UNIVERSITY CARLOS III DE MADRID
STATISTICS DEPARTMENT
MATHEMATICAL ENGINEERING PHD.

PHD. THESIS

THE MULTIVARIATE DIRECTIONAL APPROACH: HIGH-LEVEL QUANTILE ESTIMATION AND APPLICATIONS TO FINANCE AND ENVIRONMENTAL PHENOMENA

Author: Raúl Andrés TORRES DÍAZ

Advisor: Rosa Elvira LILLO RODRÍGUEZ

December 2016

A mi familia
mí tesoro más valioso.

To my family
my most valuable treasure.

AGRADECIMIENTOS

Los logros que alcanzamos siempre vienen acompañados de múltiples apoyos concedidos por diferentes personas, en diferentes etapas de nuestro proceso. Por ello deseo manifestar mi gratitud a todas las personas que fueron artifices durante mis años de doctorado y las cuales me permitieron el desarrollo y buen término de este trabajo. Y de todas ellas resaltar el apoyo fundamental que me concedieron: Mi directora de tesis Rosa Elvira Lillo, quien estuvo guiando mi investigación, pero a la vez estuvo apoyando mis decisiones, mis propuestas, pendiente de mi persona; siempre atenta, cordial e incluso maternal.

Mis padres, Adan Torres y Olga Díaz, quienes han apoyado siempre mis próposi-tos y aún en la distancia siempre aportandome el coraje para revitalizar cuando se flaquea y continuar creciendo sin conformismos. Junto a ellos también agradecer a mis hermanos Camilo, Erika y Felipe, mis sobrinos, mis abuelos y mi tía; miembros fundamentales en mi familia que han dado mucho de si mismos para apoyar este proyecto. Y formando parte de mi familia también agradecer a mi novia, Milena Ramírez una hermosa mujer que ha traido a mi vida felicidad y mucho amor. Su apoyo en la cumbre y recta final de este proyecto ha sido fundamental para mí y lo seguirá siendo en lo que a continuación decidamos alcanzar.

Mis colaboradores, Henry Laniado, Carlo De Michele y Elena Di Bernardino, quienes junto a mi directora guiaron mi formación, me entregaron parte del conocimiento adquirido por su experiencia, me apoyaron y dieron sabios consejos, así como el seguimiento y la exigencia necesarias para motivar mi desarrollo. Además de ello, me acogieron con hospitalidad y calidez durante mis visitas a EAFIT, el Politecnico Di Milano y a Le Conservatoire National des Arts et

Métiers.

Mis amigos, de quienes he aprendido grandes cosas y lecciones de vida, me han ayudado a forjarme como persona y me han acompañado de diferentes maneras. Particularmente a la distancia agradezco a mi gran amigo Andrés Torres, a quien considero otro hermano y a su familia que siempre me ha acogido como miembro de ella. A Velvet, Francisco, Lino, Willy y su esposa, Diego, Joao, María y demás amigos y compañeros de lucha en la UC3M, con los cuales se vivieron tantas experiencias. Finalmente, al Departamento de Estadística por su apoyo financiero y a los integrantes de dicho departamento, profesores y administrativos, quienes estaban siempre prestos para apoyar y asesorar, generando un gran ambiente de trabajo.

A todos ellos muchas gracias por compartir conmigo este logro. Y cerrado este capítulo, la batalla apenas comienza; lo cuál me permite citar las palabras de mi abuela, Ana María Salas: “La vida es un camino que recorremos un paso trás otro, y así vamos, uno trás otro”. En tu memoria.

ACKNOWLEDGMENTS

Our achievements are always supported by multiple persons at different stages of the process. Therefore, I wish to express my gratitude to all the people who were key during the years of my PhD., whom helped to development my knowledge and skills and led me to the happy end of this work. Out of all these people, I want to highlight the special support granted by: My advisor Rosa Elvira Lillo, she was guiding my research, but at the same time was supporting my decisions, my proposals, always taking care of me in such friendly, kindly way and even maternal.

My parents, Adan Torres and Olga DÃaz, who have always supported my goals, and even in the distance have always encouraged me in the hardest moments to continue and to grow without conformism. I am also grateful to my siblings Camilo, Erika, and Felipe; my nephews, my grandparents and my aunt. They are key members of my family who have given so much of themselves to support this project. Also, thanks to a special member of my family, my girlfriend Milena Ramírez, a beautiful woman who has brought happiness and love to my life. She has supported the main stages of this project and has been fundamental to me. More importantly, she will continue being fundamental in my future endeavors.

My collaborators, Henry Laniado, Carlo De Michele and Elena Di Bernardino, whom along with my advisor guided my training, gave me part of their experiences and knowledge, supported me and gave me wise advice. They had also kept me honest by motivating and encouraging my development. In addition, they welcomed me with warmth and hospitality during my visits to EAFIT, Politecnico di Milano and Le Conservatoire national des arts et métiers.

My friends, from whom I have learned great things and life lessons helping me to become a better person. They accompanied me in different ways. Particularly in the distance, I thank my best friend Andres Torres, whom I consider a brother and to his family who have always treated me as a member of them. I also thank Velvet, Francisco, Lino, Willy and his wife, Diego, Joao, Maria and other friends and fight comrades in the UC3M, with whom I have lived so many experiences. Finally, thanks to the Statistics Department for their financial support and to its staff, professors and administrative personnel, who were always willing to support and advise, creating a great working environment.

To all of them thank you very much for sharing this achievement. And closed this chapter, the battle is just beginning; which allows me to quote the words of my grandmother, Ana Maria Salas: "Life is a path we walk one step after another, and so we go, one after another, one after another". In your memory.

RESUMEN

El objetivo de esta tesis es el de introducir aspectos direccionales a las metodologías multivariantes utilizadas para el análisis de extremos y problemas derivados. Se explica en el documento que la utilización de direcciones en determinadas situaciones posibilitan considerar información externa o preferencias particulares del analista. El elemento matemático clave en este proyecto es la definición de cuantil direccional multivariante. Las propiedades que satisface y otras nociones relacionadas con esta definición son las bases que fundamentan los desarrollos teóricos y sus aplicaciones al análisis de riesgo, las cuales constituyen las contribuciones de esta tesis. Después de una introducción de conceptos preliminares y motivaciones dadas en el Capítulo 1, los Capítulos 2 a 4 recogen las siguientes aportaciones:

En el Capítulo 2, se introduce una extensión direccional multivariante del *Value at Risk*, el cual en dimensión uno es un referente en campos como economía, seguros y finanzas, y se define como un cuantil a nivel para la distribución de la variable de pérdidas. Nuestra propuesta describe una medida de riesgo de resultado vectorial basada en los cuantiles direccionales multivariantes. Se estudian sus propiedades como una extensión de la axiomática definida para medidas de riesgo univariantes y también se presentan relaciones entre el valor de la medida de riesgo univariante *VaR*, aplicada sobre las marginales del vector de pérdidas, y los valores de las correspondientes componentes de la medida de riesgo propuesta.

En este Capítulo se fundamenta la importancia de las direcciones, gracias a la cota conservadora (cota superior) de pérdida total que permite establecer nuestra propuesta a través del análisis en la dirección del vector de pesos de la inversión.

Se analizan expresiones cerradas de solución para la medida de riesgo direccional multivariante en modelos de copula de alta aplicación en la teoría financiera y se presenta un método de estimación no-paramétrico para el resultado de dicha medida en ámbitos generales. Finalmente, se presenta un análisis de robustez sobre los resultados obtenidos para la medida propuesta ante presencia de atípicos en la muestra, obteniendo buen comportamiento especialmente en casos de alta presencia de atípicos, en comparación con la única medida de valor vectorial encontrada en la literatura a la fecha.

El Capítulo 3 se ha centrado en la definición de extremos direccionales y en la descripción de una metodología de detección no-paramétrica de los mismos. Se presentan casos de estudio reales en el ámbito de la ingeniería ambiental, dado que en los fenómenos ambientales se requiere del análisis conjunto de variables cuya dependencia conlleva a resultados catastróficos en muchas situaciones. Debido a la necesidad de modelar estas dependencias, una de las herramientas más utilizadas en la literatura son las cópulas. Por tanto, en este Capítulo se presentan las ventajas y desventajas de los métodos de copula y direccional no-paramétrico, y se plantea la inclusión del enfoque direccional para las metodologías basadas en cópulas. Se presenta una interesante alternativa de dirección a través de la dirección de máxima variabilidad en los datos, lo cuál genera la inclusión de análisis de componentes principales a la metodología propuesta. Finalmente se analizan los casos reales de riesgo de inundación en una presa (en 3 dimensiones) y de tormentas costeras extremas (en 5 dimensiones), así como casos simulados que complementan la importancia del análisis direccional.

Por otra parte, es bien conocido que las metodologías clásicas de estimación no paramétrica fallan cuando se desea realizar análisis para niveles altos del cuantil incluso en el caso univariante, es decir, para muy cercano a o , lo cual se conoce en la literatura como estimación *out-sample* y para abordarlo es necesario recurrir a resultados asintóticos de la teoría de valores extremos. Nuestra propuesta no se encuentra exenta de esta necesidad y en el Capítulo 4 se describen las hipótesis necesarias para introducir una metodología de estimación *out-sample* para los cuantiles multivariantes direccionales. Adicionalmente, se prueban resultados que incluyen el enfoque direccional en el marco de la teoría de valores extremos multivariante y se demuestra también la propiedad de normalidad asintótica para el estimador propuesto. Finalmente, se presenta el comportamiento del estimador a través de un ejemplo basado en la distribución multivariante, para la cual los resultados teóricos de los cuantiles direccionales son conocidos, así como los valores teóricos de los elementos necesarios para el proceso de estimación.

Finalmente, en el Capítulo 5 se presentan las conclusiones de la tesis y problemas abiertos para futuros trabajos de investigación.

ABSTRACT

The aim of this thesis is to introduce a directional multivariate approach to analyze extremes. The proposal point out the importance of two factors from the dimensional world we live in, the center of reference and the direction of observation. These factors are inherent to the multivariate setting and allow us to introduce manager preferences or external information available for the system of interest. The key definition in which is based this thesis is the notion of directional multivariate quantiles. It is introduced in Chapter 1 jointly with its properties which help to develop directional risk analysis. Besides, Chapter 1 describes the background and motivation for the directional multivariate approach. The rest of the chapters are devoted to the main contributions of the thesis.

Chapter 2 introduces a directional multivariate risk measure which is a multivariate extension of the well-known univariate risk measure *Value at Risk (VaR)*, which is defined as a quantile of the distribution of the random loss at level and it has become a benchmark in fields such as Economy, Insurance and Finance. Properties for the proposed multivariate risk measure are provided as extensions of the axiomatic for univariate risk measures given in the literature. We have also proved relationships between the univariate *VaR* evaluated on the marginal loss and the component associated to that marginal loss in our vector-valued proposal.

Chapter 2 also highlights the importance of using directions thanks to a result providing a conservative bound (upper bound) of the total risk in a portfolio investment by using the direction of the weights of investment to analyze such loss. In the literature, copula models are frequently used to model the loss, thus solutions of our risk measure for some of these models are shown and a non-parametric approach to estimate the output in more general cases is also provided. Finally, a study of robustness in comparison with other vector-valued risk measure found in the literature is developed.

Chapter 3 is focused on the formal definition and estimation of the directional multivariate extremes. Given that environmental science possesses different phenomena where join behavior of variables may cause disasters, two real cases of study are analyzed. In the literature, it is possible to find copula theory to model those dependencies, which leads us to introduce the directional approach to the copula framework. Thus, advantages and disadvantages between non-parametric approaches and theoretical copula approaches are highlighted in this chapter. Moreover, it is presented a proposal to choose a suitable direction of analysis by considering the direction of the maximum variability on the data, which links the use of *Principal Component Analysis (PCA)*. Applications are performed on the real cases of study of flood risk at a dam (3 dimensional case) and sea storms (5 dimensional case).

In extreme value theory, it is known that standard non-parametric methods can not be applied to estimate quantiles at high levels. Therefore, a different approach known as *out-sample* estimation must be considered. In this sense, Chapter 4 introduces the necessary background to face the multivariate extreme value theory. Then, results including the directional approach to the multivariate extreme value theory are given. An estimator of the directional multivariate quantiles is provided and its asymptotic normality is also proved. Finally, it is presented a non-parametric methodology to accomplish the goal of estimation, with an illustration using the multivariate distribution for which are known all the theoretical elements of the estimation process.

Finally, Chapter 5 summarizes the conclusions of the thesis, open questions and future works are also commented.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

One of the main challenges of human activity is risk assessment; consequently many different scientific disciplines are devoted to model and quantify risks. Each approach leads to different stages of analysis, diverse theories and different tools describing the influence of the variables that characterize the behavior of the phenomena that imply risks. Nowadays, it is well known that the analysis can be erroneous if the variables are studied independently because their interactions may cause under-estimation or over-estimation of the risks, which respectively may imply exposure, a wasting resources, or even deaths. For instance, the insurance regulation Solvency II and the financial regulation Basel III agree in the need for a multivariate analysis of aggregated risks and the consideration of the dependencies among all the factors. Specifically, they have modified their statements motivated by the following questions: Which factors of risk are more important for the main business lines of a company or a sector? Are the models relatively stable or on the contrary are they very sensitive to small changes in one or more of the analyzed components? and requirements such as the inclusion of stress-testing methodologies to assess the stability and/or sensitivity of any proposed risk measure (see Gonzales-Rivera (2003), Longin (2000)). On the other hand, environmental risk assessment also involves political discussions raising similar questions.

Accordingly, multivariate approaches are necessary to study the components of each system of interest, including marginal aspects and correlations among the variables, location, scales and many other overall aspects. In terms of quantification, risk measure theory in the univariate framework is a very studied field (see Artzner et al. (1999)) and more recently, some extensions to the multivariate setting have been developed (see Balbas el al. (2012), Hamel and Heyde (2010), Jouini et al. (2004)). Regardless of the desirable properties of those measures, the common factor in these approaches is the use of the univariate quantile notion or an extension to the multivariate field.

It is well-known that in the univariate case, quantiles divide data according to a condition on the tail probability of either distributions or survival functions. The median, the quartiles, the deciles and the percentiles are examples of univariate quantiles. These concepts are unique because the aforementioned division is linked to the total order in the straight line, but they lose uniqueness in the multivariate setting due to the lack of a total order in . Therefore the definition of multivariate quantiles becomes a challenge. In summary, it is necessary to solve some natural and technical questions regarding the multivariate setting:

- What kind of regions or sets of points present more risk in a multivariate domain?
- In dimension , how should the " possible dependence relationships" between the variables be managed and what are the impacts of these interactions?
- Are there relationships between dimension, risk regions and to be considered in order to obtain reliability in the quantification of the risk?

In an attempt to solve these questions, we rely on characteristics of the dimensional world we live in, such as center of reference and directions. These characteristics offer the capacity of analysis under different perspectives, i.e., we can use the information of a dataset to select feasible regions of risk by the selection of a point as a center of reference and a direction to look at the data from this center of reference, considering the dependence among the variables. This thesis is devoted to analyzing risks under a multivariate perspective, but introducing the a parameter of direction.

We have reviewed the literature in this field in order to make this thesis a self-contained document. Then, this chapter is organized as follows. Section 1.1 presents some classical and recent notions of multivariate orders, including important works such as Barnett (1976), Zuo and Serfling (2000), Cascos et al. (2011), and Laniado et al. (2012) since the connection between total order and quantile in the univariate field is well known. Afterwards a thorough review of contributions on multivariate quantiles is presented in Section 1.2. Due to the absence of a total order in the multivariate framework, different definitions have been formulated aiming to extend some specific property of the univariate quantile concept. These contributions include works such as Serfling (2002), Belzunce et al. (2007), Hallin et al. (2010), Kong and Mizera (2012), and Laniado et al. (2010) among others.

Section 1.3 presents the theory of *Copulas*, which is an important theory to model dependence in the multivariate framework (see Nelsen (2006)). This theory links the joint distribution of a set of random variables with their marginals, allowing the dependence among them to be described by the *copula* tool. Moreover, using the copula concept it is possible to determine quantile surfaces in the dimensional setting. Some works that explore this idea are Nappo and Spizzichino (2009), Durante and Salvadori (2010), Chebana and Ouarda (2011), Salvadori and De Michele (2004), Grimaldi and Serinaldi (2006), among others. This chapter concludes with Section 1.4 devoted to explaining the structure of the thesis.

1.1 Multivariate Orders

1.1.1 Barnett's Type Orders

There are different ways to define multivariate data ordering. A four-fold classification of possible ordering principles is proposed in Barnett (1976): marginal ordering, reduced ordering, partial ordering and conditional ordering. Each of these multivariate ordering has their own weakness. Below we give a brief introduction to each of them.

a) **Marginal Ordering**

The extension of some statistical concepts to the multivariate field using marginal order is based on the definition of the same statistical concepts but on the marginals, i.e., using the univariate theory on the components of the vector. The most significant example of this ordering is the well-known multivariate sample mean. Even multivariate quantiles may be constructed through marginal univariate quantiles. This methodology presents some drawbacks. For instance, it cannot be extended to higher moments since the correlations among the marginals must be considered.

The dependence between the marginals generates the main problem in the methodologies that use marginal approaches. For that reason, other sophisticated procedures belonging to this class define the ordering using some linear transformations in the marginals as a previous technique that allows the dependence to be modeled properly. For instance in Galambos (1972), the author uses the principal component analysis trying to capture the overall behavior of the data in order to improve the use of the marginal techniques.

b) **Reduced Ordering**

In this type of ordering, each observation is transformed into a single value using a particular linear combination or some generalized metric. Several transformations and generalized metrics have been explored with the purpose of ordering the overall multivariate sample. In many cases, the distances or metrics can be represented as quadratic forms as for some choices of and (see Sibuya (1960)). Examples for those distances are the Euclidean distance to point , where is the identity matrix, or the Mahalanobis distance, where is the sample mean vector and is the covariance matrix. Figure 1.1 shows an example of the reduced order using the Mahalanobis distance to sort a multivariate sample. It is clear that reducing each point to a one single value it is not enough to represent information of the dataset, which is a main disadvantage of this approach.

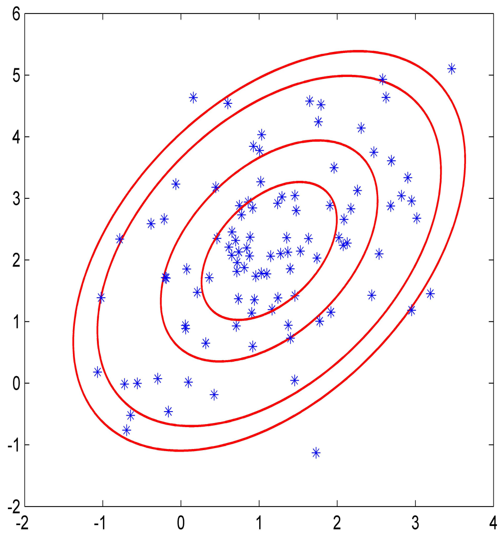


Figure 1.1: Example of Reduced Ordering

In Figure 1.1, it can be seen that the concentric ellipses define the order, i.e., points outside a fixed ellipse have higher order than those points inside the ellipse. Notice that in contrast with the marginal ordering, the processes in this category sort the data considering joint information of the multivariate sample rather than just its marginal behavior.

c) **Partial Ordering**

This category includes procedures that seek overall interrelational properties between observations in multivariate sample. These procedures usually split the multivariate sample into groups of observations and the splitting method allows an order to be defined among the groups but without distinction between observations into the same group.

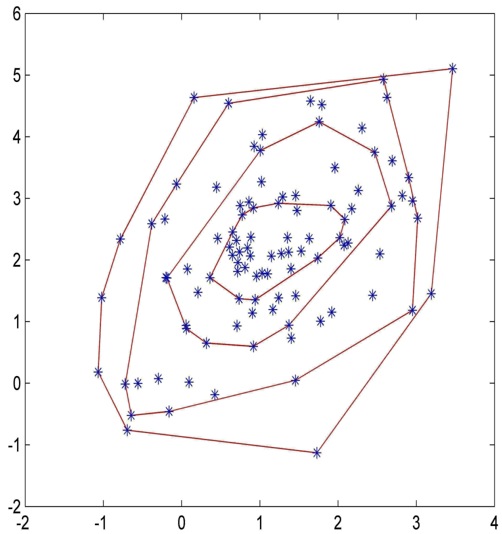


Figure 1.2: Example of Partial Ordering.

The convex hull provides an iterative method that allows ordered sets to be generated. Given the multivariate sample, the minimum convex set is built which includes all the sample points. Then those points on the perimeter are discarded and the iterative process goes on with the remaining points. Figure 1.2 shows some groups constructed using this method and it is clear that we have an order across the groups, but observations within a fixed group have no distinction between one another. The major problem of this methodology is its performance in high dimension.

d) **Conditional Ordering**

The final Barnett category for multivariate data is one in which the ordering or ranking is conducted on one of the marginal sets of observations conditional on ordering or ranking, within the data in terms of other marginal sets of observations. Examples appear in the work of Kreimerman (1975), or the use of concomitants in David (1973). The marginal samples used may be the original ones, or those derived from some preliminary coordinates transformation. The process is often repeated sequentially throughout all the marginal sets of observations.

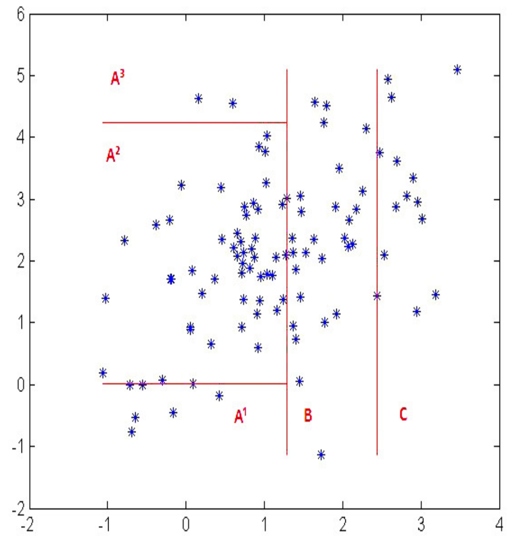


Figure 1.3: Example of Conditional Ordering.

An example of this principle is found in the notion of statistically equivalent blocks, where the conditioning is determined by real value functions called slicing or cutting functions. In a bivariate case for instance, the rectangular regions or blocks could be constructed by slicing with respect to a chosen ordered values of the first component, then internally slicing each slice with respect to the ordered values of the second component within that slice. In Figure 1.3, we present the results by partitioning the sample in three slices using ordered values of the first component and then, the first slice is cut using ordered values of the second component. Using this, we have that and . More information about this methodology can be found in Anderson (1966).

1.1.2 Ordering through Depth Functions

A special kind of multivariate ordering has attracted a good deal of attention since Barnett's paper. The idea is to sort the data from the most centered to the most outwards, Tukey (1975). These notions are better known as data depth functions. Based on this ordering, several statistical univariate techniques can be extended over the multivariate framework, such as extreme values and risk measurement, multivariate goodness-of-fit and scatter estimates. A review on depth functions and their applications can be found in Zuo and Serfling (2000), where four general structures for statistical depth functions are identified. Cascos et al. (2011) also present a review where some recent depth functions and applications are introduced. We now summarize important properties that any depth function must have.

Given a probability function in , a depth function can be defined as a bounded function that satisfies four properties.

- Affine Invariance: Depth should not depend on the underlying coordinate system, i.e., if is a no singular matrix and a vector , then , where denotes the probability distribution of the random vector .
- Vanishing at infinite: .
- Maximality at center: For a distribution having a unique center, the depth function should reach maximum value at this center, which means for any distribution function centered in .
- Monotonicity relative to the deepest point: If is the deepest point, then , for .

The sample version of each depth can be defined by replacing the probability function by its natural empirical measure . Three of the most relevant depth functions are the following.

a) **Halfspace depth function**

The halfspace depth function was presented by Tukey (1975) and measures the depth of a point as the smallest fraction of the data points in a closed halfspace containing . Both versions, population and sample are defined respectively as

close halfspace (1.1.1)

where is the sample space and denotes the scalar product.

b) **Simplicial depth function**

The simplicial depth was introduced by Liu (1990) and the depth of a point is defined as the probability that belongs to a simplice whose vertices are independent observations with probability measure . This geometric approximation uses the convex hull concept and is defined by,

(1.1.2)

where describes the convex hull and are independent observations equally distributed as .

The sample simplicial depth is given as a U-statistic, which measures the proportion of simplices that contains point in all the subsets with distinct observations as vertex,

$\mathbb{1}$ $_{1}$ $_{+1}$ (1.1.3)

$_{1}$ $_{+1}$

where $\mathbb{1}$ is the indicator function.

c) **Mahalanobis depth function**

Liu and Singh (1995) proposed this depth function as a transformation of the Mahalanobis distance, which maps the sample space into the interval. One property of this depth function is that the value is closer to one, if the point is closer to the expectation of the variable.

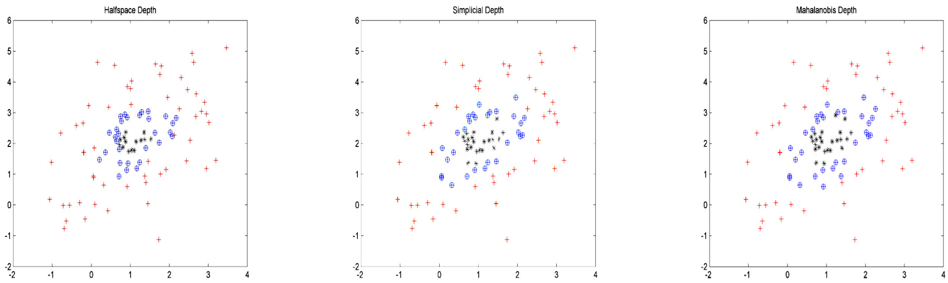
For and a distribution function , the Mahalanobis depth is defined as follows:

(1.1.4)

where and are the mean vector and the covariance matrix of the distribution probability function . The sample version is based on the sample mean vector and the sample covariance matrix,

and

The regions in this depth are always represented by ellipses centering in the mean. But the Mahalanobis depth does not work well if the distribution is not symmetric.



(A) Halfspace Depth (B) Simplicial Depth (C) Mahalanobis Depth

Figure 1.4: Examples of Ordering through Depth Functions

In Figure 1.4, we present the inner-outward order and we can say that those points in red are deeper or smaller than the blue ones and the black ones.

