210.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, CRC Finance Series, London.

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.

Sklar, A. and Schweizer, B. (1983). *Probabilistic Metric Spaces*. Dover Publications Inc, North-Holland, New York.

Terrell, G. R. (1990). The maximal smoothing principle indensity estimation. *Journal of the American Statistical Association*, 85:270–277.

Tiago de Oliveira, J. (1958). Extremal distributions. *Revista Faculdade de Ciencias de Lisboa,2 Ser., A, Mat*, 7:219–227.

Tiago de Oliveira, J. (1984). *Bivariate models for extremes*. Reidel, Dordrecht.

Van der Vaart, A. W. and Wellner, J. (2000). *Weak convergence and empirical processes*. Springer, 2 edition.

Von Mises, R. (1923). Uber die variationsbreite einer beobachtungsreihe. *Berlin. Math. Ges*, 22:3–8.

Von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Rev. Math. Union Interbalk. Reproduced in: Selected Papers of Richard Von Mises, II(1954), pp. 271–294, Amer. Math. SOC.*, pages 141–160.

Wand, P., Marron, J. S., and Ruppert, D. (1991). Transformations in density estimation. *Journal of the American Statistical Association*, 86:343–361.

Wang, S. S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, 17:43–54.

Wang, S. S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 26:71–92.

Yang, Y. and Hashorva, E. (2013). Extremes and products of multivariate ac-product risks. *Insurance: Mathematics and Economics*, 52:312–319.