SAPIENZA
UNIVERSITÀ DI ROMA

Dependence Methods For Financial Time Series With Application To Portfolio Diversification

Scuola di Dottorato in Economia

Dottorato di Ricerca in Matematica per le Applicazioni Economico-Finanziarie – XXVI Ciclo

Candidate

Hao Wang

ID number 1378359

Thesis Advisor

Prof. Fabrizio Durante

Coordinator

Prof. Maria B. Chiarolla

A thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics for Economic-Financial Applications

February 2014

Thesis defended on March 14, 2014
in front of a Board of Examiners composed by:

Prof. Sergio Scarlatti (chairman)

Prof. Sabrina Mulinacci

Prof. Massimo Costabile

Dependence Methods For Financial Time Series With Application To Portfolio Diversification
Ph.D. thesis. Sapienza – University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Version: February 26, 2014

Website: https://sites.google.com/site/haowang201011/

Author's email: hao.m.wang@gmail.com

Dedicated to
my parents

Abstract

In order to increase the diversification benefits in a bear market, we provide a two-stage portfolio selection algorithm via clustering on the basis of tail dependence. By exploiting two tail dependence-based risky measures, i.e. tail dependence coefficient and conditional Spearman's correlation, a first-step cluster analysis (hierarchical and fuzzy) is carried out for discerning between assets with the same performance during risky scenarios. After that a mean-variance efficient frontier is computed by fixing a number of assets per portfolio and by selecting only one item from each cluster. Empirical calculations on the EURO STOXX 50 index and its components prove that investing on some index components in troubled periods may improve the risk-averse investor portfolio performance significantly.

Moreover, we also propose a novel spatial contagion measure that is based on the calculation of suitable conditional Spearman's correlations extracted from the financial time series of interest. Algorithms for the numerical estimation of this measure are illustrated, together with a simulation study showing its features in relation to popular families of copulas. Finally, an illustration is presented, by showing how the symmetric contagion measure can be used to create clusters of the financial time series. The presented methodology is expected to be useful for the selection of a diversified portfolio of asset returns.

Key words: Risk Management, Portfolio Selection, Tail Dependence, Contagion, Financial Time Series, Cluster Analysis, Copulas.

Sunto

La tesi dottorale discute problemi relativi alla dipendenza tra i titoli di un portafoglio finanziario con possibili applicazioni al processo di selezione di un portafoglio di titoli che sia “diversificato”, ossia che sia composto da titoli che possono reagire diversamente al verificarsi di uno stesso (negativo) evento finanziario. Il contributo originale della tesi è costituito da due argomenti principali.

In primo luogo, al fine di aumentare i benefici di diversificazione in un mercato in crisi, proponiamo un algoritmo di selezione dei portafogli che sfrutta metodi di clustering basati sulla dipendenza di coda. L’argoritmo opera in due fasi. Nella prima fase, sfruttando due misure di rischio basate sulla dipendenza di coda, cioè il coefficiente di dipendenza di coda e la correlazione di Spearman condizionale, viene effettuata un’analisi di cluster (hierarchical e fuzzy) per discernere tra gli assets con le stesse prestazioni in scenario di rischio. Nella seconda fase, fissando il numero dei titoli per portafoglio e selezionando solo un elemento da ogni cluster, si calcola una frontiera efficiente convessa di media-varianza. Calcoli empirici sembrano indicare che investire su titoi diversificati secondo la procedura descritta può migliorare la performance del portafoglio degli investitori avversi al rischio.

In secondo luogo, si propone inoltre una nuova misura di contagio spaziale che si basa sul calcolo di adeguate correlazioni di Spearman condizionali estratte dalle serie storiche finanziarie di interesse. Paragonata alle precedenti misure di correlazione di coda, tale concetto consente di individuare se la dipendenza tra i titoli cambia in periodi crisi. Si presentano algoritmi per la stima numerica di questa misura insieme a uno studio di simulazione che mostra le sue caratteristiche in diverse condizioni di dipendenza (come descritte da copule). Alcuni illustrazioni presentano l’usa della misura di contagio per la creazione di cluster di serie storiche finanziarie e la selezione di un portafoglio diversificato di titoli finanziari.

Parole chiave: Gestione del Rischio, Selezione del Portafoglio, Dipendenza di Coda, Contagio, Serie Storiche Finanziarie, Clustering, Copula

Acknowledgments

With a lot of help, I have almost completed the transformation from a PhD student to a PhD. In three years, I have obtained not only knowledge, but also the courage and friendship. For me, the process of obtaining PhD seems like a long journey. Although you would prepare well before setting off, you never know what will happen during the journey. Now, on the occasion of the arrival to my destination, I want to acknowledge my family members and friends who gave me directions, helped during the difficult times, encouraged me going forward, and traveled with me.

First of all, I take immense pleasure to express my sincere gratitude to my supervisor, ***Dr. Fabrizio Durante****. Thank you for your permission to be my mentor. You are always gentle, kind, friendly, and available for my questions. You guided me through the research selflessly, supported and encouraged me, taught and helped me grow up.*

In addition, I would like to offer my sincere thanks to, ***Prof. Maria B. Chiarolla****, the coordinator of the PhD school of Mathematics for Economic-Financial Applications. Thank you for presenting me and helping me find my supervisor, and helping me during the whole three years of PhD life. I also want to thank* ***Dr. Emanuele Vannucci****,* ***Prof. Lucia Caramellino****,* ***Prof. Wolfgang J. Runggaldier****, and* ***Prof. Emilio Barone****, for your invaluable advice in helping me search the supervisor.*

I sincerely acknowledge my co-authors, ***Prof. Dr. hab. Piotr Jaworski****,* ***Dr. Enrico Foscolo****, and* ***PhD Roberta Pappatà****. With your valuable cooperation, I was able to develop my work and publish it. Moreover, I want to express my profound thanks to* ***Prof. Dr. hab. Piotr Jaworski*** *for accepting me to work with you for five month at University of Warsaw. The work of that period was very helpful for my following.*

In addition, I thank to the professors of the department and my colleagues. ***Prof. R. Manca****,* ***Prof. F. Tardella****, and* ***Prof. G. Bruno****. Thank you for your helpful advices.* ***Ilaria*** *and* ***Giovanni****, you gave me much help in my study. Particularly,* ***Giovanni****, you also gave me many useful suggestions on other things. I was very happy to work with you in the same room.* ***Sofia****,* ***Elisa****, and* ***Naeem****, we started together, encouraged each other during the journey and worked hard for the destination. Thank you for accompanying me on the trip.* ***Tiziano****,* ***Giorgio****,* ***Antonio****, and* ***Isabella****, your experience helped me avoid a lot of detours.* ***Laura****,* ***Denis****, and* ***Armando****, I am very happy to study with you.*

I also thank all the teachers of my school days, graduation and post-graduation for laying a solid foundation.

With great pleasure I express my sincere gratitude to the family of ***Oliveri*** *and of* ***Di Stefano****. You gave the warmth of home while I studied in a foreign land alone; it is you who gave me so much help and got me out of trouble again and again; it is you who offered me a lot of suggestions to guide me ahead.*

I also thank ***Prof. Jeff Cole*** *and* ***Eng. Christian*** *for the correction of my English and Italian grammatical errors, respectively. More over, I thank* ***Francesco Biccari*** *for developing the LATEX class sapthesis.cls.*

I am grateful to my dear friends, who gave me much selfless help and encouragement.

There may be so many others whom I may have inadvertently left out. I take this opportunity to thank all of them for their help.

I also thank the ***Sapienza University of Rome*** *for the smoothly running and providing me with a scholarship for the PhD program.*

No words can ever convey my gratitude to my dear father and mother. Without your constant support and encouragement, I could never have the strength to finish such a long journey.

Hao Wang

Contents

Chapter 1

Introduction and Motivation

Portfolio is a significant issue in risk management. In a general case, one goal of a risk manager is to find an optimal investment strategy that minimizes the total risk of the portfolio. A common practice for minimizing the whole risk consists of adopting some diversification techniques, an issue widely debated since the seminal work of [Grubel, 1968]. Namely, it has been recognized that investors can reduce the risk of their portfolios through allocating their investments in various classes of financial instruments and/or categories of assets that would move in different ways in response to the same event. In other words, diversification benefits can be achieved when the comovements among the assets are taken into account. Therefore, portfolio diversification issue naturally poses the question of investigating the relationship between financial time series and of checking whether they can be grouped together in a way that may be helpful to portfolio selection. Thus, in order to implement diversification, we need to think about the pairwise relationships among the assets and whether some of the assets can be grouped together according to their behaviors.

Usually, in a one-period portfolio model, we assume that the returns of the assets are represented by random variables. So, the relationship between two assets may be completely described by their joint probability distribution. However, the estimation of the bivariate joint distribution is not an easy task, although the marginal distributions could be estimated quite well by various methods: nonparametric, semi-parametric, or fully parametric. In fact, extensive empirical literature has found that asset returns differ from a Gaussian property. See, for instance, [Cont, 2001]. Thus, the standard modeling approach based on properties of the multivariate normal distribution often fails to reproduce some stylized facts (i.e., fat tails, asymmetry) observed in returns from financial assets.

In the 1950s, a celebrated theorem appeared in [Sklar, 1959] which made studies toward resolving this problem. Sklar's Theorem stated that the joint distribution function of a multivariate random vector can be decomposed in terms of its marginal distributions and an auxiliary function that describes the relationships among the variables. This function is called *copula*. A more precise definition will be introduced in the following chapter. In the decades after Sklar's 1959 work, the major part of copula theory was developed in the framework of Probabilistic Metric Spaces ([Schweizer and Sklar, 1983]), and gradually applied to study the dependence among random variables ([Schweizer and Wolff, 1981]). Moreover, several conferences on probability distributions with given marginal in the 1990s made the copula concept well known in the academic world. Starting from the end of the 1990s, copula became also popular in various applied fields, especially in finance, insurance, and risk management ([Embrechts, 2009]). For example, the Archimedean copula families are widely used in the risk management because of their heavy tailed property ([Joe, 1997, McNeil et al., 2005, Nelsen, 2006]). In fact, when we discuss the relationships between two variables, we are often interested in the

dependence between them in the tail of the distribution, that is the region where they have extreme values.

Traditionally, the dependence among different variables was described by numerical measures. The most common and basic dependence measure is *correlation coefficient* (or *Pearson's correlation coefficient* or *Pearson's rho*), which is the ratio of the covariance to the square root of the product of two variances. For its definition, it is invariant under linear transformation but is not comonotone-invariant. In fact, it is a linear correlation measure that only works well in the Gaussian world, or more generally, the elliptical world. In fact, it has been long recognized that the correlation coefficient cannot be used to measure the nonlinear correlations.

Other dependence measures are given by, for instance, *Kendall's tau* and *Spearman's rho*, developed in [Kendall, 1938] and [Spearman, 1904], respectively. They can play the rule of nonlinear correlation measure. These two measures are so-called *rank correlation measures*, because they are defined on the basis of the rank of the observations. Therefore, the rank correlation measures remain unchanging under the strictly increasing transformation. Interestingly, Kendall's tau and Spearman's rho do only depend on the copula between the two variables, and not on the marginal distributions. For some well-known and most-used one-parameter copula families, e.g. Gaussian, t, Frank, Clayton, Gumbel, these two rank correlation measures can be computed as the functions of the parameters of the copula, a fact that is very useful both in estimation and in simulation.

As noted previously, for the purpose of risk management, we are more interested in the behavior when they have extreme values, especially extreme losses. The extreme events are "very rare but very damaging" ([Bradley and Taqqu, 2003]). In fact, an extreme loss could make a financial institution insolvent, even in bankruptcy. Therefore, the need to cover the risk of extreme losses develops standard ways to quantify such impacts, as already written in Basel II ([BCBS, 2006]), and enhanced in Basel III ([BCBS, 2011]). So, in order to construct the portfolios, we need not only the information about the dependence among the assets, but also some specific tail dependence measures, for instance (as done in this thesis) *tail dependence coefficient* and *Spearman's rank correlation conditional on the tail*..

The tail dependence coefficient between two variables is the limit of the probability of one variable having the extreme values given the other one having the extreme values. For some one-parameter copula family, there is a closed formula for computing the tail dependence coefficients. It also does depend on the copula and not on the marginal distributions. While, conditional Spearman's rank correlation between two variables is simply the Spearman's rho when both variables assume extreme values. It depends on the tail copula, which is the copula of the tail part of the variables. Because this tail copula is too difficult and too complex to obtain, as a result this measure does not have the closed formula generally. The good news is that both these two tail dependence measures can be estimated non-parametrically.

When the random variables of interest are financial assets (often representing international markets, like market indices), another interesting phenomena is the contagion. Traditionally, this term was related to the literature about the behavior of emerging markets ([Hunter et al., 1999]). However, in recent years especially after the recent financial crisis (i.e. 1997 Asian crisis, the so-called 2008 Subprime mortgage crisis, and 2011 Sovereign debt crisis), the concept was employed at large. Roughly speaking, contagion refers to the fact that the global financial system is not stable anymore. A local-national financial crisis would affect other nations with "apparently healthy fundamentals and whose policies, 'only' a few months earlier, has been praised by market analysts and the multilateral institutions" ([Edwards, 2000, P.1]). Recognizing and managing the presence of contagion provide useful benefits when dealing with the financial risks, but on the definition and detection of contagion there is still no agreement. Most parts of the literature describes contagion as an increment of the correlation coefficient between markets (see. e.g. the discussion in [Corsetti et al., 2011]). But these tests for

contagion based on correlation coefficients may be problematic. In fact, as noticed in [Ronn, 1998], changes in market volatility can bias the estimate of correlation coefficients and, hence, the related detection of contagion may be wrong. In the early 2000s, a series of papers began to investigate how the bias affects cross-market correlations or the measurement of contagion in more detail.

[Boyer et al., 1999] proposed an adjustment to the correlation coefficient. If the conditional correlation during the turmoil (high volatility) period is greater than the unconditional correlation during the full period, they say that contagion happens between these two markets. In [Forbes and Rigobon, 2002], they argued that the conditional correlation coefficient of [Boyer et al., 1999] is still biased by heteroskedasticity, and a heteroscedasticity corrected version was proposed instead. Another contribution of [Forbes and Rigobon, 2002] was that they were able to distinguish two different phenomena: the interdependence and the contagion among financial markets. But one problem of this test was pointed out by [Bradley and Taqqu, 2004] that "the power of this test is very low due to the short crisis period". Meanwhile, they proposed a different approach to attempt to overcome this problem in [Bradley and Taqqu, 2004, 2005a,b]. In these papers, they considered a definition of contagion based on a local correlation coefficient and provided a test for contagion. In the same spirit of [Bradley and Taqqu, 2004], [Durante and Jaworski, 2010] defined and investigated a related notion of *spatial contagion* between two financial markets, by describing their dependence not by means of the local correlation coefficient, but by using the information contained in the copula of them. Specifically, a practical way to check contagion may be implemented by comparing the Spearman's correlation of the markets when they are experiencing severe losses with that of the tranquil period. The use of Spearman's correlation instead of Pearson's correlation is to detect also nonlinear (yet monotone) dependence among the data (see, e.g., [McNeil et al., 2005]).

As noted previously, different concepts and methodologies have been employed in order to extract information from the financial assets of a portfolio. In particular, a recent literature has focused on the use of such information in order to determine the number of main factors characterizing the dynamics of financial markets and the composition of the groups (clusters) in which a portfolio is intrinsically organized. To this hand, clustering is a common practice in multivariate data analysis. The purpose of clustering analysis is to obtain a meaningful partition of a set of N variables in groups according to their characteristics. In these recent years, some scholars employed this method in finance, e.g. [Mantegna, 1999] and [De Luca and Zuccolotto, 2011]. Usually, the implementation of clustering needs a dissimilarity matrix. Loosely speaking, the components of this dissimilarity matrix describe the "difference" between the assets. A great dissimilarity means that the assets are much "different" in some sense, in contrast, a small one says that they are "similar" ([Kaufman and Rousseeuw, 1990]). For example, in correlation based clustering algorithms the correlation coefficient between two time series is assumed to be a measure of the similarity between the two time series ([Mantegna, 1999, Bonanno et al., 2001, 2003]). Now, once a cluster analysis has been performed on the asset portfolio, the extracted information among the assets may be used in order to reconsider the portfolio optimization under these additional constraints. As explained by [Tola et al., 2008], for instance, the advantage of this preliminary step is that clustering algorithms are quite robust with respect to measurement noise due to the finiteness of sample size, so that they may give a more robust portfolio. Another advantage is that the grouping assets may serve as a genuine way to diversify a portfolio. Advantages of considering portfolio with cluster constraints have been exploited, for instance, by [Tola et al., 2008, Cesarone et al., 2013].

The aim of this thesis is to present some advances in the use of tail dependence measures for portfolio diversification. The main goal of the thesis is two-fold. First, we are interested in considering the effects of cluster analysis to the performance of a diversified portfolio. Here, diversification is

intended to be originated not by classical dependence measures, but by the use of novel tail dependence measures that can work well in crisis periods (Chapter 6). Specifically, we provide a two-stage portfolio selection procedure in order to increase the diversification benefits in a bear market. By exploiting tail dependence-based risk measures, a first-step cluster analysis is carried out for discerning between assets with the same performance during risky scenarios. Then a mean-variance efficient frontier is computed by fixing a number of assets per portfolio and by selecting only one item from each cluster. Empirical calculations on the EURO STOXX 50 show some evidence that investing on selected index components in trouble periods may improve the risk-averse investor portfolio performance.

Secondly, we present novel contribution to the definition and detection of analyze spatial financial contagion (Chapter 7). The new measure goes beyond the use of linear correlation and does not require, in addition, the a priori specification of crisis/non-crisis periods via suitable thresholds (as done for instance in [Durante and Jaworski, 2010]). It is completely data-driven and can be calculated via non-parametric methods. As such, it avoids possible misspecification in the dependence structure. Spatial contagion measure can be calculated empirically via simple procedures and could be also used to identify sub-groups of assets that have similar behavior in crisis periods. In fact, compared with related approaches in the literature, the spatial contagion measure detects the changes in the dependence structure among financial markets and not the presence of a persistent extreme dependence, which is due to interdependence, but not to contagion (for such viewpoint compare with the discussion in [Forbes and Rigobon, 2002]). In addition, the measure of symmetric spatial contagion is used to implement clustering procedures for financial time series, aiming at finding sub-groups of assets that have similar behavior in periods of market distress ([Durante et al., 2014]).

The thesis is organized as follows: Chapter 2-5 present some basic elements from the literature, together with some novel contributions that allow their practical application. The main novel contribution of the thesis is, instead, represented by Chapter 6 and Chapter 7. In Chapter 2, we introduce the basic elements of copula theory. We show history of the concept, its developments, the main results, properties, and theorems. Also, we present the most used families, e.g. Gaussian, t, Frank, Clayton, Gumbel, and the simulation algorithms of these mentioned copula families. This chapter will end with a section about empirical copula. Chapter 3 presents basic dependence concepts. It starts with Pearson's correlation coefficient, then it introduces two rank correlation measures, i.e. Kendall's tau and Spearman's rho, that overcome the shortcomings of Pearson's correlation. We also show the closed formulas for computing these two measures of the most used one-parameter copula families which are exhibited in Chapter 2.

In Chapter 4, we extend the measures listed in Chapter 3 to meet special need of describing the behavior in the tail region of the joint distribution. Specifically, we focus on the tail dependence coefficients and on the conditional Spearman's correlation. In particular, non-parametric estimation algorithm will be presented in detail.

A brief introduction to cluster theory will be presented in Chapter 5. Here, we will show that the data can be expressed in two kinds of matrices, i.e. data matrix and dissimilarity matrix. Another important aspect of cluster theory, i.e. clustering algorithms, are listed in Section 5.3. While, the problem of the selection of the number of clusters will be discussed in Section 5.4. In Chapter 6, we introduce a portfolio selection algorithm via tail dependence based clustering method. Section 6.2 presents the main properties of the cluster procedures that we have used, while Section 6.3 illustrates the methodology via an empirical application. It shows how a convenient diversified portfolio (using tail dependence information) may hedge against the risk of losses in crisis period. A novel spatial contagion measure for financial time series is proposed in Chapter 7. First, we will give an overview of the development about this financial contagion theory. Then, in Section 7.2, we define a new contagion

measure. Some algorithms for calculating this measure are given in Section 7.3. After a simulation study (Section 7.4), we hence present two applications in financial markets.

Chapter 2

Copula Theory: an Introduction

2.1 Introduction

What is "copula"? In linguistics, a copula (plural: copulas or copulae) is "a type of verb, of which the most common is 'be', that joins the subject of the verb with a complement"[1]. The word "copula" derives from the Latin noun for a "link" that connects two different things. Abe Sklar in 1959 used it in a mathematical or statistical sense for the first time in the theorem (which now bears his name) describing the functions that "join together" one-dimensional distribution functions to form multivariate distribution functions [Nelsen, 2006]. In [Sklar, 1996, P.5], he explained the choice of "copula":

> "Féron, in studying three-dimensional distributions had introduced auxiliary functions, defined on the unit cube, that connected such distributions with their one-dimensional margins. I saw that similar functions could be defined on the unit n-cube for all $n \geq 2$ and would similarly serve to link n-dimensional distributions to their one-dimensional margins. Having worked out the basic properties of these functions, I wrote about them to Fréchet, in English. He asked me to write a note about them in French. While writing this, I decided I needed a name for these functions. Knowing the word "copula" as a grammatical term for a word or expression that links a subject and predicate, I felt that this would make an appropriate name for a function that links a multidimensional distribution to its one-dimensional margins, and used it as such. Fréchet received my note, corrected one mathematical statement, made some minor corrections to my French, and had the note published by the Statistical Institute of the University of Paris as [Sklar, 1959]."

Roughly speaking, copulas are functions that join multivariate distribution functions to their univariate marginal distribution functions. A more precise definition will be presented in section 2.3.1.

In the first two decades of the 21 centuries, the studies about copula were implemented widely and deeply, and they have been used in various fields for constructing more flexible multivariate models, like civil engineering [Kilgore and Thompson, 2011], medicine [Onken et al., 2009a,b], weather research [Schoelzel et al., 2008, AghaKouchak et al., 2010], hydrology [Genest and Favre, 2007, Salvadori, 2007], and, in particular, statistics and quantitative finance [Schönbucher, 2003, Cherubini et al., 2004, McNeil et al., 2005, Yannick and Sornette, 2006, Genest et al., 2009].

If you search "copula" in Google Scholar, there are more than 126,000 results[2]. Before 2000, it was difficult to even locate the word "copula" in the statistical literature. As explained in [Nelsen, 2006, P.1]:

> "There is no entry for 'copula' in the nine volume Encyclopedia of Statistical Sciences, nor in the supplement volume. However, the first update volume, published in 1997, does have such an entry [Fisher, 1997]. The first reference in the Current Index to Statistics to a paper using "copula" in the title or as a keyword is in

[1] http://dictionary.cambridge.org/dictionary/british/copula?q=copula

[2] Searched in English, till October 31, 2013.

> Volume 7 (1981) [the paper is [Schweizer and Wolff, 1981]] indeed, in the first eighteen volumes (1975-1992) of the Current Index to Statistics there are only eleven references to papers mentioning copulas. There are, however, 71 references in the next ten volumes (1993-2002).]"

Although the term of "copula" was first used in 1959, Wassily Hoeffding achieved many of the basic results about copulas earlier, including the basic best-possible bounds inequality for copulas and "scale-invariant". The difference is that Hoeffding set the support of the "standardized distributions" contained in $[-1/2, 1/2]^2$ and its margins uniform on $[-1/2, 1/2]$ in [Hoeffding, 1940, 1941]. As [Schweizer, 1991] said, "had Hoeffding chosen the unit square $[0, 1]^2$ instead of $[-1/2, 1/2]^2$ for his normalization, he would have discovered copulas".

Without known the outcome of Hoeffding, [Fréchet, 1951] independently gained many of the same results, which were known as "Fréchet bounds" and "Fréchet classes". In [Nelsen, 2006], to recognize the shared responsibility for these important ideas, Nelsen referred them to "Fréchet-Hoeffding bounds" and "Fréchet-Hoeffding classes".

In the 1980s, most results about copulas were achieved in the framework of the theory of Probabilistic Metric Spaces [Schweizer and Sklar, 1983]. Copulas were gradually applied to study the dependence among random variables [Schweizer and Wolff, 1981]. In Schweizer's words from [Schweizer and Sklar, 1983],

> "After the publication of these articles and of the book ... the pace quickened as more ... students and colleagues became involved. Moreover, since interest in questions of statistical dependence was increasing, others came to the subject from different directions. In 1986 the enticingly entitled article *The joy of copulas* [Genest and Mackay, 1986], attracted more attention."

Two reasons listed in [Fisher, 1997] could explain well why the copulas can be used for studying the "statistical dependence":

> Firstly, as a way of studying scale-free measures of dependence; and secondly, as a starting point for constructing families of bivariate distributions, sometimes with a view to simulation."

From the 1990s until today, a series of conferences or workshops played an important role in spreading of the concept of "copula". In 1990 in Rome, "Probability distributions with given marginals"([Dall'Aglio et al., 1991]), which started this series, was organized by Dall'Aglio. The initials, which were also held in Seattle in 1993 [Rüschendorf et al., 1996], in Prague in 1996 [Benes and Stepan, 1997], in Barcelona in 2000 [Cuadras et al., 2002], were more or less in probability and statistics files. Two books have to be mentioned here: *Multivariate Models and Dependence Concepts* ([Joe, 1997]) and *An Introduction to Copulas*[3]([Nelsen, 1999]). In these two books, the copula theory was presented systematically and completely for the first time. One can find the precise definition of copula function, the important theorem about the existence of the copula function (Sklar 1959), the basic properties for the general copula functions, the different families of the copulas, etc.

However, at end of the nineties, probably because of David Li's model [Li, 1999], copula began to be used outside of probability and statistics, especially in finance. Another two books are necessary to mention: *Copula Methods in Finance* ([Cherubini et al., 2004])[4] and *Quantitative Risk Management: Concepts, Techniques and Tools* ([McNeil et al., 2005]). In these two books, copula as a method was extensively and systematically used to describe the "market comovement" and "tail dependence", and used to price the derivatives and employed in risk management. Embrechts commented on the widespread use of copula theory in [Embrechts, 2009]:

[3] Some new results were added in the second edition [Nelsen, 2006].

[4] A new book-*Dynamic Copula Methods in Fiance* ([Cherubini et al., 2012]) has been published.

> "As we have seen so far, the notion of copula is both natural as well as easy for looking at multivariate d.f.'s. But why do we witness such an incredible growth in papers published starting the end of the nineties (recall, the concept goes back to the fifties and even earlier, but not under that name). Here I can give three reasons: finance, finance, finance. In the eighties and nineties we experienced an explosive development of quantitative risk management methodology within finance and insurance, a lot of which was driven by either new regulatory guidelines or the development of new products; see for instance Chapter 1 in [McNeil et al., 2005] for the full story on the former. Two papers more than any others "put the fire to the fuse": the already mentioned 1998 RiskLab report [Embrechts et al., 2002] and at around the same time, the Li credit portfolio model [Li, 1999]."

From that moment, some conferences about the copula application in finance were organized: in Quebec in 2004 [Genest, 2005a,b], in Tartu in 2007 [Kollo, 2009], in Warsaw in 2009 [Jaworski et al., 2010], in Bolzano-Bozen in 2012 [5], and in Munich in 2013[6].

Now, it is too difficult to list in their entirety all the applications of copula in various fields. As long as you study dependence, copula function is the most direct, practical, simple, and convenient way. As [Schweizer, 2007] wrote:

> "The 'era of i.i.d.' is over: and when dependence is taken seriously, copulas naturally come into play."

However, in these years, some criticisms have been proposed for the copulas and their applications. For example, the very interesting discussion related to [Mikosch, 2006] (see also [de Vries and Zhou, 2006, Embrechts, 2006, Genest and Rémillard, 2006, Joe, 2006]).

From some perspective, this is logical since the copulas have been misused too much. As noted in [Durante and Sempi, 2010],

> "These criticisms were a quite natural reaction to such a wide diffusion of applications of copulas, not always in a well motivated way. It should be said that several people have wrongly interpreted copulas as the solution to "all problems of stochastic dependence". This is definitely not the case! Copulas are an indispensable tool for understanding several problems about stochastic dependence, but they are not the "panacea" for all stochastic models."

Because of the elegance, simplicity and convenience of the copula, we believe that its application in applied mathematics is still in "its infancy"[Nelsen, 2006]. There are many open problems to be resolved. The general goal of this thesis is to develop theory and practice in order to deal with one of these open problems, namely the use of copulas in understanding the link among financial time series.

This chapter is organized in the following way: Section 2.2 presents some basic notions of probability theory that will be used in the sequel. Section 2.3 is in order to grasp the basic idea of a copula and examples. Some useful parametric families of copulas, together with their properties, are presented in Section 2.4. Finally, the chapter ends with a section on copula simulations.

2.2 Preliminaries

Before starting this section, we need to introduce some notations. We will let $\mathbb{R}$ denote the ordinary real line $(-\infty, \infty)$, $\overline{\mathbb{R}}$ denote the extended real line $[-\infty, \infty]$, and for any positive integer d, we let $\overline{\mathbb{R}}^d$ denote the extended d-space $\overline{\mathbb{R}} \times \overline{\mathbb{R}} \times \ldots \times \overline{\mathbb{R}}$. We will use vector notation for points in $\overline{\mathbb{R}}^d$, e.g., $\boldsymbol{a} = (a_1, a_1, \ldots, a_d)$, and we will write $\boldsymbol{a} \leq \boldsymbol{b}$ when $a_k \leq b_k$ for all k; and $\boldsymbol{a} < \boldsymbol{b}$ when $a_k < b_k$ for all k. For $\boldsymbol{a} \leq \boldsymbol{b}$, we will let $[\boldsymbol{a}, \boldsymbol{b}]$ denote the $d - box\ B = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_d, b_d]$, the

[5] Risk and Dependence in Economics and Finance, Bolzano-Bozen, Italy, April 2012 (see details at http://www.unibz.it/en/economics/events/Documents/Workshop_13_04_2012_Final_Program.pdf)

[6] Risk Management Reloaded, München, German, September 2013 (see details at http://www.mathfinance.ma.tum.de/kpmgce/conference-2013/)

Cartesian product of d closed intervals. The *vertices* of an d-box B are the points $\mathbf{c} = (c_1, c_2, \ldots, c_d)$ where each c_k is equal to either a_k or b_k. The unit d-cube $\mathbb{I}^d$ is the product $\mathbb{I} \times \mathbb{I} \times, \ldots, \mathbb{I}$. An *d-place real function* H is a function whose domain, Dom$H$, is a subset of $\overline{\mathbb{R}}^d$ and whose range, Ran$H$, is a subset of $\mathbb{R}$. In order to save space, $\wedge$ is used to denote the function $\min$ of two real numbers, i.e. $x_1 \wedge x_2 = \min\{x_1, x_2\},\ x_1, x_2 \in \mathbb{R}$.

Definition 2.2.1. **(H-volume)** Let $S_1, S_2, \ldots, S_d$ be nonempty subsets of $\overline{\mathbb{R}}$, and let H be an d-place real function such that Dom$H = S_1 \times S_2 \times \ldots \times S_d$. Let $B = [\boldsymbol{a}, \boldsymbol{b}]$ be a d-box, whose vertices are in Dom$H$. Then the *H-volume of B* is given by

$$V_H(B) = \sum \text{sgn}(\mathbf{c}) H(\mathbf{c}),$$

where the sum is taken over all vertices $\mathbf{c}$ of B, and sgn($\mathbf{c}$ is given by

$$\text{sgn}(\mathbf{c}) = \begin{cases} 1, & \text{if } c_k = a_k \text{ for an even number of } k\text{'s}, \\ -1, & \text{if } c_k = a_k \text{ for an odd number of } k\text{'s}. \end{cases}$$

Equivalently, the H-volume of an d-box $B = [\boldsymbol{a}, \boldsymbol{b}]$ is the dth order difference of H on B

$$V_H(B) = \triangle_{\boldsymbol{a}}^{\boldsymbol{b}} H(\boldsymbol{t}) = \triangle_{a_d}^{b_d} \triangle_{a_{d-1}}^{b_{d-1}} \ldots \triangle_{a_1}^{b_1} H(\boldsymbol{t}),$$

where the d first order differences of an d-place function (such as H) is defined as

$$\triangle_{a_k}^{b_k} H(\boldsymbol{t}) = H(t_1, \ldots, t_{k-1}, b_k, t_{k+1}, \ldots, t_d) - H(t_1, \ldots, t_{k-1}, a_k, t_{k+1}, \ldots, t_d).$$

Definition 2.2.2. **(d-increasing)** An d-place real function H is *d-increasing* if $V_H(B) \geq 0$ for all d-boxes B whose vertices lie in Dom$H$.