1.1.3 Directional Order

This category contains a recently introduced directional multivariate framework developed in Laniado et al. (2012). Previous to the order it is necessary the following definition,

An oriented orthant in with vertex in the direction is defined as,

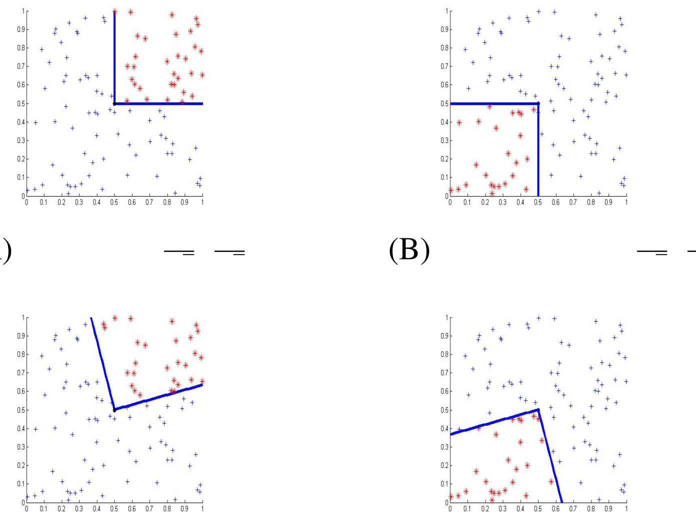
(1.1.5)

where , is an orthogonal matrix such that , with and the left hand side inequality is componentwise.

Observe that an oriented orthant is nothing else that both a translation and a rotation of the non-negative euclidean orthant toward a new vertex in the point and a rotation . Then (1.1.5) allows to define a partial data order in (denoted by) as,

if and only if, (1.1.6)

where . Note that (1.1.6) is a partial order since it satisfies reflexivity, transitivity and antisymmetry properties, but does not satisfy the trichotomy law, i.e., there are points with no comparison.



(A) (B) (C) (D)

Figure 1.5: Examples of Directional Order

Figure 1.5 shows an example of the directional order using four directions. In the plots, we have described the orthants at vertex varying the direction . Then, the points in red in each graphic hold that are lower than in the corresponding direction .

1.2 Multivariate quantiles

Once the data ordering in has been summarized, we present a review of multivariate quantiles since this concept is a usual tool to identify risk regions and extreme observations in statistics. In the univariate setting this concept can be expressed in different ways, but always attached to the total order in the real line. The usual definition for a random variable is given by

for all (1.2.1)

However all the definitions of multivariate quantiles do not necessarily attend to a definition of a multivariate order. Some of them are focused on the extension of some property of the univariate quantile concept. This fact generates the diversity of multivariate quantile notions. In Serfling (2002) has been collected some proposals. However an updated but brief review of the categories is now presented.

1.2.1 Multivariate quantiles based on inversions of mappings

It is well known that the univariate quantile can be defined as a pseudo-inverse of the associated distribution function because of the relation . Some researchers have been working to extend this concept to the multivariate framework defining a mapping from to that has an inverse, whose values may be interpreted as multivariate quantiles according to . For instance, Breckling and Chambers (1988) and Breckling et al. (2001) worked in this direction and state that given a random vector having an absolutely continuous distribution on , , a multivariate quantile can be defined as the inverse of the mapping

(1.2.2)

The function in (1.2.2) is called the *spatial rank function*, because it is a generalization of the univariate centered rank function, . Then a multivariate quantile can be obtained by solving the equation for . An example of the quantiles for a given sample varying the value of in the open unit ball centered in the origin, but fixing is presented in Figure 1.6 below.

When is determined by minimizing an objective function, the quantile can also be classified in the following category on norm minimization.

1.2.2 Multivariate quantiles based on norm minimization

Another equivalent characterization of the univariate -quantile is given by the minimization of the functional

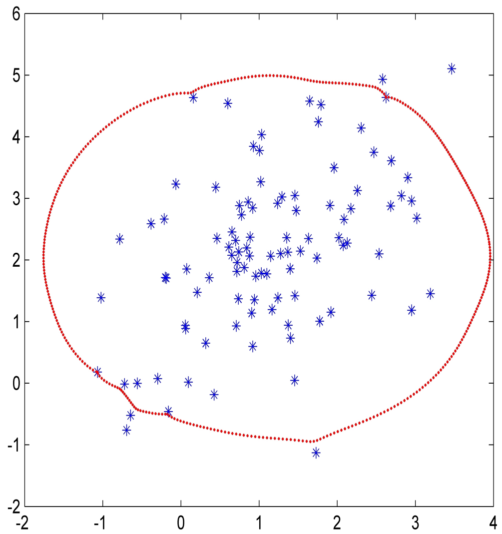


Figure 1.6: Example of Quantile based on inversion of mapping with .

for a random variable with . In this sense, there are also generalizations to the multivariate framework such as those developed by Abdous and Theodorescu (1992) and Chaudhuri (1996). In this kind of quantiles the objective is to minimize

where (1.2.3)

Each choice of the functional to be optimized leads to a different multivariate quantile function. For instance in Chaudhuri (1996),

(1.2.4)

where is the usual Euclidean inner product and . Then the -quantile is obtained minimizing over with the restriction . Moreover, the particular functional in (1.2.4) leads to the quantiles in Figure 1.6, since using (1.2.4) in (1.2.3) the function in (1.2.2) is obtained.

1.2.3 Multivariate quantiles based on gradients

In the univariate case, given a sample , its median is characterized minimizing or may be found by solving , where is the derivative of . Now to extend this equivalent univariate method on the multivariate field, many alternatives for can be considered. For instance, given the sample data in with ,

- with ,
- or
- $_1$ $_1$, with $_1$ $_{+1}$ the volume of the simplice in with vertex $_1$ $_{+1}$ and .

Obviously each choice of yields different notions of the multivariate sample median and generates its respective *gradient* that permits generalizations of the sign test statistics and also allows generalizations of the centered rank function and multivariate quantiles. For detailed information we refer to Brown and Hettmansperger (1987, 1989) or Hettmansperger et al. (1992).

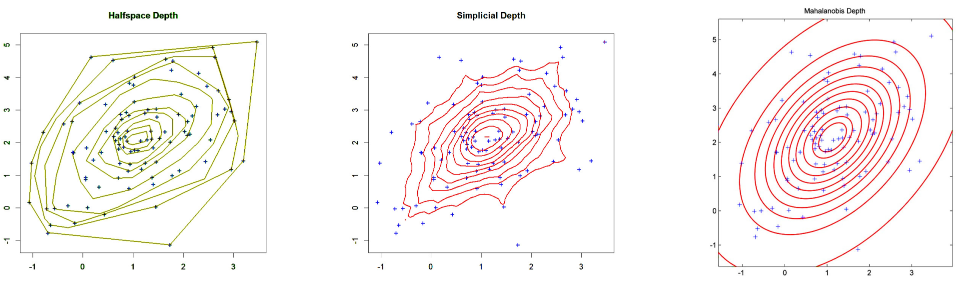
1.2.4 Multivariate quantiles based on depth functions

Statistical depth functions have an important role in multivariate analysis and provide an orderings of points in . As was pointed out in Section 1.1, there are several depth functions that allow the use of their center-outward order to construct central regions and therefore to define a multivariate quantile. To do this, the important notion is given by the central point, which is defined as the deepest point, i.e., the point that has a maximal depth function value.

Using the center-outward order, the corresponding -quantile, ( ), can be generated in a parallel way to the cumulative distribution definition in (1.2.1). by considering the points with an associated depth function greater than or equal to . In this case, it is necessary to normalize the depth measure to the interval . It is even possible to describe central regions obtaining the contours of the depth function, i.e., the central region is obtained by the points that have depth function value lower than .

Clearly, different depth functions yield different versions of quantiles. An example of the changes due to the selected depth function is presented in Figure

1.7, where the plots correspond to (a) Halfspace depth, (b) Simplicial depth, (c) Mahalanobis depth.



(A) Halfspace Depth (B) Simplicial Depth (C) Mahalanobis Depth

Figure 1.7: Example of Quantile based on depth functions with

1.2.5 Multivariate quantiles based on projections

This extension has been developed using some properties of the *inner product* in a Hilbert space. Let be a Hilbert space with inner product and induced norm . Given a random element in with distribution and such that , for , the -quantile in the direction in is defined by

(1.2.5)

where the right side quantile function in (1.2.5) is the classic univariate quantile function.

This directional approach is introduced to also be applied in infinite dimensional vector space in connection with the increasing demand of statistical tools for functional data analysis. For more information we refer to Fraiman and Pateiro-López (2012). Using the same sample as in previous examples, we presented in Figure 1.8 the multivariate quantiles obtained varying and taking .

1.2.6 Multivariate quantiles based on directions

Some multivariate quantiles described previously are linked with directions, but their characterizations remain in other properties, reason why they are include in

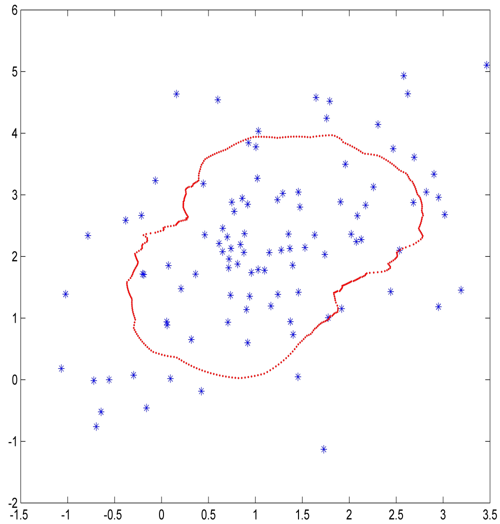


Figure 1.8: Example of Quantile based on projections with .

other categories. An example is Chaudhuri's approach which has been classified in *quantiles based on minimization*. Now, we classify here methodologies such as those developed by Tibiletti (1993), Hallin et al. (2010) and Kong and Mizera (2012), which clearly use a directional approach. For instance, a difference between the multivariate quantiles based on projections given in the previous category and the approach founded in Hallin et al. (2010) is due to the way the projections are used. They define directional quantiles associated with unit vectors in a dimensional setting, where indicates the *vertical* reference direction for the quantile regression obtained by applying the methodology proposed by Koenker and Basset (1978) on the model

(1.2.6)

where , , is a matrix whose columns are the orthonormal base of the orthogonal space of , i.e., constitutes an orthonormal basis of and are the variables to optimize in the quantile regression method. The quantiles obtained with this methodology are hyperplanes with tangent vector . Thus, this method offers a hyperplane with tangent vector as a multivariate -quantile. And by rotating the reference vector , we can construct quantile contours as the intersection of this hyperplanes.

Kong and Mizera (2012) define a multivariate directional -quantile as the vector of univariate -quantiles of the corresponding projection of over the hyperplane defined by the direction . But notice an important aspect that the authors question in their preprint (Kong and Mizera , 2008, pg. 11), and is due to behaviors in their biplot quantiles, "*Overall, quantile biplot contours appear rather counterintuitive, and their tendency to self-intersections and "mozarella" shapes probably will not win them too many friends. It seems that the question is not how to plot directional quantiles, but how to successfully incorporate this information into the plot of the data*". Such kind of comments generates active research on this topic, which can be seen on recent literature using copula theory (e.g. Grimaldi and Serinaldi (2006), Salvadori and De Michele (2004)).

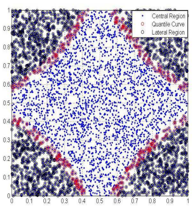
A final notion that we introduce in this category is the key element of this thesis, the **directional multivariate quantile** defined by Laniado et al. (2010). This notion is linked with the directional order described in (1.1.6) and is defined as,

(1.2.7)

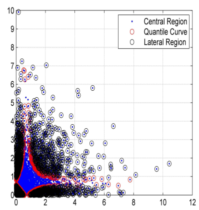
where lies in the unit ball , and was defined in (1.1.5). This definition generalizes the quantiles given in Fernández-Ponce and Suárez-Llorens (2002) and Belzunce et al. (2007) where only the following set of classical directions are considered:

(1.2.8)

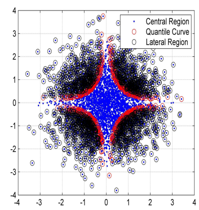
Figure 1.9 presents examples of these multivariate quantiles for in three simulated samples of bivariate distributions: uniform, exponential and normal. Top plots describe the results when the set of classical directions is selected. Meanwhile, bottom plots present the results for the set of canonical directions, i.e, . The red points define the quantile curves, the blue ones the inner region and the black ones the lateral regions in each case. Note how the direction has influence in the quantile curves.



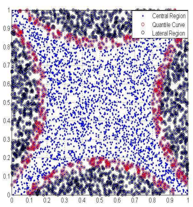
(A) Bivariate Uniform, classic directions



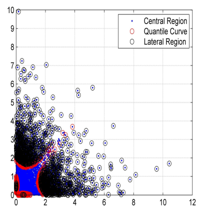
(B) Bivariate Exponential, classic directions



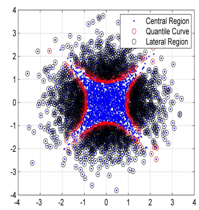
(C) Bivariate Normal, classic directions



(D) Bivariate Uniform, canonical directions



(E) Bivariate Exponential, canonical directions



(F) Bivariate Normal, canonical directions

Figure 1.9: Examples of directional multivariate quantiles

1.3 Copulas and Multivariate Quantiles

Researchers refer to *copulas* as "the multivariate distribution functions in whose one-dimensional marginal distributions are uniform in ". But their scope goes beyond that; this powerful tool allows scale-free measures of dependence to be defined and families of multivariate distributions to be constructed.

Two aspects are important in a multivariate distribution:

- The marginal distributions
- The dependence structure among them.

And the concept of copula gives a relation between these two characteristics. Copulas fully describe the overall structure of dependence between the variables of interest and provide a global model for their stochastic behavior. In order to build multivariate quantiles, the copulas become a powerful tool since it is possible to obtain closed expressions of the quantiles for specials copula families. This type of procedure to define quantiles based on copulas is implemented in works such as Nappo and Spizzichino (2009), Durante and Salvadori (2010), Chebana and Ouarda (2011) and Cousin and Di Bernardino (2013). We will give more details in subsection 1.3.2.

For an extensive presentation on copulas, we refer to Nelsen (2006). In the following section, we summarize some important definitions and their relationship with quantiles.

1.3.1 Definition and basic concepts of Copulas

The term and concept of copula was introduced by Sklar (1959) with the meaning of the word *copula* in some sense stemming from the relation with the property of joint or *couple* marginal distribution functions. This theory has a growing and very active research thanks to its capacity to model the dependence among the variables participating in a system.

A copula or dimensional copula is a multivariate distribution function in , where the margin variables are uniforms on the interval . From this concept, it is possible to develop parametric techniques to build a large number of multivariate distributions and their characteristics. In particular, quantile surfaces are one of those characteristics that could be determined through these techniques.

Now, we introduce *Sklar's theorem* which is the central theorem in this theory. This tool is the foundation for many of the applications for copulas in Statistics.

Theorem 1.3.1 (Sklar's theorem). *Let be a dimensional distribution function with marginals . Then there exists a copula such that for all ,*

(1.3.1)

If are all continuous, then is unique; otherwise, is uniquely determined on . Conversely, if is a copula and are distributions functions, then the function defined by (1.3.1) is a dimensional distribution function with marginals .

Moreover, if have quasi-inverses it is possible rewrite (1.3.1) as,

for all (1.3.2)

Thus, we have that the random vector has the joint distribution function and then it holds that

X (1.3.3)

where means equality in distribution.

There are different methods for building copulas which are provided in families having particular structures and properties that are useful in modeling and simulation (see Corbella and Stretch (2013), De Michele et al. (2007), Salvadori et al. (2011)). In the following, we summarize some of these families and present the theoretical quantile definition using this approach. Finally is presented a particular example of quantiles through copulas in a bivariate case.

I. Copula Families

Here we present some representatives and important families of bivariate copulas. The selection of these families is due to their applications in fields such as Finance, Hydrology, Economics and Medicine.

I.1. **Archimedean Family**

This kind of copulas has the property of being able to capture many structures of dependence. This family includes many types of parametric copulas with an important number of applications in real problems.

Archimedean Copulas: Let be a continuous, convex and strictly decreasing function with . Let be a pseudo-inverse function of . Then an Archimedean copula is generated by

(1.3.4)

The function is called the *generator* and this family of copulas has the following properties:

- is symmetric, i.e., for all .
- is associative, i.e., for all .
- is a generator of , for any real positive constant .
- has convex level curves.

And based on the parametric generators, this family has several subclasses are used in the literature:

a) **Gumbel-Hougaard Copula**: This Archimedean family is generated by

and (1.3.5)

Then the expression for is given by

(1.3.6)

where .

b) **Frank Copula**: The generated function of this copula is

——— and –

(1.3.7)

– ——————— (1.3.8)

where .

c) **Clayton Copula**: This family is generated by

– and (1.3.9)

(1.3.10)

where .

I.2. **Gumbel-Morgenstern Family**

This family is given by

(1.3.11)

Applications of this kind of copulas can be seen, for instance, in Bargès et al. (2009).

I.3. **Extreme Value Family**

This family is given by copulas holding the following property,

(1.3.12)

for all and .

I.4. **Elliptical Family**

This family is given by copulas generated by an elliptical distribution. For instance,

a) **Gaussian Copula:** The Gaussian copula is given by the expression,

(1.3.13)

where is a multivariate standard Gaussian distribution with correlation matrix , to are the pseudo-inverses of Gaussian univariate distributions with parameters , , ..., , respectively.

b) **t-Student Copula**: The t-student copula is given by,

(1.3.14)

where is a multivariate standard t-stundent distribution with degrees of freedom and correlation matrix , is the pseudo-inverse of the univariate t-stundent distribution with degrees of freedom.

1.3.2 Quantiles based on Copulas

As we remark in Section 1.2, multivariate quantiles have been studied extensively, extending each of the properties of the univariate quantile. But the extensions are far from being general and including all the complexity in a multidimensional framework. A theoretical way to obtain generalizations is through copulas. By using (1.3.3) and (1.3.1), the multivariate quantile based on the copula concept is defined as,

$$\mathbf{x} \qquad (1.3.15)$$

Then, there are cases where is possible to obtain closed forms to multivariate quantiles, but those cases are a small number of families combined with particular marginal distributions, which generates many restrictions. Besides, the complexity in the computation increases with the dimension. In fact, most of the references have presented their applications only in dimension two (e.g. Durante and Salvadori (2010), Nappo and Spizzichino (2009), Salvadori (2004)).

For instance, an example in the bivariate framework is a model with Gumbel-Hougaard copula and standard Gumbel marginals. Then by taking in (1.3.6), it can be deduced from (1.3.15) and (1.3.3) that,

then

and

Finally the bivariate quantile at level is described by,

$$\mathbf{x} \quad {}^{1} \quad {}^{2} \qquad (1.3.16)$$

In Figure 1.10 we can see the quantile curves given by (1.3.16) in the cases in red, in blue and in black.

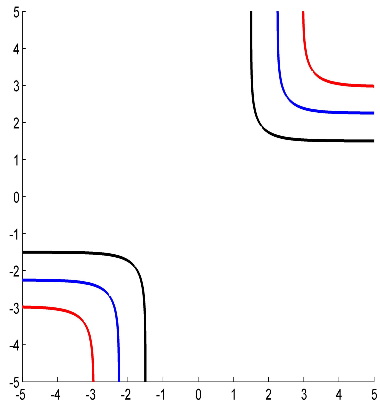


Figure 1.10: Quantile curve based on Copulas.

1.4 Structure of the Thesis

This thesis contains five chapters. In the present chapter we have reviewed multivariate orders and quantiles, as well as copulas, which is another multivariate theory related to quantile notions. We focus on the study and improvement of the *directional notion* introduced in (1.1.6) and (1.2.7). The contributions of this dissertation are developed in Chapters 2, 3, 4.

In Chapter 2, we introduce the **directional multivariate Value at Risk** as a vector-valued extension of the financial univariate risk measure. This extension is based on improvements of (1.2.7). The directional approach can be summarized as the inclusion of a free-parameter to give versatility to the multivariate exploratory analysis that allows a manager to consider external information or risk preferences in her/his analysis. For instance, we motivate the direction of the investment portfolio weights, due to the fact that the risk for the investor is highly dependent on the way he/she is investing money. We derive some properties for the directional quantile along with the properties of the directional multivariate *VaR* and we compare the univariate *VaR* over the marginals with the components of our vector-valued risk measure. We also analyze the relationship between some families of copulas, for which it is possible to obtain closed forms of the directional multivariate *VaR*. Finally, comparisons with other alternative multivariate VaR given by Cousin and Di Bernardino (2013) are provided in terms of robustness.

Chapter 3 is devoted to presenting a methodology to identify **directional multi-**

variate extremes, which has been applied to **environmental phenomena**. Several environmental phenomena can be described by different correlated variables that must be considered jointly in order to be more representative of the nature of these phenomena. For such events, identification of extremes is inappropriate if it is based on marginal analysis. Extremes have usually been linked to the notion of quantile, which is an important tool for analyzing risk in the univariate setting. We propose to identify multivariate extremes and analyze environmental phenomena in terms of the directional multivariate quantile, which allows us to analyze the data considering all the variables implied in the phenomena, as well as looking at the data in interesting directions that can best describe an environmental catastrophe. Since there are many references in the literature that propose extremes detection based on copula models (e.g. De Michele and Salvadori (2003), De Michele et al. (2007), Grimaldi and Serinaldi (2006), Salvadori and De Michele (2004)), we also generalize the copula method by introducing the directional approach. Advantages and disadvantages of the non-parametric proposal that we introduce and the copula methods are provided in the paper. We show with simulated and real data sets how by considering the first principal component direction we can improve the visualization of extremes. Finally, two case studies are analyzed: a case of flood risk at a dam (a variable case), and a case study of sea storms (a variable case).

Along with the set up of the directional method, one important question comes up, *how should the directional multivariate quantiles be estimated?* Two cases arise, estimation *in-sample* and estimation *out-sample*. In Chapter 2 and Chapter 3 the applications are based on the *in-sample* approach through a non-parametric method that we propose. However, in extreme value analysis, research is focused on the quantification of the multivariate risk outside of the observable sampling zone; that is, a region of interest located at high levels. Chapter 4 introduces an *out-sample* method to estimate *directional multivariate quantiles*, with all the necessary tools and hypothesis to formalize the estimation. The asymptotic normality of the proposed estimator is derived. Finally, the methodology is illustrated with simulated examples for which the theoretical directional multivariate quantiles are known.

The final chapter of this dissertation presents some conclusions and some possible future research lines.

CHAPTER 2

A DIRECTIONAL MULTIVARIATE VALUE AT RISK

Value at risk (*VaR*) has become a benchmark for risk management, and it is defined as the threshold quantity that does not exceed a certain probability level which is considered to be dangerous. It is commonly implemented by investment banks to measure the market risk of their asset portfolios. Although (*VaR*) has been broadly criticized from the work of Artzner et al. (1999) since it does not verify the diversification property, it has also been defended by Heyde et al. (2009) for its robustness. For univariate risks, the *VaR* is simply the quantile of the loss distribution function. Thus, the *VaR* is a risk measure easily interpretable, and it still remains the most popular measure used by risk managers. However, there is not a unique definition of *VaR* in the multivariate context because there are different possible definitions of multidimensional quantiles which are related to a specific partial order considered in , or to a property of the univariate quantile that is desirable to be extended to . Therefore, each definition of quantile could provide a potential definition of multivariate *VaR*. For instance, the proposals given by Koltchinskii (1997) of multivariate quantiles as inversions of mappings, multivariate quantiles in terms of norm minimization as in Chaudhuri (1996), multivariate quantiles as level-sets given by Fernández-Ponce and Suárez-Llorens (2002), multivariate quantiles based on depth functions developed in Serfling (2002), and finally, multivariate quantiles based on projections as in Fraiman and Pateiro-López (2012), Hallin et al. (2010), Kong and Mizera (2012).

Currently business and financial activities generate data for which it has been shown that it is insufficient to consider single real-value measures over marginal aspects, in order to quantify risks jointly associated to the data. For instance, one of the drawbacks detected in the global banking regulatory *Basel II* is the solvency and liabilities dependence among the financial institution branches, or even the domino effect in the markets that could be generated by dependence among filial products. Thus, the solvability of each individual branch may strongly be affected, not only by its activities, but also by the level of dependence among all the branches. In consequence, it is necessary to quantify the risk, considering both the multivariate nature of the data and the dependence among the marginal risks.

In *Basel III*, a new liquidity regulation was proposed in order to avoid the weakness detected in the 2007-2009 crisis; but these regulations have to be complemented by internal models in the institutions, in order to obtain better hedge results. These models have to include multivariate risk measures computable in high dimensions and also, to consider possible internal and external risks, even if the nature of those risks is strongly heterogeneous.

In Insurance, there is also interest in analyzing joint risks considering claims from different types of policies offered by the company, e.g. life, fire or health insurances, among others. Thus, allocated loss adjustment expenses play an important role in determining the expenses that are due to the processing of a specific insurance claim and they are part of the insurer expense reserves. It is one of the largest expenses that an insurer has to set aside funds for, along with contingent commissions. Insurers set aside reserves for these expenses so that they can ensure that claims are not being fraudulently made, and to process legitimate claims quickly. Since the joint behavior of the different types of policies have to be taken into consideration to determine the reserves for the insurance company, multivariate risk measures are necessary (e.g. see Frees and Valdez (1998)).

In recent decades, literature devoted to extend the *VaR* measure to the multivariate setting has been published. For instance, bivariate versions have been studied in Arbia (2002), Tibiletti (2001), Nappo and Spizzichino (2009). Also, for multivariate distributions in general, some notions of *VaR* have been introduced (e.g. Cousin and Di Bernardino (2013), Embrechts and Puccetti (2006), Lee and Prékopa (2012)). Embrechts and Puccetti (2006) linked the risk measure to the level surface defined when the distribution function of risk , or the survival function, accumulate some -value, which is considered as a quantile surface. Recently, Cousin and Di Bernardino (2013) introduced a new notion of multivariate *VaR* based on those level surfaces studied in Embrechts and Puccetti (2006). They commented that considering the whole surface as a risk measure could result in interpretation problems. Therefore, they defined the multivariate

VaR as the mean of the points belonging to the surface considered in Embrechts and Puccetti (2006) and hence, the focus should be a point with the same dimension as the random vector of losses. Specifically, they define the *upper–orthant Value–at–Risk* (*lower–orthant Value–at–Risk*) at –level (–level) as the conditional expectation of , given that belongs to the -set of its distribution (survival) function.

In this chapter, we introduce a *directional multivariate Value at Risk*, based on the extremality level sets introduced in Laniado et al. (2012), which permit the concept of directional multivariate quantile to be defined. The extremality level sets are surfaces defined by following the same idea as in Embrechts and Puccetti (2006) but linked to rotations of the multivariate distribution; that is, we consider a directional approach. We share with Cousin and Di Bernardino (2013) the idea that a multivariate *VaR* seen as a surface could bring problems with its interpretation. Hence, we introduce the idea of considering the multivariate *VaR* as a vector-valued point that defines the vertex of an oriented orthant in the direction of analysis accumulating a probability . The vertex is obtained using the mean of to establish a reference system.

The risk measure that we propose, considers the high dimension nature of the real problems, and the dependence among the risks is implied in the analysis. Finally, we admit the possibility of various manager preferences, introducing a parameter of direction . For instance, directions like the maximum variability given for the principal components in the portfolio, or the assets weight composition could be more interesting to analyze than the classic directions assumed in the information summarized in the cumulative or survival distribution functions. Besides, the directional approach allows us to give bounds for the *VaR* related to linear combination of random variables, mainly when they are statistically dependent.

We prove properties of the directional *VaR* that we consider as relevant for a multivariate risk measure, such as consistency with respect to a particular stochastic order, tail subadditivity in the mean loss direction, as well as some invariance properties. We compare the components of the directional multivariate *VaR* with the univariate *VaR* on the marginals, in order to show that the vector given by the *VaR* on the marginals provides incomplete information on the joint risk. Some of these properties can be viewed as a extension to the multivariate field of the axiomatic given by Artzner et al. (1999). Some of the properties that we prove, are implicitly related with the axiomatics introduced recently in the literature (see Balbas el al. (2012), Hamel and Heyde (2010), Jouini et al. (2004)) for coherent mulivariate measures.

We also obtain closed expressions of the *VaR* when bivariate copulas are considered or when a multivariate Archimedean's copulas govern the dependence among the components of the portfolio. Finally, we present comparisons in terms of robustness with the alternative vector-valued multivariate *VaR*, introduced by Cousin and Di Bernardino (2013).

The chapter is structured as follows. In Section 2, we introduce some preliminary concepts and notation necessary in order to understand the main contributions. In Section 3, the *directional multivariate Value at Risk* (is introduced and we provide analytic properties for this risk measure. Section 4 provides the comparisons between the univariate *VaR* over the marginals and the components of the directional multivariate *VaR*. Section 5 is devoted to theoretical results and closed forms of the multivariate *VaR* when particular families of copulas are considered. In Section 6, we develop the robustness analysis. Finally, some conclusions are summarized as well as some possible extensions are suggested for future work.

2.1 Preliminaries

The main objective of this chapter is to introduce a directional multivariate Value at Risk, based on the notion of directional multivariate quantile given in Laniado et al. (2010). We devote this section to introduce and review main concepts necessary to properly define the risk measure.

Definition 2.1.1. *An oriented orthant in with vertex in the direction is defined as,*

(2.1.1)

where and is an orthogonal matrix such that , with .

Note that given , is not unique for and thus, Definition 2.1.1 generates a family of oriented orthants. In order to simplify the definition of the risk measure introduced in this chapter, we impose conditions on the possible to guarantee uniqueness in the transformation. From now on, let be a unit vector with non-null components and let and be matrices defined as,

(2.1.2)

where , is the th component of , is the scalar sign function and is the vector with all its components equal to zero except the th component equal to one. Note that the hypothesis of , guarantees that always is a matrix of rank . Now, we consider the QR decomposition of and (see e.g. Horn and Johnson (2013), Ch. 2),

such that and are triangular matrices with positive diagonal elements, and and are orthogonal matrices. Note that these decompositions are unique due to both the full rank of and and the structure of the decomposition (see e.g. Horn and Johnson (2013), Theorem 2.1.14, p.g. 89). Also, the first columns on and are the same as in and ; that is, and respectively. Therefore, and and thus, , which motivates the following definition.

Definition 2.1.2. *The QR oriented orthant with vertex in direction is the oriented orthant as in Definition 2.1.1 but using . It is denoted by*

Based on Definition 2.1.2, a partial data order in (denoted by) can be defined by,

if and only if, (2.1.3)

where and is as in (2.1.2). Equivalently,

if and only if,

where and the order on the right side is component-wise.

Throughout the document we will use the following notation related to subsets in . Given , , and , the sets and are defined as,

(2.1.4)

We recall some results on *QR oriented orthants* that will be useful in the main sections of this dissertation. The proofs are given in the Appendix.

Lemma 2.1.3. *Given a direction and a vertex , then*

(2.1.5)

Lemma 2.1.4. *Given and , then*

(2.1.6)

We also recall some definitions of useful stochastic orders; see Shaked and Shanthikumar (2007), for more details.

Definition 2.1.5. *Given two random vectors and , is said to be smaller than in:*

(i) usual stochastic order (denoted by) if , for any increasing function with finite expectations.

(ii) upper orthant order (denoted by) if , for all , where , denote the survival functions of and , respectively.

(iii) lower orthant order (denoted by) if , for all , where , denote the cumulative distribution functions of and , respectively.

It is easy to verify that both orders, the upper orthant and the lower orthant, are implied by the usual stochastic order. The following stochastic order defined in Laniado et al. (2012) is a key tool in providing some properties of the multivariate VaR that we introduce in the next Section.

Definition 2.1.6. *Let and be two random vectors with associated probability distribution , is said smaller than in the extremality order in the direction (denoted by) if,*

for all in .

It is easy to show that . Moreover, if then , as it is proven in [Laniado et al. (2012), *Property 3.4*]. Since the multivariate VaR is based on the definition of a quantile, we also need to introduce the directional multivariate quantile given in Laniado et al. (2010).

Definition 2.1.7. *Let be a random vector with associated probability distribution . Then the directional multivariate quantile at level , in direction is defined as*

(2.1.7)

where denoted the boundary of the subset considered into brackets and .

From now on, we focus on an absolutely-continuous random vector (with respect to the Lebesgue measure on) with increasing marginal distribution functions and such that , for . These conditions are called *regularity conditions*.

We also recall the two versions of the vector-valued *VaR* introduced in Cousin and Di Bernardino (2013). They are the benchmarks of the risk measure introduced in this chapter as it is shown in Section 5 and Section 6.

- The *lower multivariate VaR* at level is defined as,

(2.1.8)

- The *upper multivariate VaR* at level is defined as,

(2.1.9)

Note that (2.1.8) and (2.1.9) are the expected value of the hyper–surfaces defined as *Upper-Orthant VaR* and *Lower-Orthant VaR* in Embrechts and Puccetti (2006).

2.2 Directional Multivariate Value at Risk

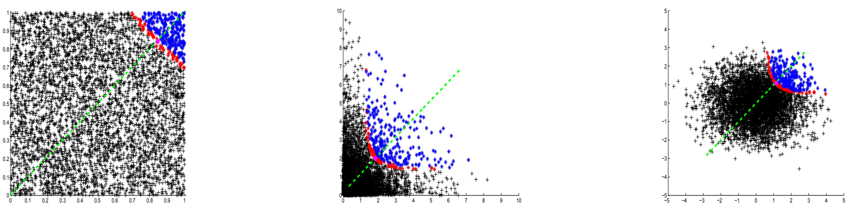
In the univariate setting, the relationship between the quantiles related to the loss distribution and the *VaR* is obvious. In this Section, we propose a definition of multivariate *VaR* for a portfolio of -dependent risks, linked with the directional multivariate quantile defined in (2.1.7). As well, the result is a point in ; that is, a vector of the same dimension as the considered portfolio of risks. Specifically, as in the univariate case, this point defines the vertex of an oriented orthant that accumulates a probability , but in the direction that the investor or the risk manager considers more convenient.

Definition 2.2.1. *Let be a random vector satisfying the regularity conditions and . Then the directional multivariate Value at Risk of in direction at probability level is given by*

(2.2.1)

We must highlight that given a direction , the is the intersection between the directional quantile at level , and the line defined by both the direction and the mean of ; that is, is a point in . Note that the *regularity conditions* on ensure that the intersection in (2.2.1) is non void. We want to point out that the central tool is chosen to be the mean as a reference point for the random vector space, i.e., for the support of the associated probability distribution. As we demonstrate, the choice of the mean in Definition 2.2.1 allows us to derive desirable and interpretable analytic properties related to the risk measure. However, other central reference points can be possible; for example the median seen as the deepest point associated with some multivariate depth measure, which may provide a more robust risk measure (e.g. Cascos et al. (2011), Zuo and Serfling (2000)).

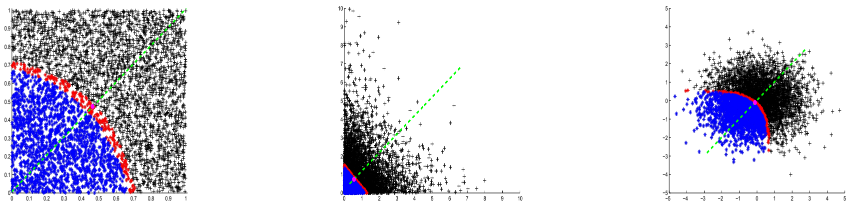
Remark 2.2.2. *Definition 2.2.1 assumes that is a vector with non-null components in order to the associated QR oriented orthant be properly defined (see Definition 2.1.2). However, this is not a restrictive condition in multivariate risk analysis since a null component in is equivalent to ignore/depreciate the information related to that specific component. Therefore, the advisable is to reduce the dimension of the problem avoiding the null componets before the evaluation of the directional risk measure.*



(A) Bivariate Uniform (B) Bivariate Exponential (C) Bivariate Normal

Figure 2.1:

Figure 2.1 displays some examples of the risk measure defined in (2.2.1), for three different bivariate distributions in the direction with . This direction makes reference to the distribution function of . Figure 2.2 presents examples with the same bivariate distributions, but in the direction and for ; that is, regarding the information contained in the survival function of . We call these two directions classical directions, but the aim of this work is to show that it could be interesting to consider other directions in the analysis of risk.



(A) Bivariate Uniform (B) Bivariate Exponential (C) Bivariate Normal

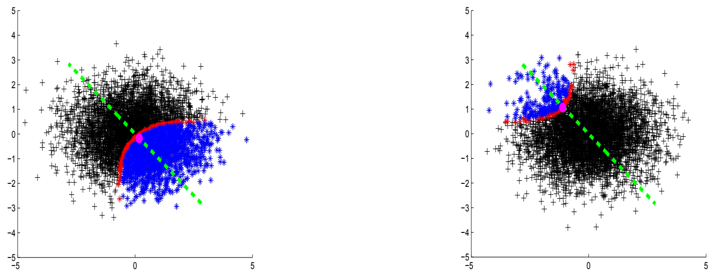
Figure 2.2:

Observe that in the figures, the line in direction crossing the mean in green is displayed while the quantile curve is displayed in red. The *VaR* that we propose is just the intersection between the line and the quantile curve. On the other hand, the points in blue are the points "below" the level of risk in the corresponding direction; meanwhile the black points are those "exceeding" the level risk. Observe Figure 2.1, if you take any point on the blue region as a vertex of an oriented orthant in direction , then the probability of that orthant will be greater than . It will be equal to or smaller than if the point is taken from the red curve or black region, respectively. From Figure 2.2 in direction , the same conclusion can be drawn. Figure 2.3 displays the risk measure for a bivariate normal distribution, but considering alternative directions. Specifically, we consider the directions corresponding to the second and fourth orthants in , which are the complementary orthants of those used by the distribution and survival functions. These orthants result interesting when it is necessary to analyze the relationships between random variables of the type or , or when the bivariate distribution in consideration has negative dependence.

It is desirable that the classical univariate *VaR* agrees with our definition of *VaR* in the case ; this fact is shown in the following. Recall that the univariate *VaR* is defined as,

(2.2.2)

where is usually considered closed to 1. Moreover, the *VaR* may also be



(A) (B)

Figure 2.3:

defined in terms of the distribution function as,

(2.2.3)

As in the univariate setting under continuity, then (2.2.2) and (2.2.3) are the same. To be consistent with the univariate *VaR*, our definition of multivariate *VaR* agrees with the classical definition for . That is, we have, in terms of , that:

where is related to definition (2.2.2) and is related to definition (2.2.3). However, this fact does not hold in the multivariate context where is not true in general, being

(2.2.4)

(2.2.5)

The remainder of this section is devoted to providing some properties of which are similar to those properties considered in the risk literature; (see Artzner et al. (1999), Burgert and Ruschendorf (2006), Cardin and Pagani (2010), Rachev et al. (2008)). Specifically, we provide properties of the multivariate in terms of Artzner et al. (1999)'s properties related to coherent risk measures in the univariate setting. In a similar way, Cascos and Molchanov (2007, 2013), Hamel and Heyde (2010), Jouini et al. (2004) propose some properties to coherent versions of multivariate risk measures defined as set-value measures.

Balbas el al. (2012) also include properties referred to vector-value measures, but we have explored other properties inherent to the vector-value output in our proposal, such as the invariance under orthogonal transformations.

Property 2.2.3 (Non-Negative Loading)**.** *For small,*

(2.2.6)

This property reflects that the risk measure is an upper-bound of the mean value of the losses, with respect to the partial order given in 2.1.3.

Property 2.2.4 (Quasi-Odd Measure)**.** *satisfies the property:*

(2.2.7)

This property shows *symmetry* with respect to the analysis of risk for positive random losses, or the analysis of negative random returns.

Property 2.2.5 (Positive Homogeneity and Translation Invariance)**.** *Let , and , then,*

(2.2.8)

Property 2.2.6 (Consistency w.r.t. extremality stochastic order)**.** *Let and be random vectors satisfying the regularity conditions. If with , and , then:*

(2.2.9)

Now, we introduce a type of orthogonal transformations before the following property.

Definition 2.2.7. *A QR rotation of a unit vector over another unit vector is characterized by the matrix , where the matrices , correspond to the orthogonal parts in the QR decompositions of the matrices and , defined in (2.1.2).*

Note that a QR rotation of over implies that .

Property 2.2.8 (Orthogonal Quasi-Invariance)**.** *Let and be two unit vectors. If is the QR rotation of over . Then,*

(2.2.10)

Property 2.2.9 (Non-Excessive Loading)**.** *Let be the orthogonal matrix described in (2.1.2). Then,*

(2.2.11)

This property shows that is upper bounded by the supremum of the losses in the direction considered. Another desirable property in the literature for risk measures is the subadditivity. As it is well-known, the classical univariate *VaR* is not a subadditivity measure. However, there are conditions that ensure the tail region subadditivity property (see Artzner et al. (1999), Daníelsson et al. (2013), Heyde et al. (2009)). In the same way, we stress that the is not subadditive in general, but we prove that this property holds under some conditions. First another definition is necessary.

Definition 2.2.10. *A random vector is multivariate regularity varying with tail index if there is a real-value function that is* [1] *regularly varying at infinity with exponent − and a non-zero measure on the Borel field such that,*

(2.2.12)

where means vague convergence and (see e.g. Jessen and Mikosh (2006), Resnick (1987)).

In this case, the measure has the property

(2.2.13)

for all and every Borel set . In [Mikosch (2003), pg. 25], it is possible to see the proof of the property. Illustrative examples of Definition 2.2.10 can be found in [Resnick (2007), pg. 192].

As it is noted in Daníelsson et al. (2013), the previous definition allows to introduce a notion of a fat-tailed multivariate distribution that induces the tail region subadditivity property of the .

Property 2.2.11 (Tail Region Subadditivity)**.** *Let and be random vectors, with the same mean . If is a regularly varying random vector with index*

[1] If a function $\phi(\,)$ holds $\lim_{x\to\infty}\frac{\phi(tx)}{\phi(x)} = t^{\frac{1}{\beta}}$, for all $t > 0$ is called *regularly varying at infinity with exponent* $\frac{1}{\beta}$.

and non-degenerate tails then, the is subadditive in the tail region in direction ——, i.e.,

(2.2.14)

The proof is provided in the Appendix following a similar approach as in Daníelsson et al. (2013). Note that Property 2.2.11 extends to the multivariate case the Proposition 1 given in Daníelsson et al. (2013) for the univariate case. As you can see, the property ensures that at least in the direction of the mean loss, it is useful to merge two risky activities in order to diversify the risk. Property 2.2.11 could be extended to random vectors with means satisfying for .

2.3 Comparison of the univariate *VaR* componentwise and the Directional Multivariate *VaR*

The aim of this section is to compare the components of with the univariate *VaR* related to each marginal distribution of . But prior to this we need to recall the definition of a multivariate quasi-concave function.

Definition 2.3.1. *A multivariate function is a quasi-concave function if the upper-level set is a convex set for all . Or equivalently, the complementary of the lower set is a convex set for all .*

We point out that both the distribution and survival functions, satisfy Definition 2.3.1 under regularity conditions. Specifically, this result is proven by Tibiletti (1995) but for elliptical distributions and Archimedean copula families.

Let us denote by the -th marginal of the random vector and by the -th component related to a point in . The following result provides comparisons between the components of the multivariate *VaR* introduced in this work and the classical univariate *VaR* on the marginals.

Proposition 2.3.2. *Consider a random vector satisfying the regularity conditions. Assume that its survival function is quasi-concave. Then, for all :*

for all

If its multivariate distribution function is also quasi-concave, then, for all , we have that

for all

The proof is given in the Appendix. As you can see, the preceding result can be extended in other directions as follows.

Corollary 2.3.3. *Let be a random variable satisfying the regularity conditions and let be a specified direction. If the survival function of is a quasi-concave function, then, for all ,*

for all

Besides, if has a quasi-concavity cumulative distribution, then

for all

with as in Definition 2.1.2.

The proof is straightforward from Proposition 2.2.8 and Proposition 2.3.2. Therefore, by linking the previous results we have the following inequality for all pairs , .

(2.3.1)

This relationship allows us to define a *directional upper VaR* and a *directional lower VaR* in a similar way to Embrechts and Puccetti (2006) and Cousin and Di Bernardino (2013), but with a unified notation that takes into consideration the directional parameter. Specifically, we introduce the following by redenoting our measure in the pairs and :

The *upper VaR in direction* is defined as,

(2.3.2)

And the *lower VaR in a direction* is defined as,

(2.3.3)

An example of these concepts is displayed in Figure 2.4, where we can see in a bivariate normal distribution, the *upper VaR in direction* for a level of risk , and the corresponding *lower VaR in direction* and level risk . Note that we describe on the figure types of asymptotes for the quantile curves, that represent the univariate quantiles for each marginal of the rotated random vector at the same , where the rotation matrix is the same as in (2.1.2). These asymptotes can be seen as a generalization of those defined in Belzunce et al. (2007) for the quantile curves in the classical directions.

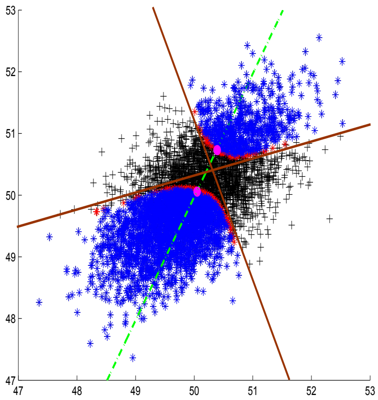


Figure 2.4: Lower and upper with and for a bivariate Normal.

There is another practical application where the link between the multivariate *VaR* and the univariate *VaR* is interesting (see e.g. Bernard et al. (2014), Embrechts and Puccetti (2006), Wang et al. (2013)). It is when is necessary to give upper-bounds of the univariate *VaR* over a linear transformation of the marginal losses. For instance, when the risk over the transformation given by the portfolio weights vector is considered, i.e., when the objective random variable is the return function given by

where is the portfolio weights vector chosen by the investor. Since it is difficult to obtain the *VaR* of mainly when the components of the portfolio can not be assumed independent, there is special interest in obtaining at least a bound for . Fortunately, we can give an upper-bound using the directional approach.

Proposition 2.3.4. *Let* — *be the unit vector in direction of the portfolio weights. If* , *then* .

The proof is given on the Appendix. As a consequence of Proposition 2.3.4 we can consider the bound given by,

$$\| \; \| \tag{2.3.4}$$

which is another justification to consider a directional approach of the multivariate *VaR*, as well as its utility in financial applications.

2.4 Directional multivariate *VaR* and copulas

Now, we recall the concepts about copulas given in Chapter 1. Then, thanks to Sklar's theorem (1.3.1), two aspects are important in multivariate distributions, the distribution of the marginals and the dependence structure among them. Therefore, the copula fully describes the overall structure of dependence between the marginal variables and provides a global model for their stochastic behavior.

Hence, the objective of this section is to analyze how the can be obtained in terms of some families of copulas. The first result shows the representation of the restricted to bivariate copulas. Let be a bivariate random vector with marginals uniformly distributed in the interval . In this case, the distribution function of is a copula with density . It is well known that – – . Note that assuming , a direction can be characterized by an angle such that , and then, . Following with the notation given by the angles, the must be a point on the line defined by,

$$\begin{array}{ll} - \quad - & \text{if} \\ - & \text{if} \end{array} \tag{2.4.1}$$

Therefore, given a direction , is characterized by its first component and the second one is obtained using (2.4.1). Now, the first component can be obtained by solving the following equation on the domain of the integral,

$$ \tag{2.4.2} $$

where is given by the intersection of the unit square and the oriented quadrant with direction determined by and vertex . Specifically, can be expressed in terms of the unknown by using the semi-lines , that bound the corresponding quadrant which are defined as,

$$ \{ \; _1 \; _2 \; _2 \; (\; -) \; _1 \; (\; -) \; _1(\; (\; -) \; (\; -)) \; - \; (\; -) \} $$

$$ \{ \; _1 \; _2 \; _2 \; (\; -) \; _1 \; (\; -) \; _1(\; (\; -) \; (\; -)) \; - \; (\; -) \} $$

For instance, if − − , we can write the equation as follows:

$$ \tag{2.4.3} $$

Figure 2.5 shows a case of the region with − − being the solution to (2.4.3), a point over the line . In summary, we can obtain for a given bivariate vector with copula density .

Now, we focus on the Archimedean family of copulas defined in 1.3.1 paragraph I.1., which has been widely used in the literature.

In this case, for a -dimensional random variable with distribution function belonging to the Archimedean family of copulas with generator , is given by the vector with all components equal to

$$ \tag{2.4.4} $$

Moreover, if has a survival copula belonging to the Archimedean family with generator , the equivalent Sklar's representation gives the relation , where is the join survival function and its marginal survival functions. Hence, we obtain that:

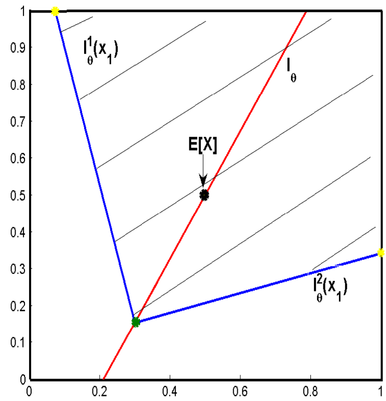


Figure 2.5: Quadrant given by – – and vertex over the line .

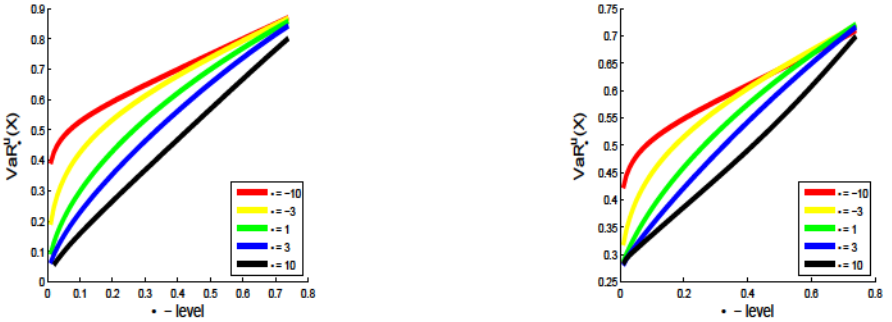
(2.4.5)

Recall that if a vector has a copula , then the survival copula of will also be . Therefore, if d , then the copula of and its survival copula are the same; for example, Frank's copula (1.3.8) in the Archimedean family holds this property, as well as the elliptical family of copulas. Then, in this case the closed expression for is the reflection point of with respect to the point – – .

Now we present some examples using some Archimedean copulas. Firstly, we use Frank's subclass (1.3.8) to present an example of for any direction in the bivariate case. Later we present some comparisons between our proposal and the notions reviewed in (2.1.8) and (2.1.9) but considering a -dimensional copula belonging to Clayton's subclass (1.3.10).

In Figure 2.6 we have drawn the first component of the directional for a bivariate random vector with density given by the Frank copula (1.3.8). The left plot is related to and the right plot is related to . Both plots present the behavior for but considering different values of the parameter in the copula density.

In the left plot where , note that if , we obtain the cases known as comonotonic and counter-monotonic, respectively. Also, it can be seen that



a) Direction $-\mathbf{e} = -\frac{\sqrt{2}}{2}[1,1]'$ b) Direction $\mathbf{u} = -\frac{3\sqrt{5}}{5}[\frac{1}{3},\frac{2}{3}]'$

Figure 2.6: Behavior for the first component in $VaR^{\mathbf{u}}_{\alpha}(\mathbf{X})$ varying α.

in the comonotonic case, the component reaches the value given by the *VaR* on the marginals, which is α in this case. In addition, it is well known that rotations over random vectors do not preserve the dependence structure in the rotated distribution. This fact is captured in the right plot.