A *probability space* is a triplet $(\Omega, \mathscr{F}, \mathbb{P})$, where Ω is a nonempty set, $\mathscr{F}$ is a $\sigma - algebra$ of subsets of Ω and $\mathbb{P}$ is a *probability measure* on $\mathscr{F}$. A d-dimensional *random vector* is a measurable mapping $\boldsymbol{X} : \Omega \to \overline{\mathbb{R}}^d$; in this case, the word measurable means that the counter image $\boldsymbol{X}^{-1}(B)$ of every Borel set B in $\mathscr{B}(\overline{\mathbb{R}}^d)$ belongs to $\mathscr{F}$. It can be proved that a random vector $\boldsymbol{X}$ can be represented in the form $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ where, for every $j \in \{1, 2..., d\}$, X_j is a 1-dimensional random vector, also called *random variable* (r.v.).

Definition 2.2.3. (Distribution Function and Generalized Inverse of a Distribution Function) A *distribution function* (d.f.) of a r.v. X defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is a function F with domain $\overline{\mathbb{R}}$ such that

$$F(x) = \mathbb{P}(X \leq x) \tag{2.1}$$

Its *generalized inverse* F^{-1}, is defined, for all $u \in (0, 1)$, by

$$F^{-1}(u) = \inf\{x : F(x) \leq u\}.$$

Proposition 2.2.1. *([McNeil et al., 2005]) Let F be a d.f. of r.v. X and let F^{-1} be its generalized inverse,*

1. ***(Probability transformation).*** *If F is a continuous univariate d.f., then*

$$F(X) \sim U(0, 1).$$

2. ***(Quantile transformation).*** *If $U \sim U(0,1)$ has a standard uniform distribution, then*

$$\mathbb{P}\left(F^{-1}(U) \leq x\right) = F(x).$$

Definition 2.2.4. (Joint Distribution Function) A *joint distribution function* (joint d.f.) of a random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is a function $F_{\boldsymbol{X}}$ with domain $\overline{\mathbb{R}}^d$ such that

$$F_{\boldsymbol{X}}(x_1, x_2, \ldots, x_d) = \mathbb{P}\left(\bigcap_{i=1}^{d}\{X_i \leq x_i\}\right). \tag{2.2}$$

From (2.2), we could get $F_{\boldsymbol{X}}(+\infty, +\infty, \ldots, +\infty) = 1$; and $F_{\boldsymbol{X}}(x_1, x_2, \ldots, x_d) = 0$, if at least one of the arguments equals to $-\infty$.

Theorem 2.2.1. *([Durante and Sempi, 2010, P.8]) The joint d.f. $F_{\boldsymbol{X}}$ of the r.v. $\boldsymbol{X} = (X_1, \ldots, X_d)$ has the following properties:*

1. *$F_{\boldsymbol{X}}$ is increasing in each component, i.e. $F_{\boldsymbol{X}}(\boldsymbol{x}) \leq F_{\boldsymbol{X}}(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \overline{\mathbb{R}}^d$, $\boldsymbol{x} \leq \boldsymbol{y}$;*
2. *for all $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d) \in \overline{\mathbb{R}}_{d-1}$, the function*

$$\mathbb{R} \ni t \mapsto F_{\boldsymbol{X}}(x_1, x_2, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_d)$$

   is right-continuous;

3. ***(d-increasing property)*** *the $F_{\boldsymbol{X}}$-volume $V_{F_{\boldsymbol{X}}}$ of every d-box $[\boldsymbol{a}, \boldsymbol{b}]$ is positive, i.e., $V_{F_{\boldsymbol{X}}}([\boldsymbol{a}, \boldsymbol{b}]) \geq 0$.*

The d-increasing property in the previous theorem simply means that the probability that the r.v. $\boldsymbol{X}$ takes values in the d-box $[\boldsymbol{a}, \boldsymbol{b}]$ is non-negative. We will use the symbol $\boldsymbol{X} \sim F_{\boldsymbol{X}}$ to denote the fact that $F_{\boldsymbol{X}}$ is thd joint d.f. of $\boldsymbol{X}$. Usually, in case of non-ambiguity, the subscript $\boldsymbol{X}$ is omitted so that $F_{\boldsymbol{X}}$ is denoted by F.

Definition 2.2.5. (Marginal Distribution) Let $d \geq 2$ and let F be a d-dimensional joint d.f.. Let $\sigma = (j_1, \ldots, j_m)$ a subvector of $(1, 2, \ldots, d)$, $1 \leq m \leq d-1$. We call *σ-marginal of F* the d.f. $F_\sigma : \overline{\mathbb{R}}^m \to \mathbb{I}$ defined by setting the remaining $d-m$ arguments of F equal to $+\infty$, namely, for every $x_1, \ldots, x_m \in \overline{\mathbb{R}}$,

$$F_\sigma(x_1, \ldots, x_m) = F(y_1, \ldots, y_d)$$

where, for every $j \in \{1, 2, \ldots, d\}$, $y_j = x_j$ if $j \in \{j_1, \ldots, j_m\}$, and $y_j = +\infty$ otherwise.

In particular, when $\sigma = \{j\}$, $F_{(j)}$ is usually called 1-dimensional marginal and it is denoted by F_j. Obviously, if F is the joint d.f. of the r.v. $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$, then the σ-marginal of F is simply the joint d.f. of the subvector $(X_{j_1}, \ldots, X_{j_m})$. As well known, if the r.v.'s $X_1, X_2, \ldots, X_d$ are independent and if F_{X_i} denotes the d.f. of $X_i (i = 1, 2, \ldots, d)$, then the d-dimensional joint d.f. of the random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ is the product of the marginals, i.e. for all $x_1, x_2, \ldots, x_d \in \overline{\mathbb{R}}$,

$$F(x_1, x_2, \ldots, x_d) = \prod_{i=1}^{d} F_{X_i}(x_i).$$

Definition 2.2.6. (Survival Distribution Function) A *survival function* of a r.v. X defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ having d.f. F is a function $\overline{F}$ with domain $\overline{\mathbb{R}}$ such that

$$\overline{F}(x) = \mathbb{P}(X > x).$$

In fact, from (2.1) we get immediately that

$$\overline{F}(x) = 1 - F(x).$$

Definition 2.2.7. (Joint Survival Distribution Function) A *joint survival distribution function* of a random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$ having joint d.f. F is a function $\overline{F}$ with domain $\overline{\mathbb{R}}^d$ such that

$$\overline{F}(x_1, x_2, \ldots, x_d) = \mathbb{P}\left(\bigcap_{i=1}^{d} \{X_i > x_i\}\right). \tag{2.3}$$

In 2-dimensional case, we have

$$\begin{aligned} \overline{F}(x_1, x_2) &= 1 - F_1(x_1) - F_2(x_2) + F(x_1, x_2) \\ &= \overline{F}_1(x_1) + \overline{F}_2(x_2) - 1 + F(x_1, x_2) \end{aligned} \tag{2.4}$$

Note that survival functions are decreasing in each argument.

2.3 Copulas

In this section, the precise definition of copula function will be given. Some important properties and copula families will be also presented. The main part of the definitions and the properties used in this section come from [McNeil et al., 2005, Nelsen, 2006, Durante and Sempi, 2010].

Definition 2.3.1. (Copula) A d-dimensional *copula* is a joint distribution function on $\mathbb{I}^d$ with standard uniform marginal distributions.

We reserve the notation $C(\boldsymbol{u}) = C(u_1, \ldots, u_d)$ for the multivariate d.f.'s which are copulas. Hence C is a mapping from $\mathbb{I}^d$ to $\mathbb{I}$, i.e. a mapping of the unit hypercube into the unit interval. Thus, the following three properties must hold:

1. $C(u_1, \ldots, u_d)$ is increasing in each component u_i;
2. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i \in 1, \ldots, d, u_i \in [0, 1]$;
3. $C(u_1, \ldots, u_d)$ is d-increasing.

By Theorem 2.2.1, we could see that any multivariate joint d.f. requires the first and third property. Meanwihile if the random vector $(U_1, \ldots, U_d)'$ has d.f. C, then

$$C(1, \ldots, 1, u_i, 1, \ldots, 1) = \mathbb{P}\left(\bigcap_{j \neq 1} \{U_j \leq 1\} \bigcap \{U_i \leq u_i\}\right) = \mathbb{P}(U_i \leq u_i) = u_i.$$

So we get the second property. These three properties characterize a copula: if a function C fulfills them, then it is a copula. Note also that, for $2 \leq k < d$, the k-dimensional marginal d.f.'s of a d-dimensional copula are themselves copulas.

Example 2.3.1. Now we introduce three fundamental copulas. For all $\boldsymbol{u}$ in $\mathbb{I}^d$,

$$\textbf{(Independence Copula)} \qquad \Pi_d(\boldsymbol{u}) = \prod_{i=1}^{d} u_i, \tag{2.5}$$

$$\textbf{(Comonotonicity Copula)} \qquad M_d(\boldsymbol{u}) = \min\{u_1, \ldots, u_d\}, \tag{2.6}$$

$$\textbf{(Countermonotonicity Copula)} \qquad W_2(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}. \tag{2.7}$$

Remark 2.3.1. $W_d(\boldsymbol{u}) = \max\{\sum_{i=1}^{d} u_i + 1 - d, 0\}$ is not a copula for $d > 2$ (See [McNeil et al., 2005, Example 5.21]).

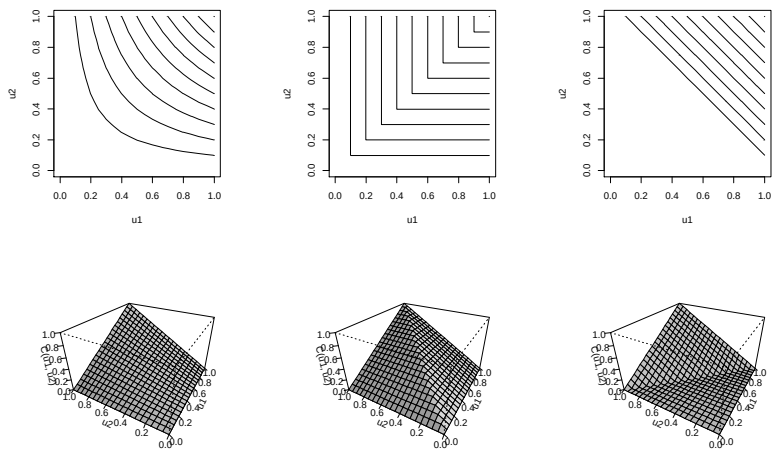


Figure 2.1. (upper) Contour plots and (lower) perspective plots of the three fundamental copulas: (left) independence, (central) comonotonicity and (right) countermonotonicity.

The most important theorem in copula theory is summarized by the following elegant theorem, which guarantees the existence of copulas for every multivariate joint d.f.. It also shows the convenience of copulas in the study of multivariate joint d.f., since copulas can be used in conjunction with univariate d.f.'s to construct multivariate joint d.f.'s.

Theorem 2.3.1. ***(Sklar 1959)*** *Let F be a joint distribution function with marginal d.f.'s $F_1, \ldots, F_d$. Then there exists a copula $C : \mathbb{I}^d \mapsto \mathbb{I}$ such that, for all $x_1, \ldots, x_d$ in $\overline{\mathbb{R}}$,*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{2.8}$$

If the marginal d.f.'s are continuous, then C is unique; otherwise C is uniquely determined on $\mathrm{Ran}F_1 \times \mathrm{Ran}F_2 \times \ldots \times \mathrm{Ran}F_d$, where $\mathrm{Ran}F_i = F_i(\overline{\mathbb{R}})$ denotes the range of F_i. Conversely, if C is a copula and $F_1, \ldots, F_d$ are univariate distribution functions, then the function F defined in (2.8) is a joint distribution function with marginal d.f.'s $F_1, \ldots, F_d$.

Notice that, Sklar's theorem has been announced in [Sklar, 1959], however its first proof for the bivariate case appeared in [Schweizer and Sklar, 1974]. See [McNeil et al., 2005, P.187] for the proof.

In fact, we could derive a copula C from any d-variate joint d.f. by using (2.8). When F_i is continuous for every $i \in \{1, 2, \ldots, d\}$, a unique C can be obtained by the formula

$$C(u_1, u_2, \ldots, u_d) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_d^{-1}(u_d)\right), \tag{2.9}$$

where $F_i^{-1}(u_i)$ denotes the generalized inverse of F_i.

Definition 2.3.2. (Copula of F) If the random vector $\boldsymbol{X}$ has joint d.f. F with continuous marginal distributions $F_1, \ldots, F_d$, then the copula of F(or $\boldsymbol{X}$) is the joint d.f. C of $(F_1(X_1), \ldots, F_d(X_d))$.

Now we recall some basic properties of copulas.

Proposition 2.3.1. *$(X_1, \ldots, X_d)$ is a random vector with continuous marginal d.f.'s and copula C. Let $T_1, \ldots, T_d$ be strictly increasing functions. Then $(T_1(X_1), \ldots, T_d(X_d))$ also has copula C.*

Proposition 2.3.2. *Let $(X_1, \ldots, X_d)$ be a continuous random vector. Then $X_1, \ldots, X_d$ are independent if and only if their copula is Π.*

Theorem 2.3.2. *For every copula C and for every $\mathbf{u} \in [0, 1]^d$, we have the bounds*

$$W_d(\boldsymbol{u}) \leq C(\boldsymbol{u}) \leq M_d(\boldsymbol{u}).$$

M_d and W_d are called the *Fréchet-Hoeffding bounds* for all the copulas.

As a consequence of Sklar's theorem, if X_1 and X_2 are random variables with a joint distribution function F, copula C, and marginal d.f.'s F_1 and F_2, respectively, then for all x_1, x_2 in $\overline{\mathbb{R}}$,

$$\max\{F_1(x_1) + F_2(x_2) - 1, 0) \leq F(x_1, x_2) \leq \min\{F_1(x_1), F_2(x_2)\}$$

Because M_2 and W_2 are copulas, the above bounds are joint distribution functions and are called the *Fréchet-Hoeffding bounds* for joint distribution functions F with marginal d.f.'s F_1 and F_2.

Definition 2.3.3. (Partial Concordance Ordering of Copula) If C_1 and C_2 are copulas, we say that C_1 is *smaller than* C_2 (or C_2 is *larger than* C_1), and write $C_1 \prec C_2$ (or $C_2 \succ C_1$) if $C_1(u_1, u_2) \leq C_2(u_1, u_2)$ for all $u_1, u_2 \in \mathbb{I}$.

In 2-dimensional case, the Fréchet-Hoeffding lower bound copula W_2 is smaller than any copula, and the Fréchet-Hoeffding upper bound copula M_2 is larger than any copula. This point-wise partial concordance ordering of the set of copulas is important in Chapter 3 when we discuss the relationship between copulas and dependence properties for random variables (at which time the reason for the name of the ordering will become apparent). It is a partial order rather than a total order because not every pair of copulas is comparable.

Now the relationship between joint funcion and marginal d.f.'s is clear. But what about the relationship between joint survival distribution and survival margin distribution functions? Take the bivariate case as an example. With Sklar's theorem, (2.4) can be restated as:

$$\begin{aligned}\overline{F}(x_1, x_2) &= \overline{F}_1(x_1) + \overline{F}_2(x_2) - 1 + F(x_1, x_2) \\ &= \overline{F}_1(x_1) + \overline{F}_2(x_2) - 1 + C(F_1(x_1), F_2(x_2)) \\ &= \overline{F}_1(x_1) + \overline{F}_2(x_2) - 1 + C(1 - \overline{F}_1(x_1), 1 - \overline{F}_2(x_2))\end{aligned}$$

Let $\widehat{C}$ be a function from $\mathbb{I}^2$ to $\mathbb{I}$, and

$$\widehat{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2). \tag{2.10}$$

Then

$$\overline{F}(x_1, x_2) = \widehat{C}(\overline{F}_1(x_1), \overline{F}_2(x_2)). \tag{2.11}$$

The function $\widehat{C}$ is a copula.

Definition 2.3.4. (Bivariate Survival Copula) Let X_1 and X_2 be random variables, having joint distribution function F and copula C. The function defined in (2.10) is called *survival copula* of X_1 and X_2.

In quantitative finance, we are always interested in some risk factors' behaviors when some others stay in a particular set. Conditional distributions of copula are good tools to study these cases. We concentrate on a 2-dimensional case and suppose that (U_1, U_2) have d.f. C. Since a copula is an increasing continuous function in each argument,

$$\begin{aligned}C_{U_2|U_1}(u_2|u_1) \overset{\text{def}}{=\!=} P(U_2 \leq u_2 | U_1 = u_1) &= \lim_{\delta \to 0} \frac{C(u_1 + \delta, u_2) - C(u_1, u_2)}{\delta} \\ &= \frac{\partial}{\partial u_1} C(u_1, u_2),\end{aligned} \tag{2.12}$$

where this partial derivative exists almost everywhere (see [Nelsen, 2006, P.41] for precise details).

Definition 2.3.5. (Bivariate Conditional Distribution of Copulas) Let X_1 and X_2 be random variables, having copula C. The function $C_{U_2|U_1}(u_2|u_1)$ defined in (2.12) is called *conditional distribution of copula* C.

The conditional distribution is a distribution on the interval $[0, 1]$ which is only a uniform distribution in the case where C is the independence copula. A risk management interpretation of the conditional distribution is the following: Suppose continuous risks (X_1, X_2) have the (unique) copula C. Then $1 - C_{U_2|U_1}(q|p)$ is the probability that X_2 exceeds its qth quantile given that X_1 attains its pth quantile.

In the study of conditional distribution functions, it is often convenient to consider absolutely continuous copulas. Although copulas do not always have joint densities: e.g. the comonotonicity (2.6) and countermonotonicity (2.7), many parametric copulas have densities given by

$$c(u_1, \ldots, u_d) = \frac{\partial C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d}, \tag{2.13}$$

and we are sometimes required to calculate them. For example, if we wish to fit copulas to data by maximum likelihood. It is useful to note that, for the implicit copula of an absolutely continuous joint d.f. F with strictly increasing, continuous marginal d.f.'s $F_1, \ldots, F_d$, we may differentiate

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

to see that the *copula density* is given by

$$\begin{aligned} c(u_1, \ldots, u_d) &= \frac{\partial C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d} \\ &= \frac{\partial F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{\partial u_1 \ldots \partial u_d} \\ &= \frac{\partial F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{\partial F_1^{-1}(u_1) \ldots \partial F_d^{-1}(u_d)} \frac{\partial F_1^{-1}(u_1)}{\partial u_1} \cdots \frac{\partial F_d^{-1}(u_d)}{\partial u_d} \\ &= \frac{f(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \ldots f_d(F_d^{-1}(u_d))}, \end{aligned} \tag{2.14}$$

where f is the joint density of F, $f_1, \ldots, f_d$ are the marginal densities, and $F_1^{-1}, \ldots, F_d^{-1}$ are the ordinary inverses of the marginal d.f.'s.

2.4 Copula Families

We provide several copula examples in this section and they are subdivided into three categories: *Elliptical copulas* are extracted from well-known multivariate distributions by using Sklar's Theorem, but do not necessarily possess simple closed form expressions; *Archimedean copulas and Plackett family copulas* have simple closed-form expressions and follow general mathematical constructions known to yield copulas. In the end of section, some figures are plotted to show the difference between different copula families. Some different families (or construction methods) are discussed in [Aas et al., 2009, Gudendorf and Segers, 2010, Hofert, 2010].

2.4.1 Elliptical Copulas

We start with the basic definitions.

Definition 2.4.1. (Elliptical Distribution) A random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ is said to have an *elliptical distribution* with mean vector $\boldsymbol{\mu} \in \mathbb{R}^d$, covariance matrix $\Sigma = (\sigma_{ij})$, and generator $g : [0, +\infty) \mapsto [0, +\infty)$, if it can be expressed in the form

$$\boldsymbol{X} = \boldsymbol{\mu} + RA\boldsymbol{U}, \tag{2.15}$$

where A is the Cholesky decomposition of Σ i.e. $AA^T = \Sigma$, $\boldsymbol{U}$ is a d-dimensional random vector uniformly distributed on the sphere $\mathbb{S}^{d-1} = \{\boldsymbol{u} \in \mathbb{R}^d : u_1^2 + \ldots + u_d^2 = 1\}$, and R is a positive random variable independent of $\boldsymbol{U}$, with density given by

$$f_g(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1} g(r^2), \text{ for } \forall r > 0. \tag{2.16}$$

One could use $\boldsymbol{X} \sim \mathscr{E}(\boldsymbol{\mu}, \Sigma, g)$ to denote an elliptical distribution. The density function (if it exists) of an elliptical distribution is given, for every $\boldsymbol{x} \in \mathbb{R}^d$, by

$$h_g(\boldsymbol{x}) = |\Sigma|^{-1/2} g((\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})). \tag{2.17}$$

For more details, see [Fang et al., 1990, 2002, Frahm et al., 2003, McNeil et al., 2005, Genest et al., 2007].

One property of the elliptical distribution is that the scaled components $X_i/\sqrt{\sigma_{ii}}, i = 1, \ldots, d$ are identically distributed according to a d.f. F_g i.e. $(X_i/\sqrt{\sigma_{ii}}) \sim F_g$. This feature results that the elliptical distribution cannot be used widely in risk management, because not all the risk factors have the same distribution after some transformation ([Durante and Sempi, 2010]). Let Y_i, for every $i \in 1, 2, \ldots, d$, be the standardized variables of X_i i.e. $Y_i \overset{d}{=} X_i/\sqrt{\sigma_{ii}} - \mu_i$. Then $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ has the elliptical distribution $\mathscr{E}(\boldsymbol{0}, \mathrm{P}, g)$, where P is the correlation matrix of $\boldsymbol{X}$.

Definition 2.4.2. (Elliptical copula) Let $\boldsymbol{X}$ be an elliptical random vector, $\boldsymbol{X} \sim \mathscr{E}_d(\boldsymbol{\mu}, \Sigma, g)$. Suppose that, for every $i \in \{1, 2, \ldots, d\}$, $(X_i/\sqrt{\sigma_{ii}}) \sim F_g$. The distribution function of the random vector

$$\left(F_g\left(\frac{X_1}{\sqrt{\sigma_{11}}}\right), F_g\left(\frac{X_2}{\sqrt{\sigma_{22}}}\right), \ldots F_g\left(\frac{X_d}{\sqrt{\sigma_{dd}}}\right)\right) \tag{2.18}$$

is called *elliptical copula*.

Since the standardized transformation is strictly increasing, by the Proposition 2.3.1, the random vector $\boldsymbol{Y}$ has the same copula which $\boldsymbol{X}$ has.

Example 2.4.1. (Multivariate Gaussian Copula) ([McNeil et al., 2005, P191]) A copula of a Gaussian random vector $\boldsymbol{X} \sim N_d(\boldsymbol{0}, \mathrm{P})$ is called *Gaussian Copula*, and that is defined by

$$\begin{aligned} C_{\mathrm{P}}^{Ga}(\boldsymbol{u}) &= \mathbb{P}(\Phi(X_1) \leq u_1, \ldots, \Phi(X_d) \leq u_d) \\ &= \mathbb{P}(X_1 \leq \Phi^{-1}(u_1), \ldots, X_d \leq \Phi^{-1}(u_d)) \\ &= \boldsymbol{\Phi}_{\mathrm{P}}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)), \end{aligned}$$

where $\boldsymbol{\Phi}_{\mathrm{P}}$ is the joint d.f. of $N_d(\boldsymbol{0}, \mathrm{P})$, P is the correlation matrix, and Φ^{-1} denotes the inverse of the standard univariate Gaussian d.f..

Note that the independence copula (2.5), comonotonicity copula (2.6), and countermonotonicity copula (2.7) are special cases of the Gaussian copula. If $\mathrm{P} = I_d$, we obtain the independence copula (2.5), and if $\mathrm{P} = \boldsymbol{1}_d$, the $d \times d$ matrix consisting entirely of ones, we obtain comonotonicity (2.6). For $d = 2$ and $\mathrm{P} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$, the Gaussian copula is equal to the countermonotonicity copula (2.7)([McNeil et al., 2005]).

In the same way that we extract a copula from the multivariate normal distribution, we can also extract an implicit copula from any other distribution with continuous marginal d.f.'s.

Example 2.4.2. (Multivariate t-Copula)

$$C^t_{\nu,\mathrm{P}} = t_{\nu,\mathrm{P}}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)),$$

where t_ν is the d.f. of a standard univariate t-student distribution with ν degrees of freedom, $t_{\nu,\mathrm{P}}$ is the joint d.f. of the vector $\boldsymbol{X} \sim t_d(\nu, \mathbf{0}, \mathrm{P})$, and P is a correlation matrix.

As in the case of the Gaussian copula, if $\mathrm{P} = \mathbf{1}_d$, then we obtain comonotonicity (2.6). However, in contrast to the Gaussian copula, if $\mathrm{P} = I_d$ we do not obtain the independence copula (assuming $\nu < +\infty$) since uncorrelated multivariate t-distributed random variables are not independent.

Meanwhile, we could use the generalized inverse of the d.f. of the generator F_g^{-1} to obtain the random variable vector $\boldsymbol{X}$, such that $\boldsymbol{X} \in \mathbb{R}^d$ has distribution $\mathscr{E}(\boldsymbol{\mu}, \Sigma, g)$, marginal d.f. F_g and copula C.

Remark 2.4.1. An elliptical copula is typically not available in closed form.

Example 2.4.3. (Bivariate Gaussian Copula)

$$C^{Ga}_\rho(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right) ds\, dt,$$

where $\rho \in (-1, 1)$, and $\Phi^{-1}(\cdot)$ denotes the inverse of the univariate normal distribution.

In two dimensions the Gaussian copula can be thought of as a dependence structure that interpolates between perfect positive and negative dependence, where the parameter ρ represents the strength of dependence. The points in picture (a) of Figure 2.3 present an example of bivariate Gaussian distribution with $\rho = 0.77$. If $\rho = 1(-1)$, the points in picture (a) of Figure 2.3 is a line with positive (negative) slope, i.e. $X_2 = aX_1 + b$ where $a > 0 (a < 0)$.

Example 2.4.4. (Bivariate t-Copula)

$$C^t_{\nu,\rho}(u_1, u_2) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \int_{-\infty}^{t_\nu^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 - 2\rho st + t^2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} ds\, dt,$$

where $\rho \in (-1, 1)$, and $t_\nu^{-1}(\cdot)$ denotes the inverse of the univariate t-distribution with ν degrees of freedom.

2.4.2 Archimedean Copulas

The so-called Archimedean copulas have been very extensively studied (e.g. Section 5.4 of [McNeil et al., 2005], Chapter 4 of [Nelsen, 2006]). Originally, they were not studied in statistics, but in probabilistic metric spaces as part of the development of a probabilistic version of the triangle inequality. For an account of the history of Archimedean copulas, see [Schweizer, 1991] and the references cited therein.

Nowadays, this family of copulas is applied in various fields for the reasons that ([Nelsen, 2006]) states:

> "(1) the ease with which they can be constructed; (2) the great variety of families of copulas which belong to this family; and (3) the many nice properties possessed by the members of this family. "

Especially in quantitative finance, it has been proved useful for modelling portfolio risk. This will be demonstrated in Chapter 7. In this subsection we will show the simple structure of Archimedean copulas, present some of the properties that we will need, and introduce several most used families. For more details, see [McNeil et al., 2005, Section 5.4], and [Nelsen, 2006, Chapter 4].

Definition 2.4.3. (Pseudo-inverse) Let φ be a continuous, strictly decreasing function from $\mathbb{I}$ to $[0,\infty]$ such that $\varphi(1)=0$. The *pseudo-inverse* of φ is the function $\varphi^{[-1]}$ with $[0,\infty]$ and $\text{Ran}\varphi^{[-1]}=\mathbb{I}$ given by

$$\varphi^{[-1]}(t)=\begin{cases}\varphi^{-1}(t), & 0\le t\le\varphi(0)\\ 0, & \varphi(0)\le t\le\infty\end{cases} \tag{2.19}$$

Note that $\varphi^{[-1]}$ is continuous and nonincreasing on $[0,\infty]$, and strictly decreasing on $[0,\varphi(0)]$. Furthermore, $\varphi^{[-1]}(\varphi(u))=u$ on $\mathbb{I}$, and

$$\begin{aligned}\varphi(\varphi^{[-1]}(t))&=\begin{cases}t, & 0\le t\le\varphi(0)\\ \varphi(0), & \varphi(0)\le t\le\infty\end{cases}\\ &=t\wedge\varphi(0)\end{aligned} \tag{2.20}$$

Finally, if $\varphi(0)=\infty$, then $\varphi^{[-1]}=\varphi^{-1}$.

Theorem 2.4.1. *Let φ be a continuous, strictly decreasing function from $\mathbb{I}$ to $[0,\infty]$ such that $\varphi(1)=0$, and let $\varphi^{[-1]}$ be the pseudo-inverse of φ defined by (2.19). Let C be the function from $\mathbb{I}^2$ to $\mathbb{I}$ given by*

$$C(u_1,u_2)=\varphi^{[-1]}(\varphi(u_1)+\varphi(u_2)). \tag{2.21}$$

Then the function C is a copula if and only if φ is convex.

Proof. See [Nelsen, 2006, PP.111-112]. □

Copula C of the form (2.21) is called an *Archimedean copula*[7], where φ is a convex, continuous, strictly decreasing function from $\mathbb{I}$ to $[0,\infty]$ such that $\varphi(1)=0$. The function φ is called an *Archimedean generator* of the copula. If $\varphi(0)=\infty$, then φ is called a *strict Archimedean generator* and $C(u,v)$ is called a *strict Archimedean copula*.

In case of a strict Archimedean generator, $\varphi^{[-1]}=\varphi^{-1}$ and $C(u,v)=\varphi^{-1}(\varphi(u)+\varphi(v))$.

Theorem 2.4.2. *Let C be an Archimedean copula with generator φ. Then:*

1. *C is symmetric; i.e., $C(u_1,u_2)=C(u_2,u_1)$ for all $u_1,u_2\in\mathbb{I}$;*
2. *C is associative, i.e., $C(C(u_1,u_2),u_3)=C(u_1,C(u_2,u_3))$ for all $u_1,u_2,u_3\in\mathbb{I}$;*
3. *If $c>0$ is any constant, then $c\varphi$ is also a generator of C.*

Proof. See the proof of Theorem 2.4.4 when $d=2$. □

Archimedean copulas can be constructed by using Theorem 2.4.1 and using (2.21), if suitable functions will serve as generators, that are continuous decreasing convex functions φ from $\mathbb{I}$ to $[0,\infty]$ with $\varphi(1)=0$.

[7] For the meaning of the term "Archimedean" for these copulas, see [Nelsen, 2006, P.115].