Let $\mathbf{X}$ be a random vector having as distribution function a Clayton copula (1.3.10). Hence, the survival function of $\mathbf{1}-\mathbf{X}$ is a Clayton survival copula. Now, we compare the first components of $VaR^{-\mathbf{e}}_{\alpha}(\mathbf{X})$ and $\underline{VaR}_{\alpha}(\mathbf{X})$ in (2.1.8), and the same for the first components of $VaR^{\mathbf{e}}_{1-\alpha}(\mathbf{1}-\mathbf{X})$ and $\overline{VaR}_{\alpha}(\mathbf{1}-\mathbf{X})$ in (2.1.9). Table 2.1 contains the explicit expressions of $\underline{VaR}_{\alpha}(\mathbf{X})$ and $\overline{VaR}_{\alpha}(\mathbf{1}-\mathbf{X})$ in dimension 2, and the generalized expressions for our proposal in terms of α and β in any dimension. Figure 2.7 shows the graphical comparison for $n = 2$; the left plot presents the results for $VaR^{-\mathbf{e}}_{\alpha}(\mathbf{X})$ in solid line and $\underline{VaR}_{\alpha}(\mathbf{X})$ in dashed line, while the right plot presents the results for $VaR^{\mathbf{e}}_{1-\alpha}(\mathbf{1}-\mathbf{X})$ in solid line and $\overline{VaR}_{\alpha}(\mathbf{1}-\mathbf{X})$ in dashed line.

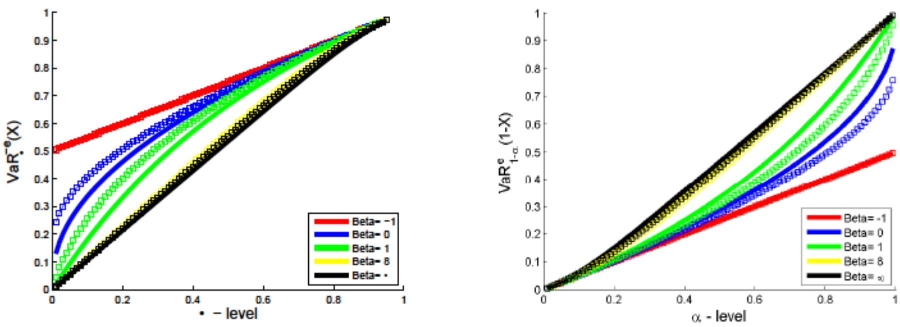
	Directional $VaR^{\mathbf{e}}_{\alpha}(\cdot)$	Cousin and Di Bernardino (2013)'s *VaR*
$\mathbf{X}$	$\left(\frac{1+\alpha^{-\beta}}{n}\right)^{-\frac{1}{\beta}}$	$\frac{\beta}{\beta-1}\frac{\alpha^{\beta}-\alpha}{\alpha^{\beta}-1}$
$\mathbf{1}-\mathbf{X}$	$1-\left(\frac{1+(1-\alpha)^{-\beta}}{n}\right)^{-\frac{1}{\beta}}$	$1-\frac{\beta}{\beta-1}\frac{(1-\alpha)^{\beta}-(1-\alpha)}{(1-\alpha)^{\beta}-1}$

Table 2.1: Clayton's Copula Case

The results in Figure 2.7 also shows us that in the case of random vectors with Clayton copula class (1.3.10), $VaR^{-\mathbf{e}}_{\alpha}(\mathbf{X})$ increases with respect to the parameter α and decreases in the parameter β. On the other side, $VaR^{\mathbf{e}}_{1-\alpha}(\mathbf{1}-\mathbf{X})$ is an increasing function of the parameter α, but also an increasing function of the dependence parameter β. These features for this class of copulas were commented and proved by Cousin and Di Bernardino (2013) and for our risk measure can be easily proved following the same scheme. In addition, we need to highlight that for each fixed pair (α, β), the following relationships hold,

$$\underline{VaR}_{\alpha}(\mathbf{X}) \leq VaR_{\alpha}^{-\mathbf{e}}(\mathbf{X}) \quad \text{and} \quad VaR_{1-\alpha}^{\mathbf{e}}(\mathbf{1}-\mathbf{X}) \leq \overline{VaR}_{\alpha}(\mathbf{1}-\mathbf{X}), \tag{2.4.6}$$

where the inequalities are componentwise. Hence, we can say that our measurement is more conservative in the upper case and it is more optimistic in the lower case. This can be taken into consideration by the manager according to her/his preferences.



a) Lower Case b) Upper Case

Figure 2.7: Comparison for Clayton's family of copulas.

2.5 Robustness

The previous section presents the analytic results for random vectors with $[0,1]$-uniform marginals distributions. However, in practical situations, it is necessary to obtain $VaR_{\alpha}^{\mathbf{u}}(\mathbf{X})$ for any random vector $\mathbf{X}$. In this case, we use a computational approach summarized in the following routine:

Input: $\mathbf{u}$, α, h and the multivariate sample $\mathbf{X}_m$.
for $i = 1$ to m
  $P_i = \mathbb{P}_{\mathbf{X}_m}\left[\mathfrak{C}_{\mathbf{x}_i}^{\mathbf{u}}\right]$,
  If $|P_i - \alpha| \leq h$
    $\mathbf{x}_i \in \hat{\mathcal{Q}}_{\mathbf{X}_m}^{h}(\alpha, \mathbf{u})$,

end
for
,
end
end

where is the sample of the random vector , the sample mean, the sample quantile hypersurface with a slack and is the empirical probability distribution of . Using this procedure, we are able to deal with high dimensional random vectors. We are aware that this procedure can be improved using more sophisticated tools of the non-parametric statistics, but they are outside the scope of this thesis.

On the other hand, it is well known that in risk theory, it is desirable that a measure be robust, (see Artzner et al. (1999), Burgert and Ruschendorf (2006), Cardin and Pagani (2010), Rachev et al. (2008)). But in general, most of the measures are sensitive to extreme outlying observations. In this section, we present a simulation study to show the sensitivity of our *upper VaR in direction* , using as a benchmark the *upper VaR* in (2.1.9) introduced in Cousin and Di Bernardino (2013). Setting , we compare and in terms of robustness using the following contamination model:

d with probability
with probability (2.5.1)

where d , d and . The parameters of are,

remains fixed in the analysis, but the parameters of the normal distribution of are changed in various ways to generate outliers. As a measure to quantify the effect of the outliers, we define the percentage variation, where $Measure(\mathbf{X}^0)$ is the risk measure evaluated in the sample without contamination, $\omega = 0$, and $Measure(\mathbf{X}^\omega)$ is a risk measure evaluated for the sample with a level of contamination $\omega\%$. We have considered the scenarios for $\mathbf{X}_2$, described in Table 2.2.

Scenarios	Parameters of $\mathbf{X}_2$ distribution
Variance Analysis	$\boldsymbol{\mu}_1, \quad \Sigma_1 + \begin{bmatrix} 4.5 & 0 \\ 0 & 6.5 \end{bmatrix}$
Covariance Matrix Analysis	$\boldsymbol{\mu}_1, \quad \Sigma_1 + \begin{bmatrix} 4.5 & 0.2 \\ 0.3 & 6.5 \end{bmatrix}$
Mean Analysis	$\boldsymbol{\mu}_1 + \Delta_\mu, \quad \Sigma_1$
Join Analysis	$\boldsymbol{\mu}_1 + \Delta_\mu, \Sigma_1 + \begin{bmatrix} 4.5 & 0.2 \\ 0.3 & 6.5 \end{bmatrix}$

Table 2.2: Simulation Stages and Parameters

The procedure is the following: firstly, we have generated a non-contaminated sample $\mathbf{X}^0$ with 5000 observations and we calculate both $VaR^{\mathbf{e}}_{0.1}(\mathbf{X})$ and $\overline{VaR}_{0.1}(\mathbf{X})$.

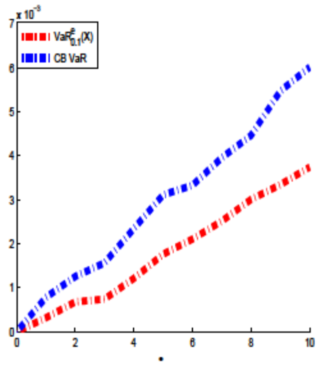


Figure 2.8: Percentage variation of the measures varying the variances

Secondly, we have used the contamination model (2.5.1) taking values for ω from 1% to 10%. Then, we have generated for each ω, 5000 observations of $\mathbf{X}_1$ containing an expected value of outliers $\omega\%$. We have evaluated the risk measure as well as the percentage variation for each level of contamination, performing this procedure 100 times and we have reported the average of PV^ω in the following plots.

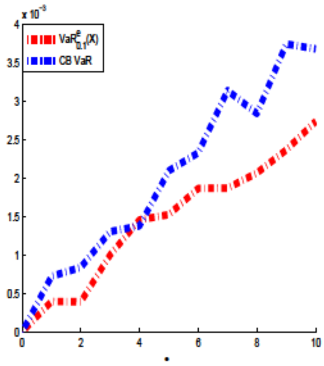
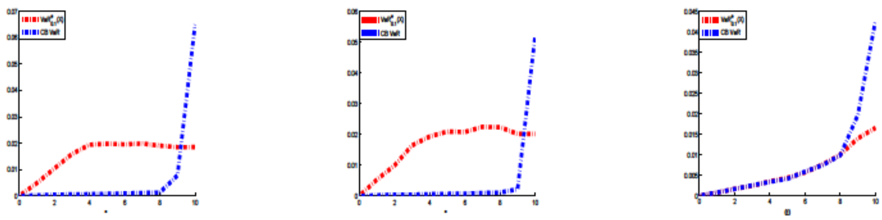


Figure 2.9: Percentage variation of the measures varying the covariance matrix

The first scenario suggests outliers result from changes on the variance of the marginals, which are difficult to detect in practice. We can see in Figure 2.8 that the behavior of $VaR^{\mathbf{e}}_{0.1}(\mathbf{X})$ is better than that corresponding to *upper-VaR* in Cousin and Di Bernardino (2013) for any level of contamination. "better", in this context, means that PV^{ω} is smaller.

The second scenario considers changes in all the components of the covariance matrix. The results are depicted in Figure 2.9, which shows again the better behavior of $VaR^{\mathbf{e}}_{0.1}(\mathbf{X})$ with respect to robustness.



a) $\Delta_{\mu} = [0, 25]'$ b) $\Delta_{\mu} = [25, 0]'$ c) $\Delta_{\mu} = [25, 25]'$

Figure 2.10: Percentage variation of the measures varying the mean

The last scenario consists of changes in the mean. Firstly, we vary the first component of the mean and then we affected the second one and finally both of them simultaneously. Figure 2.10 summarizes the results. As we can see, $VaR^{\mathbf{e}}_{0.1}(\mathbf{X})$ shows robustness in presence of a high percentage of outliers, but an extra-sensitivity under outliers in a unique component (outliers of shape type). The use of the mean of the random loss as the central point in the definition of our VaR could be the cause of this sensitivity.

To evaluate the impact of the dimension in the robustness analysis, we have carried out simulations with normal random vectors in high dimensions obtaining similar conclusions to the previous one. For instance, Figure 2.11 displays the results of the percentage variation when we consider , and the covariance matrix is modified while the mean remains fixed.

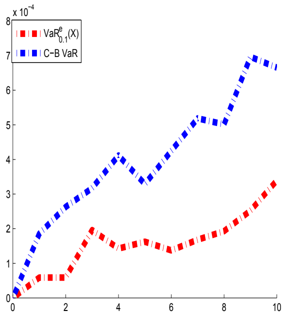


Figure 2.11: Percentage variation for a 3D-contamination model

Obviously, the robustness study can be extended by varying other aspects such as type of distributions, or changes in the "*level of risk*" given by the parameter .

2.6 Conclusions

In this chapter we have defined a multivariate extension of the classical risk measure *VaR* based on a directional multivariate quantile recently introduced in the literature. Specifically, we have proposed the *directional multivariate Value at Risk* ( ) as a tool to analyze a portfolio of heterogeneous and dependent risks considering external information or manager preferences.

We have given analytic properties of in the same way as Artzner et al. (1999)'s axiomatic. We have provided some invariance properties as well as consistency and tail subadditivity properties, which are desirable in a risk measure. We have demonstrated relations between the components of the output of with respect to the corresponding univariate *VaR* over the marginals. A interesting link between the univariate *VaR* over the linear transformation using the portfolio weights vector , and the value of this transformation over $\| \|$ is provided. We have also presented closed forms for in terms of some families of copulas, considering particular dimensions or particular directions.

Finally we have presented a simulation study of robustness comparing the behavior of with respect to the risk measure proposed in Cousin and Di Bernardino (2013). The simulations show the advantages of our proposal in relation to the presence of outliers.

We have also detected in this study an open question to be taken into consideration in future work. The idea is to consider another central point instead of the mean as the center of the reference system, in order to improve the robustness of the risk measure, but, at the same time, preserving the desirable properties demonstrated. A possibility that could bring us more robustness is to use a multivariate depth measure.

Appendix

Proof of Lemma 2.1.3. Without loss of generality, we may assume that .
From (2.1.2), we have that:

and

Then, under the constraint of positive diagonal elements in the corresponding triangular matrices in the QR decompositions, we have that:

Thus , which implies,

□

Proof of Lemma 2.1.4. The proof is straightforward using the definitions given in (2.1.4). □

Proof of Property 2.2.3. The proof is straightforward using the hypothesis of small, because this implies that in Definition 2.2.1 and hence the result. □

Proof of Property 2.2.4. Due to Lemma 2.1.3, it is easy to prove that

(2.6.1)

and hence,

Then,

□

Proof of Property 2.2.5. This property is derived using Lemma 2.1.4. □

Proof of Property 2.2.6. Since , we get:

$$L \quad (\alpha, \mathbf{u}) := \quad \mathbf{z} \quad : \quad (\quad) \quad \alpha \quad \mathbf{z} \quad : \quad (\quad) \quad \alpha \quad := L \quad (\alpha, \mathbf{u})$$

Besides, and . Therefore, using the partial order defined in (2.1.3) there are three possibilities for :

(i) ,

(ii) and ,

(iii) .

We can prove that the two first options are not possible for the points and . Suppose to the contrary that

which implies that,

Hence,

Which is a contradiction, if we assume the *regularity conditions*. Moreover, the hypothesis , for all and the result derived in [Laniado et al. (2012) (Property 3.4.)], permit us to reject the second possibility of ordering between the two points. Thus, the only option possible is,

□

Proof of Property 2.2.8. First, note that:

In addition, given the QR rotation of over , we have . Then,

Therefore, , and . Thus, we get

(2.6.2)

which proves the result. □

Proof of Property 2.2.9. Property 2.2.8 implies that,

where — . Then,

and the proof is complete. □

Proof of Property 2.2.11. It is easy to see that the equality in the mean implies that the vectors , and lie on the same line, the line with direction vector . Then, we can write:

(2.6.3)

where is the vector whose components are the value − and denotes the first component of the vector. Following a similar approach as in the proof of tail subadditivity of Daníelsson et al. (2013) for the univariate case, we develop a multivariate version. Then, for small, − , and then,

On the other hand, the Borel set − satisfies the following property:

Or equivalently,

Hence using (2.6.3), we have:

(2.6.4)

In the same way,

(2.6.5)

Now, in the case of the random variable , we have;

(2.6.6)

where the inequalities in the expression hold componentwise. As a consequence we get,

–

Then using the last equality in (2.6.6), we finally get,

1 –

(2.6.7)

It is well known that in all the norms are equivalent, i.e., for two norms and , there are positive constants such that . Then, whatever norm is taken, we use the transformation [Resnick (1987), pg. 267.], and rewrite in terms of a new measure in as , due to the property of the measure in (2.2.13). The relationship satisfying both measures for a Borel set in , it is given by,

(2.6.8)

Then the measure of the Borel sets in (2.6.4), (2.6.5) and (2.6.7) can be expressed using as:

$$- \qquad (2.6.9)$$

$$- \qquad (2.6.10)$$

$$(2.6.11)$$

Now using the Minkowski inequality we obtain:

$$(u \; + v \;) \quad \eta(d\mathbf{u}, d\mathbf{v}) \quad \frac{1}{}$$

$$u \quad \eta(d\mathbf{u}, d\mathbf{v}) \quad \frac{1}{} \quad + \qquad v \quad \eta(d\mathbf{u}, d\mathbf{v}) \quad \frac{1}{} \quad . \qquad (2.6.12)$$

Hence combining (2.6.4), (2.6.5), (2.6.7) and (2.6.12), we have the result

or equivalently, from Proposition 2.2.8 and the partial order defined in (2.1.3), we have for —— that:

$$(2.6.13)$$

□

Proof of Proposition 2.3.4. By Definition 2.1.7, if , we have . Therefore,

where (2.6.14)

Since , we obtain,

–

–

Thus, (2.6.14) and (2.2.2) imply .

□

Proof of Proposition 2.3.2. The proof follows the same outline as that of [Cousin and Di Bernardino (2013), Proposition 2.4.]. Note that in direction ,

Then we can write,

The convexity of follows from the quasi-concavity of the survival function , where denotes the complement of a set. Now, as , belongs to the set . Moreover, from the definition of survival function we have that,

for all and . Then, each component of a vector belonging to is upper bounded by the univariate *VaR* at level of the corresponding marginal. As a consequence, each component of is upper bounded by the univariate *VaR* at level of the corresponding marginal and hence, the first inequality holds. Now for the second inequality,

Then, we have,

But, if is a quasi-concave function, we have that is a convex set and . Therefore belongs to the set . Additionally, from the definition of distribution function, it is easy to show that each component of an element in is lower bounded by the univariate *VaR* at level of the corresponding marginal; hence, we obtain the result to be proved. □

CHAPTER 3

DIRECTIONAL MULTIVARIATE EXTREMES IN ENVIRONMENTAL PHENOMENA

3.1 Introduction

Serious economic and social consequences are generally associated with extreme environmental events such as floods, storms and droughts (Chebana and Ouarda (2011)), which are usually defined in terms of several correlated variables. For instance, rainfall is characterized by storm intensity and duration (e.g. De Michele and Salvadori (2003), Salvadori and De Michele (2004)); air quality is described in terms of levels of ozone and nitrogen dioxide (e.g. Chebana and Ouarda (2011), Heffernan and Tawn (2004); floods are modeled by their peak, volume and duration (e.g. Chebana and Ouarda (2011), De Michele et al. (2005), Grimaldi and Serinaldi (2006), Shiau (2003)); droughts are modeled by volume, duration and magnitude (e.g. De Michele et al. (2013), Kim et al. (2003), Salvadori and De Michele (2015)) and sea storms are represented by wave height, peak, direction and duration (e.g. De Michele et al. (2007)). Consequently, extremes detection cannot be made on the basis of a univariate analysis.

There are references in the literature that tackle multivariate extreme detection.

Some studies use copulas since the work by De Michele and Salvadori (2003), for example Salvadori (2004), Salvadori et al. (2011, 2013, 2016), whom also define multivariate versions of the return period[1], and Grimaldi and Serinaldi (2013). Another alternative is given by Chebana and Ouarda (2011) through depth functions. However, both alternatives have drawbacks when they have to be implemented in high dimensional scenarios. Copulas due to their intrinsic parametric nature are difficult to estimate in large dimensions, and depth functions are problematic due to the lack of computational implementation in most of the cases. Therefore, the first contribution of the chapter is to introduce a method to detect extremes based on a non-parametric procedure suitable for high dimensional analysis.

On the other hand, extremes have been traditionally analyzed in one dimension by considering only the probabilities of exceeding quantiles related to either the distribution function or the survival function. In other words, observations are considered extreme if they are associated to lower values or upper values of the variable, which is equivalent to looking at the data in one of the two possible directions . Some extensions of quantiles to the bivariate case have been proposed in Embrechts and Puccetti (2006), Fernández-Ponce and Suárez-Llorens (2002), Salvadori (2004), Shiau (2003) and to the dimensional setting in Cousin and Di Bernardino (2013), Di Bernardino et al. (2015), Fraiman and Pateiro-López (2012), Gupta and Manohar (2005), Salvadori et al. (2011). The generalizations of the quantile notion in all the previous references consider, as in the univariate case, the directions associated with the distribution function or the survival function.

But why not look at the data with different perspectives and take advantage of the inherent complexity of the dimensional setting the data lives in? There exist infinite directions to look at the data from a reference point that could help with the accuracy of the analysis and the interpretation of the results. Attempts have been made considering alternative directions, for instance in Chapter 2 and in Laniado et al. (2012) where have been developed financial applications to assess the risk of losses considering the direction of the investment weight composition in a portfolio; Cascos and Molchanov (2007), Hallin et al. (2010) and Kong and Mizera (2012) have applied a directional setting to define quantile trimmings. Hence, a second contribution of this chapter is to outline a general approach to detect directional multivariate extremes, which can be useful in other statistical areas apart from environmental sciences.

The definition of directional multivariate extremes is based on the directional mul-

[1]For further information of the return period we refer to Salvadori et al. (2007)

tivariate quantile introduced in Laniado et al. (2010) and improved in Chapter 2.2.1, where the free parameter of direction included can be chosen considering external information such as anthropogenic forces generating today's the environmental global-change (see Hegerl et al. (2004)). Specifically, we propose to use principal component analysis (*PCA*) in the environmental framework since the visualization of the extremes improves with respect to the use of the classical directions, as is shown in two cases of study, *PCA* is only a suitable method to select a direction of analysis. However, if prior information is available about the physical phenomenon, other directions can be more appropriated. Firstly, we use the flood model proposed in Salvadori et al. (2011) for the Ceppo Morelli dam in Italy to perform a Monte Carlo study for a time window of 1000 years. Our approach improves previous results by the reduction of the ratio of false positives (regular observations which are classified as extremes). Secondly, we perform a study of sea storms considering variables such as wave height, storm duration, storm magnitude, storm direction, and inter-arrival time which provide information about the period of calm between two successive storms. The study shows relevant differences with the work by De Michele et al. (2007) such as the computational feasibility of the method in the dimensional setting and also the visualization of the extremes with cross-sectional plots, where it is shown how the classical theory identifies an excessive number of observations as extremes.

The third contribution of the chapter is to introduce the directional approach in the copula method. We obtain results that establish the equivalence between the directional approach and the copula based methods. It is also shown with the simulations across the document how using a mixture of both settings (directional and copula approach), we can describe better a multivariate system.

The structure of the chapter is the following: Section 2 introduces the notion of directional multivariate extremes and the non-parametric procedure to carry out the identification in practice. Section 3 presents a summary of the classical methodology based on copulas, and theoretical results linking copulas and the notion of directions. In Section 4 and Section 5, we motivate the use of principal components (*PCA*) to get an interesting direction of analysis in real case studies. We also present in Section 6 some examples of the pros and cons for the extreme identification using our directional non-parametric procedure or the extended copula method. Finally, Section 7 presents some conclusions.

3.2 Methodology

In this section, we present the procedure to identify directional multivariate extremes based on the directional setting proposed in Chapter 2 (first contribution), a non-parametric algorithm for practical implementation (second contribution) and the motivation of the first *PCA* direction as a proposal of direction to be considered.

3.2.1 Directional Multivariate Extreme Value Analysis

Recall that the directional multivariate setting is defined in terms of the *oriented orthant* introduced by Laniado et al. (2012) and summarized in Definition 2.1.1. Note that an oriented orthant is a translation and a rotation of the non-negative euclidean orthant toward a new vertex in the point and a new direction . In Chapter 2 has been pointed out that is not unique for . Then, in order to guarantee uniqueness in the orthogonal transformation, the QR oriented orthant has been summarized in Definition 2.1.2. Particularly, the direction generates the rotation matrix equal to the identity matrix.

We remark that the consideration of directions with non-null components is not restrictive, because if a vector of direction has a null component, then the variable associated to the null component can be analyzed in a marginal way. Figure 3.1 shows examples of the divisions in the bivariate plane that can be performed using the concept of QR oriented orthant for different directions.

If (univariate setting), there are only two possible directions and the corresponding orthants at vertex are the intervals ,), respectively. Then, in terms of probability, they represent the valuation of the distribution and survival functions in . But, when , note that the values of the distribution and survival functions at some point correspond to the probability of the QR oriented orthants with vertexes in directions respectively. In the multivariate extremes literature, there are many studies that use those functions as a natural way to extend different procedures from the univariate extreme analysis (e.g. De Michele et al. (2005), Di Bernardino et al. (2015), Embrechts and Puccetti (2006), Salvadori and De Michele (2004)).

However, infinite directions are possible when , which motivates the directional approach, since more important than using the distribution and survival functions for a random vector , could be using directly the probability measure of the random vector to describe the extremes properly. To clarify ideas, one can

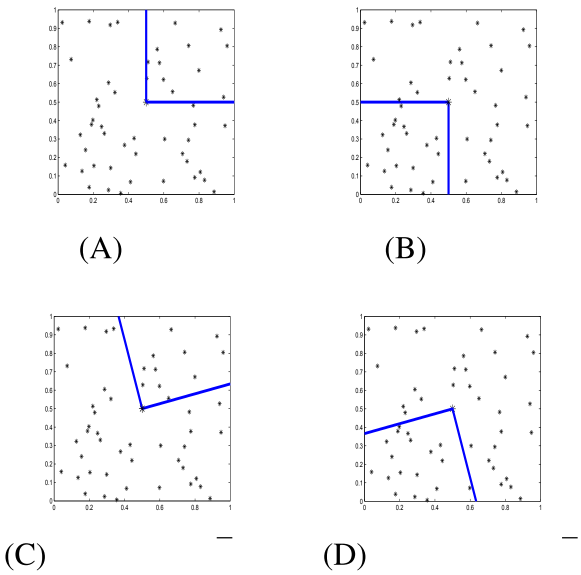


Figure 3.1: Examples of QR oriented orthants with the same vertex but different directions

think in the bivariate setting and a random vector with negative dependence. Then, it seems more convenient to use the complementary part of the division of the plane than the pair of directions , i.e., to use the directions given by (e.g. Belzunce et al. (2007), Chebana and Ouarda (2011)), hence the importance of the directional approach. Hereafter we call *classical directions* to the collection of orthants that divide naturally the hyper-plane, i.e., the collection of unitary dimensional vectors with components in . Now, we can introduce the necessary tools to attain the main purposes of our work, after motivation of the directions.

Recall also that a directional multivariate quantile of a random vector at level in direction , can be defined as,

(3.2.1)

where , and means the boundary of a set. Once a value of is fixed (near to for extreme value analysis), divides the space into two sets:

- *The upper level set in direction* :

(3.2.2)

- *The lower level set in direction* :

(3.2.3)

These sets motivate the definition of extreme related to the pair as those points exceeding the threshold given by the hyper-curve , i.e., we characterize the *extreme events* as those points belonging to the associated *upper level set*. The *risky points* are the ones belonging to *the directional multivariate quantile* and the *non-risky points* are those in the *lower level set*. That is,

is a directional extreme related to

Note that expressions (3.2.1), (3.2.2) and (3.2.3) with and values of close to zero are the multivariate extension of the univariate quantile definition based on the survival function. Now, if we rewrite those expressions in terms of the pair and reversing the inequalities, we obtain the corresponding quantile extension related to the distribution function, (we summarize the alternative set up and the results of synthetic and real cases of study in the Appendix B. of this chapter). However, these two alternatives are not equivalent for dimension unlike the univariate case. Such duality can be also seen in the approaches AND and OR defined in De Michele et al. (2007)), or the UPPER and LOWER differentiation given in Cousin and Di Bernardino (2013), Embrechts and Puccetti (2006). But, without loss of generality, we have decided to implement the extreme detection analysis in terms of the survival analogy, because a key relationship can be established between the extremes given by (3.2.2) and those associated to the arguments reversing the inequalities (see the proof of Corollary 2.3.3 in Chapter 2); that is,

(3.2.4)

Then, in terms of risks, relation (3.2.4) allows us to consider risk aversion; that is, we would expect more extreme events which corresponds to a conservative position. In Appendix B. of this chapter, one can find arguments for the selection of the survival framework. Now, we describe a non-parametric procedure to estimate the extreme thresholds, i.e., the directional multivariate quantiles, as well as, the lower and upper level sets for a dataset.

3.2.2 Non-parametric procedure

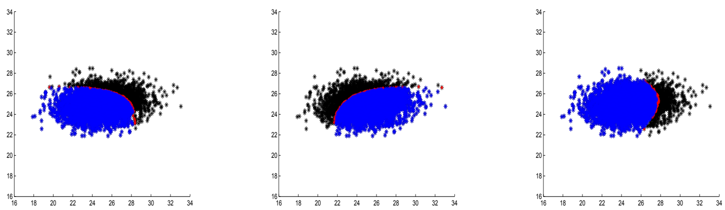
As we mentioned in the Introduction, one of the contributions of this chapter is to provide a non-parametric algorithm to estimate the quantiles. It is remarkable that most of the references that deal with the multivariate extreme identification problem are based on copula procedures that have inherent weaknesses due to the complex process of parameter estimation and the absence of computational feasibility in high dimensions. Therefore, we try to improve these issues by introducing a pseudo-algorithm based on the empirical distribution in order to get the level sets we are interested in. Firstly, we fix a preliminary notation:

- , sample data from the random vector ,
- is the empirical probability law of ,
- the sample quantile curve with a slack , avoiding an empty set of estimated quantiles.
- the sample upper level set with a slack ,
- the sample lowe level set with a slack .

Once defined the direction of analysis and the parameter , it is possible to estimate the directional multivariate quantile and the level sets through the following pseudo-algorithm:

Input: , , and the multivariate sample .
for to
,
If
,
end
If
,
end
If
,
end

As output, we get an estimation of the directional quantile, or in other words, the hyper surface of thresholds in the selected direction of analysis, . We also obtain those points belonging to the non-risky level set, , and the extreme level set, . An example is presented in Figure 3.2, where simulated data from a bivariate normal distribution with , , , is considered. We show the three sets of observations, the *directional* *quantiles* in red, the *upper* *level set* or *extreme level-set* in black and the *lower* *level set* or *non-risky level set* in blue. We have used three different directions: the classical direction (survival distribution), the complementary bivariate direction and the direction given by taken the first vector of its *PCA*. One can observe how the identification of extremes varies according to changes in the direction in which the data is analyzed, for the same level .



(A) (B) — — (C) First *PCA*

Figure 3.2: Directional Extremes at

Notice that different contexts or phenomena could lead to consider different particular directions of interest. For instance in portfolio theory, the direction given by the portfolio weights of investments is of particular interest because it takes into account the losses due to the composition of the investment in the portfolio (see Chapter 2 and Laniado et al. (2012)). On the other hand, researchers in environmental science could consider important other directions more related to the phenomenon of analysis.

In any case, we want to motivate here an interesting way to obtain a relevant direction of analysis by considering the principal component analysis (*PCA*) based on the original available data, which is an important statistical multivariate tool that describes the information about variability of the data jointly considered. It is well known that the first component provides the direction that accumulates the maximum amount of uncertainty of the data by the strongest linear combination representing the behavior of the system. This is a statistical property of the first *PCA*, but other possible choices can be more appropriated depending on the problem under consideration. We have tested this direction as a good candidate for the analysis in the applications considered in the chapter.

3.3 Extremes based on copulas and the directional approach

The importance of copulas have been recognized due to their capacity to capture the dependence structure of a set of random variables. Copulas are also a tool to construct families of multivariate distributions. In addition, copulas move in a compact support which guarantees theoretical advantages. Recall for example the capability to simulate data through copulas as we show in the case study of flood risk at a dam. Therefore, this section is devoted to introduce the directional approach to detect extremes when the dataset is modeled using copulas, which is the third contribution of the chapter.

In this chapter, we are considering the framework of survival distributions, then the survival copula will be denoted by . In general, a copula and its survival hold a relationship but it is difficult to apply when the dimension is large. For example, when , and are linked as follows,

(3.3.1)

The importance of modeling through copulas is due to Sklar's theorem 1.3.1, since any *joint survival function* (*joint distribution function*) can be obtained through its *marginal survivals* , (*marginal distributions*) and the *survival copula* (*copula*). This representation of the models makes more feasible to obtain closed or approximated expressions for in (3.2.1), in (3.2.2) and in (3.2.3). Thereby, in terms of survival copulas, equations (3.2.1), (3.2.2) and (3.2.3) for become the following,

such that

(3.3.2)

such that

(3.3.3)

such that

(3.3.4)

Most of the studies dealing with extremes detection in terms of copulas are based on definitions similar to (3.3.2), (3.3.3), (3.3.4) (e.g. Grimaldi and Serinaldi (2013), Salvadori and De Michele (2004)), which are focused on the direction given by the survival function (or those associated to the parameters , which are considering distribution functions). However, Chebana and Ouarda (2011) used the directions when negative dependent bivariate models are considered. Indeed, they consider copulas associated to a random vector, but rotated and degrees, which can be done taking advantage of the relationships between the corresponding copula and the following expressions,

and

These considerations highlight the need to include directions in the copula approach. Thus, the goal of this section is to include the general directional setting to the copula approach and to describe a directional extreme detection method based on copulas, although we will also show the drawbacks of the procedure with some illustrative simulations. The following result shows how the directional approach can be implemented using copulas.

Property 3.3.1. *Let be fixed, then the directional quantiles and the associated upper and lower level sets of a random vector (defined in* (3.2.1)*,* (3.2.2) *and* (3.2.3)*) are the same as those obtained by applying the copula method (summarized in* (3.3.2)*,* (3.3.3) *and* (3.3.4)*) to the random vector , where is the rotation matrix in (2.1.1).*

Proof. First note that any analysis using the information of the survival or the distribution functions for a random vector through copulas is equivalent to the analysis in the set of classic directions , i.e., the copula quantile analysis is always done in those directions. Moreover, once is fixed, Sklar's theorem (1.3.1) provides the following relationships between the random vector and the copulas , for any pre-rotation ,

(3.3.5)

(3.3.6)

where , , are respectively the marginal survival and distribution functions of the rotated random vector .

Hence we get the directional level sets (3.2.1), (3.2.2) and (3.2.3) by applying the inverse of the rotation to the elements belonging to the sets defined in equations (3.3.2), (3.3.3) and (3.3.4) where the copula modeling of has been used. All this thanks to Property 2.2.8 in Chapter 2 and relationship 3.3.5, (the result also holds through (3.3.6) when the alternative definition based on joint distributions is used). □

As a conclusion, the directional analysis can be done theoretically using copula models but over the pre-rotated random vector. An example to illustrate Proposition 3.3.1 on the bivariate field is provided below. Indeed, can be considered as the foundation of the nesting copula procedures used in the literature to confront the problem of large dimensions: *nested copula method* (see De Michele et al. (2007), Grimaldi and Serinaldi (2006)) and *Pair-copula construction*, also called the *Vine copula method* (see Grimaldi and Serinaldi (2013)), (see Appendix A. of this chapter for an introduction to these concepts).

Let be a bivariate vector with Gaussian survival marginals , with parameters , and , respectively. We also assume that satisfies a Gaussian survival copula with Pearson's correlation coefficient . Note that,

(3.3.7)

where , for all and denote the covariance matrix of by,

It is well known that the Gaussian copula is closed under orthogonal transformations. Then, for any direction , also holds a Gaussian survival copula with Pearson's correlation coefficient given by

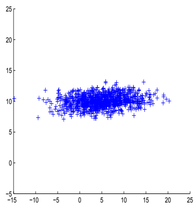
(3.3.8)

and Gaussian survival marginals , with parameters

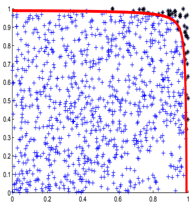
$$\mu_1^{\mathbf{u}} = \left(R_{\mathbf{u}} \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}\right)_{11}, \sigma_1^{2\,\mathbf{u}} = [R_{\mathbf{u}}\Sigma R'_{\mathbf{u}}]_{11} \quad \text{and} \quad \mu_2^{\mathbf{u}} = \left(R_{\mathbf{u}} \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}\right)_{21}, \sigma_2^{2\,\mathbf{u}} = [R_{\mathbf{u}}\Sigma R'_{\mathbf{u}}]_{22}, \tag{3.3.9}$$

where is the , position in a matrix.

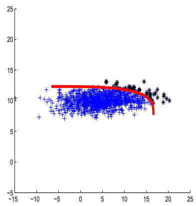
Now, we fix the parameters , , , and to illustrate the extreme detection through copulas. Figure 3.3 summarizes the results. The three top plots describe the procedure in the classical direction for and the three bottom plots describe the results for the same considering the first *PCA* direction given by the model, which in this case refers to the direction representing the principal axis of the elliptical distribu-



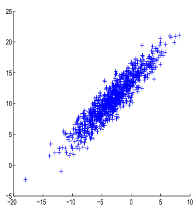
(A) Original data



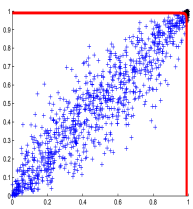
(B) Original data in the copula space



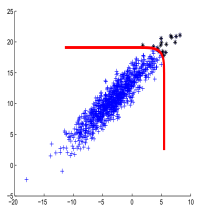
(C) Original data



(D) Rotated data



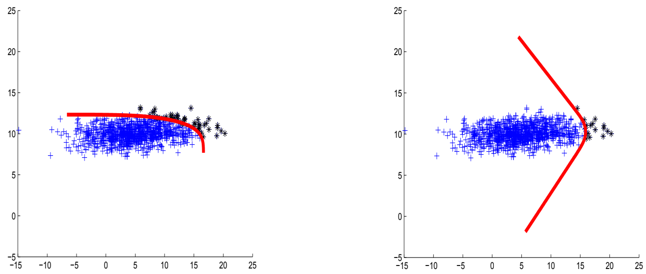
(E) Rotated data in the copula space



(F) Rotated data

Figure 3.3: Gaussian model. Top: theoretical results in direction ; Bottom: theoretical results in direction for the rotation of the data given by the first *PCA* direction

tion. Figure 3.3(A) shows the simulated data from the Gaussian model previously described, Figure 3.3(B) plots the copula space of the data (Gaussian copula) and the theoretical quantile (red), the lower (blue) and upper (black) level sets following the equations (3.3.2), (3.3.3) and (3.3.4). Finally, Figure 3.3(C) shows the corresponding results once the original space of the data is recovered through the inverse of the marginal survivals (all the colors have the same meaning as in Figure 3.3(B)).



(A) Classical direction (B) First *PCA* direction

Figure 3.4: Gaussian model. Comparison of the identification of extremes in directions and first *PCA* (black points)

In a similar way, but for the first *PCA* direction, Figure 3.3(D) shows the pre-rotated data due to the given direction , Figure 3.3(E) plots the copula space of the rotated data and the extreme detection in the copula space, and Figure 3.3(F) displays the extremes in the rotated space after applying the inverse of the rotated survival marginals. In order to compare the results in both directions, Figure 3.4(A) shows the extremes considering the direction and Figure 3.4(B) shows extremes in the first *PCA* direction undoing the rotation . Graphically, the differences in the two directions are obvious and the extremes detected using the first *PCA* direction look more realistic since they are more congruent with the shape of the data, (the results considering the extension through distribution functions can be found in Appendix B. of this chapter).

3.4 Real Case of study: flood risk at a dam

Salvadori et al. (2011) presented a dimensional model to describe floods occurring at Ceppo Morelli dam, located in Piedmont region, north-western Italy. In that work, the following three variables are considered: maximum annual *Peak*

(Q in), maximum annual *Volume* (V in) and initial *Water Level* (L in) in the reservoir before the flood event. The model that links all the variables was estimated using a copula approach to capture the correlation structure and generalized extreme value distributions (*GEV*) to describe the marginal behavior of and , while a non-parametric Normal kernel for . However, for simplicity in the calculations, the simulation has been made using GEV for all the marginals. Then, the model was finally completed through Sklar's theorem (1.3.1) and nested copula procedures, (see Appendix A. of this chapter for an overview of these elements). Figure 3.5 shows the scatterplot and the plot of the dataset used in Salvadori et al. (2011).

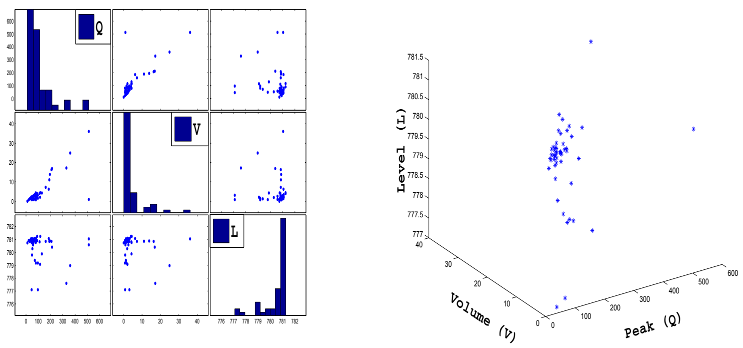


Figure 3.5: Flood risk at a dam: Cross-sections and plot of the dataset from Ceppo Morelli dam

The water level in Ceppo Morelli dam, L, before the flood event is the result of the rules used by the manager at the dam. Generally, they try to keep the water level as high as they can, in order to maximize the energy production. In any case, it is independent from the flood occurrence, and consequently flood peak and flood volume. This is the motivation from the authors to use the specifications of the model that we provide in Table 3.1, there we can find the *GEV* distributions fitted for each variable with the corresponding parameters of location , scale and shape and the copula model to recover the joint distribution of . The pair has associated a Gumbel copula with positive dependence, the pairs, and are modeled using the copula product. Finally, the flood copula model is given by after a nesting procedure.

The authors have used the quantile surfaces associated to this model to extend the notion of return period to the multivariate setting. Assuming the previous model for the random vector as appropriate, we now perform a Monte Carlo simulation with a large sample size to compare the multivariate extreme detection between

Q	$v = F (q)$	*GEV* with $\epsilon = 59.358m\ /s$, $\beta = 36.203m\ /s$, $\gamma = 0.368$
V	$v = F (v)$	*GEV* with $\epsilon = 1.7231m\ $, $\beta = 1.5246m\ $, $\gamma = 0.6149$
L	$v = F (l)$	*GEV* with $\epsilon = 780.6261m$, $\beta = 0.7623m$, $\gamma = \quad 1.5476$
QV	$C \quad (v\ , v\ )$	Gumbel copula with $\theta = 3.1378$
QVL	$C \quad = v\ C \quad (v\ , v\ )$	Nesting using copula product

Table 3.1: Model description given by Salvadori et al. (2011), changing to a *GEV* distribution the modelization of

the classical direction (direction of the survival function) and the direction given by the first *PCA*.

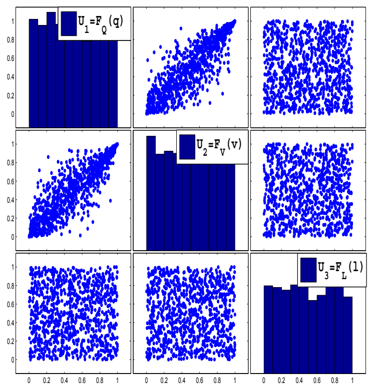


Figure 3.6: Flood risk at a dam: Cross-sections of the simulated sample from the copula model

Figure 3.6 presents the cross-sections of observations simulated from the copula model and Figure 3.7 shows the corresponding scatterplot and plot of the simulated data using the *GEV* distributions for the marginals and Sklar's theorem (1.3.1) to reconstruct dam behavior. Then, once the sample is generated, the extreme identification is made following the non-parametric approach at level in the two directions previously mentioned. Figure 3.8 (A) illustrates the analysis considering the classical direction , and Figure 3.8 (B) presents the extremes obtained considering the first *PCA* direction given by the original dataset collected since 1937 to 1994 at Ceppo Morelli dam. Both plots draw the lower or non-risky level sets in blue, the directional quantiles in red and the upper or extreme level sets in black.

Note that the number of extremes identified in direction is significantly greater

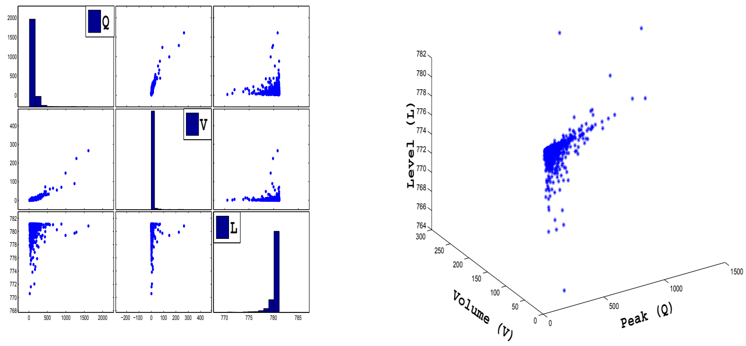
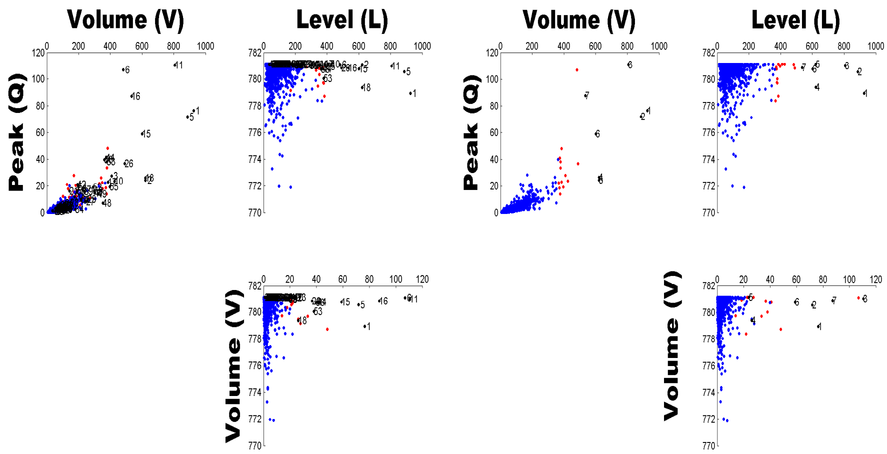


Figure 3.7: Flood risk at a dam: Cross-sections and plot of the simulated data for the model of floods

than using the first direction. Such a number of extremes seems excessive when a small value of is considered. The improvements obtained in the first *PCA* direction are remarkable graphically.



(A) (B) First *PCA*

Figure 3.8: Flood risk at a dam: Directional Extremes at

To obtain more evidence of the advantages of the directional approach, we generate triplets as inputs to operate the reservoir routing, analyzing the stress and reliability of the dam after long-time horizons of years long. This was done similarly to De Michele et al. (2005). In particular, each couple (Q, V) is transformed into a triangular flood hydrogragh of volume V and maximum peak Q, with base time , time of rise , and time of recession , (see e.g., [Chow et al. (1988), pg. 229]). is the water level in the dam at the beginning of the flood event. We operate the reservoir routing of flood hydrographs (see for details Bras (1990)[pg. 475-478]) considering as outlets only the uncontrolled spillways, and checking if the spillways are capable of disposing the flood events without overtopping the dam crest.

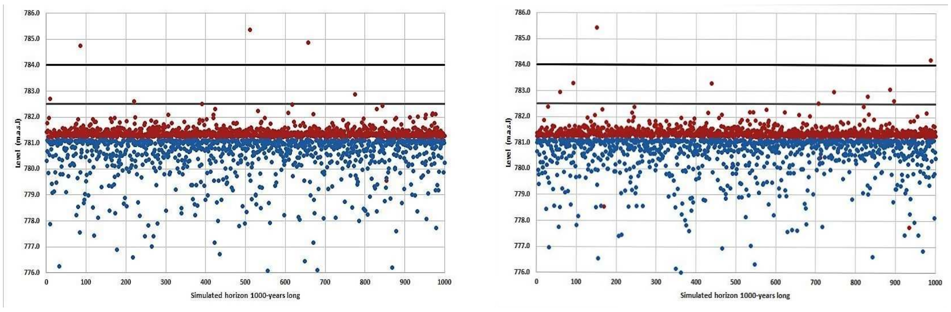


Figure 3.9: Flood risk at a dam: Two (out of 100) examples of dam simulation triplets ( )

Figure 3.9 presents two examples of the results after the simulation of dam behavior. In the images, it is possible to see the level of the dam spillway which is the virtual line drawn between the maximum levels occurred (red points) and the initial levels (blue points). Also shown are the lines defining the maximum level and the dam crest . Therefore, all the points between the maximum regulation level and the dam crest are considered as risky events and those points above the dam crest are considered catastrophic events. We have done simulations where each simulation spans years and the conclusion in all of them is that the *PCA* directional analysis captures better the critical events, i.e., the union of the sets of points given by the risky events and the extreme or catastrophic events. Meanwhile the classical direction identifies a huge number of such events, (the results considering the extension through distribution functions can be found in Appendix B. of this chapter).

Table 3.2 summarizes average indexes over the simulated samples analyzed in the two directions with . Specifically, the table describes: 1) The false positive ratio, which is the number of observations bad identified as critical over the total number of critical identifications. 2) The true positive ratio, which is the number of critical values correctly identified over the total number of real critical values from the dam routing simulation. 3) The extremes detection ratio, which is the number of observations identified as critical over the total number of observations. 4) The true extremes ratio, which is the number of real critical values over the total number of observations.

	Classical Direction	PCA Direction
False Positives Ratio		
True Positives Ratio		
Extremes Detection Ratio		
True Extremes Ratio		

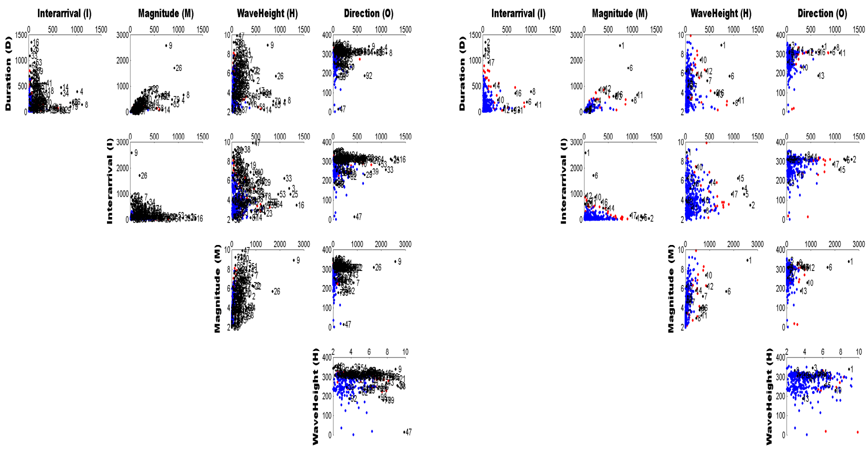
Table 3.2: Flood risk at a dam: Results of the Directional Extreme Analysis

The table shows that both directions identify correctly all the critical values with a of *true positives ratio*, but the analysis using the first *PCA* direction reduces significantly the *false positives ratio* of detection to , compared to in the classic direction. Also observe the small exceeding in the *extreme detection ratio* given by the first *PCA* direction with respect to the *true extremes ratio* , in comparison with the critical detection in the classic direction, which has a huge number of exceedances with a *extremes detection ratio*.

3.5 Real case of study: Sea storms

This case study is based on a dataset of sea storms which are described by five variables. This dataset has been studied in De Michele et al. (2007) and was collected by a wave buoy at Alghero (Sardinia, Italy) for a period of 12 years: from July 1, 1989 to October 31, 2001. The variables considered in the study are: wave height (in), storm duration (in), storm magnitude (in), storm direction (in) and storm inter-arrival time (in), which records the period of calm between two successive storms. It is assumed that sea storms can be considered independent and homogeneous events. A sea storm occurs when the wave height crosses upwards of meters and ends when the wave height stays below meters for at least consecutive hours. Specifically, the dataset counts a total of sea storms during the considered period.

Our objective in this case study is to identify those risky events with our directional proposal in this dimensional setting, comparing the analysis in the two directions proposed in the previous case study, the classical direction and the first *PCA* direction, which in this case is equal to . It indicates that *Magnitude* and *Duration* are the more relevant variables while *inter-arrival time* has negative value due to the fact that lower values of the variable increases the risk of storms. Figure 3.10 shows the cross-sections of the sea storms dataset, where the left plot presents the identification of extremes in



(A) (B) First *PCA*

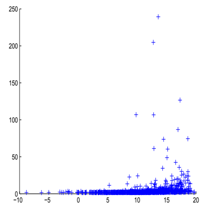
Figure 3.10: Sea storms: Directional Extremes for the case study sea storms at

direction for (black points) and right plot shows the extremes associated with the first *PCA* direction for the same (black points). In the same way as in the previous section, the visualization of extremes is more acceptable when the first *PCA* direction is used, (the results considering the extension through distribution functions can be found in the Appendix B. of this chapter).