Example 2.4.5. (Bivariate Frank Copula)

$$C_\theta^{Fr}(u_1,u_2) = -\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{(e^{-\theta}-1)}\right), \quad \theta\in\mathbb{R}.$$

C_θ^{Fr} is always strict. For $\theta = 0$ we obtain the *independence copula*, as a special case. The limit of C_θ^{Fr} for $\theta \to -\infty$ is the countermonotonicity copula; and for $\theta \to \infty$ is the comonotonicity copula. The Archimedean generator of this class is given by

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

Example 2.4.6. (Bivariate Clayton Copula)

$$C_\theta^{Cl}(u_1,u_2) = \left[\max\left\{u_1^{-\theta}+u_2^{-\theta}-1,0\right\}\right]^{-\frac{1}{\theta}}, \quad \theta\in[-1,\infty).$$

When θ is positive, C_θ^{Cl} is strict and absolutely continuous. For $\theta = -1, 0$ we obtain the *countermonotonicity and independence* respectively, as special cases. The limit of C_θ^{Cl} for $\theta \to \infty$ is the comonotonicity copula. The Archimedean generator of this class is given by

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

Example 2.4.7. (Bivariate Gumbel Copula)

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

For $\theta = 1$ we obtain the independence copula as a special case, and the limit of C_θ^{Gu} for $\theta \to \infty$ is the comonotonicity copula. The Archimedean generator of this class is given by

$$\varphi_\theta(t) = (-\ln t)^\theta.$$

Remark 2.4.2. The 2-dimensional results and examples given above could be generalized to d-dimensional cases. But one should keep in mind that some assumptions and the range of the parameters will be changed. In this section, we just introduced some basic definitions, propositions and examples. [Nelsen, 2006] performed a full investigation and more advanced studies.

Definition 2.4.4. (Completely Monotonic) A decreasing function $f(t)$ is *completely monotonic* on an interval $[a,b]$ if it satisfies

$$(-1)^k f^{(k)}(t) \geq 0, k\in\mathbb{N}, t\in(a,b). \tag{2.22}$$

Theorem 2.4.3. ***(Multivariate Archimedean Copula)*** *Let φ be a continuous, strictly decreasing function from $\mathbb{I}$ to $[0,\infty]$ such that $\varphi(0)=\infty$ and $\varphi(1)=0$, and let φ^{-1} be the inverse of φ and $\varphi^{[-1]}$ be the corresponding generalized inverse. If C^d [8] is the function from $\mathbb{I}^d$ to $\mathbb{I}$ given by*

$$C^d(\boldsymbol{u}) = \varphi^{[-1]}(\varphi(u_1)+\ldots+\varphi(u_d)). \tag{2.23}$$

Then C^d is a d-copula for all $d\geq 2$ if and only if φ^{-1} defined in (2.19) is completely monotonic on $[0,\infty]$.

[8] The superscript d on C denotes dimension.

For more details, see [Nelsen, 2006, McNeil and Nešlehová, 2009].

Theorem 2.4.4. *Let C^d be an Archimedean copula with generator φ. Then for all $d \geq 2$ and $\boldsymbol{u} \in \mathbb{I}^d$:*

1. *C^d is symmetric; i.e. for any permutation $(j(1), \ldots, j(d))$ of $(1, \ldots, d)$,*
$$C^d(\boldsymbol{u}) = C^d(u_{j(1)}, \ldots, u_{j(d)});$$
2. *C^{d+1} is associative, i.e. for all u_i in $\mathbb{I}$, $i = 1, \ldots, d, d+1$,*
$$C^{d+1}(u_1, \ldots, u_d, u_{d+1}) = C^2(C^d(u_1, \ldots, u_d), u_{d+1});$$
3. *If $c > 0$ is any constant, then $c\varphi$ is also a generator of C.*

Note: Usually, in case of non-ambiguity, the superscript d on C^d is omitted.

Example 2.4.8. (Multivariate Frank Copula)

$$C_\theta^{d(Fr)}(\boldsymbol{u}) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u_1} - 1) \ldots (e^{-\theta u_d} - 1)}{(e^{-\theta} - 1)^{d-1}} \right), \quad \theta \in (0, \infty).$$

The Archimedean generator of this class is given by

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

Example 2.4.9. (Multivariate Clayton Copula)

$$C_\theta^{d(Cl)}(\boldsymbol{u}) = \left[u_1^{-\theta} + \ldots + u_d^{-\theta} - d + 1 \right]^{-\frac{1}{\theta}}, \quad \theta \in (0, \infty).$$

The Archimedean generator of this class is given by

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

Example 2.4.10. (Multivariate Gumbel Copula)

$$C_\theta^{d(Gu)}(\boldsymbol{u}) = \exp \left(-\ln \left[(-\ln u_1)^\theta + \ldots + (-\ln u_d)^\theta \right]^{-\frac{1}{\theta}} \right), \quad \theta \in [1, \infty).$$

The Archimedean generator of this class is given by

$$\varphi_\theta(t) = (-\ln t)^\theta.$$

2.4.3 Plackett Family

Plackett's family of bivariate distributions is also a parameter copula and was discussed in [Plackett, 1965]. For more details, see [Nelsen, 2006, Section 3.3].

Example 2.4.11. (Bivariate Plackett Copula)

$$C_\theta^{Pl}(u_1, u_2) = \begin{cases} \dfrac{1 + (\theta - 1)(u_1 + u_2) - \sqrt{[1 + (\theta - 1)(u_1 + u_2)]^2 - 4u_1 u_2 \theta(\theta - 1)}}{2(\theta - 1)}, & \theta > 0, \text{ and } \theta \neq 1, \\ u_1 u_2, & \theta = 1. \end{cases}$$

The limit of C_θ^{Pl} for $\theta \to 0$ is the countermonotonicity copula; and for $\theta \to \infty$ is the comonotonicity copula. For this reason, Plackett family copulas have been widely used both in modeling and as alternatives to the bivariate normal for studies of power and robustness of various statistical tests [Conway, 1986, Hutchinson and Lai, 1990].

2.4.4 Extreme-Value Copula

Definition 2.4.5. (Extreme-value Copula) A copula C is called an *extreme-value copula* if there exists a copula C_F such that

$$C_F\left(u_1^{\frac{1}{n}}, \ldots, u_d^{\frac{1}{n}}\right)^n \xrightarrow[n\to\infty]{} C(u_1, \ldots, u_d) \tag{2.24}$$

for all $(u_1, \ldots, u_d) \in \mathbb{I}^d$. The copula C_F is said to be in the *domain of attraction* of C.

The extreme value copula is most used in risk management. The following theorem simplifies the extreme value copula.

Theorem 2.4.5. *A bivariate copula C is an extreme-value copula if and only if*

$$C(u_1, u_2) = (u_1 u_2)^{A\left(\frac{\ln u_1}{\ln(u_1 u_2)}\right)}, (u_1, u_2) \in (0,1]^2 \setminus \{(1,1)\}, \tag{2.25}$$

where $A : [0,1] \to [1/2, 1]$ is convex and satisfies $\max\{t, 1-t\} \leq A(t) \leq 1$ for all $t \in [0,1]$.

By reference to the work, the function A is often referred to as the *Pickands dependence function*. For more details, see [Nelsen, 2006, McNeil and Nešlehová, 2009, Gudendorf and Segers, 2010].

Proof. See [Nelsen, 2006, PP.97-98]. □

2.4.5 Empirical Copula

In practice, usually one needs to estimate some measures that are associated with some observed data, e.g. the VaR, CVaR, rank correlations (Spearman's ρ and Kendall's τ)[9], λ_L (lower tail dependence coefficient[10]) etc. One method is to choose a "good" copula whose parameters are estimated by the data, e.g. via maximum likelihood techniques or inversion of Kendall's tau (or other measures of association). Although there is some helpful criterion, the choice of a "good" copula is a critical problem. Another method is obtained by finding the empirical copula associated with the data. For more details about empirical copulas, see [Nelsen, 2006, Section 5.6].

Definition 2.4.6. (Empirical Copula) Let $\{(x_k^1, x_k^2)\}_{k=1}^n$ denote a sample of size n from a continuous bivariate distribution. The *empirical copula* is the function C_n given by

$$C_n(\frac{i}{n}, \frac{j}{n}) = \frac{\text{number of pairs}(x^1, x^2)\text{in the sample with } x^1 \leq x_{(i)}^1, x^2 \leq x_{(j)}^2}{n},$$

where $x_{(i)}^1$ and $x_{(j)}^2$, $1 \leq i, j \leq n$, denote order statistics from the sample. The *empirical copula frequency* c_n is given by

$$c_n(\frac{i}{n}, \frac{j}{n}) = \begin{cases} \frac{1}{n}, & \text{if } (x_{(i)}^1, x_{(j)}^2) \text{ is an element of the sample,} \\ 0, & \text{otherwise.} \end{cases}$$

Note that C_n and c_n are related via

$$C_n(\frac{i}{n}, \frac{j}{n}) = \sum_{p=1}^{i} \sum_{q=1}^{j} c_n(\frac{p}{n}, \frac{q}{n})$$

[9] They will be introduced in Chapter 3.
[10] It will be introduced in Chapter 4.

and

$$c_n(\frac{i}{n},\frac{j}{n}) = C_n(\frac{i}{n},\frac{j}{n}) - C_n(\frac{i-1}{n},\frac{j}{n}) - C_n(\frac{i}{n},\frac{j-1}{n}) + C_n(\frac{i-1}{n},\frac{j-1}{n}).$$

Empirical copulas were introduced and first studied by [Deheuvels, 1979], who called them *empirical dependence functions*.

2.4.6 Comparison among Several Copula Families

In this section, 5 different types of plots from 6 bivariate copulas are shown. Through comparing these plots we can understand well the characters of each copula.

The copulas are (a) Gaussian, (b) t-Student, (c) Frank, (d) Clayton, (e) Gumbel, (f) Plackett. The corresponding parameter(s), which is (are) listed in Table 2.1, of each copula is (are) chosen such that the associated Spearman's ρ is 0.75.

Five different types of plots are comprised of five thousand simulated points (Figure 2.2), five thousand simulated points (Figure 2.3) with standard normal margins constructed using the copula data from Figure 2.2, perspective plot (Figure 2.4) of the distribution function, contour plot (Figure 2.5) of the distribution function, and perspective plot (Figure 2.6) of the density function from (a) Gaussian, (b) t-Student, (c) Frank, (d) Clayton, (e) Gumbel, (f) Plackett with parameter(s) listed in Table 2.1.

Copulas	**Parameter(s)**	
Gaussian	$\rho = 0.77$	
t-Student	$\rho = 0.77$	$\nu = 2$
Frank	$\theta = 6.73$	
Clayton	$\theta = 2.58$	
Gumbel	$\theta = 2.29$	
plackett	$\theta = 17.14$	

Table 2.1. The parameter(s) of different copulas such that the Spearman's ρ of each copula is 0.75.

In Figure 2.2, 5000 simulated points from: (a) Gaussian with $\rho = 0.77$, (b) t-Student with $\rho = 0.77, \nu = 2$, (c) Frank with $\theta = 6.73$, (d) Clayton with $\theta = 2.58$, (e) Gumbel with $\theta = 2.29$, (f) Plackett with $\theta = 17.14$. All these points are generated such that the associated Spearman's $\rho = 0.75$, which results that the associated correlation coefficient is 0.77 for Gaussian and t-Student copula, and more or less 0.75 for other copulas.

In Figure 2.3, 5000 points are transformed to get the realizations from the points simulated in Figure 2.2 by using the quantile function of the standard normal distribution from six copulas: (a) Gaussian, (b) t, (c) Frank, (d) Clayton, (e) Gumbel, (f) Plackett. The Gaussian plot shows the points generated from a standard bivariate Gaussian distribution with the correlation coefficient 0.77. The other plots show the points sampled from the copulas constructed by Sklar's Theorem with the marginal distribution of univariate standard Gaussian distribution. Notice that the shape of the points generated from elliptical copulas (Gaussian and t) looks like an ellipsis. However, the points generated from the Frank and Plackett copulas are in the central region, like a rectangle, and that of Plackett is more narrow and long with respect to Frank's.

It can be seen that the Gaussian copula almost does not exhibit fat tails. As do Frank and Plackett copulas. In fact, these two families of copula could be used or considered similar to the Gaussian copula. By contrast, t, Clayton and Gumbel do have a great probability of extreme value pairs. The Clayton copula may have negative, large absolute value pairs. The Gumbel copula performs the

opposite, that may have positive, large absolute value pairs. However, t-Student has both negative and positive extreme values. In probability and statistics, the distribution like these three copulas is said to have *left(right) heavy tail*. In risk management, such heavy-tailed distributions are of interest. Because if the plots of loss-profits (returns) of two stocks is one of these three pictures presented in Figure 2.3((d), (e), (b)), this means that the two stocks have major chances to have big losses ((d)), big profits ((e)), or both big losses and big profits ((d)). The character of the distribution of the loss-profit is so important that it relates to the strategies of the hedging or the investment. Heavy tails are sometimes underconsidered when the copulas are used to simulate the data for various motivations e.g. the model testing (See Chapter 7), stress testing, and etc.

In Figure 2.4 and Figure 2.5, we can see that there is not much difference among perspective plots of these six copulas, and among contour plots of them.

In Figure 2.6, the conclusion obtained from Figure 2.3 is again confirmed that the Gaussian ((a)), the Frank ((c)) and Plackett ((f)) copulas almost do not exhibit strong tail dependence. However, the Clayton ((d)) copula may have negative, large absolute value pairs; the Gumbel ((e)) copula may have positive, large absolute value pairs; and the t ((b)) has both negative and positive extreme values.

2.5 Copula Simulation

Nowdays, it is easy to generate the pseudo-observations from the well-known copulas using computer programing, e.g. **C, C++, MATLAB**, R and so on. In my work, R[11] has been used.

In general, if we can generate a vector $\boldsymbol{X}$ with the d.f. F, we can transform each component with its own marginal d.f. to obtain a vector $\boldsymbol{U} = (U_1, \ldots, U_d)' = (F_1(X_1), \ldots, F_d(X_d))'$ with d.f. C, the copula of F. The theoretical algorithms for some particular examples in [McNeil et al., 2005] are given in the following.

Algorithm 2.5.1. (**Simulation of Elliptical Copula**)

1. Generate $\boldsymbol{X} \sim \mathscr{E}_d(\boldsymbol{\mu}, \Sigma, g)$[12].
2. Return $\boldsymbol{U} = (F_g(X_1), \ldots F_g(X_d))$, where F_g is the standard maginal d.f.. The random vector U has d.f. of elliptical copula .

Algorithm 2.5.2. (**Simulation of Archimedean Copula**)

1. Generate a variate V with d.f. G such that $\widehat{G} = \int_0^\infty e^{-tx} dG(x), t \geq 0$, the Laplace-Stieltjes transform of G, is the inverse of the generator ϕ of the required copula.
2. Generate independent uniform variates $X_1, \ldots, X_d$.
3. Return $\boldsymbol{U} = \left(\widehat{G}\left(-\frac{\ln X_1}{V}\right), \ldots, \widehat{G}\left(-\frac{\ln X_d}{V}\right)\right)'$.

(a) For the special case of the Frank copula, a discrete r.v. V with probability mass function $p(k) = P(V = k) = \frac{(1-\exp(-\theta))^k}{k\theta}$ for $k = 1, 2, \ldots$ and $\theta > 0$ needs to be generated. This can be achieved by standard simulation methods for discrete distributions (see [Ripley, 2009]).

[11] Now there are some R packages available, even more complicate copula simulation. For example, **copula**[Hofert et al., 2013], **VineCopula**[Schepsmeier et al., 2013] and etc.

[12] The algorithm for generating multivariate normal distribution and multivariate t-distribution can be found in [McNeil et al., 2005] on Page 66 and on Page 76, respectively.

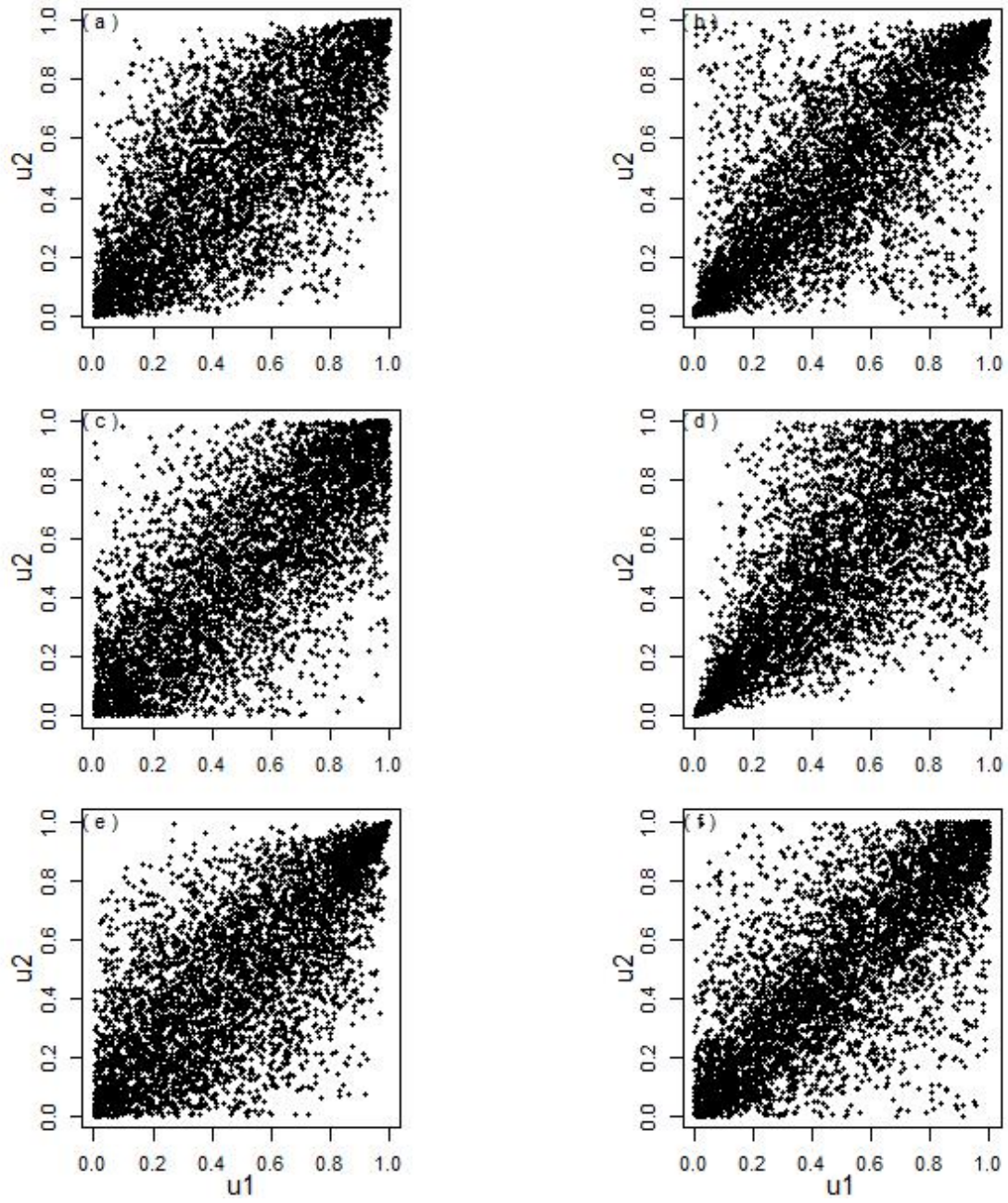


Figure 2.2. Five thousand simulated points from various copulas with Spearman's $\rho = 0.75$: (a) Gaussian with $\rho = 0.77$, (b) t-Student with $\rho = 0.77, \nu = 2$, (c) Frank with $\theta = 6.73$, (d) Clayton with $\theta = 2.58$, (e) Gumbel with $\theta = 2.29$, (f) Plackett with $\theta = 17.14$. All these show the Spearman's $\rho = 0.75$.

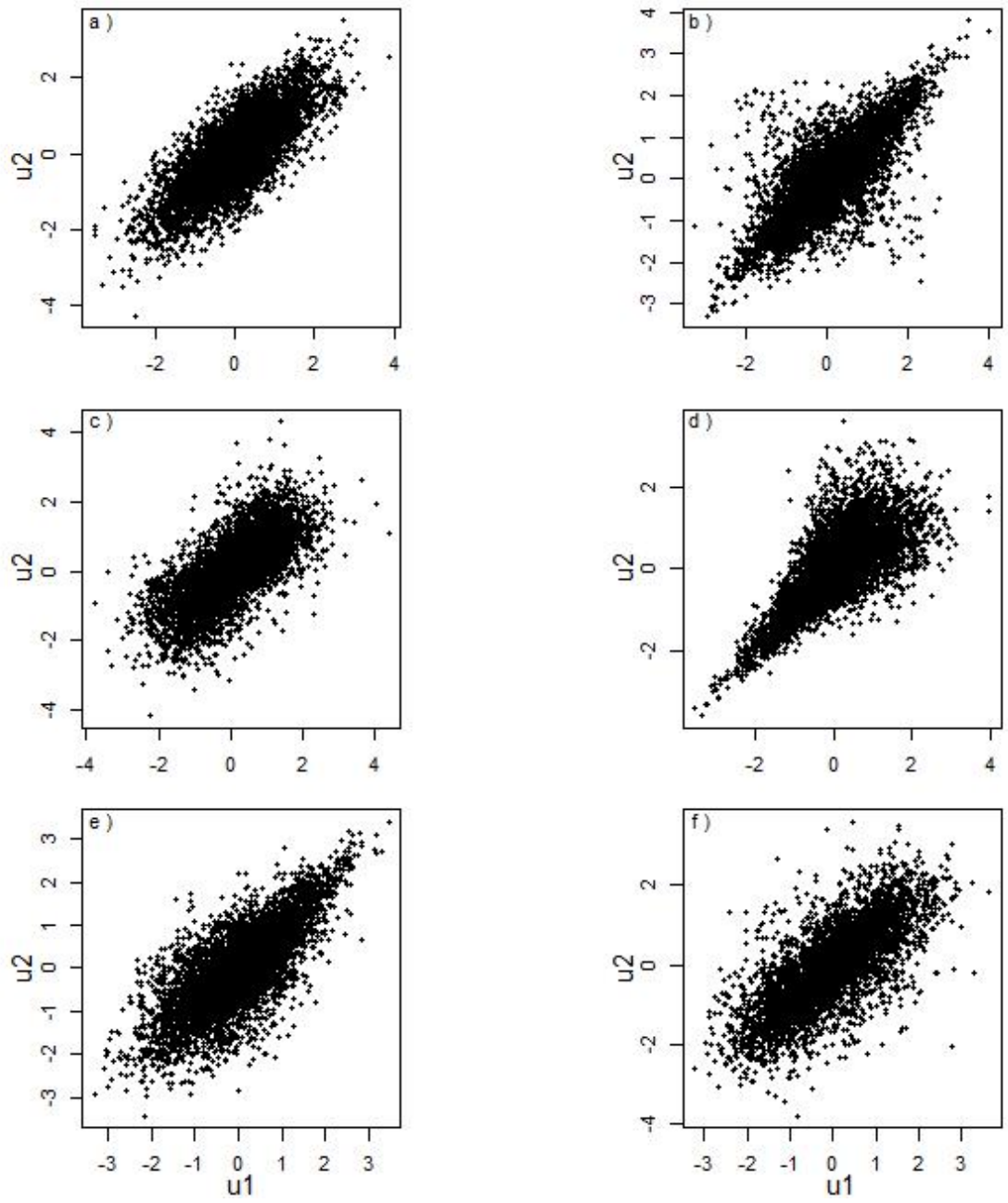


Figure 2.3. Five thousand simulated points from six distributions with standard normal margins, constructed using the copula data from Figure 2.2. (a) Gaussian, (b) t-Student, (c) Frank, (d) Clayton, (e) Gumbel, (f) Plackett.

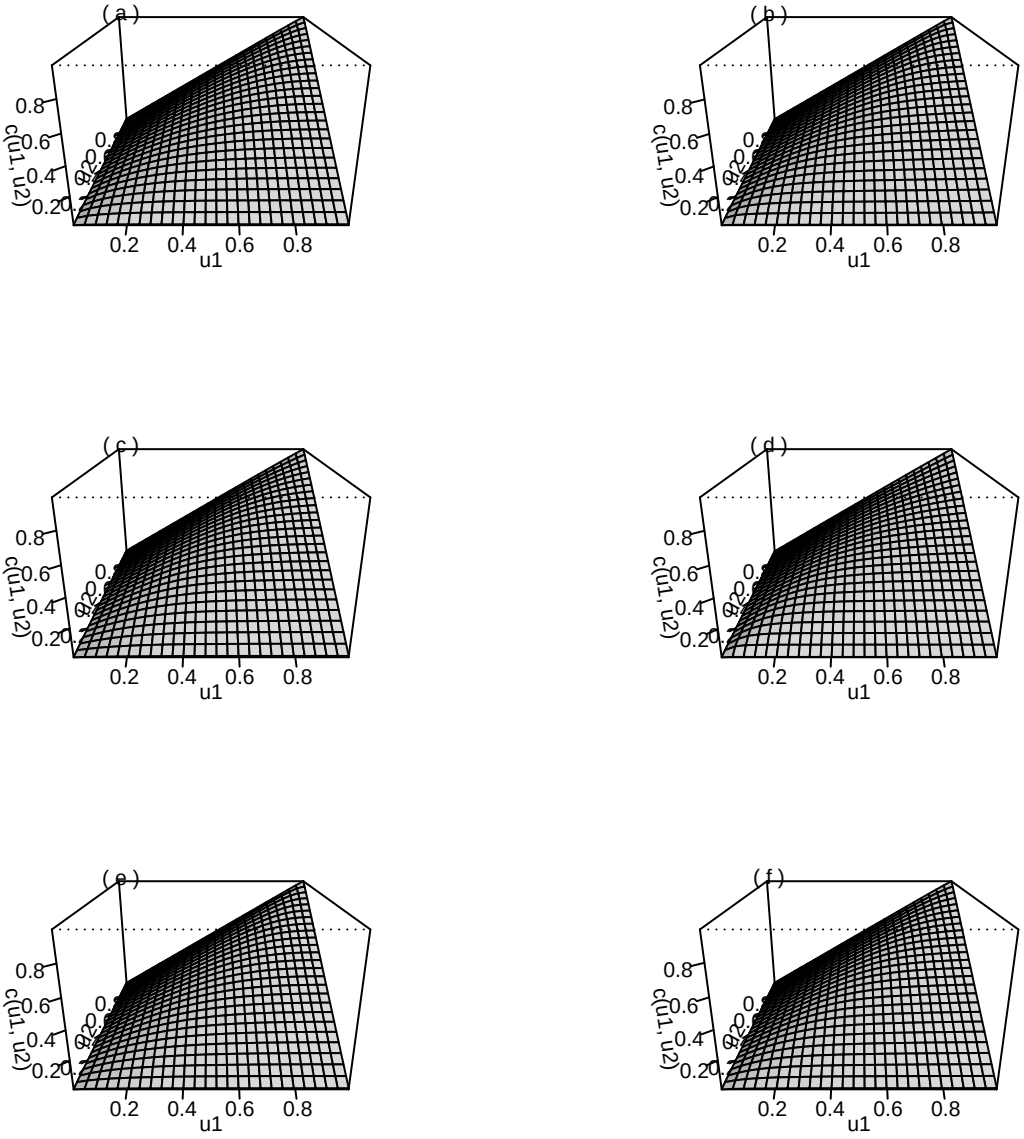


Figure 2.4. Perspective plot of the distribution function of the six copulas: (a) Gaussian with $\rho = 0.77$, (b) t-Student with $\rho = 0.77, \nu = 2$, (c) Frank with $\theta = 6.73$, (d) Clayton with $\theta = 2.58$, (e) Gumbel with $\theta = 2.29$, (f) Plackett with $\theta = 17.14$.

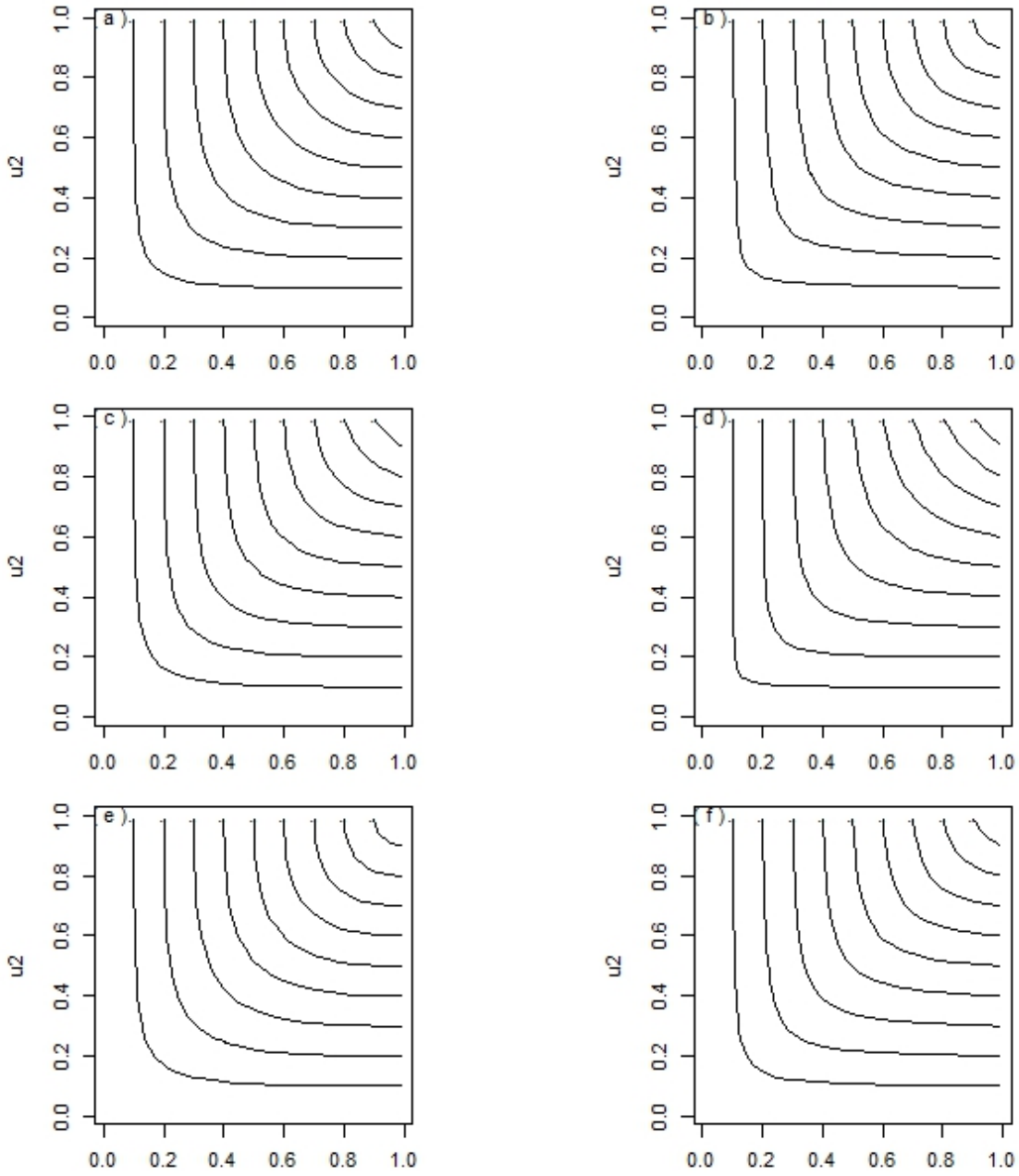


Figure 2.5. Contour plot of the distribution function of the six copulas: (a) Gaussian with $\rho = 0.77$, (b) t-Student with $\rho = 0.77, \nu = 2$, (c) Frank with $\theta = 6.73$, (d) Clayton with $\theta = 2.58$, (e) Gumbel with $\theta = 2.29$, (f) Plackett with $\theta = 17.14$.