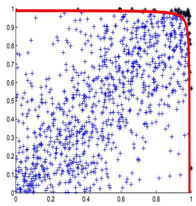
3.6 Weak points of the directional approach using copulas

The Gaussian copula is a toy example where the directional approach can be theoretically extended to the classical copula procedure. However, the usual fact is that the knowledge of the copula and the marginals associated to a random vector does not imply knowing the copula and the marginals over a rotation of the random vector. Therefore, a disadvantage of the directional copula approach is that it increases the computational cost when one decides to consider another direction of analysis different from .

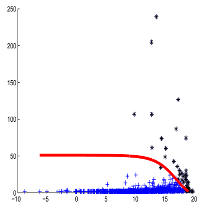
For example, let us consider a *Frank* copula, and marginal distributions belonging to the *GEV* family. Firstly, we have assumed positive dependence in the model by setting a Frank survival copula with dependence parameter and *GEV* marginals with parameters , , , , and



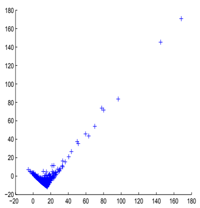
(A) Original data



(B) Original data in the copula space

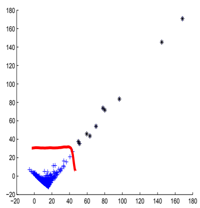


(C) Original data



(D) Rotated data

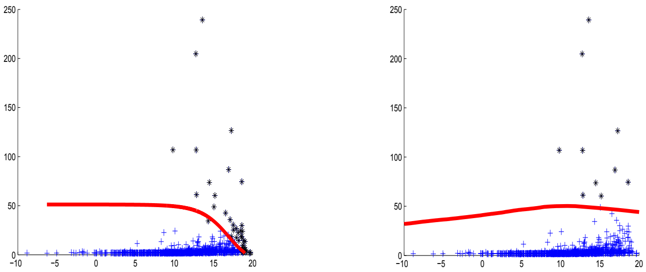
(E) Rotated data in the copula space



(F) Rotated data

Figure 3.11: Frank copula model with positive dependence. Top: theoretical results in direction ; Bottom: non-parametric approach in direction for the rotation of the data given by the first *PCA* direction

.



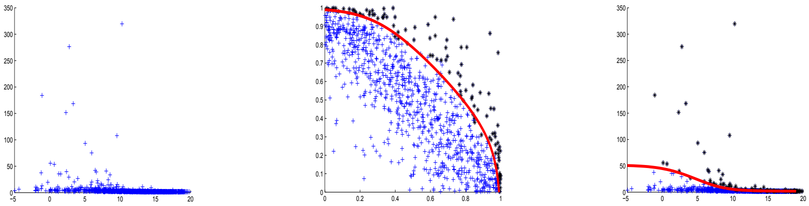
(A) Classical direction (B) First *PCA* direction

Figure 3.12: Frank copula model with positive dependence. Comparison of the identification of extremes in directions and first *PCA* (black points)

Figure 3.11(A, B, C) show the classical theoretical procedure used with copulas for and direction with the same meaning as in Figure 3.3(A,B,C). However, Figure 3.11(D, E, F) plot the analysis for the same , but using the non-parametric approach in direction over the pre-rotated data under the rotation given by the first *PCA* direction . Figure 3.11(D) shows the data in the rotated space, Figure 3.11(E) is empty due to the absence of theoretical evidence of the copula after the rotation of the data. Note that a possibility to fill the empty figure is to apply the non-parametric directional procedure presented in Section 3.2.2, but to the non-parametric copula of the rotated data (see Capéraà et al. (1997)), since Proposition 3.3.1 guarantees the theoretical equivalence. However, the directional approach has the advantage that the extremes can be obtained without considering the copula space of the rotated data as is shown in the identification presented in Figure 3.11(F).

To compare the detected extremes, Figure 3.12 displays in black those points considered as extremes in both directions, once the rotation of the data is undone in the case of the first *PCA* direction. The large number of points identified as extremes in the case of the classical direction with can be observed, when many of these identified observations could be considered as regular observations. Meanwhile using the first *PCA* direction, the number of extremes is considerably reduced and they appear more reasonable.

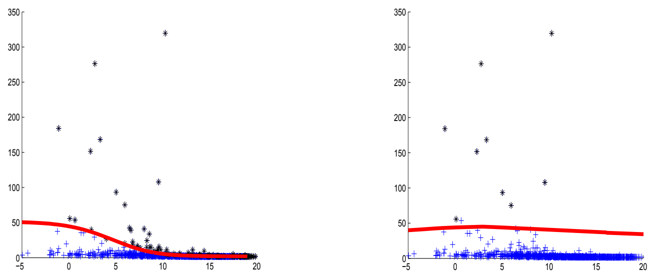
To conclude this section, we consider a model with negative dependence. In this case, is again , but the parameter of dependence in the Frank survival copula is and we use the same *GEV* marginals as in the previous example. Figure 3.13 shows the outputs in the same framework as Figure 3.11, and Figure 3.14



(A) Original data (B) Original data in the copula space (C) Original data

(D) Rotated data (E) Rotated data in the copula space (F) Rotated data

Figure 3.13: Frank copula model with negative dependence. Top: theoretical results in direction ; Bottom: non-parametric approach in direction for the rotation of the data given by the first *PCA* direction



(A) Classical direction (B) First *PCA* direction

Figure 3.14: Frank copula model with negative dependence. Comparison of the identification of extremes in directions and first *PCA* (black points)

shows the contrast between the classical and the first *PCA* directions for the detection of extremes. Once again, we can observe a better pattern of extreme recognition by considering the alternative direction of analysis , the first *PCA* direction, (all the examples described in this Section have been analyzed in Appendix B. of this chapter, but considering the extension through distribution functions).

3.7 Conclusions

In this chapter we propose a directional multivariate extreme identification procedure based on the notion of directional multivariate quantile. A non-parametric implementation feasible in high dimensions is also presented. We have proposed a directional inclusion to the classical extreme detection procedure based on copulas. We have highlighted the advantages and disadvantages of the directional non-parametric approach and the directional copula procedure, and we have analyzed simulated and real scenarios where the advantages of using different directions to detect extremes is evident.

Specifically, *Principal Component Analysis* has been tested as a method to choose a suitable direction of analysis that offers a reasonable number of points identified as extremes, but more importantly, the locations of those identifications are more in the "*atypical zone*", if one looks at the cloud of observations and its shape. However, it is well known that the *PCA* is very sensitive to skewed data, data with heavy univariate tails or outliers. It would be very interesting to carry out a sensitivity analysis of the directional extreme detection method respect to these scenarios or to consider a more appropriate directions such as robust *PCA* (see e.g. Candès et al. (2011)). Anyway if the dimension of the data is high, Donoho (2000) states that the usual *PCA* is better than any robust alternative due to the *blessing of dimensionality*.

Appendix A

In this appendix, we summarize the elements of the copula modeling presented in the case of study of floods at a dam. Firstly we point out the copula product, which is a simple way to couple variables when there is independence among them, its expression in the bivariate setting is . Now we proceed to recall basics from Archimedean nesting procedures and Vine copulas, also called

Pair copula methods.

Archimedean nesting procedures: Archimedean copulas are characterized by the representation,

where is the generator and must be monotone in each component on . This representation allows to define a nesting procedure (not necessarily in bivariate terms) through a hierarchical structure,

(3.7.1)

where the argument ( ) in (3.7.1) is replaced by another Archimedean copula (again, not necessarily bivariate one), such as,

Then, the nested Archimedean copula is obtained as follows

(3.7.2)

Vine copulas: Under the hypothesis of Sklar's theorem (1.3.1), for every bivariate copula , a bivariate copula density exists given by,

This implies,

- joint density,

(3.7.3)

- conditional density,

(3.7.4)

Then, a density can be represented as a product of a pair copula densities and marginal densities. For instance, in dimension , and replacing properly (3.7.3) and (3.7.4) we get the representation. Vine method is more flexible than Archimedean nesting procedures, since we can select bivariate copulas from a wide range of parametric families.

Finally, we summarize the family of *Generalized Extreme Value Distributions* (*GEV*) (see Kotz and Nadarajah (2000)). Then a distribution belongs to the *GEV* family if it follows the structure:

GEV Distributions:

$$\text{— } ^{1} \text{ if} \qquad \text{— if} \tag{3.7.5}$$

where , are the scale, location and shape parameters respectively. We refer to Salvadori et al. (2007) for a thorough review of univariate extreme analysis and multivariate approaches using copulas, with applications.

Appendix B: Alternative Methodology

The purpose of this appendix is to show the differences of detecting multivariate extremes when the *distribution function* is used instead of the *survival function*. Recall that in the notion QR *oriented orthant* given in Definition 2.1.2, the value of the distribution function of a random vector evaluated at some point is the same as the probability of the oriented orthant in direction and vertex . On the other hand, the value of its survival function at agrees with the probability of the oriented orthant in direction and vertex . In the directional framework, the same discussion holds for .

We also highlight that upper and lower sets are strictly related with directions due to geometrical aspects. For instance, in the univariate setting when we are interested in minima of the variable, we focus on the left tail of the density function and the interpretation of an upper set is related to values less than a chosen quantile. However, when the interest is in maxima, one look to the right tail of the density function and the upper set corresponds to values greater than the quantile, i.e., the upper set depends on the chosen direction (*distributions*) or (*survivals*). The same happens in the multivariate setting, but the complexity increases because there are infinite directions to analyze extremes and it is possible to get both approaches for upper sets for each chosen direction, due to the duality on the extension.

In the chapter, we have carried out all the analysis using survival function (see equations (3.2.1), (3.2.2) and (3.2.3)), but we can also consider distribution function easily if we consider the pair , and the inequalities also changed in the following way,

— (3.7.6)

— (3.7.7)

— (3.7.8)

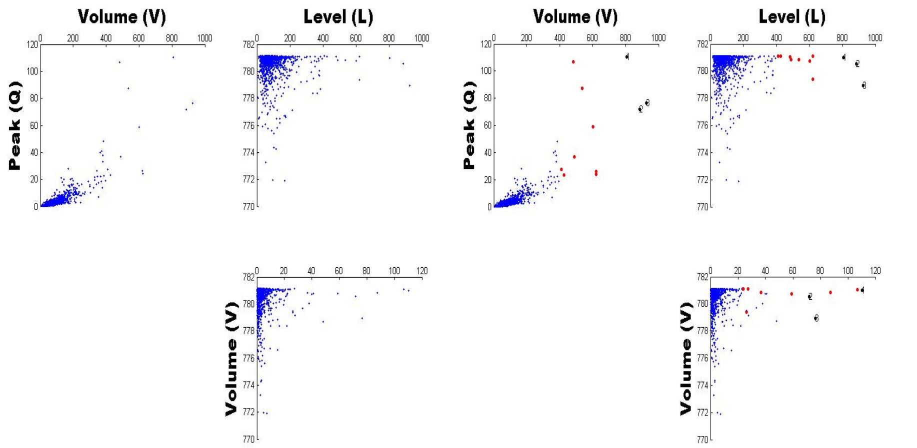
Equation (3.2.4) establishes an important relationship between upper sets in both approaches, which induces that an approach through distribution functions is less conservative, because the number of points identified as extremes in each chosen direction is lower.

However, this appendix shows the results under the distribution function approach in each of the scenarios proposed in the chapter, real and simulated ones, in order to present the alternative and compare the approaches. Hereafter, we refer to the analysis using equations (3.7.6), (3.7.7) and (3.7.8) evaluated at as the analysis through distribution functions at level .

3.7.1 Case study: flood risk at a dam

The analysis through distributions at level for Ceppo Morelli dam model can be performed through simulation as we did in the chapter. Figure 3.15 is an example of the result for the classical and the first *PCA* directions. Notice that Figure 3.15 (A) displays the results in the classical direction , which is an empty identification of extremes. On the other hand, the improvement of using the first *PCA* direction is evident.

Now, using the stress and reliability method over the behavior of the reservoir routing in horizons of years long, that we have proposed in the chapter, it is possible to highlight again the advantages of the directional approach. Table 3.3 summarizes average indexes from the analysis through distribution functions at level over simulated samples in both directions. Specifically, the table describes: 1) The false positive ratio, which is the number of observations bad identified as critical over the total number of critical identifications. 2) The true



(A) (B) First *PCA*

Figure 3.15: Directional Extremes through distributions at , ( )

positive ratio, which is the number of critical values correctly identified over the total number of real critical values from the dam routing simulation. 3) The extremes detection ratio, which is the number of observations identified as critical over the total number of observations. 4) The true extremes ratio, which is the number of real critical values over the total number of observations.

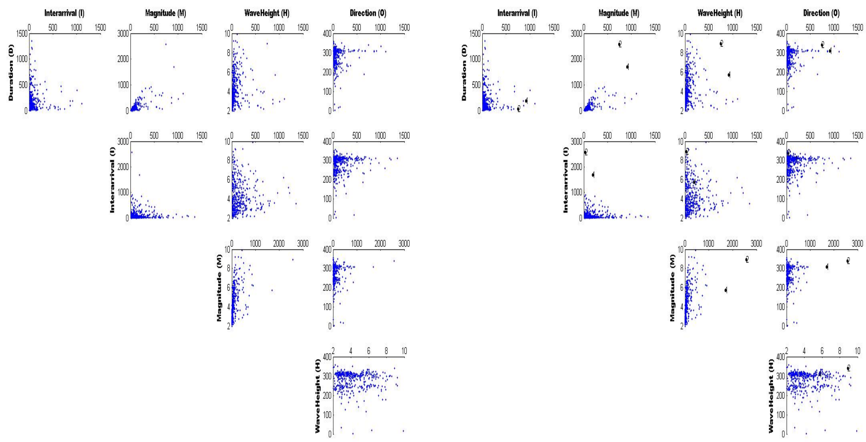
	Classical Direction	PCA Direction
False Positives Ratio		
True Positives Ratio		
Extremes Detection Ratio		
True Extremes Ratio		

Table 3.3: Results of the Directional Extreme Analysis through distributions at

The table shows better indexes using the first *PCA* direction than the classical. Note that results between *upper sets* in both approaches are consistent according to equation (3.2.4).

3.7.2 Case study: Sea storms

If we analyze the sea storms dataset collected at Alghero, Italy. Figure 3.16 (A) presents the results of the classical approach through distributions at level and Figure 3.16 (B) displays the results in the first *PCA* direction. The results are similar to those in the flood risk analysis at a dam showing better performance using the directional approach over the first *PCA* direction.



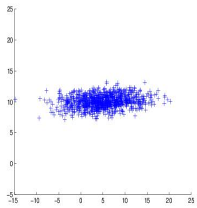
(A) (B) First *PCA*

Figure 3.16: Directional Extremes throug distributions at level , ( )

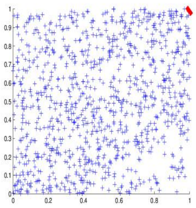
3.7.3 Extremes based on copulas and the directional approach

Finally, in this section we recall the copula examples used in the chapter to show the extreme detection through distribution functions. Specifically, we consider: 1. The elliptical copula example, 2. The positive dependent Frank copula, and 3. The negative dependent Frank copula.

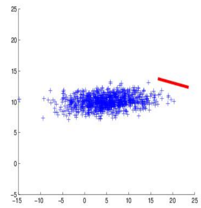
Firstly, we recall the Gaussian copula example with parameters , , , and . Figure 3.17 summarizes the results of the extreme detection through distribution function approach at level ( ). The three top plots describe the procedure in the classical direction for the original data, meanwhile the three bottom plots describe the results also in the classical direction, but for the data rotated according to the corresponding when is the first *PCA* direction. The quantiles (red) in the copula space are located in the upper-right corner, very close to , and they can be calculated theoretically in



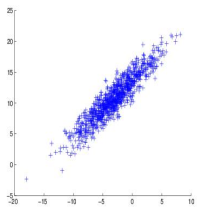
(A) Original data



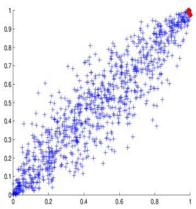
(B) Original data in the copula space



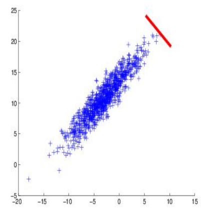
(C) Original data



(D) Rotated data

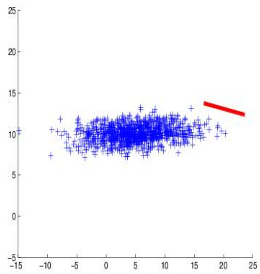


(E) Rotated data in the copula space

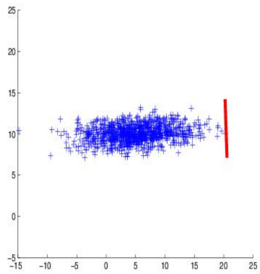


(F) Rotated data

Figure 3.17: Top: theoretical results through the classical distribution approach; Bottom: theoretical results through the classical distribution approach for the rotation of the data in the first *PCA* direction



(A) Classical direction

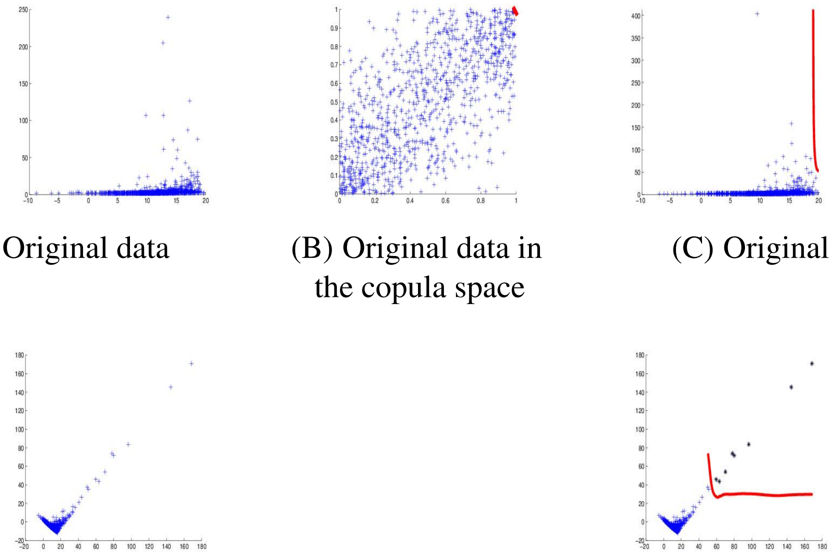


(B) First *PCA* direction

Figure 3.18: Comparison of the identification of extremes in the directions and first *PCA* (black points)

both cases thanks to the elliptical properties. Figure 3.18 shows the comparison of the results by undoing the previous rotation (right plot), where can be observed an improvement in the identification of extremes according to the shape of the data.

Now, we describe the results in a model considering *Frank* copula with dependence parameter , and marginals given by *GEV* distributions with parameters , , , , and .



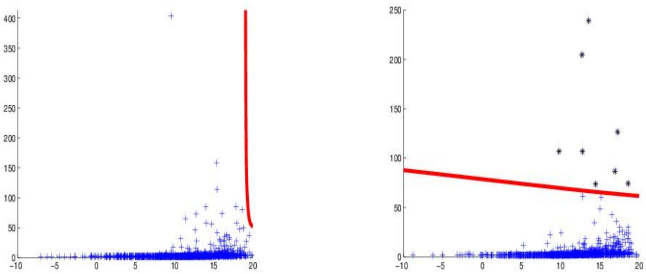
(A) Original data (B) Original data in the copula space (C) Original data

(D) Rotated data (E) Rotated data in the copula space (F) Rotated data

Figure 3.19: Top: theoretical results in direction ; Bottom: non-parametric approach in direction for the rotation of the data given by the first *PCA* direction

Figure 3.19(A, B, C) show the classical approach through distributions at level ( ) which can be obtained theoretically. However, Figure 3.19(D, E, F) plot the analysis for the same level, but using the non-parametric approach in direction over the pre-rotated data under the rotation given by the first *PCA* direction. Figure 3.19(D) shows the data in the rotated space, Figure 3.19(E) is empty due to the absence of theoretical evidence of the copula after the rotation of the data and finally, Figure 3.19(F) indicates the extremes in the rotated space but using the non-parametric approach.

Figure 3.20 displays the results to compare and it is noted that the classical approach through distributions does not identify extremes, even when graphically there exist some of them. On the other hand, the use of the first *PCA* direction in

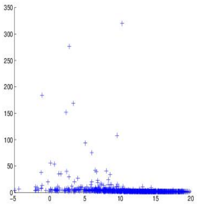


(A) Classical direction (B) First *PCA* direction

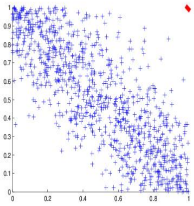
Figure 3.20: Comparison of the identification of extremes in the directions and first *PCA* (black points)

the analysis is more accurate.

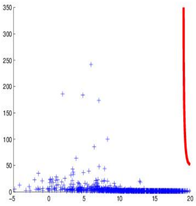
Finally, we perform the study in the model considering the *Frank* copula with dependence parameter , and marginals given by the same *GEV* distributions. Figure 3.21 shows the outputs in the same framework as Figure 3.19, and Figure 3.22 shows the contrast between the classical and the first *PCA* directions for the detection of extremes. Once again, we can observe a better pattern of extreme recognition by considering the first *PCA* direction as an alternative in the analysis.



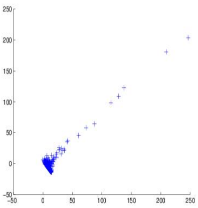
(A) Original data



(B) Original data in the copula space

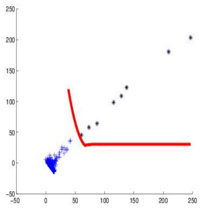


(C) Original data



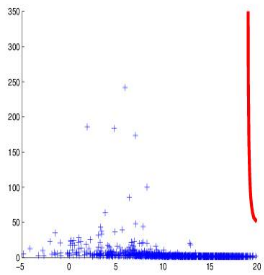
(D) Rotated data

(E) Rotated data in the copula space

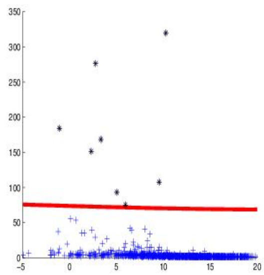


(F) Rotated data

Figure 3.21: Top: theoretical results in direction ; Bottom: non-parametric approach in direction for the rotation of the data given by the first *PCA* direction



(A) Classical direction



(B) First *PCA* direction

Figure 3.22: Comparison of the identification of extremes in the directions and first *PCA* (black points)

CHAPTER 4

ESTIMATION OF DIRECTIONAL EXTREME RISK REGIONS AT HIGH LEVELS

4.1 Introduction

The estimation of extreme level curves is important for identifying extreme events and for characterizing the joint tails of multidimensional distributions. They are usually considered as quantiles at high levels; that is, they are linked with a probability of occurrence of a certain event, where is a very small number. This proposal considers values of lower or equal than , where denotes the sample size, which implies that the number of data points that fall beyond the quantile curve is small and can even be zero; thus we are outside of the observable region, or in other words, in the framework of *out-sample* estimation. This lack of relevant data points makes the estimation difficult, making it necessary to introduce tools from the multivariate extreme value theory.

The main purpose of this chapter is to provide an *out-sample* estimation method for the directional multivariate quantiles used in previous chapters. In these chapters, the directional setting refers to the inclusion of a parameter of direction that allows the analysis of data by looking at the cloud of observations from different perspectives. Accurate assessments of these quantiles are sought in a diversity of applications from financial risk management (e.g. Chapter 2, Laniado et al. (2012)) to environmental impact assessment (e.g. Chapter 3). A non-parametric estimation method was developed in Chapter 2 to estimate the directional quantile based on the empirical distribution function, which is valid just for the *in-sample* scenario; that is .

Both scenarios, *in-sample* and *out-sample*, have been widely studied in the univariate setting and recently the literature has focused on the extension to the multivariate context. Some relevant references in this area can be grouped into three categories as follows. Firstly, estimation under optimization processes; for instance, those proposals with estimation methods based on linear quantile regression (see, e.g., Chaudhuri (1996), Girard and Stupfler (2015), Hallin et al. (2010), He and Einmahl (2016), Kong and Mizera (2012), Mukhopadhyay and Chatterjee (2011), Serfling (2002)). This category contains estimation methods *in-sample* such as the proposal by Chaudhuri (1996) for his notion of geometric quantiles. Recently Girard and Stupfler (2015) has also proposed an *out-sample* estimation method for these geometric quantiles.

A second category contains methods determining level curves of joint density functions in such a way that the set of points outside those contours has a probability equal to a given level . This method easily describes inner and outer regions at the given level through these curves (e.g., Cai et al. (2011), Einmahl et al. (2013)). The estimators proposed in this category have been developed mainly for the *out-sample* framework. For instance, Cai et al. (2011) have provided estimation of bivariate contour levels for some joint densities with elliptical and non-elliptical distributions, considering cases with asymptotic dependence and asymptotic independence.

Finally, the third category considers level curve estimations using either joint distribution or survival functions (e.g. Belzunce et al. (2007), Chebana and Ouarda (2009), De Haan and Huang (1995), Di Bernardino et al. (2011), Fernández-Ponce and Suárez-Llorens (2002)). Works based on copulas are also classified in this group (e.g. Binois et al. (2015), Chebana and Ouarda (2011), Durante and Salvadori (2010), Salvadori et al. (2011)). These works have introduced proposals in both contexts *in-sample* and *outsample*, but most of them present the theory or have applications only in the bivariate case.

The proposal developed in this work is inspired in the work of De Haan and Huang (1995), where an *out-sample* estimator for bivariate level curves of a distribution function was proposed. Since our approach is based on distributions, it belongs to the third category presented above. The contributions of the chapter are two fold: the inclusion of the directional framework provided in previous chapters to the estimation of multivariate high level sets, and the extension of the *out-sample* estimation method in De Haan and Huang (1995) to any dimension.

The chapter is organized as follows. In Section 4.2 we summarize the main definitions and results related to the directional multivariate framework used in this thesis. Section 4.3 introduces definitions from the multivariate extreme in order to fix conditions over the random vector to ensure the results of this chapter. In Section 4.4, the characterization of the elements of the directional multivariate quantiles at high levels based on the heuristic ideas in De Haan and Huang (1995) are described. Section 4.5.1 presents an *out-sample* estimator for these quantiles and also the asymptotic normality of this estimator. Section 4.6 illustrates the estimation procedure comparing both theoretical and estimated results using a multivariate distribution to model . Finally, in Section 4.7 some conclusions are provided.