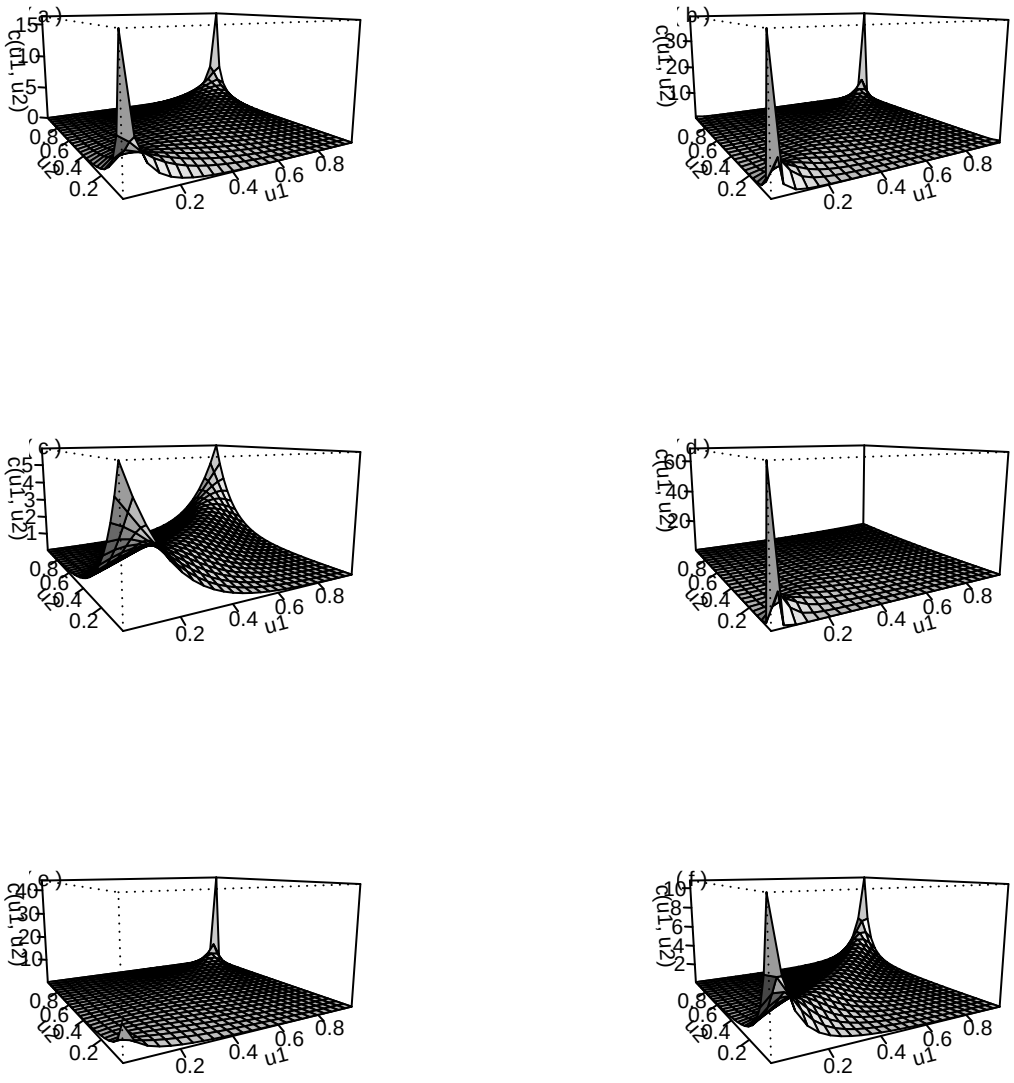


Figure 2.6. Perspective plot of the density of the six copulas: (a) Gaussian with $\rho = 0.77$, (b) t-Student with $\rho = 0.77, \nu = 2$, (c) Frank with $\theta = 6.73$, (d) Clayton with $\theta = 2.58$, (e) Gumbel with $\theta = 2.29$, (f) Plackett with $\theta = 17.14$.

(b) For the special case of the Clayton copula, a gamma variate $V \sim Ga\left(\frac{1}{\theta}, 1\right)$[111] with $\theta > 0$ needs to be generated. The d.f. of V has Laplace transform $\widehat{G}(t) = (1+t)^{-\frac{1}{\theta}}$. Note that the inverse $\widehat{G}^{-1}(t) = t^{-\theta} - 1$ differs from the generator in Example 2.4.6 by a constant factor that is unimportant[13].

(c) For the special case of the Gumbel copula, a positive stable variate $V \sim St\left(\frac{1}{\theta}, 1, \gamma, 0\right)$[111], where $\gamma = \cos^{\theta}\left(\frac{\pi}{2\theta}\right)$ and $\theta > 1$ needs to be generated. This d.f. has Laplace transform $\widehat{G}(t) = \exp\left(-t^{\frac{1}{\theta}}\right)$ as desired.

2.6 Conclusion

In this chapter, we have introduced some basic concepts of copula theory, the most used copula families, and the simulation algorithms of the copula. It is a convinient and elegant method to model the joint distribution of the variables. Also we could use some copula families to simulate the data with specific dependence structure, e.g. the pictures shown in Figure 2.2, 2.3, and 2.6, for furture use.

However, we should pay much attention to the use of this method. When we choose several candidate copula families to do the model fitting, although there are some criteria, e.g. AIC, BIC, and so on, we just take the best one among the supposed copulas. But, no one knows the true model. This will be problematic particularly when we use this chosen model to forecast. Another problem is that the copulas presented here are static. Naturally, the dependence structure between the variables does not remain unchanged, e.g. more correlated in the crisis period and less in the normal period. So, it is debatable if we use a constant correlation measure for all the time. In other words, dynamic copulas should be more reasonable. For more details of the copulas, see the very interesting discussion related to the [Mikosch, 2006] (see also [de Vries and Zhou, 2006, Embrechts, 2006, Genest and Rémillard, 2006, Joe, 2006]).

[13] This is granteed by (3) of Theorem 2.4.2

[111] The Gamma distribution and the stable distribution are discussed in Appendix 2 of [McNeil et al., 2005].

Chapter 3

Dependence Concepts

3.1 Introduction

The word "*dependence*" means "*the situation in which you need something or someone all the time, especially in order to continue existing or operating*"[1]. When we talk about the dependence between two or more variables in mathematics, in fact we are talking how another variable reacts when one variable changes.

The mechanism of reaction is studied as the dependence structure and the degree of interaction is studied as the dependence measure. Dependence is a fundamental characteristic between random variables. As [Jogdeo, 1982] notes,

> "Dependence relations between random variables is one of the most widely studied subjects in probability and statistics. The nature of the dependence can take a variety of forms and unless some specific assumptions are made about the dependence, no meaningful statistical model can be contemplated."

Not only in applied mathematics, but also in investment and risk management, dependence is an important issue. Investors choose strategies based on the dependence between assets under consideration and on risk attitude. Thus, risk managers have to think about the dependence between risk factors and also their motivations.

Practical applications suggest that the dependence measure must have good properties. "Scale-invariant" ([Hoeffding, 1940, 1941]) is one of them and sometimes comonotone-invariant is also needed, especially in risk management. From this perspective, copula is an intuitive instrument in the study of dependence. As introduced in Chapter 2, the copula between random variables remains unchanged under strictly increasing transformation. In this chapter, we focus on copula's role in the dependence study and on the dependence measure expressed in terms of the copula.

In this chapter, we start with the well-known Pearson's correlation that plays a central role in financial theory. Moreover, we will discuss its possible pitfalls and limitations. Then we will introduce a kind of rank correlation which is calculated based on the ranks of random variables, namely Kendall's tau and Spearman's rho. These two kinds of dependence measure are copula-based dependence measures.

In this chapter, the properties of different dependence measures will be also presented. The inclusion of properties is important because, in a given situation or application, the choice of appropriate measures can depend on the properties one wants to preserve or satisfy (See [Nelsen, 2006, Section 5.7] for multivariate dependence).

[1] http://dictionary.cambridge.org/dictionary/british/dependence?q=dependence

Other surveys concerning dependence or/and dependence measures can be found in the literature mentioned in this chapter.

3.2 Linear Correlation

Definition 3.2.1. (Pearson's Correlation Coefficient) Let X_1 and X_2 be two random variables with finite variance and covariance. The *Pearson's correlation coefficient* ρ of them are defined by

$$\rho_{X_1,X_2} = \rho(X_1, X_2) = \frac{\mathrm{cov}(X_1, X_2)}{\sqrt{\mathrm{var}(X_1)}\sqrt{\mathrm{var}(X_2)}} \tag{3.1}$$

For d-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$, the correlation matrix is denoted by $\mathrm{P}(\boldsymbol{X}) = (\rho_{ij})$, where $\rho_{ij} = \rho_{X_i,X_j}$.

ρ is a measure of linear dependence and takes values in $[-1, 1]$. If $\rho_{X_1,X_2} = 0$, they are said to be *uncorrelated*. If $|\rho_{X_1,X_2}| = 1$, this means that X_2 and X_1 are perfectly linearly dependent, i.e. $X_2 = aX_1 + b$ almost surely for some $a, b \in \mathbb{R}$. Moreover, $\beta > 0 (< 0)$ for $\rho_{X_1,X_2} = 1(-1)$ means the positive (negative) linear dependence.

Remark 3.2.1.

Independence $\Rightarrow$ **Uncorrelatedness**. But, in general, **Uncorrelatedness** $\not\Rightarrow$ **Independence**, see Example 3.2.1. It is true only when the variables are Gaussian, see [McNeil et al., 2005, P.74].

Example 3.2.1. Consider two independent random variables X, Y and two dependent variables X, X^2, where $X, Y \sim N(0, 1)$. Then we have

$$\rho(X, Y) = 0 \quad \text{and} \quad \rho(X, X^2) = 0$$

since

$$\mathrm{cov}(X, X^2) = \mathbb{E}(X^3) - \mathbb{E}(X)\mathbb{E}(X^2) = 0.$$

Obviously, X and X^2 are not independent. See Figure 3.1.

Proposition 3.2.1.

$$\rho(a_1X_1 + b_1, a_2X_2 + b_2) = \rho(X_1, X_2), \qquad a_1, a_2 > 0.$$

"However, correlation is not invariant under nonlinear strictly increasing transformations $T : \mathbb{R} \to \mathbb{R}$, i.e. in general, $\rho(T(X_1), T(X_2)) \neq \rho(X_1, X_2)$"([McNeil et al., 2005]).

Example 3.2.2. Let X_1, X_2 be two random variables, where $X_1 \sim N(0, 1)$ and $X_2 \overset{\text{def}}{=\!=} \sigma X_1, \quad \sigma > 0$. Then we have

$$\rho(X_1, X_2) = 1$$

Now consider the correlation coefficient between random variables e^{X_1}, e^{X_2}. By [Tarmast, 2001], we have

$$\rho\left(e^{X_1}, e^{X_2}\right) = \frac{e^{\sigma} - 1}{\sqrt{(e-1)(e^{\sigma^2} - 1)}}.$$

In this case, ρ is a function of σ (see Figure 3.2), and

$$\rho\left(X_1, X_2\right) \neq \rho\left(e^{X_1}, e^{X_2}\right) \text{ for } \sigma \neq 1.$$

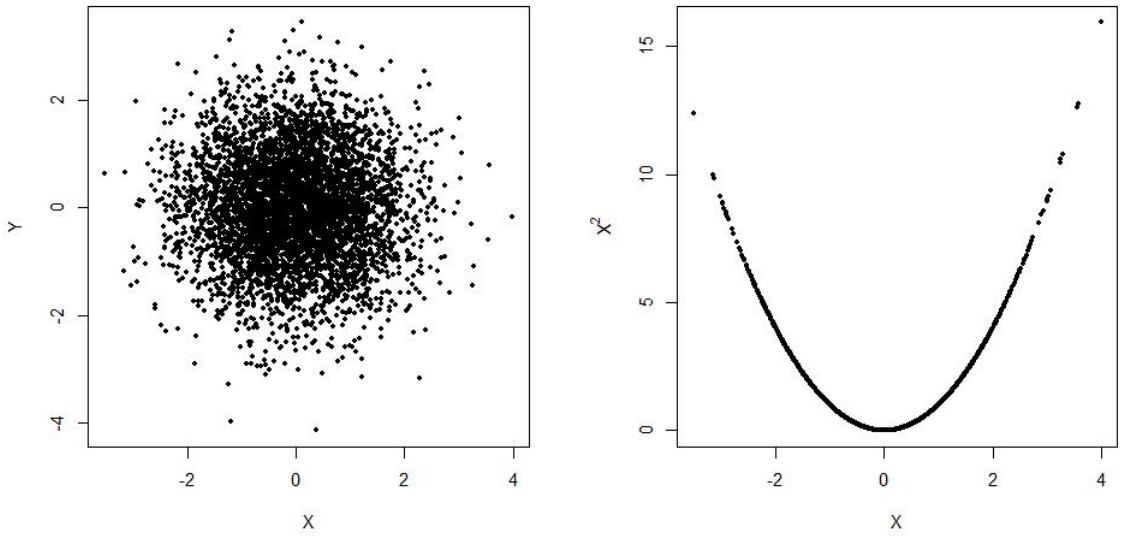


Figure 3.1. Plots of two independent variables (X, Y) and two dependent variables $\left(X, X^2\right)$ with the same correlation coefficient $\rho = 0$, where $X, Y \sim N(0, 1)$.

3.3 Rank Correlation

As stated in the previous section, the correlation coefficient is a measure of linear dependence. When X_2 is a non-linear increasing function of X_1, ρ_{X_1,X_2} may be a very small positive number (See Figure 3.2. $\rho(e^{X_1}, e^{\sigma X_1}) \approx 0$ when $\sigma = 5$.). This would lead to wrong decisions. However, rank correlation could overcome this problem well. It is a kind of correlation which is calculated according to the *"ranks"* as its name. As a result, the rank correlation is invariant under both linear and nonlinear strictly increasing transformations T which do not change the original ranks at all. Rank correlation depends only on the copula of bivariate random vector, not on the marginal distributions. As a contrast, that correlation coefficient depends on both. Therefore, the formula of rank correlation of some parametric families can be included. This is useful for copula. The most used rank correlations are Kendall's tau and Spearman's rho, which will be discussed below.

First of all, we need to define the concordance of two points in $\mathbb{R}^2$.

Definition 3.3.1. (Concordant and Discordant) ([Nelsen, 2002]) We say that two observations (x_1^1, x_2^1) and (x_1^2, x_2^2) of a pair (X_1, X_2) of continuous random variables are *concordant* if $(x_1^1 - x_1^2)(x_2^1 - x_2^2) > 0$ and are *discordant* if $(x_1^1 - x_1^2)(x_2^1 - x_2^2) < 0$.

An intuitive concordance measure between two random variables is the difference of the probability between concordance and discordance, that is defined as following.

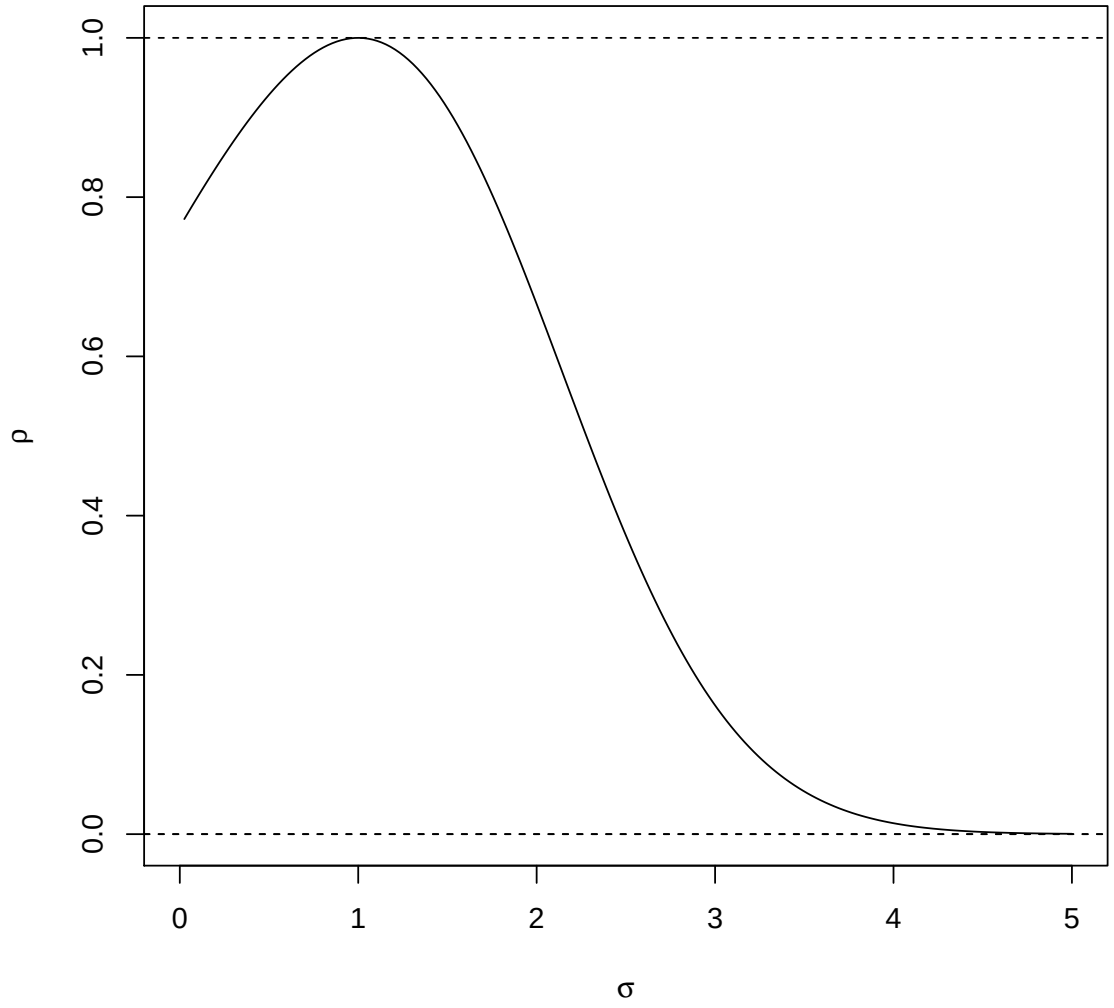


Figure 3.2. Correlation coefficient between two log-normal random variables $\left(e^{X_1}, e^{\sigma X_1}\right)$ against σ.

3.3.1 Kendall's tau

Definition 3.3.2. (Kendall's tau) Let (X_1^1, X_2^1) and (X_1^2, X_2^2) be independent and identically distributed random vectors, each with the same joint distribution function F. Then *Kendall's tau* between X_1 and X_2, τ_{X_1,X_2}, is defined as the probability of concordance minus the probability of discordance:

$$\tau_{X_1,X_2} = \tau(X_1, X_2) = \mathbb{P}((X_1^1 - X_1^2)(X_2^1 - X_2^2) > 0) - \mathbb{P}((X_1^1 - X_1^2)(X_2^1 - X_2^2) < 0). \quad (3.2)$$

Sometimes, τ is also defined as

$$\mathbb{E}\left(\operatorname{sgn}\left((X_1^1 - X_1^2)(X_2^1 - X_2^2)\right)\right). \text{ ([McNeil et al., 2005])}$$

In higher dimensions, the Kendall's tau matrix of a random vector may be written as

$$\tau(\boldsymbol{X}) = \operatorname{cov}(\operatorname{sgn}(\boldsymbol{X} - \widetilde{\boldsymbol{X}})),$$

where $\widetilde{\boldsymbol{X}}$ is an independent copy of $\boldsymbol{X}$; since it can be expressed as a covariance matrix, $\tau(\boldsymbol{X})$ is obviously positive semidefinited.

In fact, τ between two continuous random variables (X_1, X_2) can be also defined via copulas and depends only on the copula of (X_1, X_2) not on the marginal distributions. In order to prove this, the following theorem proposed in [Nelsen, 2006, P. 159] is presented here.

Theorem 3.3.1. *Let (X_1^1, X_2^1) and (X_1^2, X_2^2) be independent vectors of continuous random variables with joint distribution functions F^1 and F^2, respectively, with common margins F_1 (of X_1^1 and X_1^2) and F_2 (of X_2^1 and X_2^2). Let C^1 and C^2 denote the copulas of (X_1^1, X_2^1) and (X_1^2, X_2^2) , respectively, so that $F^1(x_1, x_2) = C^1(F_1(x_1), F_2(x_2))$ and $F^2(x_1, x_2) = C^2(F_1(x_1), F_2(x_2))$. Let Q denote the difference between the probabilities of concordance and discordance of (X_1^1, X_2^1) and (X_1^2, X_2^2), i.e.*

$$Q = \mathbb{P}\left((X_1^1 - X_1^2)(X_2^1 - X_2^2) > 0\right) - \mathbb{P}\left((X_1^1 - X_1^2)(X_2^1 - X_2^2) < 0\right). \tag{3.3}$$

Then

$$Q = Q\left(C^1, C^2\right) = 4 \int\int_{\mathbb{I}^2} C^2(u_1, u_2)\, dC^1(u_1, u_2) - 1. \tag{3.4}$$

Proof. See [Nelsen, 2006, P. 159]. □

As a consequence of Theorem 3.3.1, we have the following corollary.

Corollary 3.3.1. *Let C^1, C^2 and Q be as given in Theorem 3.3.1. Then*

1. *Q is symmetric in its arguments: $Q\left(C^1, C^2\right) = Q\left(C^2, C^1\right)$.*
2. *Q is nondecreasing in each argument: if $C^1 \prec C^{1\prime}$ and $C^2 \prec C^{2\prime}$ for all $(u, v) \in \mathbb{I}^2$, then $Q\left(C^1, C^2\right) \leq Q\left(C^{1\prime}, C^{2\prime}\right)$.*
3. *Copulas can be replaced by survival copulas in Q, i.e., $Q\left(C^1, C^2\right) = Q\left(\widehat{C^1}, \widehat{C^2}\right)$.*

Now we return to the Definition 3.3.2 of Kendall's tau. Since (X_1^1, X_2^1) and (X_1^2, X_2^2) are independent and identically distributed random vectors, each with same joint distribution function F, then (X_1^1, X_2^1) and (X_1^2, X_2^2) have the same copula, C. As a result, we could get the following theorem.

Theorem 3.3.2. *Let X_1 and X_2 be two continuous random variables with copula C. Then the Kendall's tau of (X_1, X_2) is given by*

$$\tau_{X_1, X_2} = \tau_C = Q(C, C) = 4 \int\int_{\mathbb{I}^2} C(u_1, u_2)\, dC(u_1, u_2) - 1. \tag{3.5}$$

Since τ depends only on the copula of (X_1, X_2) and not on the marginal distributions, it is invariant under both linear and nonlinear strictly increasing transformations $T : \mathbb{R} \to \mathbb{R}$, i.e. $\tau(T(X_1), T(X_2)) \equiv \tau(X_1, X_2)$.

Note that the integral part in (3.5), $\int\int_{\mathbb{I}^2} C(u_1,u_2)\,dC(u_1,u_2)$, is the expected value of distribution C of uniform random variables (U_1,U_2) on $\mathbb{I}$ ([Nelsen, 2006, P.162]). Thus, τ can also be written as

$$\tau_C = 4\mathbb{E}(C(U_1,U_2)) - 1.$$

When the copula C is a member of a parametric family of copulas (e.g., if C is denoted C_θ or $C_{\alpha,\beta}$), τ_{C_θ} or $\tau_{C_{\alpha,\beta}}$ will be denoted as τ_θ and $\tau_{\alpha,\beta}$, respectively.

Some most used parametric copula families admit a closed formula for Kendall's tau. As a consequence of Theorem 3.3.2, the following two corollaries show how to compute Kendall's tau of Elliptical copulas and of Archimedean copulas.

Corollary 3.3.2. *([McNeil et al., 2005]) Let $\boldsymbol{X} \sim \mathscr{E}_2(\mathbf{0},\mathrm{P},g)$ for a correlation matrix P with off-diagonal element ρ, and assume that $\mathbb{P}(\boldsymbol{X}=\mathbf{0})=0$. Then the relationship*

$$\tau(X_1,X_2) = \frac{\pi}{2}\arcsin\rho$$

holds.

Proof. See [McNeil et al., 2005, P.217]. □

Example 3.3.1. (Kendall's tau for Gaussian (t-) copula) [McNeil et al., 2005, P. 215] Let $\boldsymbol{X}$ have a bivariate Gaussian (t-Student) distribution with copula C^{Ga}_ρ ($C^t_{\nu,\rho}$) and continuous margins. Then Kendall's tau is

$$\tau(X_1,X_2) = \frac{\pi}{2}\arcsin\rho.$$

Proof. See [McNeil et al., 2005, P. 215]. □

Corollary 3.3.3. *Let X_1 and X_2 be random variables with an Archimedean copula C generated by φ in Ω. Then τ_C for X_1 and X_2 is given by*

$$\tau_C = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}\,dt.$$

Proof. See [Nelsen, 2006, P.163]. □

Example 3.3.2. Consider the Frank family presented in Example 2.4.5. It can be shown that (see e.g. [Genest, 1987]) Kendall's tau is

$$\tau_\theta = 1 - 4\frac{1-D_1(\theta)}{\theta},$$

where $D_k(x)$ is the Debye function, given by

$$D_k(x) = \frac{k}{x^k}\int_0^x \frac{t^k - t}{e^t - 1}\,dt \tag{3.6}$$

for any positive integer k. For Archimedean copulas of Frank family, we could not get a closed form of $\theta = f(\tau_\theta)$, where f is a suitable function. It is obtained numerically.

Example 3.3.3. Consider the Clayton family with generator $\varphi_\theta(t) = \frac{1}{\theta}(t^{-\theta}-1)$, for $\theta \in [-1,\infty)\backslash\{0\}$. Then

$$\frac{\varphi(t)}{\varphi'(t)} = \frac{t^{\theta+1}-t}{\theta}.$$

Using Corollary 3.3.3 we can calculate Kendall's tau for the Clayton family.

$$\tau_\theta = 1 + 4\int_0^1 \frac{t^{\theta+1}-t}{\theta}\,dt = \frac{\theta}{\theta+2}.$$

As a consequence, we could also get

$$\theta = \frac{2\tau_\theta}{1-\tau_\theta}\text{, for } \tau_\theta > 0.$$

Example 3.3.4. Consider the Gumbel family with generator $\varphi_\theta(t) = (-\ln t)^\theta$, for $\theta \geq 1$. Then

$$\frac{\varphi(t)}{\varphi'(t)} = \frac{t\ln t}{\theta}.$$

Using Corollary 3.3.3 we can calculate Kendall's tau for the Gumbel family.

$$\tau_\theta = 1 + 4\int_0^1 \frac{t\ln t}{\theta}\,dt = 1 - \frac{1}{\theta}.$$

As a consequence, we could also get

$$\theta = \frac{1}{1-\tau_\theta}\text{, for } \tau_\theta > 0.$$

3.3.2 Spearman's rho

As with Kendall's tau, Spearman's rho is another rank correlation which is also based on concordance and discordance.

Definition 3.3.3. (Spearman's rho) ([Nelsen, 2006, P. 167]) Let (X_1^1, X_2^1), (X_1^2, X_2^2), and (X_1^3, X_2^3) be three independent random vectors with a common joint distribution function F (whose continuous margins are again F_1 and F_2) and copula C.

The *Spearman's rho* of (X_1, X_2), ρ_{S,X_1X_2}, is defined to be proportional to the probability of concordance minus the probability of discordance for the two vectors (X_1^1, X_2^1) and (X_1^2, X_2^3)–i.e., a pair of vectors with the same margins, but one vector has distribution function F, while the components of the other are independent:

$$\rho_S(X_1,X_2) = 3\left(\mathbb{P}\left((X_1^1 - X_1^2)(X_2^1, X_2^3) > 0\right) - \mathbb{P}\left((X_1^1 - X_1^2)(X_2^1, X_2^3) < 0\right)\right), \tag{3.7}$$

(the pair (X_1^3, X_2^2) could be used equally as well).

Note that in Definition 3.3.3 while the joint distribution function of (X_1^1, X_2^1) is $F(x_1, x_2)$, the joint distribution function of (X_1^2, X_2^3) is $F_1(x_1)F_2(x_2)$ (because X_1^2 and X_2^3 are independent). Thus the copula of X_1^2 and X_2^3 is the independence copula Π, and using Theorem 3.3.1 and part 1 of Corollary 3.3.1, we immediately have

Theorem 3.3.3. *Let X_1 and X_2 be continuous random variables whose copula is C. Then the Spearman's rho for X_1 and X_2 (which we will denote by either ρ_{S,X_1X_2}, or ρ_{SC}) is given by*

$$\rho_S(X_1, X_2) = \rho_{S,C} = 3Q(C, \Pi), \tag{3.8a}$$

$$= 12 \int\int_{\mathbb{I}^2} u_1, u_2 \, dC(u_1, u_2) - 3, \tag{3.8b}$$

$$= 12 \int\int_{\mathbb{I}^2} C(u_1, u_2) \, du_1 \, du_2 - 3. \tag{3.8c}$$

In other textbooks (e.g. [McNeil et al., 2005]), Spearman's rho is sometimes definited as the linear correlation of the random variables' d.f.'s:

$$\rho_S(X_1, X_2) = \rho_{F_1(X_1),F_2(X_2)} = \rho(F_1(X_1), F_2(X_2)). \tag{3.9}$$

In fact, the definition via concordance measure (3.7) and the definition via correlation of d.f.'s (3.9) are equivalent. As explained in [Nelsen, 2006, P. 170], set the uniform [0,1] random variables $U_1 = F_1(X_1)$ and $U_2 = F_2(X_2)$ whose joint distribution function is C. Since U_1 and U_2 each has mean $1/2$ and variance $1/12$, then $\rho_{S,C}$ in (3.8b) can be re-written in the following form:

$$\begin{aligned}\rho_S(X_1, X_2) = \rho_{S,C} &= 12 \int\int_{\mathbb{I}^2} u_1 u_2 \, dC(u_1, u_2) - 3 = 12\mathbb{E}(U_1U_2) - 3, \\ &= \frac{\mathbb{E}(U_1U_2) - 1/4}{1/12} = \frac{\mathbb{E}(U_1U_2) - \mathbb{E}(U_1)\mathbb{E}(U_2)}{\sqrt{\mathrm{Var}(U_1)}\sqrt{\mathrm{Var}(U_2)}} \\ &= \rho(F_1(X_1), F_2(X_2)).\end{aligned}$$

Example 3.3.5. [McNeil et al., 2005, P. 215] Consider $\boldsymbol{X}$ as is bivariate Gaussian r.v. with copula C_ρ^{Ga} and continuous margins. Then Spearman's rho is

$$\rho_S(X_1, X_2) = \frac{6}{\pi} \arcsin \frac{\rho}{2}. \tag{3.10}$$

Remark 3.3.1. The relationship (3.10) between Spearman's rho and linear correlation of the elliptical distributions is not like the relationship (Corollary 3.3.2) between Kendall's tau and linear correlation, which holds for all elliptical distributions. A counterexample is found in [Hult and Lindskog, 2002]. Simple formulas for elliptical distributions other than the Gaussian, such as the bivariate t-copula , are not known to us.

Example 3.3.6. [Cherubini et al., 2004, P. 126] Consider the Frank family presented in Example 2.4.5. It can be shown that Spearman's rho is

$$\rho_{S,\theta} = 1 - 12\frac{D_2(-\theta) - D_1(-\theta)}{\theta},$$

where $D_k(x)$ is the Debye function, given by (3.6), for any positive integer k.

Remark 3.3.2. Spearman's rho of an Archimedean copula has in general no closed form, and it could be estimated numerically.

Until now, two measures on the basis of concordance have been addressed. But no definition of "measure of concordance" has been presented. For this, we show the work of Scarsini in 1984[2], in which a precise definition of measure of concordance was defined and some related theorems were also proved. We use the denotation of the measure κ in [Nelsen, 2006] which is I in the original work.

[2] [Scarsini, 1984].

Definition 3.3.4. (Measure of Concordance) A measure κ is called a *measure of concordance* for pairs of continuous random variables (X_1, X_2) if the following hold:

1. $\kappa(X_1, X_2)$ is defined for any random pair (X_1, X_2) with continuous d.f.;
2. $\kappa(X_1, X_2) = \kappa(X_2, X_1)$ for all random pairs (X_1, X_2);
3. $\kappa(X_1, X_2)$ is monotone in $C_{X_1X_2}$, i.e., if $C_{X_1X_2} \succeq C_{W,Z}$, then $\kappa(X_1, X_2) = \kappa(W, Z)$;
4. $\kappa(X_1, X_2) \in [-1, 1]$ for all random pairs (X_1, X_2);
5. $\kappa(X_1, X_2) = 0$ when X_1, X_2 are independent;
6. $\kappa(-X_1, X_2) = -\kappa(X_1, X_2)$ for all random pairs (X_1, X_2);
7. if $(X_1, X_2) \sim F$ and $(X_n^1, X_n^2) \sim H_n$ $n \in \mathbb{N}$, and if $H_n \to H$, then $\lim_{n\to\infty} \kappa(X_n, Y_n) = \kappa(X, Y)$.

Then the measure defined in Definition 3.3.4 satisfies the following theorem.

Theorem 3.3.4. *Let κ be a measure of concordance for continuous random variables X_1 and X_2:*

1. *if g_1 and g_2 are both strictly increasing or decreasing on $RanX_1$ and $RanX_2$, respectively, then $\kappa(g_1(X_1), g_2(X_2)) = \kappa(X_1, X_2)$.*
2. *$\kappa(X_1, X_2) = 1(-1)$ if $X_2 = g(X_1)$ with f almost surely strictly increasing (decreasing) on $RanX_1$.*

[Nelsen, 2006, Theorem 5.1.9] proved that Kendall's tau (3.2) and Spearman's rho (3.7) are two *measures of concordance*. As a consequence of Theorem 3.3.4, we know that Kendall's τ and Spearman's ρ stay in $[-1, 1]$. If X_2 increases with X_1 almost surely, both τ and ρ equal to 1; if X_2 decreases with increasing X_1 almost surly, both τ and ρ equal to -1.

For the relationship between Kendall's tau and Spearman's rho and some other concordance measures, see [Nelsen, 2006] Section 5.

3.4 Conclusion

In this chapter, we have introduced one linear correlation measure, i.e. Pearson's correlation coefficient, and two rank correlation measures, i.e. Kendall's tau and Spearman' s rho. As said that the correlation coefficient is easy to compute and scale-invariant, but not comonotone-invariant. For this reason, it can not be used for nonlinear correlation. Kendall's tau and Spearman's rho are calculated based on the concordance, and could play the role for measuring the nonlinear correlation. For the most used copula families, the closed formulas of these two measures can be obtained.

Chapter 4

Tail Dependence Measures

4.1 Introduction

In finance, extreme values need much attention. As stated in [Bradley and Taqqu, 2003], these events are "very rare but very damaging". A huge loss to a financial institution could make it insolvent or bankrupt and this bankruptcy will result to defaults to his creditor(s). In a bad case and by chain reaction, the defaults will be passed on and it even might cause a troubles in related institutions or in the whole sector, as happened in the last global financial crisis. From the viewpoint of confidence to the market, a financial firm's fall could produce the market fear and panic, which will cause a loss of investor confidence. As a result, underselling and a run on a bank make the market fell further. The adverse impact of the financial blow to the market will be amplified greatly.

To avoid this happening, the financial institutions are required to reserve the capital to cope with the extreme losses. The minimum of capital requirement has been already written down in regulations. For example, Basel II ([BCBS, 2006]) requires that the total capital ratio, that is the percentage of a bank's capital to its risk-weighted assets, "must be no lower than 8%" (read the document for more details). After outbreak of the last financial crisis, the minimum capital requirement to the banks is strengthened in Basel III ([BCBS, 2011]), e.g. "4.5% of common equity (up from 2% in Basel II) and 6% of Tier I capital (up from 4% in Basel II) of 'risk-weighted assets'"[1].

In fact, any financial institution or individual investor have to pay attention to the correlation between the extreme losses of the risky assets. One should avoid to construct a portfolio composing the financial assets which have a great probability of having co-extreme values. For this reason, the study of the dependence of tail, particular the left tail, dependence of the returns appears to be especially important for the portfolio risk management purpose.

Extreme value theory (EVT) is a good approach to model the extreme events in a statistically sound way.It was first used in hydrology, see, for example, [Gumbel and Goldstein, 1964]. In recent years, it has been applied in finance gradually. For some early works, see, for example, [Embrechts, 1997], [Embrechts, 2000], and [Bradley and Taqqu, 2003]. There are several methods are used in univariate EVT, and *Block Maxima Method* and *Peaks Over Threshold Method* are the most employed (for other methods, see [McNeil et al., 2005, Section 7]).

In this chapter, we will show two different kinds of tail (extreme) dependence measures: the coefficient of tail dependence which is the limit of the probability that one random variable assumes extreme values given that the other random variable assumes extreme values, and condition Spearman's rank correlation which is the Spearman's rho between two random variables given that both two assume

[1] http://en.wikipedia.org/wiki/Basel_III

the extreme values. These two measures of the five most frequently used copula families will be calculated in each section. Moreover, we will present the algorithms using the extreme values to estimate the tail dependence measures non-parametrically.

4.2 Tail Dependence Coefficient

The definition of tail dependence that we use originated from [Joe, 1993, 1997]. There are a number of alternative definitions of tail dependence measures, as discussed, for example, in [Coles et al., 1999].

In the case of upper tail dependence, we consider the limit of the probability that X_2 exceeds its q-quantile, given that X_1 exceeds its q-quantile. Obviously the roles of X_1 and X_2 are interchangeable and we will see that it depends only on the copula not on the margins. Formally, we have the following.

Definition 4.2.1. (Coefficient of Upper and Lower Tail Dependence) [McNeil et al., 2005, P.209] Let X_1 and X_2 be r.v.'s with d.f.'s F_1 and F_2. The *coefficient of upper tail dependence* between X_1 and X_2 is defined as

$$\lambda_u := \lambda_u(X_1, X_2) = \lim_{q \to 1^-} \mathbb{P}\left(X_2 > F_2^{-1}(q) | X_1 > F_1^{-1}(q)\right),$$

if the limit exists.

Analogously, the *coefficient of lower tail dependence* is defined as

$$\lambda_l := \lambda_l(X_1, X_2) = \lim_{q \to 0^+} \mathbb{P}\left(X_2 \leq F_2^{-1}(q) | X_1 \leq F_1^{-1}(q)\right),$$

if the limit exists.

The general, *tail dependence coefficient* is denoted by *TDC* for short notation.

X_1 and X_2 are said to *show upper (lower) tail dependence* if $\lambda_u(\lambda_l) > 0$, and *asymptotically independence in the upper (lower) tail* if $\lambda_u(\lambda_l) = 0$.

λ_l and λ_u can be expressed in terms of the unique copula C of the bivariate r.v.'s, if F_1 and F_2 are continuous. By elementary conditional probability, we have

$$\begin{aligned}
\lambda_l &= \lim_{q \to 0^+} \frac{\mathbb{P}\left(X_2 \leq F_2^{-1}(q), X_1 \leq F_1^{-1}(q)\right)}{\mathbb{P}\left(X_1 \leq F_1^{-1}(q)\right)} \\
&= \lim_{q \to 0^+} \frac{F\left(F_1^{-1}(q), F_2^{-1}(q)\right)}{F_1\left(F_1^{-1}(q)\right)} \\
&= \lim_{q \to 0^+} \frac{C\left(F_1\left(F_1^{-1}(q)\right), F_2\left(F_2^{-1}(q)\right)\right)}{F_1\left(F_1^{-1}(q)\right)} \\
&= \lim_{q \to 0^+} \frac{C(q,q)}{q}. \qquad (4.1)
\end{aligned}$$

For upper tail dependence we use (2.10) to obtain

$$\begin{aligned}
\lambda_u &= \lim_{q\to 1^-} \frac{\mathbb{P}\left(X_2 > F_2^{-1}(q), X_1 > F_1^{-1}(q)\right)}{\mathbb{P}\left(X_1 > F_1^{-1}(q)\right)} \\
&= \lim_{q\to 1^-} \frac{\bar{F}\left(F_1^{-1}(q), F_2^{-1}(q)\right)}{\bar{F}_1\left(F_1^{-1}(q)\right)} \\
&= \lim_{q\to 1^-} \frac{\bar{F}_1\left(F_1^{-1}(q)\right) + \bar{F}_2\left(F_2^{-1}(q)\right) - 1 + C\left(F_1\left(F_1^{-1}(q)\right), F_2\left(F_2^{-1}(q)\right)\right)}{\bar{F}_1\left(F_1^{-1}(q)\right)} \\
&= \lim_{q\to 1^-} \frac{1-q+1-q-1+C(q,q)}{1-q} \\
&= \lim_{q\to 1^-} \frac{\widehat{C}(1-q,1-q)}{1-q} \\
&= \lim_{q\to 0^+} \frac{\widehat{C}(q,q)}{q}.
\end{aligned} \tag{4.2}$$

where $\widehat{C}$ is the survival copula of C.

Notice that, $\lambda_l = \lambda_u$ if $C(x_1, x_2) = \widehat{C}(x_1, x_2)$, i.e. if C is radially symmetric .

Tail dependence between two continuous random variables is a copula property, hence the coefficient of tail dependence is invariant under strictly increasing transformations of r.v.'s and does not depend on whatever margins we choose.

The calculation of these coefficients is straightforward if the copula has a simple closed form.

Tail Dependence Coefficients of Normal Variance Mixture Copula[2]

Consider a pair of uniform r.v.'s (U_1, U_2) whose distribution function $C(u_1, u_2)$ is a normal variance mixture copula.

By (4.1) and L'Hôpital's rule, we have

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

By the definition of total derivative, we know that

$$\frac{\mathrm{d}C(q,q)}{\mathrm{d}q} = \lim_{\delta\to 0} \frac{C(q+\delta,q) - C(q,q)}{q} + \lim_{\delta\to 0} \frac{C(q,q+\delta) - C(q,q)}{q}$$

By (2.12), we obtain that

$$\frac{\mathrm{d}C(q,q)}{\mathrm{d}q} = \mathbb{P}\left(U_2 \le q | U_1 = q\right) + \mathbb{P}\left(U_1 \le q | U_2 = q\right).$$

Since a normal variance mixture copula is symmetric and we have that $\mathbb{P}\left(U_2 \le q | U_1 = q\right) = \mathbb{P}\left(U_1 \le q | U_2 = q\right)$. Then,

$$\lambda_l = 2 \lim_{q\to 0^+} \mathbb{P}\left(U_2 \le q | U_1 = q\right). \tag{4.3}$$

[2] For the concepts of the normal variance mixture copula, see [McNeil et al., 2005, Section 5.3].

Due to the radial symmetry of C (see [McNeil et al., 2005, Section 5.1.5]),

$$\lambda_l = \lambda_u.$$

Now we compute the TDC's of two well-known normal variance mixture copulas: Gaussian and t-copulas. We will see that the Gaussian copula shows asymptotically independence in the tail, whereas the t-copula is tail dependent.

Example 4.2.1. **(TDC's of Gaussian copula)** [McNeil et al., 2005, P.211] Let (U_1, U_2) have d.f. C_ρ^{Ga} and

$$(X_1, X_2) \overset{\text{def}}{=\!=} \left(\Phi^{-1}(U_1), \Phi^{-1}(U_2)\right).$$

Therefore (X_1, X_2) has a bivariate normal distribution with correlation ρ and standard normal marginal d.f.. By (4.3), we have that

$$\begin{aligned}\lambda_l &= 2 \lim_{q \to 0^+} \mathbb{P}\left(\Phi^{-1}(U_2) \leq \Phi^{-1}(q) | \Phi^{-1}(U_1) = \Phi^{-1}(q)\right) \\ &= 2 \lim_{x \to -\infty} \mathbb{P}\left(X_2 \leq x | X_1 = x\right).\end{aligned}$$

Following the concepts of conditional normal distribution, we know that

$$X_2 | X_1 = x \sim N(\rho x, 1 - \rho^2).^{3}$$

As a result,

$$\lambda_l = 2 \lim_{x \to -\infty} \Phi\left(x\sqrt{\frac{1-\rho}{1+\rho}}\right) = 0, \quad \rho < 1.$$

Because of the radial symmetry of C_ρ^{Ga}, the Gaussian copula is asymptotically independent in both tails, i.e. $\lambda_u = \lambda_l = 0$, however high a correlation we choose. "Hence, ..., if we go far enough into the tail, extreme events appear to occur independently in each margin"[McNeil et al., 2005].

Example 4.2.2. **(TDC's of t-copula)** Let (U_1, U_2) have d.f. $C_{\nu,\rho}^t$ and

$$(X_1, X_2) \overset{\text{def}}{=\!=} \left(t_\nu^{-1}(U_1), t_\nu^{-1}(U_2)\right),$$

where t_ν denotes the d.f. of a standard univariate t-distribution with ν degrees of freedom.

Thus $(X_1, X_2) \sim t_2(\nu, \mathbf{0}, \mathrm{P})$, where P is a correlation matrix with off-diagonal element ρ. By [Kotz and Nadarajah, 2004] and [Roth, 2013], we know that

$$X_2 | X_1 = x \sim t\left(\nu + 1, \rho x, \frac{\nu + x^2}{\nu + 1}\left(1 - \rho^2\right)\right),$$

or

$$\sqrt{\frac{\nu+1}{\nu+x^2}} \frac{X_2 - \rho x}{\sqrt{1-\rho^2}} \sim t_{\nu+1}. \tag{4.4}$$

By (4.3), we find that

$$\lambda_l = 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right). \tag{4.5}$$

By the radial symmetry of $C_{\nu,\rho}^t$, $\lambda = \lambda_u = \lambda_l$ and always positive for $\rho > -1$. Hence, the copula of the bivariate t-copula always shows the asymptotically dependence in both the upper and lower tail.

[3] https://onlinecourses.science.psu.edu/stat414/node/118

Tail Dependence Coefficients of Archimedean Copula

Now we try to calculate the three most used Archimedean copula families: Frank, Clayton, and Gumbel copulas. Since they are one-parametric, we follow the (4.1) and (4.2) to calculate the TDC's.

Example 4.2.3. **(TDC's of Frank Copula)** Take the Archimedean Frank family copula presented in Example 2.4.5, i.e.

$$C_\theta^{Fr}(u_1,u_2) = -\frac{1}{\theta}\ln\left(1+\frac{\left(e^{-\theta u_1}-1\right)\left(e^{-\theta u_2}-1\right)}{(e^{-\theta}-1)}\right), \quad \theta\in\mathbb{R}.$$

By (4.1) and L'Hôpital's rule, we have that

$$\begin{aligned}
\lambda_l &= \lim_{q\to 0^+}\frac{C_\theta^{Fr}(q,q)}{q}\\
&= \lim_{q\to 0^+}\frac{\mathrm{d}C_\theta^{Fr}(q,q)}{\mathrm{d}q}\\
&= \lim_{q\to 0^+} -\frac{1}{\theta\left(1+\frac{\left(e^{-\theta q}-1\right)^2}{\left(e^{-\theta}-1\right)}\right)}\frac{1}{e^{-\theta}-1}2\left(e^{-\theta q}-1\right)e^{-\theta q}(-\theta)\\
&= \lim_{q\to 0^+}\frac{2e^{-\theta q}\left(e^{-\theta q}-1\right)}{(e^{-\theta}-1)+\left(e^{-\theta q}-1\right)^2}\\
&= 0.
\end{aligned}$$

Before calculating λ_u, we give the following proposition.

Proposition 4.2.1. *C_θ^{Fr} is radially symmetric i.e. $\widehat{C_\theta^{Fr}}(u_1,u_2) = C_\theta^{Fr}(u_1,u_2)$.*

Proof. By (2.10),

$$\begin{aligned}
\widehat{C_\theta^{Fr}}(u_1,u_2) &= u_1+u_2-1-\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{(e^{-\theta}-1)}\right)\\
&= -\frac{1}{\theta}\ln\left(\frac{e^{-\theta u_1-\theta u_2}+e^{-\theta}-e^{-\theta u_1}+e^{-\theta u_2}}{(e^{-\theta}-1)}\right)\\
&= -\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{(e^{-\theta}-1)}\right)\\
&= C_\theta^{Fr}(u_1,u_2)
\end{aligned}$$

□

As a consequence of Proposition 4.2.1,

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

Therefore, the Frank copula shows the asymptotic independence in the tails.

Example 4.2.4. **(TDC's of Clayton Copula)** Take the Archimedean Clayton family copula presented in Example 2.4.6, i.e.

$$C_\theta^{Cl}(u_1,u_2)=\left(u_1^{-\theta}+u_2^{-\theta}-1\right)^{-\frac{1}{\theta}},\quad \theta>0.$$

As a consequence, we know that $C_\theta^{Cl}(q,q)=\left(2q^{-\theta}-1\right)^{-\frac{1}{\theta}}$.

Then the lower and upper tail dependence coefficient can be calculated by (4.1)

$$\lambda_l=\lim_{q\to 0^+}\frac{C_\theta^{Cl}(q,q)}{q}=\lim_{q\to 0^+}\frac{\left(2q^{-\theta}-1\right)^{-\frac{1}{\theta}}}{q}=\lim_{q\to 0^+}\left(2-q^\theta\right)^{-\frac{1}{\theta}}=2^{-\frac{1}{\theta}},$$

and by (4.2), (2.10)

$$\lambda_u=2-\lim_{q\to 0^+}\frac{1-C_\theta^{Cl}(1-q,1-q)}{q}=2-\lim_{q\to 0^+}\frac{1-\left(2(1-q)^{-\theta}-1\right)^{-\frac{1}{\theta}}}{q}.$$

By L'Hôpital's rule,

$$\lambda_u=2-2\lim_{q\to 0^+}\left(2(1-q)^{-\theta}-1\right)^{-\frac{1}{\theta}-1}(1-q)^{-\theta-1}=0.$$

Provided that $\theta>1$, the Clayton copula has lower tail dependence and upper tail independence. The strength of lower tail dependence tends to 1 as $\theta\to\infty$, which is to be expected since the Clayton copula tends to the comonotonicity copula as $\theta\to\infty$ ([Nelsen, 2006]).

Example 4.2.5. **(TDC's of Gumbel Copula)** Take the Archimedean Gumbel family copula presented in Example 2.4.7, i.e.

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

As a consequence, we know that $C_\theta^{Gu}(q,q)=q^{2^{\frac{1}{\theta}}}$.

Then the lower and upper tail dependence coefficient can be calculated by (4.1)

$$\lambda_l=\lim_{q\to 0^+}\frac{C_\theta^{Gu}(q,q)}{q}=\lim_{q\to 0^+}q^{2^{\frac{1}{\theta}}-1}=0,$$

and by (4.2), (2.10), and L'Hôpital's rule,

$$\begin{aligned}\lambda_u=2-\lim_{q\to 0^+}\frac{1-C_\theta^{Gu}(1-q,1-q)}{q}&=2-\lim_{q\to 0^+}\frac{1-(1-q)^{2^{\frac{1}{\theta}}}}{q}\\&=2-\lim_{q\to 0^+}2^{\frac{1}{\theta}}(1-q)^{2^{\frac{1}{\theta}}-1}\\&=2-2^{\frac{1}{\theta}}.\end{aligned}$$

Provided that $\theta>1$, the Gumbel copula has upper tail dependence and lower tail independence. The strength of this tail dependence tends to 1 as $\theta\to\infty$, which is to be expected since the Gumbel copula tends to the comonotonicity copula as $\theta\to\infty$ ([Nelsen, 2006]).

To this point, we have discussed the most used Elliptical and Archimedean copulas: Gaussian, t, Frank, Clayton, and Gumbel. Among them, the Gaussian and Frank copulas do non present the tail dependence at all; however, the Clayton copula has lower tail dependence, and the Gumbel copula shows the upper one, and t-copula is dependent in both lower and upper tails. This characteristic demonstrates the conclusion shown in Figure 2.3 and in Figure 2.6. The characteristic of the tail dependence is useful in risk management. For example, one could choose a suitable copula to generate the extreme values in the corresponding corners.

Tail Dependence Coefficients of Extreme Value Copula

In general, the tail dependence coefficients of the extreme value copulas has a closed formula to compute. It depends on the Pickands dependence function.

Example 4.2.6. (Upper TDC of Extreme Value Copula) Take the extreme value copula given by (2.25), i.e.

$$C(u_1,u_2)=(u_1u_2)^{A\left(\frac{\ln u_1}{\ln(u_1u_2)}\right)},(u_1,u_2)\in(0,1]^2\setminus\{(1,1)\},$$

where A is the Pickands dependence function. Then, the upper tail dependence coefficient is given in [Gudendorf and Segers, 2010] by

$$\lambda_u=2\left(1-A\left(\frac{1}{2}\right)\right). \tag{4.6}$$

In general, λ_u can be estimated by plugging in the Pickands dependence function estimator $\widehat{A}$ into (4.6). The following non-parametric algorithm has been developed in [Frahm et al., 2005, Genest and Segers, 2009]for estimating A with Block Maxima Method(for more details of this method, see[Embrechts, 1997, Durante et al., 2013d]).

Algorithm 4.2.1. (Non-Parametric Estimation of Pickands Dependence Function)

1. Select the extreme value series $(\widetilde{M}_j^1,\widetilde{M}_j^2)_{j=1,\dots,m}$ from $(X_t^1,X_t^2)_{t=1,\dots,T}$ by using Block Maxima Method (See [Embrechts, 1997, Durante et al., 2013d] for details of this method).

2. Calculate the scaled ranks for every $i=1,\dots,m$ by

$$\widehat{U_i}=\frac{1}{m+1}\sum_{j=1}^{m}\mathbf{1}(\widetilde{M}_j^1\le\widetilde{M}_i^1),\qquad \widehat{V_i}=\frac{1}{m+1}\sum_{j=1}^{m}\mathbf{1}(\widetilde{M}_j^2\le\widetilde{M}_i^2).$$

3. For every $i=1,\dots,m$, let

$$\widehat{S_i}=-\log\widehat{U_i},\qquad \widehat{T_i}=-\log\widehat{V_i}.$$

4. Let

$$\widehat{\xi}_i(t)=\begin{cases}\widehat{S_i}, & t=0,\\ \dfrac{\widehat{S}_i}{1-t}\wedge\dfrac{\widehat{T}_i}{t}, & t\in(0,1),\\ \widehat{T_i}, & t=1.\end{cases}$$

Rank-based version of $\widehat{A}_{m,r}$ is then given by

$$\widehat{A}_{m,r}(t) = \frac{m}{\sum_{j=1}^{m} \widehat{\xi}_i(t)}. \tag{4.7}$$

Therefore, the non-parametric upper tail dependence coefficient estimator is

$$\widehat{\lambda_u} = 2\left(1 - \widehat{A}_{m,r}\left(\frac{1}{2}\right)\right). \tag{4.8}$$

The lower tail dependenc coefficient λ_l can be estimated in the same way. In fact, the lower tail set of (U_1, U_2) can be transformed to the upper tail set of $(g(U_1), g(U_2))$, where $g(u)$ is set equal to $1 - u$.

4.3 Conditional Spearman's Rank Correlation

In the previous section, we talked about one tail dependence measure: the coefficients of tail dependence that is the probability of $X2$ having the extreme values given that $X1$ has the extreme values. In this section we propose another measure of dependence, namely the conditional version of Spearman's correlation ρ_B introduced by [Dobric et al., 2014]. ρ_B is the Spearman's rank correlation coefficient ρ of the pair series $[X_1, X_2|(X_1, X_2) \in B]$ (where B is a suitable Borel set included in $\mathbb{R}^2$).

First of all, we define the *tail set* of a bivariate random vector.

Definition 4.3.1. (Tail Set) Let (X_1, X_2) be a bivariate random vector. For every $\alpha_1, \alpha_2 \in (0, \frac{1}{2})$, the *tail set* of (X_1, X_2) is defined as

$$T_{\alpha_1\alpha_2} = [-\infty, q_{X_1}(\alpha_1)] \times [-\infty, q_{X_2}(\alpha_2)],$$

where, q_{X_1} and q_{X_2} are the quantile functions associated with X_1 and X_2, respectively.

The tail set represents the "risky scenario", since it includes the observations related to X_1 (respectively, X_2) that are less than a given threshold. The rank correlation defined on the tail set could be considered as a bivariate tail dependence measure.

In our definition here $B = T_{\alpha_1\alpha_2}$ and $\rho(B)$ is denoted by $\rho(T_{\alpha_1\alpha_2})$ (or $\rho_{T_{\alpha_1\alpha_2}}$).

By Sklar's theorem ([Sklar, 1959]), we know that

$$\mathbb{P}\left((X_1, X_2) \in T_{\alpha\alpha}\right) = F\left(q_{X_1}(\alpha_1), q_{X_2}(\alpha_1)\right) = C(\alpha, \alpha) > 0.$$

Now it is necessary to think about the joint distribution and the marginal distributions conditional on the tail set. We use the subscript T to denote these corresponding distribution functions.

By the conditional probability, we have

$$\begin{aligned}
F_T(x_1, x_2) &\overset{\text{def}}{=\!=} \mathbb{P}(X_1 \le x_1, X_2 \le x_2 | (X_1, X_2) \in T_{\alpha\alpha}) \\
&= \frac{F(x_1 \wedge q_{X_1}(\alpha), x_2 \wedge q_{X_2}(\alpha))}{F(q_{X_1}(\alpha), q_{X_2}(\alpha))} \\
&= \frac{C(F_1(x_1 \wedge q_{X_1}(\alpha)), F_2(x_2 \wedge q_{X_2}(\alpha)))}{C(\alpha, \alpha)}, \quad \forall x_1, x_2 \in \mathbb{R}; \\
F_{1T}(x_1) &\overset{\text{def}}{=\!=} \mathbb{P}(X_1 \le x_1 | (X_1, X_2) \in T_{\alpha\alpha}) = F_T(x_1, +\infty) \\
&= \frac{C(F_1(x_1 \wedge q_{X_1}(\alpha)), \alpha)}{C(\alpha, \alpha)}, \quad \forall x_1 \in \mathbb{R}; \\
F_{2T}(x_1) &\overset{\text{def}}{=\!=} \frac{C(\alpha, F_2\left(x_2 \wedge q_{X_2}(\alpha)\right))}{C(\alpha, \alpha)}, \quad \forall x_2 \in \mathbb{R}.
\end{aligned}$$

By Sklar's theorem again, there exists a copula $C_T : \mathbb{I}^2 \to \mathbb{I}$ such that

$$F_T(x_1, x_2) = C_T(F_{1T}(x_1), F_{2T}(x_2)), \quad \forall x_1, x_2 \in \mathbb{R},$$

or

$$C_T(u_1, u_2) = F_T\left(F_{1T}^{-1}(u_1), F_{2T}^{-1}(u_2)\right), \quad \forall u_1, u_2 \in \mathbb{I}.$$

Since F_{1T} and F_{2T} are continuous, C_T is unique and is called *extreme tail dependence copula* (also called *threshold copula*) relative to C at the level α ([Juri and Wüthrich, 2002]).

Using the lower extreme tail dependence copula C_T and by (3.8b), Spearman's rank correlation coefficient conditional on the tail set $\rho_{T_{\alpha\alpha}}$ is given by

$$\rho_{T_{\alpha\alpha}} = 12 \int \int_{\mathbb{I}^2} u_1 u_2 \, dC_T(u_1, u_2) - 3. \tag{4.9}$$

Notice that the tail set we used before is the lower tail set. The upper tail set could be defined in the same way as the lower one. The concepts related to the upper tail set can be obtained, including the upper tail dependence measure.

In fact, the upper tail set of (X_1, X_2) can be transformed to the lower tail set of $(g(X_1), g(X_2))$ by using a similar transformation which is mentioned in the previous section, i.e. $g(x) = -x$.

After introducing the conditional Spearman's ρ, it is useful to discuss some computational aspects that are related to its practical implementation. Suppose that $(X_t^1)_{t=1,\dots,T}$ and $(X_t^2)_{t=1,\dots,T}$ are iid observation from a continuous random pair (X_1, X_2).

$\rho_{T_{\alpha\alpha}}$ could be approximated by following the algorithm developed in [Dobric et al., 2014].

Algorithm 4.3.1 (Calculation of $\rho(T_{\alpha,\alpha})$)**.**

1. Set $\alpha \in (0, 0.5)$.
2. Calculate the empirical cumulative distribution functions $\widehat{F_{X_1}}$ and $\widehat{F_{X_2}}$ associated with $(X_t^1)_{t=1,\dots,T}$ and $(X_t^2)_{t=1,\dots,T}$, respectively.
3. For any $t = 1, \dots, T$, let $(R_t, S_t) = (\widehat{F_{X_1}}(X_t^1), \widehat{F_{X_2}}(X_t^2))$.
4. Select all the observations in the sets
$$\widehat{T} = \{(R_t, S_t) \mid R_t \leq \alpha, S_t \leq \alpha\}.$$
5. Denote by $\mathscr{I}_T$ (respectively, $\mathscr{I}_M$) the set of all indices t's such that $(R_t, S_t) \in \widehat{T}$.
6. Calculate the univariate empirical cumulative distribution functions $\widehat{F_{T,X_1}}$ and $\widehat{F_{T,X_2}}$ associated with all the observations $(X_t^1, X_t^2)_{t \in \mathscr{I}_T}$.
7. For any index $t \in \mathscr{I}_T$ (respectively, $\mathscr{I}_M$), let
$$(R'_t, S'_t) = (\widehat{F_{T,X_1}}(X_t^1), \widehat{F_{T,X_2}}(X_t^2))$$
8. Calculate Spearman's correlation $\widehat{\rho}(T_{\alpha,\alpha})$ and $\widehat{\rho}(M_{\alpha,\alpha})$ given by
$$\widehat{\rho}(T_{\alpha,\alpha}) = \frac{12}{n_T} \sum_{t \in \mathscr{I}_T} R'_t S'_t - 3$$
where n_T is the cardinality of $\widehat{T}$.

Notice that, by using similar arguments as in [Schmid and Schmidt, 2007], it can be proved that the conditional versions of Spearman's rho described above are consistent and asymptotically normally distributed; i.e.,

$$\sqrt{n_T}(\widehat{\rho}(T_{\alpha,\alpha}) - \rho(T_{\alpha,\alpha})) \xrightarrow{d} N(0, \sigma_T^2)$$

as n_T tend to ∞, provided that the threshold tail copulas exist and satisfy some regularity assumptions. Here σ_T^2 depends on the threshold copulas associated to $T_{\alpha,\alpha}$.

Furthermore, in previous algorithm, the calculation of Spearman's correlation depends on the number of points n_T in tail. If such number is small, the estimated correlations would be affected by this small sample size ([Dobric et al., 2014]). Typically, a convenient sample size may be reached by selecting an appropriate threshold α (in this chapter, if not otherwise specified, we consider the point 0.05 as lower bound for L).

4.4 Conclusion

In this chapter, we have discussed about two measures that can be used to describe the dependence in the tails, particularly the left tail. For tail dependence coefficients, the closed formulas can be developed for the most used copula families, e.g. Gaussian, t, Frank, Clayton, and Gumbel. While, for conditional Spearman's rho, they are still difficult to obtain for these mentioned families. Also, the algorithms for estimating these two measures non-parametrically are presented. One problem to these algorithms is that the time series should be long enough so that there will be adequate number of points in the tails. In other words, for short time series, the algorithms presented will not give accurate estimation of the tail dependence measures.

Chapter 5

Cluster Theory

5.1 Introduction

Cluster analysis is a fundamental, important, and essential method for data mining, particularly for large data sets. By using this method, the data set is divided into several *clusters* (or *groups* or *classes*), such that the objects in the different clusters are dissimilar, but the objects in the same cluster are alike. The dissimilarity between objects can be defined by some kind of distance between them, for instance Euclidean, or by the measure of the correlation, for example correlation coefficient (see, for instance, [Mantegna, 1999]).

[Mirkin, 2005] made a historical review about the development of cluster analysis. It was first used "to study the question of the territorial structure of bird species and its determinants". With the proliferation of computers, high-dimensional matrix operations became possible, resulting that "cluster analysis research grew fast in many disciplines simultaneously". Nowadays, cluster analysis is studied and applied in various fields: data mining, computer science, statistics, genetics, biology, knowledge management, social sciences, economics, finance and so on. Many computer programs have been designed for clustering tasks, e.g. the package *cluster*[1] in R 3.0.1 (64-bit)[2]. Usually these programs carrying out the cluster analysis are based on the dissimilarity matrix which is used to represent a data set. These concepts will be addressed in Section 5.2 of this chapter.