4.2 Directional multivariate quantiles

This section recalls necessary notions of the directional framework to understand the contributions of the chapter. As commented in the introduction, the directional multivariate setting was introduced in Laniado et al. (2012) based on Definition 2.1.1. In Chapter 2 was pointed out that is not unique for . Then, in order to guarantee uniqueness Definition 2.1.2 was introduced.

In the univariate setting, extremes are analyzed considering the possibilities of exceeding from either distribution or survival functions and most of the extensions of these analyses to the multivariate setting have also been concentrated on these two types of exceeding. For instance, in the bivariate case Embrechts and Puccetti (2006), Fernández-Ponce and Suárez-Llorens (2002), Salvadori (2004), Shiau (2003), and generalized multivariate versions are presented in Cousin and Di Bernardino (2013), Di Bernardino et al. (2015), Fraiman and Pateiro-López (2012), Gupta and Manohar (2005). Our proposal includes both options by taking into consideration the directional multivariate quantiles and level-sets in (3.2.1), (3.2.2), (3.2.3) and (3.7.6), (3.7.7), (3.7.8).

To develop the contributions of this chapter, we have selected directional multivariate quantiles based on distributions. Then, we recall (3.7.6),

Definition 4.2.1. *Let be a random vector with associated probability distribution . Then the directional multivariate quantile at level in direction is* *defined as*

(4.2.1)

where denotes the boundary of the subset considered into brackets and .

Once the framework has been fixed, note that implies the high level notion. Therefore, asymptotic analysis is necessary to characterize the directional multivariate quantiles, which implies the introduction of univariate and multivariate extreme value concepts. Hereafter this chapter has a self-contained structure of these concepts, thus a few definitions introduced previously in this dissertation will be repeated.

4.3 Probabilistic assumptions

In this section, we introduce conditions over that must be satisfied in order to introduce an estimator properly defined of when . We also include some constraints to obtain asymptotic properties of the estimator that we propose.

Definition 4.3.1. *A random vector is absolutely continuous if the underlying probability measure is an absolutely continuous measure with respect to the Lebesgue measure on .*

Assumption 1 (A1). *The random vector must be absolutely continuous with increasing marginal distribution functions and such that , for . Hereafter, these conditions are called regularity conditions.*

Note that **A1** avoids jump processes. It was also considered in Chapter 2 to define a multivariate value at risk, where the existence of the first moment of the random vector is important to fix a center of location as a reference point. Hereafter, we can suppose for simplicity that is a random vector with zero mean.

Assumption 2 (A2). *Given , possesses positive upper-end points of the marginal distributions.*

This assumption was introduced in De Haan and Huang (1995) for the marginals distributions of , but **A2** is more general and establishes the condition for the correspondent rotation associated to the direction of analysis .

Definition 4.3.2. *A distribution function with non-degenerate marginals is called a multivariate extreme value distribution if given , . . . , independent and identically distributed random vectors, there exist sequences of vectors with real components , , such that for all continuity points of ,*

(4.3.1)

The set of multivariate extreme value distributions is called the class of *multivariate max-stable distributions*. Let be the distribution function of the i.i.d. random vectors , . . . , , then (4.3.1) can be written as,

(4.3.2)

and is considered in the domain of attraction of , denoted by . Now, we recall Definition 2.2.10 introduced in Chapter 2 to point out an assumption over .

Definition 4.3.3. *A random vector has first order multivariate regular variation with tail index , denoted by , if there exists a real-value function that is regularly varying at infinity*[1] *with exponent and a non-zero measure on the Borel field such that for every Borel set B,*

(4.3.3)

where means vague convergence and (see, e.g., Jessen and Mikosh (2006), Resnick (1987)).

Assumption 3 (A3). *has first order multivariate regular variation with tail index .*

The following result establishes **A3** for all orthogonal transformations .

Proposition 4.3.4. *If has first order multivariate regular variation with tail index , then the random vector has first order multivariate regular variation with tail index , for any orthogonal transformation .*

Proof. An orthogonal transformation is a measurable function. Hence for each Borel set , we have that is a Borel set. On the other hand,

[1]A function $\phi(\ )$ is called *regularly varying at infinity with exponent* $1/\beta$, if it holds $\lim_{x\to\infty} \frac{\phi(tx)}{\phi(x)} = t^{\frac{1}{\beta}}$, for all $t > 0$.

we denote and as the probability measures of and , respectively. For a Borel set , we have that,

or analogously

Therefore, we obtain that the random vector is also regularly varying with tail index since,

□

Corollary 4.3.5. *Under the hypothesis of Proposition 4.3.4, the marginals of are regularly varying with tail index .*

Proof. This corollary is a straightforward consequence of the marginal implications of Proposition 4.3.4 (see De Haan and Ferreira (2006)[Chapter 6]). □

Definition 4.3.6. *A random vector has second order multivariate regular variation if there exist functions and , such that , ; satisfying for all relatively compact rectangles ,*

(4.3.4)

where is finite and not identically zero.

Definition 4.3.6 was given in Resnick (2002)[Section 3] in terms of relatively compact rectangles in , but we have rewritten it for all relatively compact rectangles in .

Assumption 4 (A4). *has second order multivariate regular variation.*

Proposition 4.3.7. *If has second order multivariate regular variation, then the random vector has second order regular variation for any orthogonal transformation .*

Proof. As before, we denote and as the probability measures of and , respectively. Then, we get for any Borel set , Hence,

□

Hereafter, we consider that a random vector satisfies Assumptions A1-A4.

4.4 Characterization of the directional multivariate quantiles at high levels

The objective of this section is to characterize the points of defined in (4.2.1) at high levels ( ). Our proposal is based on: the relationship provided in Property 2.2.8 in Chapter 2,

(4.4.1)

and the heuristic ideas of the bivariate quantile parametrization given in De Haan and Huang (1995), but extended to a general multivariate context. Therefore, we assume that verifies **A1** **A4**. Thus, Proposition 4.3.4 implies that the distribution function of the random vector belongs to the domain of attraction of a non-degenerate multivariate extreme value distribution $_{\mathbf{u}}$. Moreover, if are independent random vectors distributed as , there exist two sequences , satisfying (4.3.2), which also implies,

(4.4.2)

where and being the floor function. In addition, a direct consequence of (4.4.2) is that each marginal of has the form , for , and is called the tail index of the marginal (see De Haan and Ferreira (2006)[Chapter 6]). Then it is possible to write,

(4.4.3)

where is the marginal of .

Remark 4.4.1. *If , then the expression for the marginal of has the form .*

Thus, (4.4.3) implies that at high levels the quantile related to verifies the following relationship,

for all (4.4.4)

However, we want to estimate (see (4.4.1)), which according to Definition 4.2.1 is the set of points such that . To this end, we generalize the heuristic idea developed for the bivariate case by De Haan and Huang (1995) of introducing a parametrization of the quantile.

Recall that any point can be written alternatively in polar coordinates as , where and belonging to the unit dimensional ball (for a further discussion see Driver (2003)[pg. 217]). Note also that **A2** in the polar coordinates is equivalent to the assumption of upper-end points when is such that for all .

From now on, any point of under the new parametrization will be denoted by , . Then, the heuristic ideas in De Haan and Huang (1995) applied to combined with the characterization of the marginal quantiles at high levels given in (4.4.4) provide the following characterization of the elements of ,

for all

(4.4.5)

Therefore, given that all the elements in (4.4.5) are known or can be estimated, except , the problem of estimating turns into the problem of finding a solution for the scalar function and its estimation. Then, from (4.4.2) and (4.4.5), we obtain that,

(4.4.6)

Last equality in (4.4.6) is due to the following homogeneity property of for all (see De Haan and Ferreira (2006) [Theorem 6.1.9]),

(4.4.7)

Hence, from (4.4.6), we achieve a solution of by approximation. This solution will be denoted as,

(4.4.8)

which implies the approximation of by,

for all

(4.4.9)

Thus, is approximated at high levels by the parametrization,

(4.4.10)

where . This characterization allows an estimator *out-sample* for to be established based on extreme value theory, which is the objective of next section.

Remark 4.4.2. *As we have commented in Section 4.2, it is also possible to define directional multivariate quantiles based on joint survival functions, for which, the following definition has to be used,*

(4.4.11)

This implies that any point in should satisfy the equation . Then, we can characterize (4.4.11) adapting all the results of this section to joint survivals.

4.5 Inference for extreme-level directional multivariate quantiles

It is well known that extreme value theory is devoted to the asymptotic theory of sample maxima (minima). The word asymptotic refers to the assessment of level sets *out-sample*, i.e., when is near to zero. Our interest in this section is to find directional extreme sets based on a random sample. Then, given a direction , let be *i.i.d.* random vectors distributed as ; denote by the collection of th order statistics, for each marginal .

Marginal order statistics are important since they allow (4.4.2) (see Cai et al. (2011), De Haan and Ferreira (2006)[Section 7.2]) to be written in terms of a subsample that provides significant information about the behavior of the distribution at high levels. This subsample is related to intermediate order statistics starting from an order such that , when . Then, we obtain,

(4.5.1)

However, the key question in practice is how to select the optimal value of . Methods to establish this subsequence are still matter of research and discussion. Nevertheless, this problem is far beyond the scope of the present chapter, but we will use heuristic guidelines found in the literature (e.g. Cai et al. (2011, 2015), Di Bernardino and Palacios-Rodríguez (2016)) to select this tuning parameter adapted to each application, as we will explain in the following sections.

Therefore, an estimator of (4.4.10) can be given if we have the estimators of all the elements in (4.4.9). First, we state the estimator for the marginal tail indexes, , in terms of the moments estimator given in Dekkers et al. (1989).

(4.5.2)

Now, the estimators for the components of the sequences , are based on the ideas in De Haan and Huang (1995),

(4.5.3)

(4.5.4)

Finally, the estimator of the scalar function is defined by,

$$- \qquad (4.5.5)$$

Hence, equations (4.5.6) to (4.5.5) and (4.4.9) ensure an estimator for the elements of by,

for all

(4.5.6)

Finally, we obtain,

(4.5.7)

4.5.1 Asymptotic normality for x

The aim now is to prove the asymptotic normality of following a similar structure to the one presented by De Haan and Huang (1995). An important condition to complete this objective is **A4**, which allows a general multivariate version of the equations (1.8), (1.9), (1.10) and Theorem 2.1 provided in De Haan and Huang (1995) to be written in the directional framework. Therefore, we can establish asymptotic normality for our estimator in the case of such that , when . The referred equations become,

(4.5.8)

Then, we have that (see De Haan and Resnick (1977)),

$$\bar{k} \quad \log \hat{G} \ (\mathbf{x}) + \log G \ (\mathbf{x}) \qquad V \ (\mathbf{x}) := W \ (\mathbf{x}) + (\mathbf{B} \ + \mathbf{x} \ \ \mathbf{A} \ ) \ \nabla(\ \ \log G \ (\mathbf{x})), \tag{4.5.9}$$

where means convergence in distribution, means a componentwise product and ∇ means the vector of partial derivatives of a function;
, , and is a zero-mean random field for with covariance function,

(4.5.10)

and is a finite measure such that,

Theorem 4.5.1. *Let . Suppose that is a second order multivariate regularly varying random vector and suppose that,*

–

and

– 1

Then,

–

converges in distribution to

(4.5.11)

for all .

To prove this, we need to introduce the directional multivariate versions of the four lemmas, Lemma 2.1-2.4 in De Haan and Huang (1995).

Lemma 4.5.2. *If has continuous first-order derivatives , then*

–

converges to

Lemma 4.5.3. *Under the conditions of Theorem 4.5.1,*

converges in distribution to

for all .

Lemma 4.5.4. *Under the conditions of Theorem 4.5.1,*

locally uniformly.

Lemma 4.5.5. *Under the conditions of Theorem 4.5.1,*

locally uniformly for all .

The proofs of these Lemmas work in a similar way as in De Haan and Huang (1995) but considering the arrangements due to the directional multivariate framework, for which they are omitted.

Proof of Theorem 4.5.1. Lemma 4.5.2 proves the asymptotic normality of the standardized difference . This implies the asymptotic normality of the standardized difference in Lemma 4.5.3.

Also, Lemma 4.5.4 proves the convergence to zero of the standardized difference , which helps to prove Lemma 4.5.5 where the convergence to zero of the standardized difference is given. Thus, Lemma 4.5.3 and Lemma 4.5.5 complete the result in (4.5.11). ☐

Finally, the asymptotic property of is derived using that orthogonal transformations preserve the result in Theorem 4.5.1.

Corollary 4.5.6. *The asymptotic normality property of the estimator is preserved under orthogonal transformations. Therefore* (4.5.7) *implies the asymptotic normality of .*

4.6 An illustrative example

In this section, we calculate for a particular random vector (holding a multivariate distribution) both theoretical and estimations of the directional multivariate quantiles in two different directions. Recall that any multivariate distribution belongs to the family of multivariate elliptical distributions. Thereby, we have considered the classical direction and the direction given by the main axis of the elliptical random vector, which is equivalent to the vector characterizing the principal component analysis (*PCA*). It is also well known that a random vector with a multivariate distribution holds **A1** **A4** (see Arslan (2004)). Moreover, the -distribution holds the heavy tailed property by choosing a small parameter of degrees of freedom , because the tail index of in Definition 4.3.3 is equal to . Now, we recall a lemma for elliptical distributions, which is necessary to derive the theoretical calculation method for the directional multivariate quantiles (see Hult and Lindskog (2002)[Lemma 3.1]).

Lemma 4.6.1. *If has an elliptical distribution and decomposition given by*

where is a random variable independent from the random vector , which is uniformly distributed in the unit circle of dimension , a location parameter and a matrix indicating scale. Then has an elliptical distribution with associated decomposition given by

Moreover, its marginals are the associated univariate elliptical distributions with parameters of location and scale given by and , for .

Then, this lemma establishes that any rotation of is again a multivariate distribution with the same degrees of freedom , but with location and scale given by , . We also obtain that the univariate marginals are distributions with degrees of freedom, which implies a marginal tail index of for all .

For the illustration developed in the chapter, we choose the following parameters for a bivariate distribution,

This implies that the first *PCA* is . Therefore,

Figure 4.1 shows in black the curves holding for three different 's ( ). These curves are the theoretical directional quantiles of in the classical direction , obtained by applying optimization methods in MATLAB. Figure 4.1 also displays in red the curves holding for the same 's, but after applying the inverse of the rotation . In this case, the curves correspond to . Visual improvements of the extremes captured using the directional approach against the classical one can be observed since extremes captured in the first *PCA* directional analysis take into account the shape of the data.

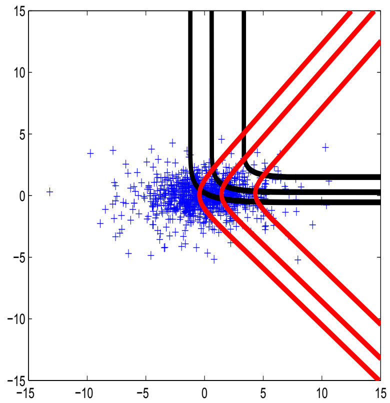
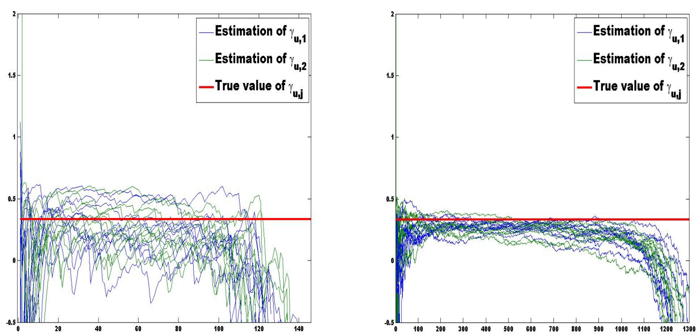


Figure 4.1: Classic and first *PCA* directional quantiles for , and .

Now, we proceed to present step by step all the necessary elements for the estimation of the directional multivariate quantiles. First, we present the procedure in the classical direction , when is equal to . We have performed this procedure by Monte Carlo simulation for different sample sizes, but we present two important cases: (1) "small", and (2) "large". In the process, we have replicated the simulation of times to obtain the results.

1. **Marginal tail index and tuning parameter selection**, and : Under the assumptions of the bivariate distribution, the theoretical value of the tail index for both marginals , for any direction , is equal to . Now, we perform the estimation of this index through (4.5.2), i.e., the moments estimator also called modified Hill estimator (see Dekkers et al. (1989)). This step is key to our estimation procedure, because it is used to select the value of the tuning parameter .

   We have used a heuristic approach similar to those given in Cai et al. (2011, 2015), Di Bernardino and Palacios-Rodríguez (2016)) to select suitable values of as the values with a common region of convergence in the estimation of all the marginal tail indexes, , . Figure 4.2 presents the results of vs. after simulations for each sample size. The straight line in red is the true value of the tail index, the behavior of the estimation for the first marginal is plotted in blue and the second marginal is in green. Hence, Figure 4.2(A) suggests that could be an integer value in when and 4.2(B) indicates the interval when .

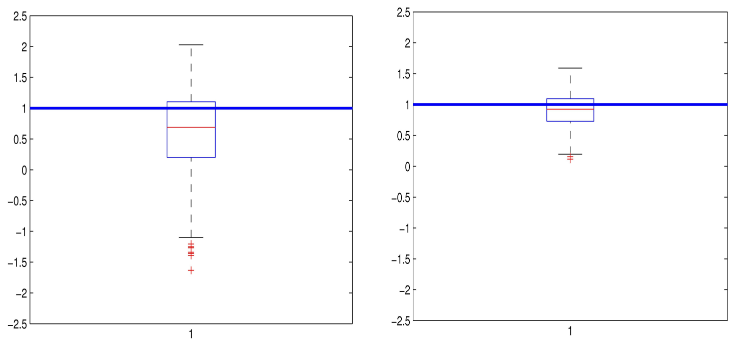


(A) (B)

Figure 4.2: k vs.

Once the intervals for are decided, we run simulations up to to construct Figure 4.3. In each simulation, the value has been calculated as the average of the estimations with varying in either or . Hereafter, all the estimations are the average of the results varying in its corresponding interval.

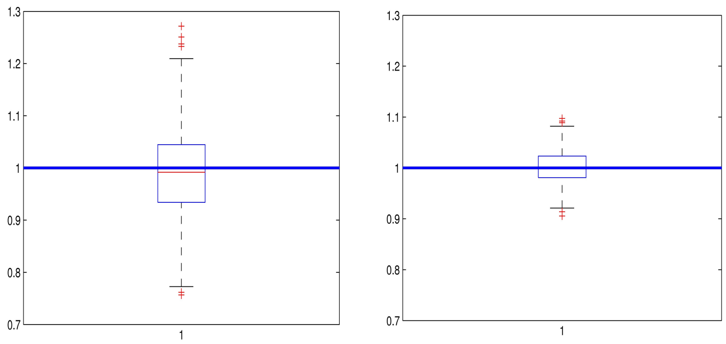
Figure 4.3(A) displays the boxplot of the quotients with and Figure 4.3(B) also presents the results for the marginal with . We can observe that the estimation decreases when the sample size is small, as can be expected. We have decided to present the behavior of through the illustration because this marginal possesses more variance in the model. However, the behavior of marginal is similar for both sample sizes. Although Figure 4.3 shows that the performance in this step is not totally accurate for both cases, but this relies on the selection of . Therefore, any improvement in that selection will also improve the whole procedure.



(A) (B)

Figure 4.3: Boxplot of

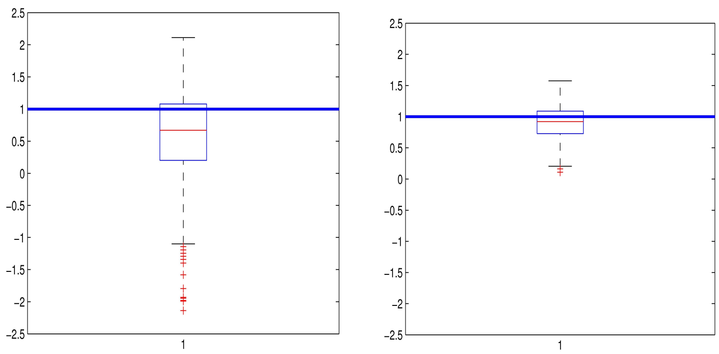
2. **The sequence** : This sequence depends on the theoretical quantiles of each marginal and its estimation is given by (4.5.4). Theoretically for a sample size , takes the value —— , where $_2$ is the univariate distribution with location , scale and degrees of freedom . Then, we study the behavior of the ratio between the estimated and the theoretical values. Figure 4.4 displays the results of for both sample sizes, (the behavior of is very similar for each case). It is easy to appreciate the good behavior of the estimation.



(A) (B)

Figure 4.4: Boxplots of

3. **The sequence** : This sequence depends on the previous elements. For a general direction , the theoretical value is and its estimation is obtained by (4.5.3). Therefore, by using the information on the previous steps, we obtain the results plotted in Figure 4.5, where we found the boxplots of and the loss of precision in the estimation arises from the estimation of .



(A) (B)

Figure 4.5: Boxplots of

4. **The scalar function** : Note that (4.4.8) uses the function , which is the tail function of the multivariate extreme value distribution .

Then, we remark the results in Nikoloulopoulos et al. (2009) in order to express the theoretical tail function for a multivariate distribution.

Remark 4.6.2 (Nikoloulopoulos et al. (2009) Theorem 2.3)**.** *The theoretical tail function of , a dimensional distribution with degrees of freedom, location parameter and scale parameter , is given by,*

(4.6.1)

where are the correlations between the components , is a distribution in dimension (removing the component), with degrees of freedom, location parameter and scale parameter given by,

$$\vdots \quad \ddots \quad \vdots \quad \vdots \quad \vdots$$

$$\vdots$$

$$\vdots$$

$$\vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots$$

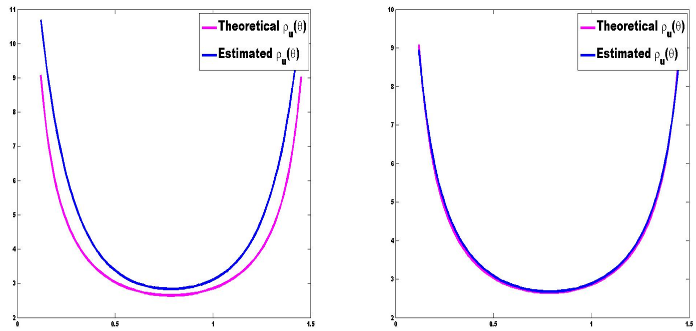
$$\vdots$$

where $\frac{\quad}{\sqrt{\quad}_{2}\sqrt{\quad}_{2}}$*, for* .

Therefore for any direction , we can calculate and by using Remark 4.6.2, Lemma 4.6.1 and (4.5.5). Figure 4.6 shows the theoretical curves (magenta) and the estimated ones (blue) of . We can appreciate a good performance of the estimation for both sample sizes.

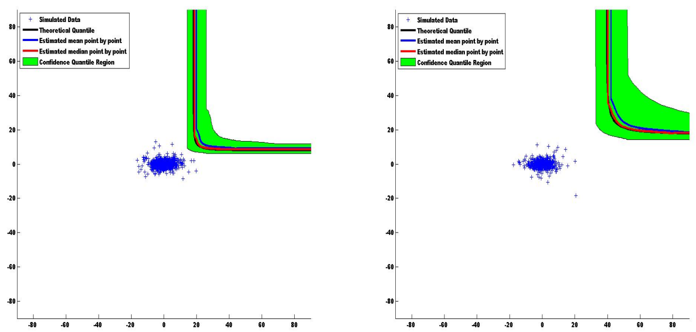
5. **The directional quantile curve** : Having all the previous theoretical and estimated ingredients, we proceed to calculate both theoretical and estimated results for . Figure 4.7 presents these results for both sample sizes . Theoretical quantiles are plotted in black, means point by point of the estimated curves are in blue, medians point by point of the estimated curves are in red, and confidence regions for the estimated curves from to are shaded in green. We can appreciate the accuracy of the estimations of . It is also possible to see from the confidence regions that uncertainty grows when the extreme level approaches zero.

Finally, we repeat the same procedure but using the first *PCA* direction. Thereby, we calculate theoretical and estimated solutions for



(A) (B)

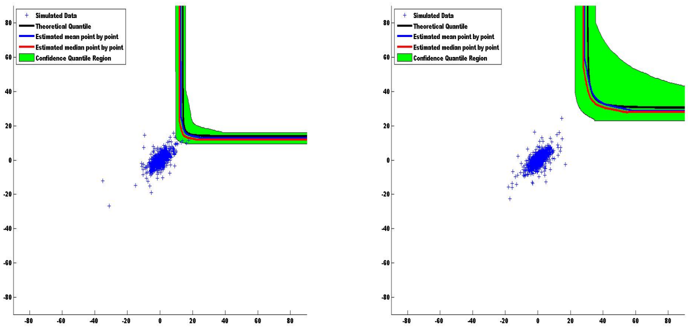
Figure 4.6: Theoretical and estimated curves



(A) (B)

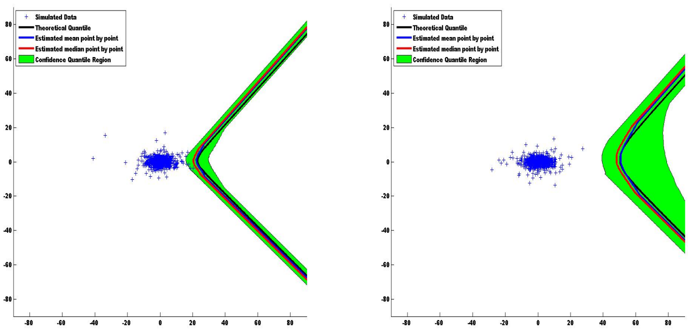
Figure 4.7: Estimations for

. We have considered the same sample sizes in the Monte Carlo simulations and Figure 4.8 presents the results for , in the same colors as before. Thus using (4.5.7), Figure 4.9 presents the results for , where the rotation has been applied according to (4.5.7). We can appreciate the good performance of the estimators and the improvements of the identification of the extremes based on the shape of the data.



(A) (B)

Figure 4.8: Estimations for ,



(A) (B)

Figure 4.9: Estimations for

4.7 Conclusions

This chapter has presented an *out-sample* estimator of the directional multivariate quantiles , introduced in previous chapters. Necessary conditions over to ensure a feasible estimation at high levels have also been presented. The proposed estimator integrates different results from the univariate and the multivariate extreme value theory through a heuristic parametrization in polar coordinates in and the asymptotic normality of the estimator has also been proved.