In addition to the dissimilarity matrix, we also need to consider the clustering algorithms. There are many clustering algorithms proposed in the literature. Basically, two kinds of clustering algorithms, namely *partitioning* and *hierarchical* are applied. For each algorithm, there are also several methods, e.g. k-means, k-medoids, fuzzy analysis for partitioning algorithm, and divisive, agglomerative for hierarchical algorithm. The choice of a clustering algorithm depends both on the type of data under consideration and on the goal of the investigations. Sometimes, more than one method can be implemented, hence one possible way is to run all of these algorithms and then select one on the basis of some criterion. The various cluster methods will be introduced in Section 5.3 of this chapter.

One important issue in cluster analysis is the determination of the optimal number of different clusters. The silhouette introduced by [Rousseeuw, 1987] may be helpful in determining what such a number is. The silhouette is usually represented via a plot showing the classification goodness of the objects belonging to one cluster. The number which maximizes the average silhouette width for the entire data set is usually taken as the best number of clustering. More details will be presented in last section of this chapter.

[1] [Maechler et al., 2013]
[2] [R Core Team, 2013]

5.2 Expression of the Data

Assume there is a data set containing n objects to be clustered. These objects may be persons, companies, countries, and so on. There are p variables to be measured for each object. These variable are the characters of the objects, such as gender, profession, nationality, and etc.. Let x_{if} be the measurement of object i to variable f,where $i = 1, \ldots, n$ and $f = 1, \ldots, p$.

Two kinds of matrices can be used to present the data set and are typically operated by main clustering algorithms: data matrix and dissimilarity matrix.

5.2.1 Data Matrix

The *data matrix* is an $n \times p$ order matrix, which shows the most original results without any treatment and demonstrates visually the value of each object to each variable. Usually, it looks like

$$\begin{pmatrix} x_{11} & \ldots & x_{1f} & \ldots & x_{1p} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_{i1} & \ldots & x_{if} & \ldots & x_{ip} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_{n1} & \ldots & x_{nf} & \ldots & x_{np} \end{pmatrix}. \tag{5.1}$$

Sometimes, if the p variables are interval variables, the data matrix can be used to do clustering analysis directly.

5.2.2 Dissimilarity Matrix

A *dissimilarity matrix*, namely is an $n \times n$ matrix $\Delta = (\delta_{ij})$, where its entry δ_{ij} is the measurement of the *dissimilarity* between objects i and j.

Sometimes, some kind of distance between object j and object j is used as the dissimilarity, e.g. Euclidean distance

$$\delta_{ij} = \sqrt{\sum_{f=1}^{q} (x_{if} - x_{jf})^2} \tag{5.2}$$

or Manhattan distance

$$\delta_{ij} = \sum_{f=1}^{q} |x_{if} - x_{jf}|, \tag{5.3}$$

or some other measures else.

Basically, the dissimilarity δ_{ij} usually satisfies the following properties. For every $i, j = 1, \ldots, n$:

(i) $\delta_{ij} \geq 0$ (*Non Negative*);

(ii) $\delta_{ij} = \delta_{ji}$ (*Symmetric*);

(iii) $\delta_{ii} = 0$ (*Identity*).

The assumption of non-negativity of the entries is logical since there is either no or some difference between object i and object j. δ_{ij} is small (close to zero) when object i and object j are very similar or "near" to each other, and becomes more larger when i and j are more different. We shall usually assume that dissimilarities are symmetric and that an object is not different from itself. Therefore, the dissimilarity matrix is a symmetric nonnegative matrix with the zero main diagonals.

5.3 Clustering Algorithms

Basically, for every cluster analysis, we need to consider two fundamental questions: one is how to determine k, i.e. the number of the clusters; the other one is, for given k, how to carry out the partition. In this section, we focus on the second question, i.e. clustering algorithms, and will leave the first question to next section.

Many clustering algorithms exist in the literature. The choice of a clustering algorithm depends both on the type of data under consideration and on the aim of the investigation. However, the final goal is to find "good" partitions, in the sense that "objects of the same cluster should be close or related to each other, whereas objects of different clusters should be far apart or very different"([Kaufman and Rousseeuw, 1990]).

In this section, we consider two kinds of clustering algorithms, namely *partitioning* and *hierarchical*, together with their most important characteristics. For the details of other cluster methods, please see [Han et al., 2006, Chapter 7] and [Kaufman and Rousseeuw, 1990, Chapter 3].

5.3.1 Partitioning Method

Given a dataset of n objects, a partitioning method constructs $k(\leq n)$ clusters, such that

(1) each cluster must contain at least one object;

(2) each object must belong to exactly one cluster.

The commonly used partitioning methods are k-means, k-medoids, fuzzy clusters, and their variations. We will introduce them in turn.

k-Means Method

In general, a *k-Means* cluster method groups the n objects under consideration into k clusters, such that the objects belonging to the same cluster are "similar" and the objects in different clusters are "different". The similarity between the objects are measured by the dissimilarity or the distance between them.

The k-means method is carried out as following

Algorithm 5.3.1. **(k-Means)**

1. Choose k objects randomly as the cluster means.
2. For each of the remaining objects, select the nearest (smallest dissimilarity) cluster mean and assign this object to the corresponding cluster.
3. Calculate the mean value of the objects in a cluster as the new means.
4. Compute the square-error criterion, which is given by

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} |\boldsymbol{p} - \boldsymbol{m_i}|^2, \tag{5.4}$$

   where $\boldsymbol{m_i}$ is the mean of cluster C_i.
5. If E converges, stop. If not, return to 2.

The criterion in Algorithm 5.3.1 is the sum of the square distances between the cluster mean to the objects belonging to this cluster. It makes the objects compact "within" the clusters and separate "between" clusters.

k-Medoids Method

The *k-medoids method* is similar with the k-means method, in sense that the implementation of k-medoids method also follows the Algorithm 5.3.1. One difference between these two methods is that the k-medoids method does not take the mean of the objects but a real object to represent a cluster. These selected objects are called *medoids* of the corresponding clusters. Moreover, k-medoids use the sum of the dissimilarities not of the square dissimilarities.

The k-medoids method is carried out as following

Algorithm 5.3.2. (k-Medoids)

1. Choose k objects randomly as the cluster medoids.
2. For each of the remaining objects, select the nearest (smallest dissimilarity) cluster medoid and assign this object to the corresponding cluster.
3. Choose one of the objects belonging to the same cluster as the new medoid such that the sum of the dissimilarities between the other objects of the cluster and this object is minimum.
4. Compute the absolute-error criterion, which is given by

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} |\boldsymbol{p} - \boldsymbol{o_i}|, \tag{5.5}$$

   where $\boldsymbol{o_i}$ is the medoid of cluster C_i.
5. If E converges, stop. If not, return to 2.

Note that the k-medoids method is more robust than the k-means algorithm in the presence of outliers, because the mean is more influenced by an object with an extremely large value. This effect is particularly exacerbated due to the use of the square-error function.

Fuzzy Analysis

A fuzzy analysis ia a "soft" clustering algorithm, which does not clearly indicate to which cluster an object belongs. Instead, it provides the membership degree of each object to one cluster, e.g. "object i belongs for 80% to cluster 1, for 10% to cluster 2, and for 10% to cluster 3". These percentages are exactly the *membership coefficients*.

An important advantage of the fuzzy analysis is to describe the "ambiguous situations". Sometimes, a "hard" cluster algorithm forcing to assign an object to some cluster may mislead the judgment, in contrast, the fuzzy analysis shows that the object most likely belongs to which cluster, e.g. cluster 1, but also to tell you all the favors of other clusters, e.g. cluster 2 and cluster 3.

However, the calculation of membership coefficients is too complex to explain here (for more details, see [Kaufman and Rousseeuw, 1990, Han et al., 2006]). Another disadvantage is the heavy outcome. The fuzzy analysis provides the membership coefficients of each object to each cluster. However, if we always assign an element to a cluster based on its maximal membership coefficient, we may use it as a hard (crisp) cluster algorithm.

5.3.2 Hierarchical Method

Roughly speaking, the hierarchical method does the cluster analysis without specifying a priori the number of clusters. It starts from $k = 1$ cluster (all objects are together in the same cluster) or $k = n$ clusters (each object forms a separate cluster with only a single element). Then, at each step of the algorithm, either one of the r clusters is divided, or two of the $r + 1$ clusters are merged. The constantly splitting process is called *divisive*, and the constantly merging process is called *agglomerative*. Therefore, the outcome the hierarchical algorithm looks like a tree, and *dendrogram* is commonly used to display this process, e.g. Figure 5.1.

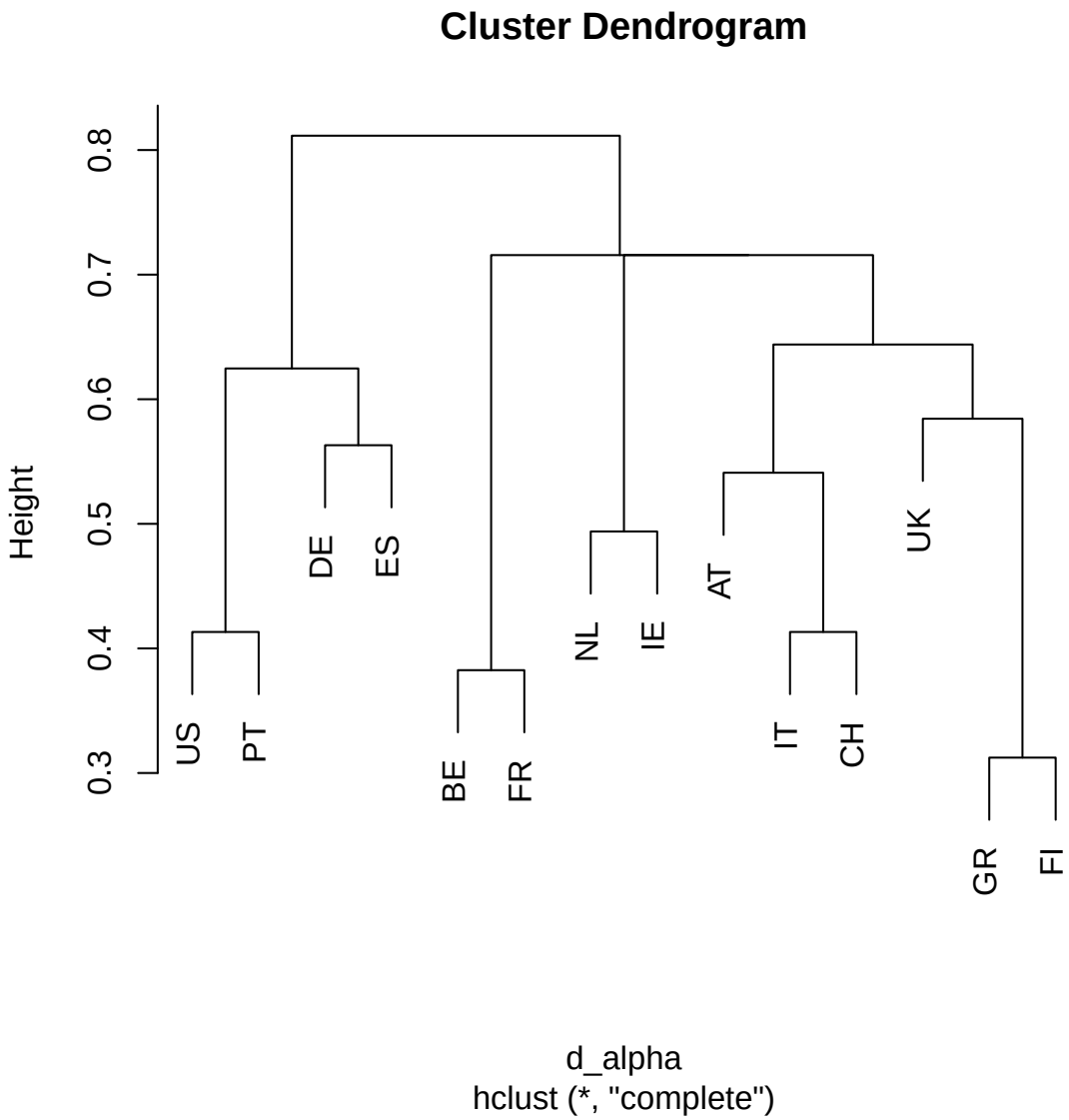


Figure 5.1. A dendrogram of hierarchical algorithm. By the height, we can conclude the order of merging of the objects. Moreover, we can also obtain the cluster compositions.

Agglomerative Hierarchical Method

The agglomerative method is a bottom-up strategy, which starts when all objects are apart (that is, at step 0 we have n clusters). In the first step, two nearest objects, that have the smallest dissimilarity, are grouped in one cluster. In the next steps, two "closest" clusters are merged to one larger cluster, until all of the objects are in a single cluster.

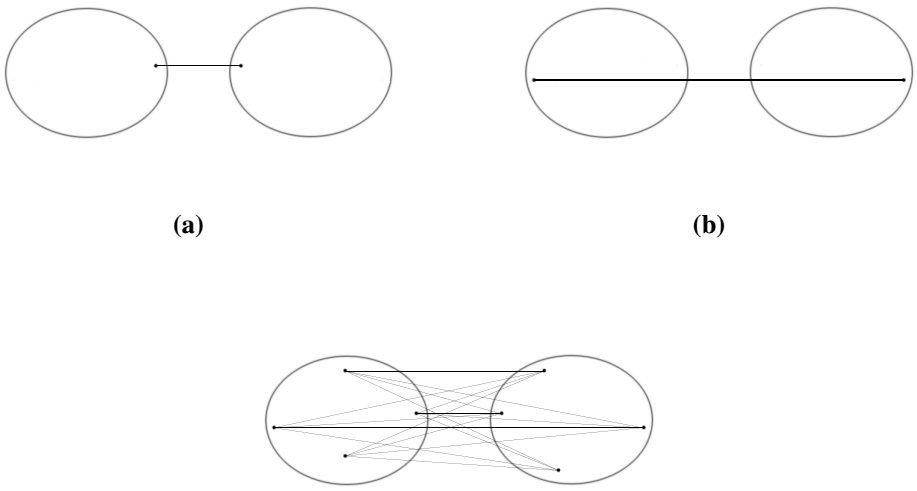


Figure 5.2. Representation of some definitions of cluster-distance based on the dissimilarities between their objects: (a) single linkage (nearest neighbor). (b) complete linkage (farthest neighbor). (c) average linkage.

Three of the most used definitions of cluster-distance based on the dissimilarities between their objects are listed in the following.

- **Single Linkage (Nearest Neighbor)**
  The distance between two clusters A and B is defined by the minimum distance between the objects in A and the objects in B, i.e.

$$D(A, B) = \min_{i \in A, j \in B} \delta_{ij}, \tag{5.6}$$

  where δ_{ij} is the dissimilarity between objects i and objects j.

- **Complete Linkage (Farthest Neighbor)**
  The distance between two clusters A and B is defined by the maximum distance between the objects in A and the objects in B, i.e.

$$D(A, B) = \max_{i \in A, j \in B} \delta_{ij}. \tag{5.7}$$

- **Average Linkage**
  The distance between two clusters A and B is defined by the average distance between all the

objects in A and all the objects in B, i.e.

$$D(A,B) = \frac{1}{n_A n_B} \sum_{i \in A} \sum_{j \in B} \delta_{ij}. \tag{5.8}$$

where n_A and n_B are the number of objects in A and B, respectively.

Divisive Hierarchical Method

The divisive method is a top-down strategy which starts from one cluster containing all n objects. In each of the next steps, one cluster is split up two until every object forms a cluster. This method is more complex than agglomerative. It needs to decide which cluster will be split up, also how to be split in sense of the arrangement of the objects. For more details, see [Kaufman and Rousseeuw, 1990].

Summarizing, hierarchical methods use a computationally efficient technique to attempt to find a "good" partition without examining all possible clustering possibilities for a data set, especially a large one ([Seber, 1984]). The agglomerative and the divisive construct their hierarchy in the opposite direction, possibly yielding quite different results.

However, this is also the disadvantage. "A hierarchical method suffers from the defect that it can never repair what was done in previous steps"([Kaufman and Rousseeuw, 1990]). Once a group is divided into two parts, they cannot be merged together any more. Moreover, the two groups combined together cannot be separated any longer.

5.4 Number of Clusters

In the previous sections, we have introduced the dissimilarity matrix and two clustering algorithms. Notice that, whichever algorithm will be implemented, the number of clusters, k, needs to be decided.

The silhouette introduced by [Rousseeuw, 1987] may be helpful to determine what the k is. According to his work, each cluster is represented by one silhouette, which is a plot showing the "within–between dissimilarity" ratio s of the objects contained in this cluster. This ratio is defined as following.

Assume that object i is in Cluster A. The *within dissimilarity* of object i is the average of the dissimilarities between itself and the objects belonging to A, i.e.

$$a(i) = \frac{1}{n_A} \sum_{j \in A} \delta_{ij},$$

where n_A is the number of objects contained in Cluster A. The *between dissimilarity* of object i is the minimum average of the dissimilarities between itself and the objects belonging to any other cluster which is different from A, i.e.

$$b(i) = \min_{C \neq A} \left\{ \frac{1}{n_C} \sum_{j \in C} \delta_{ij} \right\}.$$

where n_C is the number of objects contained in Cluster C. Then the *within–between dissimilarity ratio* s of object i is defined by

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}. \tag{5.9}$$

This ratio describes the appropriateness of object i included in Cluster A. When $s(i)$ is close to 1, this means that the within dissimilarity $a(i)$ is much smaller than the smallest between dissimilarity $b(i)$, and probably we could say that the classification of i is correct. If $s(i)$ is near to 0, i.e. $a(i) \approx b(i)$, putting object i in Cluster A is not very different from putting it in the between cluster. The worst case is that $s(i)$ is almost equal to -1, i.e. $a(i) > 0$ and $b(i) = 0$, probably i should be in the between cluster but not in A.

Then the silhouette of A is built by plotting $s(i)$'s in decreasing order. Through this plot, which object is classified well in this cluster, which one is the "intermediate" object, and which is should be assigned in another cluster are clearly demonstrated.

The average of the within–between dissimilarity ratio of all the objects belonging to one cluster is called *average silhouette width*, given by

$$s(A) = \frac{1}{n_A} \sum_{i \in A} s(i),$$

where n_A is the number of the number of the objects containing in Cluster A.

The average of within–between dissimilarity of all the objects of the data set is called *average silhouette width for the entire data set*, denoted by $\bar{s}$ and given by

$$\bar{s} = \frac{1}{n} \sum_{i=1}^{n} s(i),$$

where n is the number of all the objects of the dataset. In fact, this average depends on k, i.e. the number of the clusters. Whether the clustering algorithm was partitioning or hierarchical, different k will lead to different clusters, that, in general, will result different $\bar{s}$. Therefore, strictly speaking, it should be $\bar{s}(k)$.

Usually, this average silhouette width can be used to determine the best number of clusters. In general, a suitable k should produce the "good" cluster, in sense that the objects belonging to the same cluster are "similar" and the objects in different clusters are "different". As a result, the within–between dissimilarity ratio for each object should be a large positive number, and also $\bar{s}$ is. Therefore, the value k, which maximizes $\bar{s}$, should be taken as the optimal number of cluster, i.e.

$$\widetilde{k} = \max_{k} \bar{s}(k).$$

5.5 Conclusion

In this chapter, we have given a brief introduction to the cluster theory, including the expression of the data set, the clustering algorithms, and the determination of the optimal number of cluster. The clustering method is good at finding the commons among the objects, and dividing them into groups according to these commons. For this advantage, it is most applied in data mining, particularly in computer science and social science.

When we employ these algorithms, especially the hierarchical, we need to keep in mind that the cluster analysis is done in the same run and "it can never repair what was done in previous steps." Moreover, we should find some other evaluation criterion of the cluster analysis and of the determination of the cluster numbers for different clustering algorithms. One rule for all is not appropriate.

Chapter 6

A Portfolio Diversification Strategy via Tail Dependence Measures

6.1 Introduction

In recent years financial markets have been characterized by an increasing globalization and a complex set of relationships among asset returns. Moreover, it has been recognized that the linkages among different assets vary across time and that their strength tends to increase, especially during crisis periods [Billio et al., 2012]. The presence of a stronger dependence when markets are experiencing losses is of utmost interest from a risk management perspective. In fact, in portfolio risk analysis a usual practice for minimizing the whole risk consists of adopting some diversification techniques, an issue widely debated since the seminal work of [Grubel, 1968]. Namely, it has been recognized that investors can reduce the risk of their portfolios through allocating their investments in various classes of financial instruments and/or categories of assets that would move in different ways in response to the same event. In other words, diversification benefits can be achieved when the comovements among the assets is taken into account. Therefore, the portfolio diversification issue naturally poses the question of investigating the relationship between financial time series and of checking whether they can be grouped together in a way that may be helpful to portfolio selection.

To provide a suitable way to diversify a portfolio, taking into account the occurrence of extreme scenarios, clustering techniques for multivariate time series have been proposed in the literature (see, e.g., [Levy and Sarnat, 1970, Panton et al., 1976]), mainly based on indices of global dependence like the Pearson correlation coefficient (see, e.g., [Mantegna, 1999, Tola et al., 2008]). However, it has been recognized that the presence of dependence in the tail of the joint distribution of the markets can mitigate the diversification effects precisely when they are needed most [Forbes and Rigobon, 2002, Bradley and Taqqu, 2004]. Thus, different clustering techniques have been recently applied in order to group financial time series that are similar in extreme scenarios. Such an approach includes clustering techniques based on the tail dependence coefficient (see, e.g., [De Luca et al., 2010, De Luca and Zuccolotto, 2011] and [Durante et al., 2013d]), or conditional measure of association, like Spearman's correlation, as done in [Durante et al., 2013c].

The aim of this chapter is to exploit such recently introduced clustering methods in order to develop a procedure to select a weighted portfolio in a group of assets. The whole methodology is intended to be used by an investor to have more insights into the relationships among different assets in crisis periods. In particular, it may serve to warn against the automatic use of standard portfolio selection procedures that may not work when the markets are expected to experience periods of high volatility.

The chapter is organized as follows. Section 6.2 presents the main properties of the used cluster procedures, while Section 6.3 illustrates the novel methodology via an empirical application. Section 6.4 concludes.

6.2 The Clustering Procedure

As it has been discussed in Chapter 5, clustering procedures are based on the choice of a suitable dissimilarity matrix that expresses the relations among the time series under consideration. Traditionally, correlation–based clustering has been used to find groups in a portfolio of financial assets from the correlation matrix: see, for instance, [Mantegna, 1999, Bonanno et al., 2004]. However, as known, Pearson's correlation is not a convenient dependence measure to be used outside the Gaussian (elliptical) distributions and, in particular, it does not give an accurate indication and understanding of the real dependence between risk exposures [Embrechts et al., 2003].

In order to give an accurate estimation of the link between assets in the tail of the distribution, suitable measures of tail dependence may be used instead, as suggested, for instance, in [De Luca and Zuccolotto, 2011]. Following these ideas, different clustering procedures have been implemented in [Durante et al., 2013c,d] in order to group time series according to their tail behaviour. These procedures are summarized below (for more details, please refer to the original papers).

Consider a matrix of d financial time series $(x_{it})_{t=1,\dots,T}$ $(i = 1, 2, \dots, d)$ representing the log–returns of different financial assets. We assume that each time series $(x_{it})_{t=1,\dots,T}$ is generated by the stochastic process $(\mathbf{X}_t, \mathscr{F}_t)$ such that, for $i = 1, \dots, d$,

$$X_{it} = \mu_i(\mathbf{Z}_{t-1}) + \sigma_i(\mathbf{Z}_{t-1})\varepsilon_{it}, \tag{6.1}$$

where $\mathbf{Z}_{t-1}$ depends on $\mathscr{F}_{t-1}$, the available information up to time $t-1$, and the innovations ε_{it} are distributed according to a distribution function F_i for each t. Moreover, the innovations ε_{it} are assumed to have a constant conditional distribution F_i (with mean zero and variance one, for identification) such that for every t the joint distribution function of $(\varepsilon_{1t}, \dots, \varepsilon_{dt})$ can be expressed in the form $C(F_1, \dots, F_d)$ for some copula C. Such a general model includes many multivariate time series models presented in the literature (see, for instance, [Patton, 2013]).

Then the following steps are implemented in order to group the time series into sub-groups such that elements in each sub-group have strong tail dependence.

1. Choose a copula-based time series model (e.g. ARMA-GARCH copula model) in order to model separately the marginal behavior of each time series and the link between them.

2. Estimate the (pairwise) tail dependence measure among the time series.

3. Define a dissimilarity matrix by using the information contained in the tail dependence measures and apply a suitable cluster algorithm for grouping time series according to the tail behavior.

The steps are described below in detail.

6.2.1 Univariate Time Series Fitting

We fit an appropriate ARMA-GARCH model to each univariate time series by obtaining the estimates $\widehat{\mu}_i$ and $\widehat{\sigma}_i$ of the conditional mean and variance of these processes according to Eq. (6.1). Classical model selection procedures (e.g., Akaike Information Criteria) and goodness-of-fit tests of homoscedasticity

and uncorrelatedness of the residuals may confirm the adequacy of the fit. Moreover, using the estimated parametric models, we construct the standardized residuals given, for each $i = 1, \ldots, d$, by

$$\widehat{\varepsilon}_{it} = \frac{x_{it} - \widehat{\mu}_i(\mathbf{Z}_{t-1})}{\widehat{\sigma}_i(\mathbf{Z}_{t-1})}.$$

The standardized residuals are converted to the pseudo–observations $z_{it} = F_i(\widehat{\varepsilon}_{it})$, where F_i may be estimated from a parametric model (Gaussian, Student–t, etc.) or by using the empirical distribution function.

As a result, $(z_{1t}, \ldots, z_{dt})_{t=1,\ldots,T}$ contains the information about the link (i.e. the copula) among the time series under consideration and, as such, it can be used in order to compute dependence measures or make inference about the copula of the time series. Notice that, if the marginal model is correctly specified, $(z_{1t}, \ldots, z_{dt})_{t=1,\ldots,T}$ forms (asymptotically) a random sample generated by the copula C. As such, dependence measures calculated from this sample are not biased by serially dependence and/or heteroscedasticity (see, for instance, [Patton, 2013]).

6.2.2 Tail Dependence Measure Estimation

Given the pseudo–observations from the original time series, in order to quantify the degree of tail dependence of a random vector (X, Y), we adopt two different measures:

- the lower tail dependence coefficient λ_L (shortly, TDC) [Joe, 1997], which, as been seen in chapter 4, is a measure that depends only on the copula C linking the variables under consideration via the formula
$$\lambda_L(C) = \lim_{t \to 0^+} \frac{C(t,t)}{t}. \tag{6.2}$$
- the conditional Spearman's correlation ρ_α that expressed the Spearman's correlation of (X, Y) given that X and Y are both under their α–quantile (for instance, $\alpha = 0.10$).

These measures express two different ways of looking at tail dependence since they focus, respectively, on asymptotic tail dependence (λ_L) and finite tail dependence (ρ_α). As regards the estimation of these quantities we rely on two specific techniques:

- The estimator of tail dependence coefficient is based on the procedure proposed in [Frahm et al., 2005, Section 3.5]). Specifically, by using the results of [Genest and Segers, 2009], we obtain an estimate of the extreme-value copula in the domain of attraction of C, from which the estimate of the TDC of C is derived.
- The estimator of conditional Spearman's ρ_α is based on the procedure described in [Dobric et al., 2014, Durante and Jaworski, 2010] that is related to the calculation of the Spearman's correlation in a sub-sample extracted from the pseudo–observations and dependent on the threshold α.

Both these estimations are obtained via non-parametric procedures and do not require any parametric estimation of the unknown copula linking the time series of interest. For more detail about the practical calculations of these coefficients, we refer to [Durante et al., 2013c,d].

6.2.3 The Clustering Algorithm

Once a measure of tail dependence has been computed for all pairs extracted from the time series, we have to transform it through a monotonic function f in such a way that the obtained dissimilarity

between two time series is small when their tail dependence is high, and monotonically increases when their tail dependence decreases. Thus, for $i, j = 1, \ldots, d$, we define $\Delta = (\Delta_{ij})$ whose elements are given by

$$\Delta_{ij} = \sqrt{2(1 - \widehat{m}_{ij})}, \tag{6.3}$$

where $\widehat{m}_{ij}$ is the tail dependence measure between time series i and j, that is estimated via one of the two above procedures.

Starting from the dissimilarity matrix defined in (6.3) we can perform a cluster analysis of the time series by different techniques. Here, for a comparative analysis, we focus on two methods:

- The hierarchical agglomerative algorithms start from the finest possible partition (i.e. each observation forms a cluster) and, hence, each level merges a selected pair of clusters into a new cluster according to the definition of the distance between two groups. Among all the agglomerative strategies we apply the complete linkage, which defines the distance between two clusters as the maximum distance between their individual components, and give overall good performances.

- Fuzzy clustering algorithm is a partitioning method that takes into account some ambiguity in the data, which often occurs. In fact, it allows each object to belong to one or more than one cluster according to a membership coefficient, that quantifies the degree of belonging of each object to a specific group. The main advantage of fuzzy clustering is that it yields much more detailed information on the data structure compared to other partitioning techniques. In order to perform fuzzy cluster analysis we can consider FANNY algorithm [Kaufman and Rousseeuw, 1990], which handles either interval-scaled measurements or dissimilarities. The algorithm aims at minimizing an objective function which is a kind of total dispersion, depending on dissimilarities and membership coefficients. Once the number of clusters is chosen, the algorithm returns some general information on the type of data and the actual membership for each object in each cluster are listed. Moreover, as in our case, a crisp partition of the financial assets can be determined from the membership value of each time series.

In both methods, the optimal number of clusters is chosen by the silhouette index [Kaufman and Rousseeuw, 1990, Hastie et al., 2009], which reflects the within-cluster compactness and between-cluster separation of a clustering.

A remark is needed here. In several cases (see, for instance, [Nanda et al., 2010]) the clustering algorithms are applied directly to the time series data without any previous step (pre-filtering). For our purposes, such an approach could not be applied since it can produce biased results. In fact, it has been recognized that, in order to obtain a realistic description of the dependence among assets, univariate heteroscedastic effects should be taken into account and treated in a suitable way (see, for instance, [Forbes and Rigobon, 2002]).

6.3 The Portfolio Selection

Once the clustering procedure is completed, the assets have been grouped into k clusters, where k is a predefined number. Then our possible portfolio will be selected on the basis of the following strategy:

1. Determine all possible portfolios composed by k assets such that each asset belongs to a different cluster.

2. For these portfolios, calculate the optimal weight assigned to each of its assets with the classical Markowitz portfolio selection procedure [Markowitz, 1952]. We recall that this procedure provides a general way to maximize investor's expected utility under certain conditions, namely to produce portfolios that are able to minimize the total portfolio variance.

3. Given all possible portfolios composed in such a way, plot the graph of their standard deviation against their expected return.

4. Determine the portfolios that are the vertices in the convex efficient frontier of the standard deviation/expected return graph.

According his/her preference the investor could hence choose one of the portfolios that are on the convex frontier.

In other words, the procedure has the following aspects:

- It suggests to select the assets of the portfolio by taking into account the grouping structure given by the clustering algorithms. Thus, two assets from the same group (cluster) cannot be chosen in the same portfolio.

- Once the assets have been selected, their weights are determined by classical methods, like the Markowitz approach.

- All the portfolios composed in the previous two steps provide a graphical representation of the possible choices of the investor (see, for instance, Figure 6.4). Based on his/her information, one investment strategy could be selected.

- If no preference is required by the investor, the point with the smallest risk on the convex frontier, namely the global minimum variance portfolio, can be chosen.

A remark is needed here. As clarified above, the idea of diversification by grouping assets is not new [Panton et al., 1976]. The further step of selecting assets taking into account group constraints and determining weights via Markowitz's approach has been used, for instance, in [Hui, 2005]. In the latter reference, however, the groups are determined by a different methodology, that is factor analysis. Finally, [De Luca and Zuccolotto, 2011] suggested the idea of calculating all possible portfolios with group constraints; however, again, their clustering procedure is different since it assumes a parametric form of the dependence structure (compare with [Durante et al., 2013d]).

[Cesarone et al., 2013] provided another way to solve the portfolio selection problem discussed in this chapter. In their work, they proposed a new method for the "Limited Asset Markowitz" model, where the assets are limited with the introduction of quantity and cardinality constrain, i.e. the number of the assets constructing the portfolios is not greater than $k(\leq N)$, and the weights of these assets are

in a range of $[l, u]$. The formulation of this model is

$$\begin{aligned} \min \quad & \sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}x_i x_j && \text{(P1)} \\ \text{st.} \quad & \sum_{i=1}^{n}\mu_i x_i = \rho && (6.4) \\ & \sum_{i=1}^{n} x_i = 1 \\ & x_i = 0 \quad \text{or} \quad l_i \leq x_i \leq u_i, \quad i = 1, \ldots, n \\ & |\text{supp}(x)| \leq k, \end{aligned}$$

where $\text{supp}(x) = \{i : x_i > 0\}$. This model can be reformulated as a Mixed Integer Quadratic Program (MIQP) ([Bienstock, 1996, Chang et al., 2000, Jobst et al., 2001]). [Cesarone et al., 2013] reduced P1 (Eq. (4) in their work)to a standard quadratic programming problem and resolved this problem by proposing a new algorithm: "INCREASING SET ALGORITHM".

In our case, the portfolio selection problem can also be extended to a MIQP by adding several binary variables and constraints, assuming that the n assets are divided in k clusters, the length of Cluster ℓ is k_ℓ, and the indices of the assets in Cluster ℓ is I_ℓ, i.e.

$$\begin{aligned} \min \quad & \sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}x_i x_j && \text{(P1)} \\ \text{st.} \quad & \sum_{i=1}^{n}\mu_i x_i = \rho && (6.5) \\ & \sum_{i=1}^{n} x_i = 1 \\ & \sum_{i=1}^{n} y_{il} = 1, \qquad i = 1, \ldots, n, l \in I_\ell \\ & 0 \leq x_i \leq u_i y_{il}, \quad i = 1, \ldots, n \\ & y_{il} \in \{1, 0\}, \qquad i = 1, \ldots, n, l = 1, \ldots, k \end{aligned} \tag{6.6}$$

This problem could also be reduced to a standard quadratic problem. With the algorithm proposed in [Cesarone et al., 2013], we could obtain the global minimum risk portfolio and the efficient frontier.

Here, instead of going directly to the implementation of the optimization procedure described before, we compute all possible portfolios with the constraints determined by the clustering procedure and, hence, we point the optimal portfolio in the convex frontier. In fact, such an (admittedly longer) approach not only allows to point out the minimal risk portfolio, but also has the advantage to leave the risk manager the freedom to make the decision. Obviously, it needs much time with respect to the method of [Cesarone et al., 2013]. The same choice was adopted by [De Luca and Zuccolotto, 2011], who showed for the first time the advantage of adopting clustering techniques in portfolio selection. It is worth stressing, however, that if we want to compare the portfolios of different indices or construct a very large portfolio (e.g Russell 3000 (2151 assets)), we could employ Cesarone et al.'s methodology which is strongly preferred.

6.3.1 Illustration of the Portfolio Selection Procedure

We consider time series related to EURO STOXX 50 stock index and its components in the period from January 2, 2003 to July 31, 2011. Moreover, as out-of-sample period, we will also show the performance of our procedure in the period from August 1, 2011 to September 9, 2011. The period has been selected due to the fact that EURO STOXX 50 was experiencing severe losses in the period (see Figure 6.8). This data is available from the database ***Datastream***.

"The EURO STOXX 50 Index, Europe's leading Blue-chip index for the Eurozone, provides a Blue-chip representation of supersector leaders in the Eurozone. The index covers 50 stocks from 12 Eurozone countries: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain, traded in 6 Eurozone stock exchange: Brussels Stock Exchange (BSE), Paris Bourse, Frankfurt Stock Exchange(FWB), Amsterdam Stock Exchange, Madrid Stock Exchange, Irish Stock Exchange, Milan Stock Exchange." [1]

Model Fitting to the Marginal Series

We preliminary apply a univariate t-Student ARMA(1,1)–GARCH(1,1) model:

$$\begin{aligned} X_t &= \mu + a_1 r_{t-1} + \epsilon_t + b_1 \epsilon_{t-1}, \\ \epsilon_t &= \sigma_t \gamma_t, \\ \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \\ &\text{where } \gamma_t \text{ is white noise.} \end{aligned}$$

to each time series of log–returns of 50 constituents of the index to remove autocorrelation and heteroscedasticity from the data and we computed the standardized residuals in order to check the adequacy of the fit.

Table 6.1. Estimation of the parameters of the t-Student ARMA(1,1)–GARCH(1,1) model

Asset	μ	a_1	b_1	ω	α_1	β_1	γ_1	DF
E.IND	0.00	0.70	-0.76	0.00	2.2E-02	0.94	0.07	5.70
H.ASML	0.00	0.73	-0.75	0.00	1.2E-04	0.97	0.05	7.65
I.ENEL	0.00	-0.25	0.21	0.00	2.9E-02	0.89	0.13	5.70
B.ABI	0.00	0.77	-0.83	0.00	7.0E-03	0.94	0.07	5.48
F.BNP	-0.00	-0.15	0.17	0.00	3.6E-03	0.93	0.14	8.64
E.REP	-0.00	-0.88	0.91	0.00	4.0E-02	0.88	0.11	6.06
H.ING	-0.00	-0.34	0.41	0.00	1.8E-02	0.89	0.19	9.75
D.DAIX	0.00	0.24	-0.21	0.00	4.4E-02	0.92	0.06	6.55
E.SCH	-0.00	-0.91	0.92	0.00	1.7E-02	0.88	0.20	10.13
M.NOK1	0.00	0.66	-0.67	0.00	5.0E-02	0.96	-0.03	3.79
F.GOB	0.00	-0.51	0.47	0.00	1.3E-02	0.91	0.15	10.51
F.SGE	-0.00	-0.54	0.63	0.00	3.3E-02	0.90	0.12	8.40
F.DG.F	0.00	0.84	-0.87	0.00	5.5E-06	0.90	0.18	11.53
E.BBVA	-0.00	-0.34	0.42	0.00	7.1E-03	0.91	0.15	9.07
D.ALVX	0.00	0.27	-0.24	0.00	1.6E-03	0.91	0.15	7.62
D.BASX	0.00	-0.07	0.03	0.00	5.9E-02	0.90	0.06	6.79
D.BAYX	0.00	0.88	-0.92	0.00	3.3E-02	0.86	0.10	7.05
D.BMWX	0.00	-0.47	0.51	0.00	4.1E-02	0.92	0.05	8.08
D.DBKX	-0.00	-0.35	0.40	0.00	1.7E-02	0.92	0.12	7.64
D.RWEX	0.00	0.29	-0.24	0.00	2.8E-02	0.92	0.04	5.05

[1] http://www.stoxx.com/indices/index_information.html?symbol=sx5E

Continued 6.2

Asset	μ	a_1	b_1	ω	α_1	β_1	γ_1	DF
D.SIEX	0.00	0.88	-0.91	0.00	2.4E-02	0.91	0.09	5.73
D.EONX	0.00	-0.05	0.07	0.00	6.7E-02	0.89	0.05	5.37
I.ENI	0.00	-0.19	0.14	0.00	2.2E-02	0.90	0.11	7.19
D.MU2X	0.00	-0.72	0.70	0.00	2.2E-02	0.89	0.12	6.05
D.DTEX	-0.00	0.82	-0.83	0.00	2.0E-02	0.96	0.04	3.92
F.FTEL	-0.00	0.82	-0.84	0.00	1.9E-02	0.96	0.02	4.57
D.VO3X	0.00	-0.27	0.34	0.00	8.0E-02	0.88	0.04	5.16
D.SAPX	0.00	0.15	-0.14	0.00	4.3E-05	0.97	0.04	4.46
H.MT	0.00	0.24	-0.23	0.00	9.2E-03	0.94	0.09	23.44
H.UNIL	0.00	0.60	-0.67	0.00	3.8E-02	0.93	0.05	4.70
CRGI	0.00	-0.28	0.32	0.00	6.8E-02	0.90	0.05	8.66
F.BSN	0.00	0.74	-0.79	0.00	1.1E-02	0.91	0.10	7.69
F.LVMH	0.00	0.75	-0.80	0.00	1.2E-02	0.93	0.10	8.40
F.CRFR	-0.00	0.85	-0.87	0.00	1.8E-02	0.94	0.07	5.05
F.EX.F	0.00	0.61	-0.65	0.00	3.0E-03	0.93	0.09	7.15
F.AIR	0.00	0.87	-0.93	0.00	1.2E-02	0.87	0.14	11.52
I.G	-0.00	0.62	-0.60	0.00	9.0E-03	0.89	0.16	7.64
F.OR.F	0.00	0.46	-0.54	0.00	2.4E-02	0.92	0.07	7.87
I.UCG	-0.00	-0.98	0.97	0.00	1.1E-02	0.92	0.14	9.66
I.ISP	-0.00	-0.31	0.34	0.00	1.4E-02	0.94	0.09	5.91
E.TEF	0.00	0.80	-0.82	0.00	3.5E-02	0.86	0.13	6.72
F.UBL	0.00	0.79	-0.80	0.00	5.9E-02	0.87	0.10	8.18
H.PHIL	-0.00	0.79	-0.82	0.00	1.2E-02	0.93	0.10	6.00
F.EI	0.00	-0.54	0.47	0.00	5.6E-02	0.90	0.05	5.84
F.MIDI	-0.00	-0.46	0.49	0.00	2.2E-02	0.89	0.16	11.26
F.SQ.F	0.00	0.95	-0.96	0.00	3.3E-02	0.95	0.01	4.90
F.QT.F	0.00	0.79	-0.84	0.00	6.4E-07	0.92	0.12	10.70
E.IBE	0.00	-0.45	0.51	0.00	5.2E-02	0.84	0.15	5.02
F.GSZ	0.00	0.89	-0.91	0.00	1.9E-02	0.90	0.12	10.91

Table 6.2. Box-Pierce, Ljung-Box, Arch Test, and Kolmogorov–Smirnov Test to residuals of the t-Student ARMA(1,1)–GARCH(1,1) model fitting

Asset	BP (1)	BP (4)	LB (1)	LB (4)	ARCH (1)	ARCH (4)	KS
E.IND	0.88	0.88	0.88	0.88	0.69	0.87	0.00
H.ASML	0.53	0.51	0.53	0.51	0.85	0.97	0.01
I.ENEL	0.60	0.88	0.60	0.88	0.34	0.59	0.00
B.ABI	0.84	0.77	0.84	0.77	0.39	0.94	0.00
F.BNP	0.98	0.54	0.98	0.54	0.70	0.02	0.04
E.REP	0.61	0.96	0.61	0.96	0.96	0.15	0.00
H.ING	0.71	0.81	0.71	0.81	0.88	0.73	0.04
D.DAIX	0.95	0.83	0.95	0.83	0.11	0.23	0.00
E.SCH	0.77	0.92	0.77	0.92	0.22	0.05	0.06
M.NOK1	0.73	0.98	0.73	0.98	0.89	0.97	0.00
F.GOB	0.92	0.94	0.92	0.94	0.09	0.25	0.09
F.SGE	0.38	0.93	0.38	0.93	0.58	0.01	0.03
F.DG.F	0.85	0.93	0.85	0.93	0.41	0.76	0.15
E.BBVA	0.77	0.98	0.77	0.98	0.64	0.01	0.02
D.ALVX	0.44	0.89	0.44	0.89	0.52	0.01	0.02
D.BASX	1.00	0.81	1.00	0.81	0.13	0.27	0.01
D.BAYX	0.67	0.83	0.67	0.83	0.42	0.49	0.00
D.BMWX	0.89	0.87	0.88	0.87	0.17	0.17	0.02
D.DBKX	0.70	0.70	0.70	0.70	0.95	0.74	0.04
D.RWEX	0.63	0.83	0.63	0.83	0.04	0.20	0.00
D.SIEX	0.99	0.63	0.99	0.63	0.27	0.85	0.00
D.EONX	0.76	0.98	0.76	0.98	0.53	0.66	0.00
I.ENI	0.73	0.65	0.72	0.65	0.29	0.80	0.00
D.MU2X	0.64	0.95	0.64	0.95	0.95	1.00	0.00
D.DTEX	0.22	0.37	0.22	0.37	0.85	1.00	0.00
F.FTEL	0.67	0.91	0.66	0.91	0.29	0.87	0.00
D.VO3X	0.65	0.98	0.65	0.98	0.56	0.85	0.00
D.SAPX	0.85	0.74	0.85	0.74	0.53	0.77	0.00
H.MT	0.65	0.97	0.65	0.97	0.12	0.51	0.57
H.UNIL	0.46	0.84	0.46	0.83	0.57	0.94	0.00
CRGI	0.97	0.43	0.97	0.43	0.35	0.84	0.01
F.BSN	0.30	0.65	0.30	0.65	0.69	0.97	0.02
F.LVMH	0.56	0.39	0.56	0.39	0.11	0.37	0.01
F.CRFR	0.50	0.52	0.50	0.52	0.06	0.39	0.00
F.EX.F	0.39	0.69	0.39	0.69	0.72	0.84	0.01
F.AIR	0.22	0.77	0.22	0.77	0.36	0.87	0.27
I.G	0.77	0.62	0.77	0.62	0.83	0.22	0.01
F.OR.F	0.87	0.95	0.87	0.95	0.37	0.01	0.02
I.UCG	0.41	0.84	0.41	0.84	0.68	0.64	0.05
I.ISP	0.52	0.40	0.52	0.40	0.18	0.60	0.00
E.TEF	0.18	0.58	0.18	0.58	0.73	0.45	0.00
F.UBL	0.39	0.53	0.39	0.53	0.55	0.78	0.03
H.PHIL	0.65	0.88	0.65	0.88	0.21	0.70	0.00
F.EI	0.74	0.96	0.74	0.96	0.15	0.48	0.00
F.MIDI	0.75	0.92	0.75	0.91	0.82	0.14	0.17
F.SQ.F	0.27	0.59	0.27	0.59	0.44	0.68	0.00
F.QT.F	0.76	0.96	0.76	0.96	0.64	0.93	0.07
E.IBE	0.38	0.84	0.37	0.83	0.10	0.49	0.00
F.GSZ	0.32	0.82	0.32	0.82	0.61	0.43	0.08

Construction of the Dissimilarity Matrix

Given the standardized residuals, we compute two kinds of tail dependence measures, namely the TDC λ_L and the conditional Spearman's ρ_α (here we select $\alpha = 0.10$) for all pairs of variables. Then we plug in the tail dependence measures as $\hat{m}$ in (6.3) to calculate the dissimilarity measure Δ_{ij} between the pair of variables for constructing the dissimilarity matrix $\Delta = (\delta_{ij})_{i,j=1,\ldots,p}$. Here in our example, given $p = 50$ time series, the number of each type of bivariate tail dependence measure is $1225 (= p(p-1)/2)$.

Hierarchical and Fuzzy Clustering Procedure

Then, we apply the hierarchical clustering algorithm (complete linkage) and the fanny algorithm[2] to determine the cluster compositions. As stated previously, there are two methods to perform hierarchical cluster: the agglomerative method and the divisive method. In package *cluster*, agglomerative method has been adopted: at first, each asset is a cluster by itself. Clusters are merged until only one cluster remains which contains all assets. At each stage the two "nearest"[3] clusters are combined to form one larger cluster ([Kaufman and Rousseeuw, 1990]). The output of *hclust* and *fanny* contains the silhouette information, which can be used to determine the number of clusters, that is the number k such that it gives the maximum average silhouette width $\bar{s}(k)$ for the entire data set.

We present the average silhouette width for the entire data set against the number of clusters obtained by carrying out both two clustering algorithms based on both two tail dependence measures in Figure 6.1: hierarchical clustering on λ_L in upper-left corner, fuzzy clustering on λ_L in upper-right corner, hierarchical clustering on $\rho_{0.10}$ in lower-left corner, and fuzzy clustering on $\rho_{0.10}$ in right-left corner. In our case, according to $\bar{s}(k)$, we should choose $k = 2$ and $= 3$ for λ_L used in hierarchical and fuzzy algorithms, respectively; and $k = 2$ and $= 5$ for $\rho_{0.10}$ used in hierarchical and fuzzy algorithms, respectively. However, in the portfolio risk management, we consider not only the silhouette width information, but also the diversification effect and the management cost of the portfolios. If there are too few assets in a portfolio, then the diversification effect is not sufficient; and if there are too many assets in a portfolio, then it is too costly to manage. For this reason, we make a constraint that $5 \leq k \leq 10$[4]. But, in our example we should recognize a serious problem, that whichever k we choose, the average silhouette width is always near zero. This means that probably it is a very poor partition. With the constraint of $5 \leq k \leq 10$, we can see that 5-cluster partition has the greatest average silhouette width for both λ_L and $\rho_{0.10}$ used by both hierarchical and fuzzy algorithm.

As stated in Chapter 5, the fuzzy algorithm is a "soft" partitioning method that allows each object to belong to one or more than one cluster according to a membership coefficient, So, the direct result from the fuzzy clustering procedure is a matrix which tells the membership of every object belonging to each cluster. But usually, for the convenience of the decision, the cluster which gives the maximum membership to one object is selected as its containing cluster.

Intuitively, we present the clustering analysis of the hierarchical and the fuzzy algorithm based on the λ_L and $\rho_{0.10}$ in Figure 6.2 and Figure 6.3, respectively. In the hierarchical dendrogram, we can know the agglomerative hierarchical process: smaller height earlier the merging happens. While, in Figure 6.3, for saving the space and making the picture clear, we omit the names. We can see that it is confirmed that some objects can be grouped in more than one cluster. To make confrontation

[2]The hierarchical and fuzzy algorithm can be proposed in **R** package *cluster* by using command *hclust* and *fanny*, respectively.

[3]Three most used cluster-distance are single linkage, farthest linkage, and average linkage.

[4]The constraint should be made with some criteria(take under some criteria). Here we take this constraint randomly and just want to show how it works.

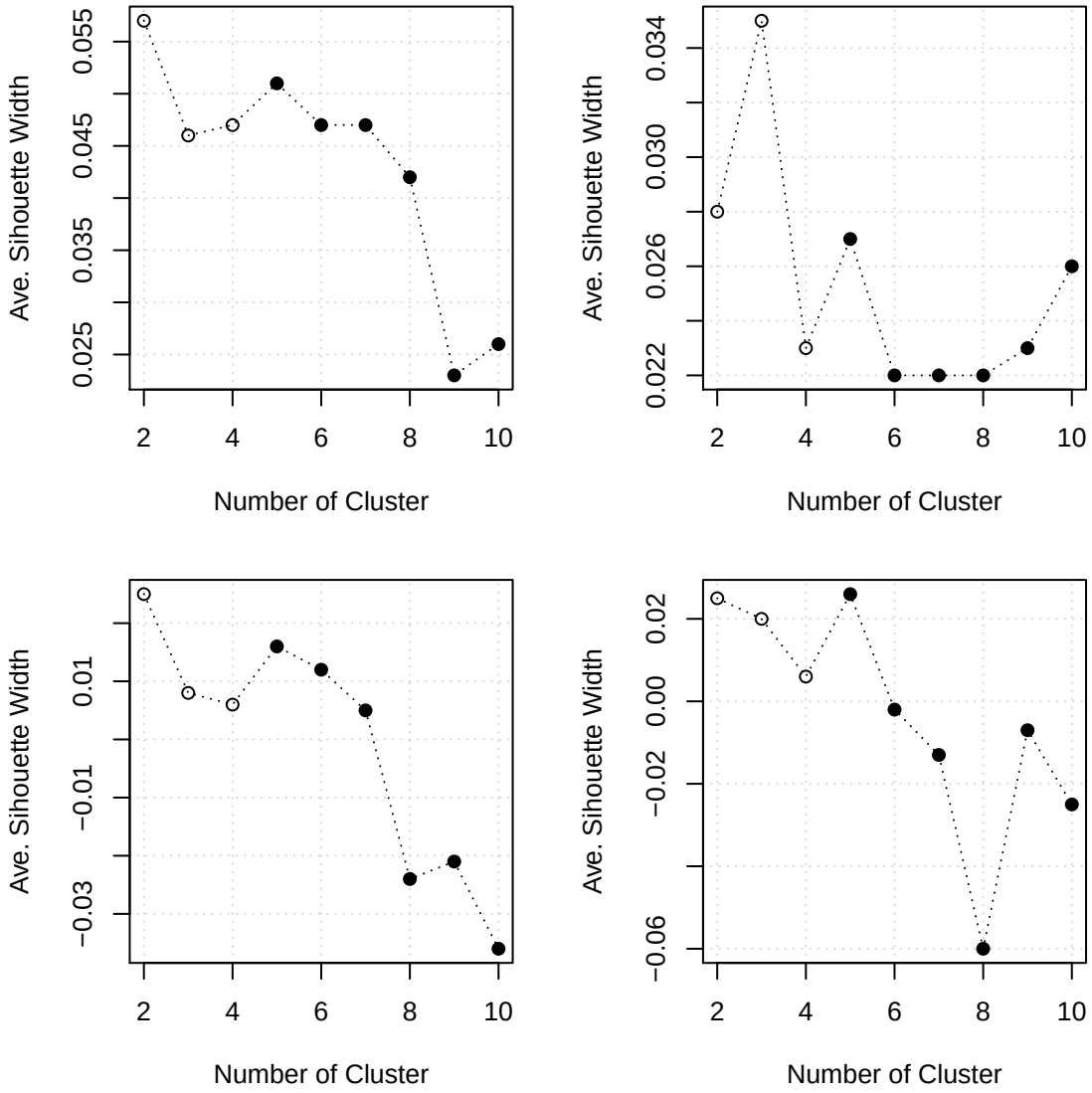
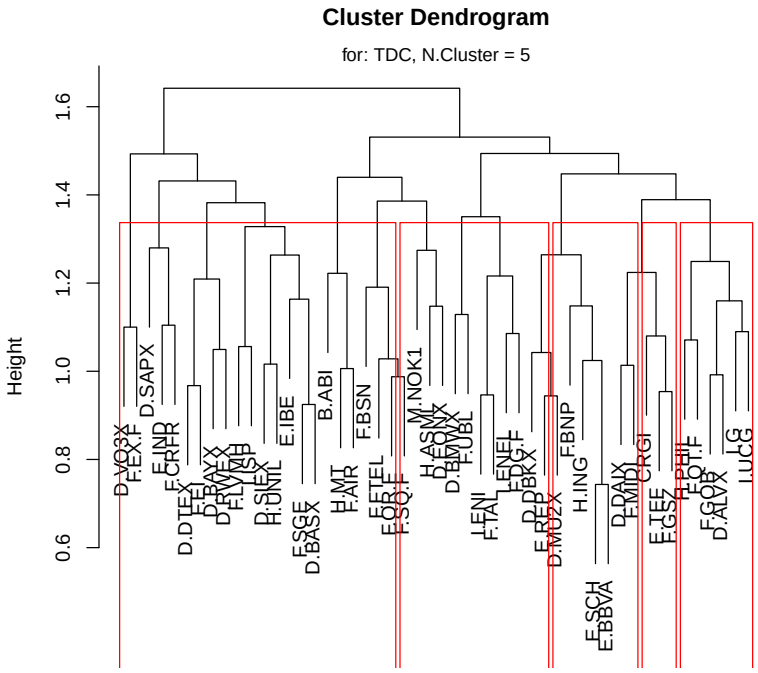
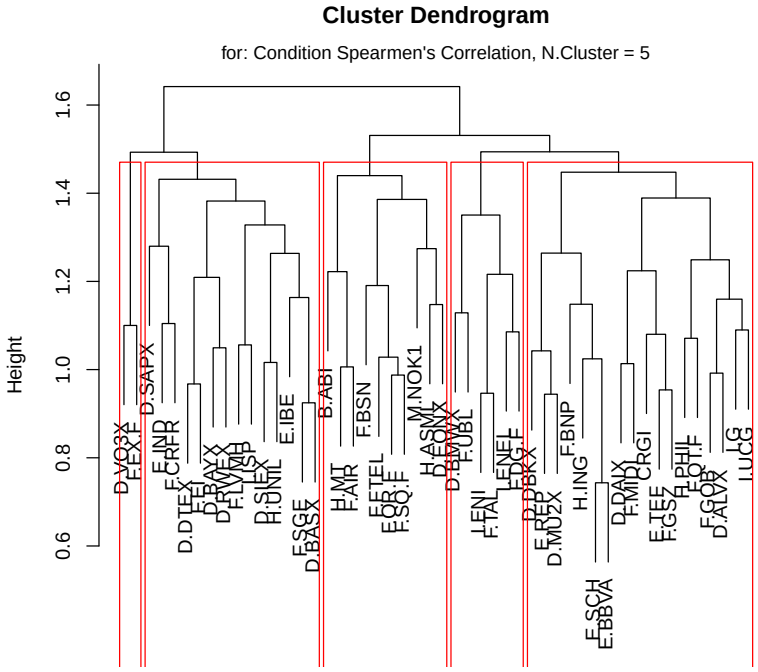


Figure 6.1. Average silhouette width for the entire data set against the number of clusters based both on the different tail dependence measure and on the different cluster algorithms. The picture in the upper-left corner is obtained by carrying out hierarchical clustering algorithm based on λ_L; the upper-right one is from fuzzy cluster based on λ_L; the lower-left presents the case of hierarchical based on ρ_α, and the lower-right is for fuzzy cluster based on ρ_α.

between two algorithms, and also between two tail dependence measures, we also present the cluster composition in Tables 6.3–6.6. As can be seen, the compositions of the sub–groups seem to differ with respect to both the tail dependence measure and the clustering algorithms.

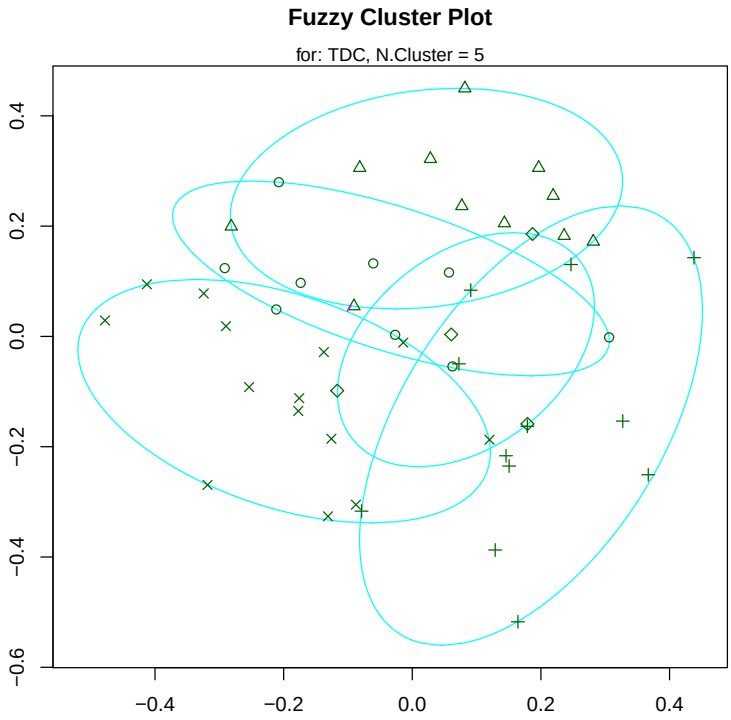


(a)

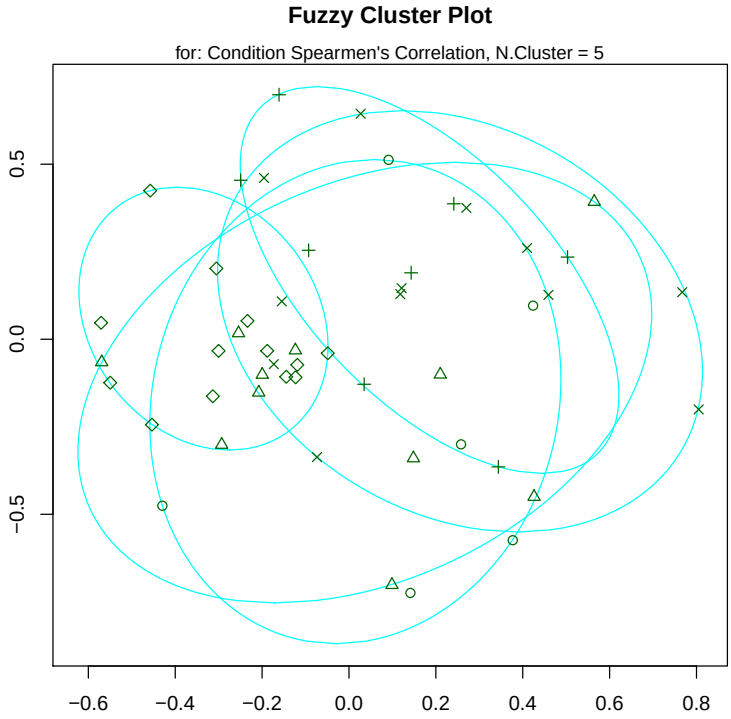


(b)

Figure 6.2. Hierarchical clustering based on (upper) λ_L and (lower) $\rho_{0.10}$.



(a)



(b)

Figure 6.3. Fuzzy clustering based on (upper) λ_L and (lower) $\rho_{0.10}$. For saving the space and making the picture clear, we omit the names.

Cluster	Asset							
1	E.IND	D.DAIX	D.BMWX					
2	H.ASML	B.ABI	D.RWEX	D.EONX	D.DTEX	D.SAPX	H.UNIL	F.CRFR
	F.EX.F	F.SQ.F	F.QT.F	F.GSZ				
3	I.ENEL	M.NOK1	F.GOB	D.BAYX	F.FTEL	CRGI		
4	F.BNP	E.REP	H.ING	E.SCH	F.SGE	F.DG.F	E.BBVA	D.ALVX
	D.BASX	D.SIEX	H.MT	F.BSN	F.LVMH	F.AIR	I.G	F.OR.F
	I.UCG	I.ISP	E.TEF	F.UBL	F.MIDI	E.IBE		
5	D.DBKX	I.ENI	D.MU2X	D.VO3X	F.TAL	H.PHIL	F.EI	

Table 6.3. Cluster composition of EURO STOXX 50 constituents by using TDC measure and hierarchical clustering.

Cluster	Asset							
1	E.IND	F.DG.F	D.BMWX	H.MT	F.LVMH	I.ISP	E.TEF	E.IBE
	F.GSZ							
2	H.ASML	B.ABI	F.SGE	D.EONX	D.DTEX	H.UNIL	F.BSN	F.EX.F
	F.AIR	F.UBL	F.QT.F					
3	I.ENEL	M.NOK1	D.BAYX	I.ENI	F.FTEL	D.SAPX	CRGI	F.TAL
	F.CRFR	H.PHIL	F.EI	F.SQ.F				
4	F.BNP	E.REP	H.ING	E.SCH	F.GOB	E.BBVA	D.ALVX	D.BASX
	D.DBKX	D.SIEX	I.G	F.OR.F	I.UCG	F.MIDI		
5	D.DAIX	D.RWEX	D.MU2X	D.VO3X				

Table 6.4. Cluster composition of EURO STOXX 50 constituents by using TDC measure and fuzzy clustering.

Cluster	Asset							
1	E.IND	F.SGE	D.BASX	D.BAYX	D.RWEX	D.SIEX	D.DTEX	D.SAPX
	H.UNIL	F.LVMH	F.CRFR	I.ISP	F.EI	E.IBE		
2	H.ASML	B.ABI	M.NOK1	D.EONX	F.FTEL	H.MT	F.BSN	F.AIR
	F.OR.F	F.SQ.F						
3	I.ENEL	F.DG.F	D.BMWX	I.ENI	F.TAL	F.UBL		
4	F.BNP	E.REP	H.ING	D.DAIX	E.SCH	F.GOB	E.BBVA	D.ALVX
	D.DBKX	D.MU2X	CRGI	I.G	I.UCG	E.TEF	H.PHIL	F.MIDI
	F.QT.F	F.GSZ						
5	D.VO3X	F.EX.F						

Table 6.5. Cluster composition of EURO STOXX 50 constituents by using conditional Spearman's correlation ρ_α with $\alpha = 0.1$ and hierarchical clustering.