Finally, based on the multivariate distribution, an illustration has been shown. This family of distributions possesses properties such as heavy tails and closure under rotations, which provide a good example for comparing theoretical and estimated solutions. After estimation through Monte Carlo simulations, we have appreciated the good performance of the estimation process.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

This chapter summarizes the main contributions of the thesis. The work is based on the notion of *directional multivariate quantile* described in (1.2.7) and introduced previously in Laniado et al. (2010). We have studied in depth this concept from a theoretical point of view, as well as, its applications. The applications are linked to extremes detection in different fields such as finance, insurance and environmental sciences. Historically, the notion of quantile has been important in extreme value theory and its extension to the multivariate field has also been studied in many different ways. A extensive review of the generalizations of the univariate quantile to the multidimensional is provided in Chapter 1.

We find that the main contribution of the thesis is the inclusion of a parameter of direction in the multivariate setting to give more flexibility to the analysis and the capability of looking at the data considering different perspectives. This general purpose is achieved considering special Borel sets defined as QR *directional orthants* which allow us to define the directional multivariate quantiles .
In Chapter 2, we have introduced a vector-valued risk measure, the **directional multivariate Value at Risk** ( ) using the previous concepts. We have proved properties for this risk measure in the same way as Artzner et al. (1999)'s axiomatic, such as consistency and tail subadditivity, which are desirable for any

risk measure. We have proved a relationship between the components of the output of and the components of the vector of marginal univariate *VaR*'s. We have also motivated this measure as a tool to analyze a portfolio of heterogeneous and dependent risks considering the direction of the portfolio weights. Finally, we have presented a non-parametric implementation to perform a simulation study of robustness comparing the behavior of with respect to the risk measure proposed in Cousin and Di Bernardino (2013). The simulations show the advantages of our proposal in relation to the presence of outliers.

In Chapter 3, we proposed a directional multivariate extreme identification based on a non-parametric implementation feasible in high dimensions, as well as, a directional inclusion to copula based methods of extreme detection. In addition, we have analyzed simulated and real scenarios where the advantages of using different directions to detect extremes is evident. Specifically, *Principal Component Analysis* has been tested as a method to choose a suitable direction of analysis that offers a reasonable number of points identified as extremes. The most significant characteristic is that the locations of those extremes are in a more "*atypical zone*", if one looks at the cloud of observations and its shape. Finally, we have highlighted the advantages and disadvantages of the directional non-parametric approach and the directional copula procedure.

In Chapter 4, we have considered the problem of the estimation of the directional multivariate quantile at high levels of . This chapter deals with the methods of estimation *in-sample* and *out-sample* in the statistical framework, which generate not only theoretical differences but also methodological differences in the estimation process. In Chapter 2 and Chapter 3, we have presented applications based on a non-parametric method of the *in-sample* estimation of the directional multivariate quantile. However, Chapter 4 is focused on the estimation of outside of the observable sampling zone; that is for . We have provided all the necessary tools and hypothesis to formalize a valid estimation. We have also proved the asymptotic normality of the estimator. Finally, the methodology was illustrated with simulated examples for which the theoretical directional multivariate quantiles are known.

5.1 Future Research Lines

We now present some future research lines and extensions of the work presented in this thesis.

- The first point concerns Chapter 2. The definition depends on a center of reference in the space of the data, which implies that a centrality point has to be given. We use the mean of the random vector , but another possibility is to choose a median based on a depth measure. This selection could improve the robustness of the measure and it would be of interest to explore the properties satisfied for this new risk measure.

- A depth measure in terms of the QR *directional orthants* can be defined, which would provide inner and outer regions of the data. However, dimensionality would be a challenge to implement this depth properly.

- In Finance and Insurance different univariate risk measures apart from *Value at Risk* have been introduced because of their properties, advantages and disadvantages. For instance, *CVaR* and *Expected Shortfall*. Therefore, extensions of these measures to the multivariate setting considering the directional framework could be explored.

- An important concept related to Chapter 3 is the return period, which is the expected number of occurrences of an extreme event in a large window of time. This measurement is very important in fields such as environmental sciences, where the impact of mistakes count losses not only in terms of money and resources but also in terms of lives. However, in the multivariate setting it is not straightforward to establishe an unique risk region to work with. Therefore, we consider the directional approach to be useful to define more reliable risk regions that can help to make decisions based on the value of the return period.

- *Principal Component Analysis* (*PCA*) has been tested as a method to choose a suitable direction for extremes analysis, because if one looks at the cloud of observations and its shape then these extremes are located in a more "*atypical zone*". However, it is well known that *PCA* is very sensitive to skewed data, data with heavy univariate tails or outliers. Therefore, it would be interesting to carry out a sensitivity analysis under these scenarios or to consider more appropriate directions selected by other methods such as robust *PCA* (see e.g. Candès et al. (2011)).

- Non-parametric methodologies are important in the multivariate setting. Therefore, it would be of interest to improve the algorithms presented in this thesis using more advanced techniques, such as splines.

- In Economics it is usual to assess the overall state of the economy through an unobserved variable . However, a vector of measurable factors is available to describe and they are usually analyzed by techniques such as dynamic factor models (e.g. Camacho and Perez-Quiros (2010)). After the 2008 economic crisis, risk assessment on variables such as has been matter of discussion, and more importantly, the possibility to perform stress-testing on any model over , (see the new financial and insurance regulations, *Basel III* and *Solvency II*). Therefore, there is a growing interest in dynamic factor techniques including risk factor analysis, which give us the opportunity to develop directional dynamic risk factor analysis.

- Directional multivariate quantile regression is another research line of interest, where the conditional quantile could be completed by adding directions.

- In relation to the characterization of the directional multivariate quantiles at high levels through multivariate extreme value theory, two concepts could have importance: asymptotic dependence and asymptotic independence (see Cai et al. (2011), Wadsworth et al. (2016)). These two concepts have not been dealt with in Chapter 4, but it could be of interest to analyze the implications of the asymptotic dependence/independence in the directional framework.

BIBLIOGRAPHY

Abdous B. and Theodorescu R., *Note on the spatial quantile of a random vector*. Statistics and Probability Letters 13, 333-336, 1992.

AghaKouchak A., Easterling D., Hsu K., Schubert S. and Sorooshian S., *Extremes in a Changing Climate: Detection, Analysis and Uncertainty*. Springer-Verlag: New York, 2013.

Agostinelli C. and Romanazzi M., *Local depth*. Journal of Statistical Planningand Inference 141, 817-830, 2011.

Anderson T., *Some nonparametric multivariate procedures based on statistically equivalent blocks*. Krishnaiah 1, 5-27, 1966.

Andrews D.F., *Plots of high-dimensional data*. Biometrics 28, 125-136, 1972.

Arbia, G., *Bivariate value at risk*. Statistica LXII, 231-247, 2002.

Arslan O., *Family of multivariate generalized t distributions*. Journal of Multivariate Analysis 89(2), 329-337, 2004.

Artzner P., Delbaen F., Heath J. Eber D., *Coherent measures of risk*. Mathematical Finance 3, 203-228, 1999.

Balbas A., Balbas R. and Jimenez-Guerra P., *Vector risk functions*. Mediterranean Journal of Mathematics 9, 563-574, 2012.

Bargès M., Cossette H., and Marceau È. *TVaR-based capital allocation with copulas*. Insurance: Mathematics and Economics 45(3), 348-361, 2009.

Barnett V., *The ordering of multivariate data (with discussion)*. JRSS, Series A 139, 318-354, 1976.

Bernard C., Jiang X. and Wang R., *Risk aggregation with dependence uncertainty*. Insurance: Mathematics and Economics 54(1), 93-108, 2014.

Bras R., *Hydrology: an introduction to hydrologic science*. Addison-Wesley, 1990.

Burgert C. and Ruschendorf L., *Consistent risk measures for portfolio vectors*. Insurance: Mathematics and Economics 38, 289-297, 2006.

Belzunce F., Castaño A., Olvera-Cervantes A. and Suárez-Llorens A., *Quantile curves and dependence structure for bivariate distributions*. Computational Statistical Data Analysis 51, 5112-5129, 2007.

Binois M., Rullieèrec D. and Roustant O., *On the estimation of Pareto fronts from the point of view of copula theory*. Information Sciences 324, 270-285, 2015.

Breckling J. and Chambers R., *M-quantiles*. Biometrika 75(4), 761-771, 1988.

Breckling J., Kokic P. and Lübke O., *A note on multivariate M-quantiles*. Statistics Probability Letters 55, 39-44, 2001.

Brown B.M. and Hettmansperger T.P., *Affine invariant rank methods in the bivariate location model*. Journal of the Royal Statistical Society, Series B 49, 301-310, 1987.

Brown B.M. and Hettmansperger T.P., *An affine invariant bivariate version of the sign test*. Journal of the Royal Statistical Society, Series B 51, 117-125, 1989.

Cai J., Einmahl J., and De Haan L., *Estimation of extreme risk regions under multivariate regular variation*. Annals of Statistics 39(3), 1803-1826, 2011.

Cai J., Einmahl J., de Haan L., and Zhou C. *Estimation of the marginal expected shortfall: the mean when a related variable is extreme*. Journals of the Royal Statistical Society. Series B. Statistical Methodology, 77(2), 417-442, 2015.

Camacho M. and Perez-Quiros G. *Introducing the euro-sting: Short-term indicator of euro area growth*. Journal of Applied Econometrics, 25(4), 663-694, 2010.

Candès E. J., Li X., Ma Y. and Wright J. *Robust principal component analysis?*. Journal of the ACM (JACM), 58(3), 11, 2011.

Capéraà P., Fougères A. and Genest C., *A nonparametric estimation procedure for bivariate extreme value copulas.* Biometrika 84(3), 567-577, 1997.

Cardin M. and Pagani E., *Some classes of multivariate risk measures*. Mathematical and Statistical Methods for Actuarial Sciences and Finance, 63-73, 2010.

Cascos I., López Á. and Romo J., *Data Depth in Multivariate Statistics*. Boletin de Estadística e Investigación Operativa, CEIO 27(3), 151-174, 2011.

Cascos I. and Molchanov I., *Multivariate risks and depth-trimmed regions*. Finance and stochastics 11(3), 373-397, 2007.

Cascos I. and Molchanov I., *Multivariate risk measures: a constructive approach based on selections*. Mathematical Finance 0(0), 1-34, 2014.

Chaudhuri P., *On a geometric notion of quantiles for multivariate data*. Journal of the American Statistical Association 91, 862-872. 1996.

Chakraborty B., *On multivariate quantile regression*. J. Statist. Plann. Inference 110, 109-132, 2003.

Chebana F. and Ouarda T., *Index flood-based multivariate regional frequency analysis*. Water Resour. Res. 45, W10435, 2009.

Chebana F. and Ouarda T., *Multivariate extreme value identification using depth functions*. Envirometrics 22, 441-455, 2011.

Chebana F. and Ouarda T., *Multivariate quantiles in hydrological frecuency analysis*. Envirometrics 22, 63-78, 2011.

Corbella S. and Stretch D., *Simulating a multivariate sea storm using Archimedean copulas*. Coastal Engineering 76, 68-78, 2013.

Chow V., Maidment D. and Mays L., *Applied Hydrology*, McGraw-Hill, Singapore, 1988.

Cousin A. and Di Bernardino E., *On multivariate extensions of Value-at-Risk*. Journal of Multivariate Analysis 119, 32-46, 2013.

Daníelsson J., Jorgensen B., Samorodnitsky G., Sarma M. and De Vries C., *Fat tails, VaR and subadditivity*. Journal of econometrics 172(2), 283-291, 2013.

David H., *Concomitants of order statistics*. Bull. Int. Statist. Inst. 45, 295-300, 1973.

De Haan, L. and Ferreira, A., *Extreme Value Theory, An Introduction*. Springer, New York, 2006.

De Haan L., and Huang X., *Large Quantile Estimation in a Multivariate Setting*, Journal of Multivariate Analysis, 53, 247-263, 1995.

de Haan L. and Resnick S. *Limit theory for multivariate sample extremes*. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 40(4), 317-337, 1977.

De Michele C., Salvadori G. *A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas*. Journal of Geophysical Research 108, D2, 4067, 2003.

De Michele C., Salvadori G., Canossi M., Petaccia A. and Rosso R., *Bivariate statistical approach to check adequacy of dam spillway*. Journal of Hydrologic Engineering 10, 50-57, 2005.

De Michele C. and Salvadori G,. *Statistical characterization of temporal structure of storms*. Advances in Water Resources 29, 827-842, 2006.

De Michele C., Salvadori G., Passoni G. and Vezzoli R., *A multivariate model of sea storms using copulas*. Coastal Engineering 54, 734-751, 2007.

De Michele C., Salvadori G., Vezzoli R. Pecora S., *Multivariate assessment of droughts: Frequency analysis and dynamic return period*. Water Resour. Res. 49, 6985-6994, 2013.

Dekkers, A. Einmahl J. and de Haan L., *A moment estimator for the index of an extreme-value distribution*. Annals of Statistics 17, 1833-1855, 1989.

Di Bernardino E., Laloe T., Maume-Deschamps V. and Prieur C. (2011). *Plug-in estimation of level sets in a non-compact setting with applications in multivariate risk theory*. ESAIM: Probability and Statistics, 17, 236-256.

Di Bernardino E. and Prieur C., *Estimation of Multivariate Conditional Tail Expectation using Kendall's Process*. Journal of Nonparametric Statistics 26(2), 241-267, 2014.

Di Bernardino E., Fernández-Ponce J., Palacios-Rodríguez F., Rodíguez-Griñolo M., *On multivariate extensions of the conditional Value-at-Risk measure*. Insurance: Mathematics and Statistics 61, 1-16, 2015.

Bernardino E. and Palacios-Rodríguez, F. *Estimation of extreme quantiles conditioning on multivariate critical layers*. Environmetrics 27(3), 158-168, 2016.

Donoho D. L. *High-dimensional data analysis: The curses and blessings of dimensionality*. AMS Math Challenges Lecture, 1-32, 2000.

Driver, B., *Analysis Tools with Applications*. E-Book. Springer, Berlin, 2003.

Dyckerhoff R., *Computing zonoid trimmed regions of bivariate data sets*. In COMPSTAT 2000: Proceedings in Computational Statistics (ed. Bethelehem and Heijden), Physica-Verlag, Heidelberg, 295-300, 2000.

Durante F. and Salvadori G., *On the construction of multivariate extreme value models via copulas*. Envirometrics 21, 143-161, 2010.

Efron B., *The convex hull of a random set of points*. Biometrika 52, 331-343, 1965.

Einmahl J., de Haan L. and Krajina A., *Estimating extreme bivariate quantile regions*. Extremes 16(2), 121-145, 2013.

Einmahl, J. H. and Mason, D. M., *Generalized quantile processes.* Annals of Statistics 20, 1062-1078, 1992.

Embrechts P. and Puccetti G., *Bounds for functions of multivariate risks*. Journal of Multivariate Analysis 97, 526-547, 2006.

Embrechts P. Puccetti G., Rüschendorf L., Wang R. and Beleraj A., *An Academic Response to Basel 3.5*. Risks 2(1), 25-48, 2014.

Fernández-Ponce J. and Suárez-Llorens A., *Central regions for bivariate distributions*. Austrian Journal of Statistics 31, 141-156, 2002.

Fraiman R. and Pateiro-López B., *Quantiles for finite and infinite dimensional data*. Journal Multivariate Analysis 108, 1-14, 2012.

Frees E. and Valdez E., *Understanding relationships using copulas*. North Amer. Actuar. J. 2, 1?25, 1998.

Galambos J., *On the distribution of the maximum of random variables*. Ann. Math. Statist. 43, 516-521, 1972.

Girard S. and Stupfler G., *Extreme geometric quantiles in a multivariate regular variation framework*. Extremes 18, 629-663, 2015.

Gonzales-Rivera G., *Value in stress: A coherent approach to stress testing*. The Journal of fixed Income, 7-17, 2003.

Grimaldi S. and Serinaldi F., *Asymmetric copulas in multivariate flood frequency analysis*. Adv. Water Resour. 29, 1155-1167, 2006.

Grimaldi S. and Serinaldi F., *Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation*. Hydrol. Earth Syst. Sci. 17, 1281-1296, 2013.

Gupta S. and Manohar C. *Multivariate extreme value distributions for random vibration applications*. Journal of engineering mechanics 131(7), 712-720, 2005.

Hallin M., Paindaveine D. and Šiman M., *Multivariate quantiles and multiple-output regression quantiles: from optimization to halfspace depth*. Annals of Statistics 38, 635-669, 2010.

Hamel A. and Heyde F., *Duality for set-valued measures of risk*. SIAM Journal of Financial Mathematics 1, 66-95, 2010.

He Y. and Einmahl J., *Estimation of extreme depth-based quantile regions*. J. R. Stat. Soc. B., doi:10.1111/rssb.12163, 2016.

Heffernan J. and Tawn J., *A conditional approach for multivariate extreme values*. Journal of the Royal Statistical Society. Series B: Statistical Methodology 66(3), 497-530 2004.

Hegerl G., Zwiers F., Kharin S. and Stott, P. *Detectability of anthropogenic changes in temperature and precipitation extremes*. J. Climate 17, 3683-3700, 2004.

Hettmansperger T., Nyblom J. and Oja H., *On multivariate notions of sign and rank*. Y. DODGE (ed.), L1-Statistical analysis and related methods, North-Holland, Amsterdam, 267-278, 1992.

Heyde C., Kou S. and Peng X., *What is a good external risk measure: bridging the gaps between robustness, subadditivity, and insurance risk measure*. (Preprint), 2009.

Hill B., *A simple general approach to inference about the tail of a distribution*. Ann. Statist 3,1163-1174, 1975.

Horn R. and Johnson C., *Matrix Analysis*. Cambridge University Press, 2nd Ed., 2013.

Hult, H. and Lindskog, F., *Multivariate extremes, aggregation and dependence in elliptical distributions*. Advances in Applied Probability 34, 587-608, 2002.

Jessen A. and Mikosh T., *Regularly varying functions*. Publications de l'institute mathematique 79, 2006.

Jouini E., Meddeb M. and Touci N., *Vector-valued coherent risk measures*. Finance Stochastics 8, 531-552, 2004.

Kim T., Valdés J. and Yoo C., *Nonparametric approach for estimating return periods of droughts in arid regions*. Journal of Hydrologic Engineering 8, 237-246, 2003.

Koenker R. and Basset G., *Regression Quantiles*. Econometrica 46, 33-50, 1978.

Koltchinskii V., *M-estimation, convexity and quantiles*. Annals of Statistics 25(2), 435-477, 1997.

Kong L. and Mizera I., *Quantile tomography: Using quantiles with multivariate data*. ArXiv preprint arXiv:0805.0056v1, 2008.

Kong L. and Mizera I., *Quantile tomography: Using quantiles with multivariate data*. Statistica Sinica 22(4), 1589-1610, 2012.

Kotz, S. and Nadarajah, S. *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London, 2000.

Kreimerman J., *A bivariate test of goodness of fit based on a gradually increasing number of order statistics*. Tech. Report No. 250, Department of Operations Research, College of Engineering, Cornell University. 1975.

Laniado H., *Extremality in Multivariate Statistics*, Ph.D. Thesis, Universidad Carlos III de Madrid, 2012.

Laniado H., Lillo R., Pellerey F. and Romo J., *Portfolio selection through an extremality stochastic order*. Insurance: Mathematics and Economics, 51, 1-9, 2012.

Laniado H., Lillo R. and Romo J., *Multivariate extremality measure*. Working Paper, Statistics and Econometrics Series 08, Universidad Carlos III de Madrid, 2010.

Lee J. and Prékopa A., *Properties and calculations of multivariate risk measures: MVaR and MCVaR*. Rutcor Research Report 25, 2012.

Liu, R., *On a notion of data depth based on random simplices*. Annals of Statistics 18, 405-414, 1990.

Liu R. and Singh K., *A quality index based on data depth and multivariate rank tests*. J. Amer. Statist. Assoc. 88, 252-260, 1995.

Longin F., *From value at risk to stress testing: The extreme value approach*. Journal of Banking and Finance 24, 1097-1130, 2000.

Marshall, A. W. and Olkin, I. *Domains of attraction of multivariate extreme value distributions*. Ann. Probab., 11, 168-177, 1983.

Matthews K. and Taylor M., *Nonparametric Methods for Multivariate Analysis Using Statistically Equivalent Blocks*. Army Research Laboratory. 1996.

Mikosch T., *Modeling dependence and tails of financial time series*. In: Finkenstädt, B., Rootzén, H. (eds.) Extreme Values in Finance, Telecommunications, and the Environment, 187-286. Chapman Hall, Boca Raton 2003.

Mukhopadhyay N. and Chatterjee S., *High dimensional data analysis using multivariate generalized spatial quantiles*. Journal of Multivariate Analysis 102 (4), 768-780, 2011.

Nappo G. and Spizzichino F., *Kendall distributions and level sets in bivariate exchangeable survival models*. Information Sciences 179, 2878-2890, 2009.

Nelsen R., *An Introduction to copulas*. Springer-Verlag: New York, 2nd Ed., 2006.

Nikoloulopoulos A., Joe H. and Li H., *Extreme value properties of multivariate t copulas*. Extremes 12, 129-148, 2009.

Pickands J. *Statisticalinferenceusingextremeorderstatistics*. Annals of Statistics 3,119-131, 1975.

Rachev S., Ortobelli S., Stoyanov S. and Fabozzi F., *Desirable Properties of an Ideal Risk Measure in Portfolio Theory*. International Journal of Theoretical and Applied Finance 11(1), 19-54, 2008.

Resnick S., *Extreme values, regular variation and point process*. Springer-Verlag, 1987.

Resnick, S., *Hidden regular variation, second order regular variation, and asymptotic independence*. Extremes, 5, 303-336, 2002.

Resnick, S. *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York, 2007.

Rockafellar R.T., *Convex analysis*. Princeton Mathematical Series. Princeton University Press, 1997.

Rousseeuw P.J. and Ruts I., *Bivariate location depth*. Applied Statistics 45, 516-526, 1996.

Rousseeuw P.J. and Ruts I., *Computing depth contours of bivariate points clouds*. Computational Statistics and Data Analysis 23, 153-168, 1996.

Salvadori, G., *Bivariate return periods via 2-copulas*. Statist. Methodol. 1, 129-144, 2004.

Salvadori G. and De Michele C., *Frequency analysis via copulas: theoretical aspects and applications to hydrological events*. Water Resour. Res. 40 (W12511), 2004.

Salvadori G., De Michele C., Kottegoda N. and Rosso R., *Extremes in Nature: An Approach Using Copulas*. Springer-Verlag, 2007.

Salvadori G., De Michele C. and Durante F., *Multivariate design via copulas*. Hydrology and Earth System Sciences Discussions 8, 5523-5558, 2011.

Salvadori G., De Michele C. and Durante F., *On the return period and design in a multivariate framework*. Hydrol. Earth Syst. Sci. 15, 3293-3305. 2011.

Salvadori G., Durante F. and De Michele C., *Multivariate return period calculation via survival functions*. Water Resour. Res., 49, 2308-2311, 2013.

Salvadori G., Durante F., De Michele C., Bernardi M., and Petrella L., *A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities*, Water Resour. Res., 52, doi:10.1002/2015WR017225, 2016.

Salvadori G. and De Michele C., *Multivariate real-time assessment of droughts via copula-based multi-site Hazard Trajectories and Fans*, Journal of Hydrology 526, 101-115, 2015.

Serfling R., *Quantile functions for multivariate analysis: approaches and applications*. Statistica Neerlandica 56, 214-232, 2002.

Serfling R., *Generalized quantile processes based on multivariate depth function, with applications in nonparametric multivariate analysis*. Journal of Multivariate Analysis 83, 232-247, 2002.

Shaked M. and Shanthikumar J., *Stochastic Orders*. Springer Series in Statistics, 2007.

Shiau J., *Return period of bivariate distributed extreme hydrological events*. Stochastic Environmental Research and Risk Assessment 17, 42-57, 2003.

Sibuya M., *Bivariate extremal statistics*. I. Ann. Inst. Stat. Math., Tokyo, 11, 195-210, 1960.

Sklar, A., *Fonctions de Repartition á n-dimensions et Leur Marges*. Publ. Inst. Statist. Univ. París.8, 229-231, 1959.

Taboga M., 2012. Lectures on Probability Theory and Mathematical Statistics. CreateSpace Amazon.

Tibiletti L., *Quasi-concavity property of multivariate distribution functions*. Ratio Mathematica 9, 27-36, 1995.

Tibiletti, L., *On a new notion of multidimensional quantile.* Metron 51, 77-83, 1993

Tibiletti L., *Incremental value at risk: traps and misinterpretations*. Birkhauser Verlag, Basel (Switzerland), 355-364, 2001.

Torres R., Lillo R. and Laniado H., *A Directional Multivariate Value at Risk*. Insurance: Mathematics and Economics 65, 111-123, 2015.

Torres R., De Michele C., Laniado H. and Lillo R. *Directional Multivariate Extremes in Environmental Phenomena*, Environmetrics in press, 2016.

Tukey, J.W., *Mathematics and the picturing of data*. In R.D. James editor, Proceedings of the International Congress of Mathematics (Vancouver, 1974), Volume 2, 523-531, 1975.

Wang R., Peng L. and Yang J., *Bounds for the sum of dependent risks and worst value at risk with monotone marginal densities*. Finance and Stochastics 17(2), 395-417, 2013.

Wadsworth J., Tawn J., Davison A., and Elton D. *Modelling across extremal dependence classes*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), DOI: 10.1111/rssb.12157, 2016.

Yue S., Ouarda T., Bobée B., Legendre P. and Bruneau P., *The Gumbel mixed model for flood frequency analysis*. Journal of Hydrology 226, 88-100, 1999.

Zhou C., *Why the micro-prudential regulation fails? The impact on systemic risk by imposing a capital requirement*. Netherlands Central Bank, Research Department, 2010.

Zhang L. and Singh V., *Bivariate flood frequency analysis using the copula method*. Journal of Hydrologic Engineering 11, 150-164, 2006.

Zuo Y. and Serfling R., *General notions of statistical depth function*. Annals of Statistics 28(2), 461-482, 2000.