Cluster	Asset							
1	E.IND	E.REP	M.NOK1	H.MT	F.UBL	F.EI		
2	H.ASML	F.BNP	H.ING	E.SCH	E.BBVA	D.BAYX	D.BMWX	D.RWEX
	I.ENI	D.MU2X	D.DTEX					
3	I.ENEL	D.EONX	CRGI	F.BSN	F.CRFR	E.TEF	E.IBE	F.GSZ
4	B.ABI	F.SGE	F.DG.F	D.BASX	F.FTEL	D.VO3X	D.SAPX	H.UNIL
	F.EX.F	F.AIR	F.OR.F	F.SQ.F				
5	D.DAIX	F.GOB	D.ALVX	D.DBKX	D.SIEX	F.TAL	F.LVMH	I.G
	I.UCG	I.ISP	H.PHIL	F.MIDI	F.QT.F			

Table 6.6. Cluster composition of EURO STOXX 50 constituents by using conditional Speaman's correlation ρ_α with $\alpha = 0.1$ and fuzzy clustering.

In order to provide a more objective comparisons among the obtained clusters, a measure of agreement is needed. To this aim, we consider the Rand Index (RI) [Rand, 1971] and the Adjusted Rand Index (ARI) [Hubert and Arabie, 1985]. The RI lies between 0 and 1, where the maximum value is taken when two partitions agree perfectly. Instead, the ARI takes a wider range of values and, hence, increases the sensitivity of the index as a measure of agreement.

As is noted in Table 6.7, if we fix the clustering method and compare the results obtained by changing the tail dependence measure, the obtained grouping compositions seem to be similar, although they do not coincide. In fact, as discussed above, the two tail dependence measures underline different aspects of tail dependence (finite and asymptotic tail behavior).

Analogously, in view of Table 6.8, if we fix the the tail dependence measure but use different clustering procedures, the obtained grouping compositions seem to be similar (even more than in the previous case). In other words, the effects of a changing clustering procedure seem to be less evident than those obtained by changing the tail dependence measure.

	Hierarchical Clustering	Fuzzy Clustering
RI	0.62	0.68
ARI	0.03	0.02

Table 6.7. Rand Index and Adjusted Rand Index between cluster compositions obtained from a hierarchical clustering (respectively, fuzzy clusteing) applied on different tail dependence measures.

	$\rho_{0.10}$	TDC
RI	0.68	0.73
ARI	0.08	0.26

Table 6.8. Rand Index and Adjusted Rand Index between cluster compositions obtained, respectively, by hierarchical clustering and by fuzzy clustering, according to different tail dependence measures.

Portfolio Selection

The assets composing the selected portfolios depend on the use of two tail dependence measures, namely λ_L and $\rho_{0.10}$ used in the hierarchical and fuzzy clustering algorithm are listed in Table 6.9 to Table 6.12, respectively. The standard deviation against expected return of all possible portfolios corresponding to λ_L and $\rho_{0.10}$ used in the hierarchical and fuzzy clustering algorithm are plotted in Figure 6.4 to Figure 6.7, respectively.

P.folio	Assets				
1	F.BSN	D.SAPX	F.EI	E.IND	F.FTEL
2	E.TEF	H.UNIL	F.EI	E.IND	D.BAYX
3	F.BSN	B.ABI	F.EI	E.IND	D.BAYX
4	F.AIR	B.ABI	F.EI	E.IND	D.BAYX
5	H.MT	B.ABI	F.EI	E.IND	D.BAYX
6	D.BASX	B.ABI	F.EI	E.IND	F.GOB
7	D.BASX	B.ABI	D.VO3X	E.IND	F.GOB
8	H.MT	B.ABI	D.VO3X	E.IND	F.GOB
9	D.BASX	F.QT.F	D.VO3X	E.IND	F.GOB

Table 6.9. Assets Composition of the portfolios lying on the efficient frontier in Figure 6.4 obtained by λ_L and by hierarchical clustering. The order from top to bottom in the table is the order from left to right in the figure.

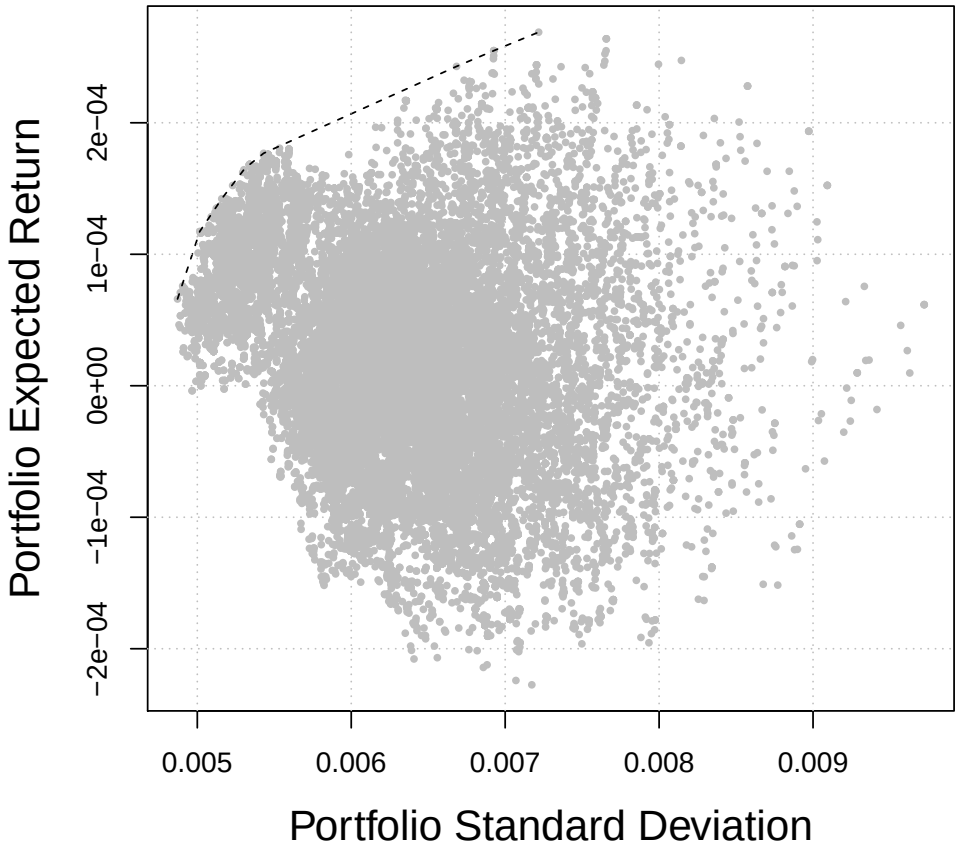


Figure 6.4. Standard deviation-Expected return plot of 5-asset portfolios generated from λ_L and hierarchical clustering.

P.folio	Assets				
1	E.TEF	H.UNIL	F.EI	I.G	D.RWEX
2	E.TEF	H.UNIL	F.EI	I.G	D.VO3X
3	E.TEF	H.UNIL	F.EI	F.OR.F	D.VO3X
4	E.IND	H.UNIL	F.EI	F.OR.F	D.VO3X
5	E.IND	B.ABI	F.EI	F.OR.F	D.VO3X
6	E.IND	B.ABI	F.EI	D.BASX	D.VO3X
7	E.IND	B.ABI	F.EI	D.DBKX	D.VO3X
8	E.IND	F.QT.F	F.EI	E.BBVA	D.VO3X
9	E.IND	F.QT.F	D.BAYX	D.BASX	D.VO3X
10	E.IND	F.QT.F	CRGI	D.BASX	D.VO3X

Table 6.10. Assets Composition of the portfolios lying on the efficient frontier in Figure 6.5 obtained by λ_L and by fuzzy clustering. The order from top to bottom in the table is the order from left to right in the figure.

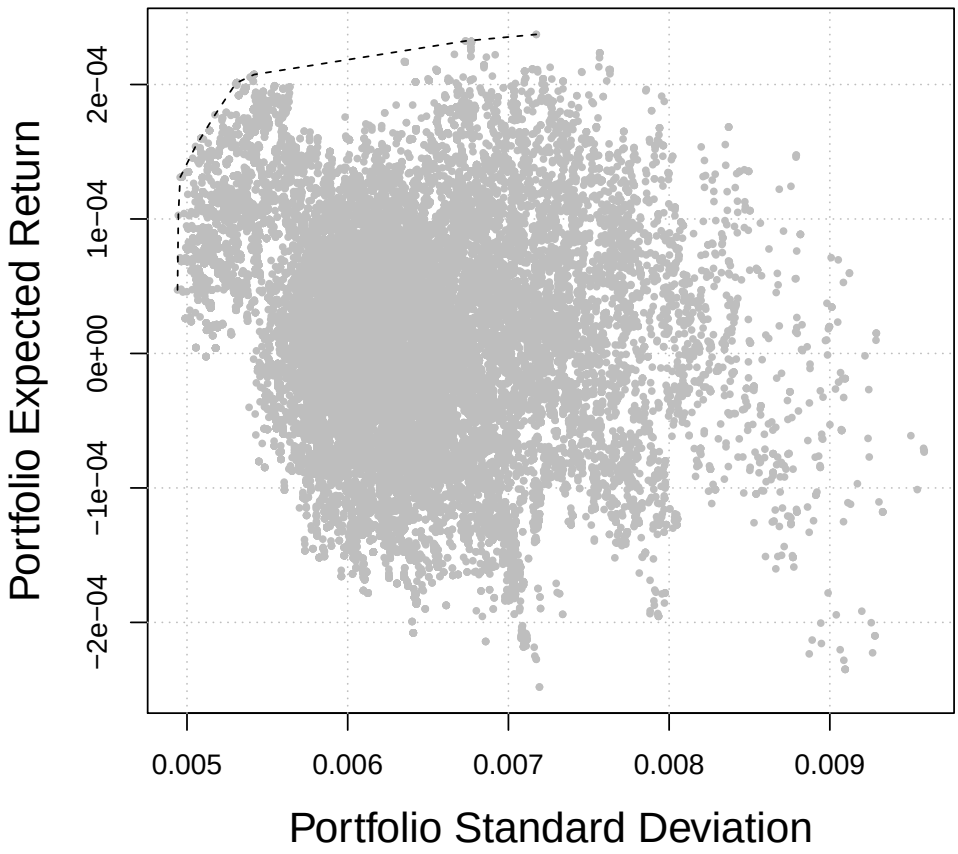


Figure 6.5. Standard deviation-Expected return plot of 5-asset portfolios generated from λ_L and fuzzy clustering.

P.folio	Assets				
1	D.VO3X	F.EI	F.FTEL	I.ENEL	E.TEF
2	D.VO3X	F.EI	F.BSN	I.ENEL	E.TEF
3	D.VO3X	F.EI	F.BSN	F.UBL	E.TEF
4	D.VO3X	F.EI	B.ABI	F.UBL	E.TEF
5	D.VO3X	F.EI	B.ABI	F.UBL	E.SCH
6	D.VO3X	F.EI	B.ABI	D.BMWX	F.QT.F
7	D.VO3X	E.IND	H.MT	D.BMWX	F.GOB

Table 6.11. Assets Composition of the portfolios lying on the efficient frontier in Figure 6.6 obtained by $\rho_{0.10}$ and by hierarchical clustering. The order from top to bottom in the table is the order from left to right in the figure.

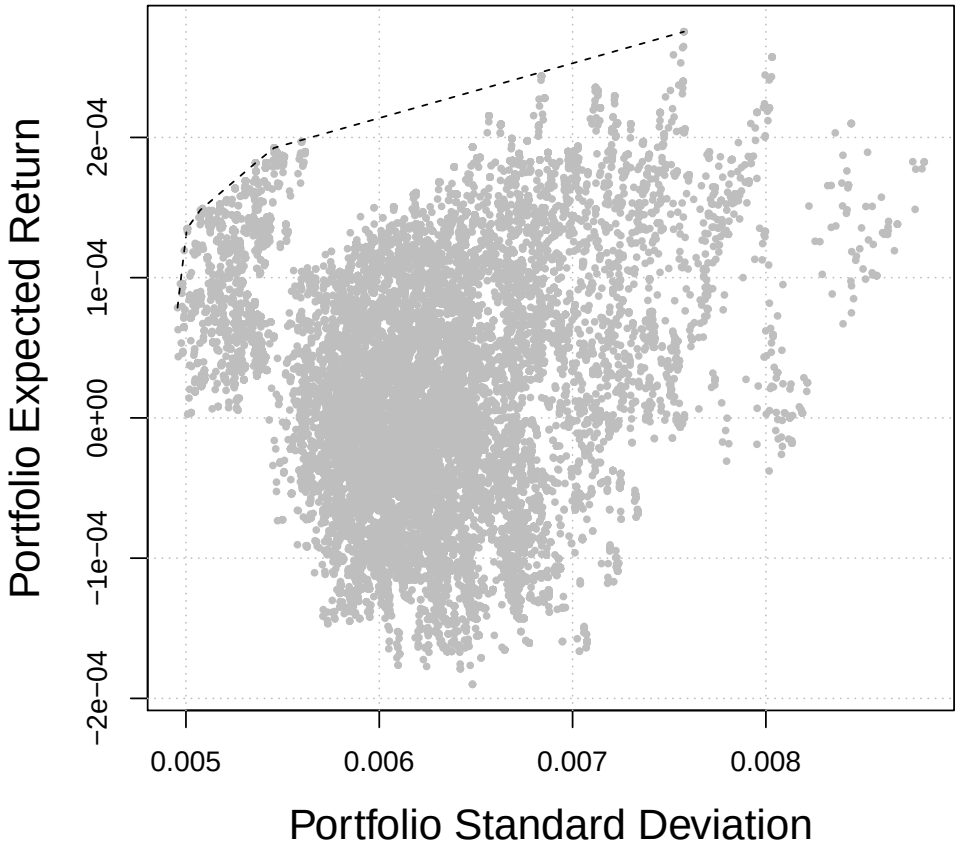


Figure 6.6. Standard deviation-Expected return plot of 5-asset portfolios generated from ρ_α with $\alpha = 0.10$ and hierarchical clustering.

P.folio	Assets				
1	F.EI	D.RWEX	F.BSN	F.FTEL	I.G
2	F.EI	D.DTEX	F.BSN	D.VO3X	I.G
3	F.EI	D.DTEX	F.BSN	D.VO3X	I.ISP
4	F.EI	D.BAYX	E.TEF	D.VO3X	D.ALVX
5	F.EI	D.BAYX	F.BSN	D.VO3X	F.MIDI
6	F.EI	D.BAYX	CRGI	D.VO3X	F.LVMH
7	E.IND	D.BMWX	CRGI	D.VO3X	D.DAIX

Table 6.12. Assets Composition of the portfolios lying on the efficient frontier in Figure 6.7 obtained by $\rho_{0.10}$ and by fuzzy clustering. The order from top to bottom in the table is the order from left to right in the figure.

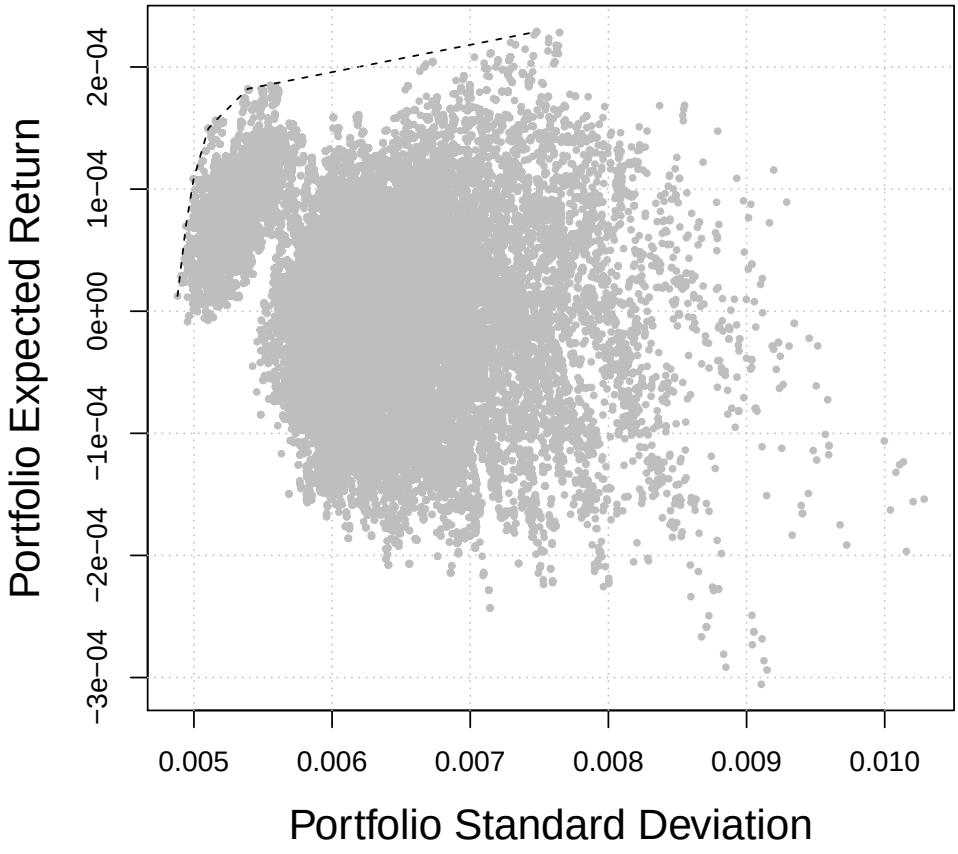


Figure 6.7. Standard deviation-Expected return plot of 5-asset portfolios generated from ρ_α with $\alpha = 0.10$ and fuzzy clustering.

Remark 6.3.1. The portfolio selection depends on the tail dependence measure λ_{ij}. Different measures lead to a different dissimilarity matrix, which leads to different partition solutions: the number of clusters and the composing assets.

Portfolio Performance

We investigated the performance of the portfolios constructed by previously described methodology during the period from August 1, 2011 to September 9, 2011, that is the period in which the index was experiencing severe losses. The returns of the portfolios lying on the efficient frontier are compared with the returns of naive minimum variance portfolio composed by all fifty assets and with the benchmark index *EURO STOXX 50* in Figure 6.8 to Figure 6.11. The return of the global minimum variance portfolio is also compared with the returns of 15000 minimum variance portfolios built with k randomly selected assets (*upper panel* of Figure 6.8 to Figure 6.11.). As it can be seen, the performance of the portfolios lying on the efficient frontier is generally better than the benchmark and, in several cases, outperforms the global minimum variance portfolio. This seems to confirm the idea that, when markets are experiencing a period of losses, a diversification strategy could be beneficial.

In order to see and to understand better, we also calculated some common risk measures of these selected portfolios during the period under consideration: *mean*, *standard deviation*, *skewness*, *5% VaR*, *5% Expected Shortfall(ES)*, and *Sharpe Ratio(SR)*. For these measures, the uses and estimations in the risk management are given in the following and the values of this example are listed in Table 6.13 to Table 6.16.

mean is taken for measuring the average level of the return, estimated by the sample mean:

$$\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i.$$

standard deviation is calculated as the risk of portfolio, estimated by the sample standard deviation:

$$\sigma_r = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(r_i - \bar{r})^2}.$$

skewness can be used to measure of the probability of extreme values, estimated by the third standardized moment:

$$\mathrm{Ske} = \frac{n^{\frac{3}{2}}}{(n-1)^{\frac{3}{2}}(n-2)}\frac{\sum_{i=1}^{n}(x-\bar{r})^3}{\sigma_r^3} = \frac{n^{-\frac{1}{2}}}{(n-1)(n-2)}\frac{\sum_{i=1}^{n}(x-\bar{r})^3}{\sqrt{\sum_{i=1}^{n}(x-\bar{r})^2}^3}$$

5% VaR is widely used to measrue the risk of loss, estimated by the 5% quantile of the distribution function:

$$\mathbb{P}\left(r \leq \mathrm{VaR}_r(5\%)\right) = 5\%.$$

5% Expected Shortfall (ES) is an alternative of VaR, but it is more sensitive to the shape of the loss distribution in the tail, and estimated by the sample mean of the values which are less than $\mathrm{VaR}_r(5\%)$.

Sharpe Ratio(SR) is a measure of earnings against per unit of deviation, and is defined by the expected value of the ratio of excess return to per unit of deviation and estimated by

$$\frac{1}{n}\frac{r - r_f}{\sigma_r},$$

where r_f is the return of risk free. Here we assum that $r_f = 0$, that the investors won't loss any money for sure if they hold cash.

	Mean	S.D.	Skewness	5% VaR	5% E.S.	Sharpe Ratio
1	0.000	0.005	-0.043	-0.008	-0.012	0.012
2	0.000	0.005	-0.064	-0.008	-0.012	0.024
3	0.000	0.005	-0.159	-0.008	-0.012	0.032
4	0.000	0.005	-0.042	-0.009	-0.012	0.037
5	0.000	0.005	-0.037	-0.009	-0.012	0.038
6	0.000	0.005	-0.025	-0.009	-0.013	0.039
7	0.000	0.007	-0.102	-0.011	-0.016	0.041
8	0.000	0.007	-0.064	-0.011	-0.016	0.044
9	0.000	0.007	0.026	-0.012	-0.017	0.037
Naive MVP	0.000	0.005	-0.153	-0.008	-0.011	0.015
EURO STOXX 50	-0.000	0.007	0.064	-0.011	-0.017	-0.013

Table 6.13. Basic statistics related to the log–returns of the portfolios lying on the convex frontier (Figure 6.4) by λ_L and by hierarchical clustering. Period: August 1, 2011 – September 9, 2011. The order from top to bottom in the table is the order from left to right in the figure.

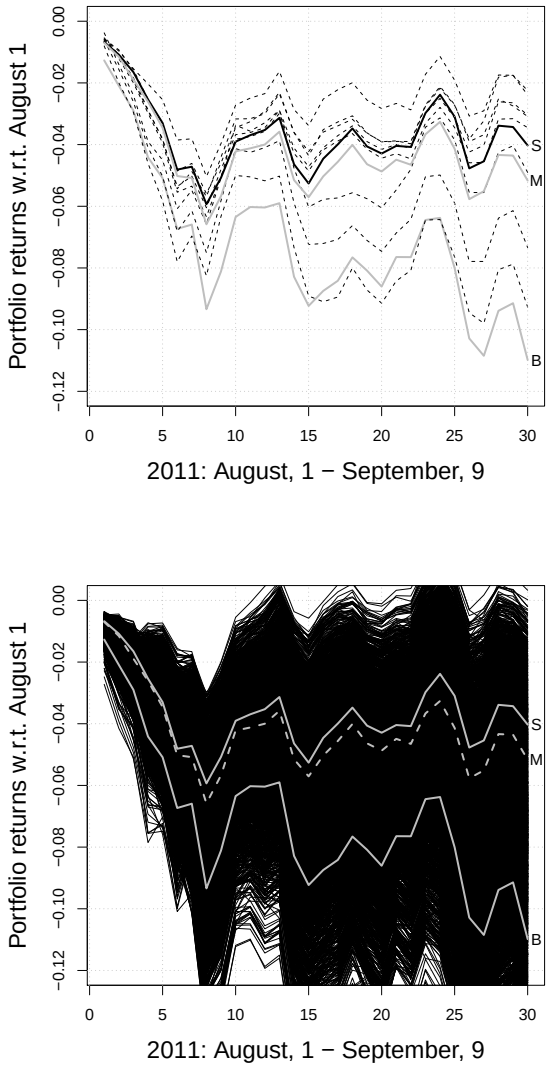


Figure 6.8. 9 minimum variance portfolios on the efficient frontier of all the possible 5-asset portfolios obtained by λ_L and by hierarchical clustering. The returns of the minimum variance portfolio on the frontier denoted by '*S*') and other 8 portfolios (*black dotted lines*) are compared (*upper*) with the returns of EURO STOXX 50 denoted by '*B*') and with the returns of the naive minimum variance portfolio denoted by '*M*') composed by all 50 assets. They are also compared (*lower*) with the returns of 15000 minimum variance portfolios built with 5 randomly selected assets, from August 1, 2011 to September 9, 2011.

	Mean	S.D.	Skewness	5% VaR	5% E.S.	Sharpe Ratio
1	0.000	0.005	0.030	-0.008	-0.012	0.011
2	0.000	0.005	-0.159	-0.008	-0.012	0.020
3	0.000	0.005	-0.179	-0.008	-0.012	0.025
4	0.000	0.005	-0.124	-0.008	-0.012	0.036
5	0.000	0.005	-0.107	-0.009	-0.012	0.039
6	0.000	0.005	-0.214	-0.009	-0.012	0.043
7	0.000	0.005	-0.205	-0.009	-0.012	0.043
8	0.000	0.005	-0.167	-0.009	-0.013	0.042
9	0.000	0.007	-0.104	-0.011	-0.016	0.036
10	0.000	0.007	0.033	-0.012	-0.017	0.034
Naive MVP	0.000	0.005	-0.153	-0.008	-0.011	0.015
EURO STOXX 50	-0.000	0.007	0.064	-0.011	-0.017	-0.013

Table 6.14. Basic statistics related to the log–returns of the portfolios lying on the convex frontier (Figure 6.5) by λ_L and by fuzzy clustering. Period: August 1, 2011 – September 9, 2011. The order from top to bottom in the table is the order from left to right in the figure.

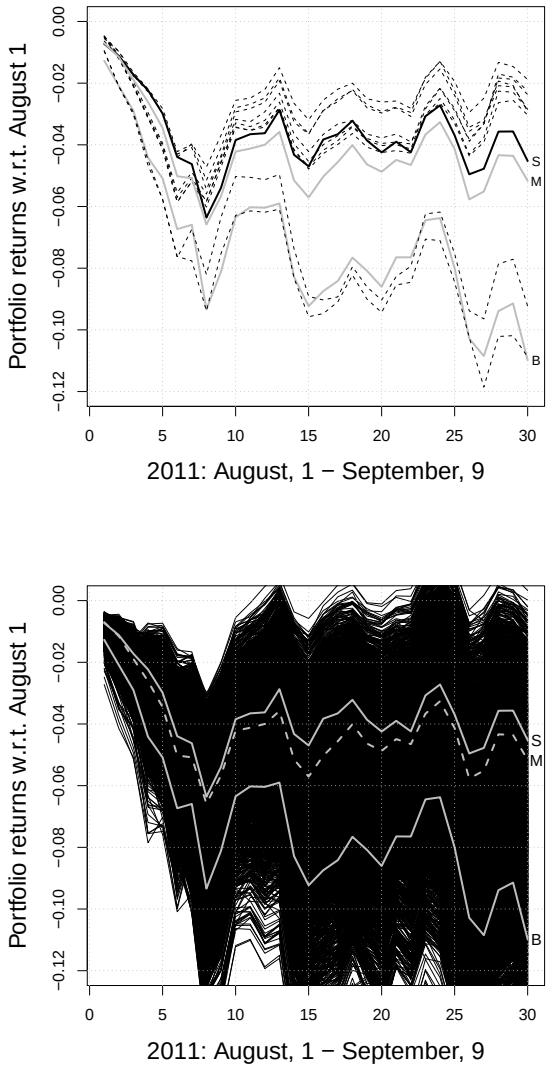


Figure 6.9. 10 minimum variance portfolios on the efficient frontier of all the possible 5-asset portfolios obtained by λ_L and by fuzzy clustering. The returns of the minimum variance portfolio on the frontier denoted by '*S*') and other 9 portfolios (*black dotted lines*) are compared (*upper*) with the returns of EURO STOXX 50 denoted by '*B*') and with the returns of the naive minimum variance portfolio denoted by '*M*') composed by all 50 assets. They are also compared (*lower*) with the returns of 15000 minimum variance portfolios built with 5 randomly selected assets, from August 1, 2011 to September 9, 2011.

	Mean	S.D.	Skewness	5% VaR	5% E.S.	Sharpe Ratio
1	0.000	0.005	-0.064	-0.009	-0.012	0.007
2	0.000	0.005	-0.262	-0.008	-0.012	0.018
3	0.000	0.005	-0.297	-0.008	-0.012	0.022
4	0.000	0.005	-0.281	-0.009	-0.012	0.025
5	0.000	0.005	-0.257	-0.009	-0.013	0.038
6	0.000	0.005	-0.232	-0.009	-0.013	0.039
7	0.000	0.008	0.044	-0.012	-0.017	0.037
Naive MVP	0.000	0.005	-0.153	-0.008	-0.011	0.015
EURO STOXX 50	-0.000	0.007	0.064	-0.011	-0.017	-0.013

Table 6.15. Basic statistics related to the log–returns of the portfolios lying on the convex frontier (Figure 6.6) by $\rho_{0.10}$ and by hierarchical clustering. Period: August 1, 2011 – September 9, 2011. The order from top to bottom in the table is the order from left to right in the figure.

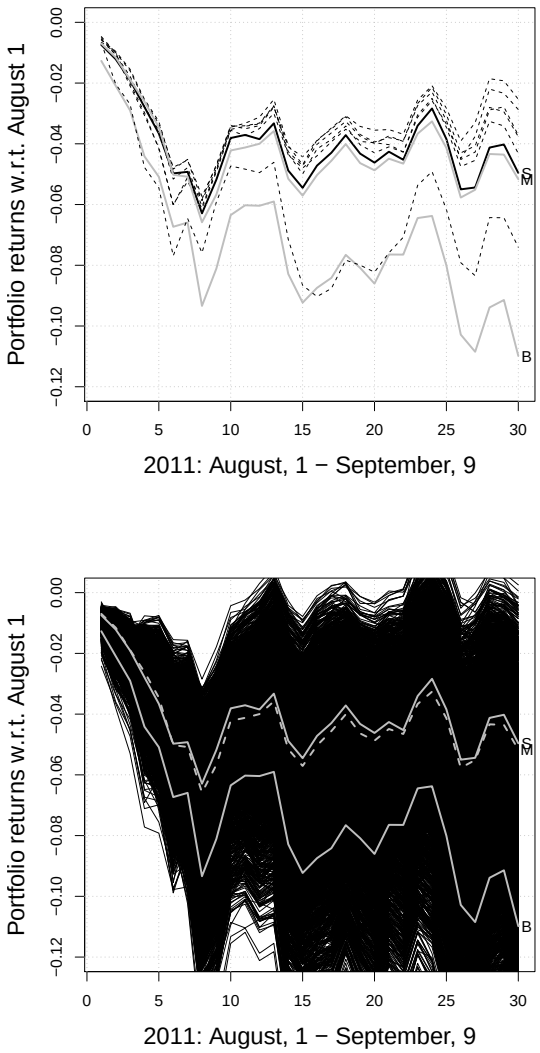


Figure 6.10. 7 minimum variance portfolios on the efficient frontier of all the possible 5-asset portfolios obtained by $\rho_{0.10}$ and by hierarchical clustering. The returns of the minimum variance portfolio on the frontier denoted by '*S*') and other 6 portfolios (*black dotted lines*) are compared (*upper*) with the returns of EURO STOXX 50 denoted by '*B*') and with the returns of the naive minimum variance portfolio denoted by '*M*') composed by all 50 assets. They are also compared (*lower*) with the returns of 15000 minimum variance portfolios built with 5 randomly selected assets, from August 1, 2011 to September 9, 2011.

	Mean	S.D.	Skewness	5% VaR	5% E.S.	Sharpe Ratio
1	-0.000	0.005	0.062	-0.008	-0.012	-0.000
2	0.000	0.005	-0.166	-0.008	-0.012	0.014
3	0.000	0.005	-0.190	-0.008	-0.012	0.021
4	0.000	0.005	-0.235	-0.009	-0.013	0.024
5	0.000	0.005	-0.285	-0.008	-0.012	0.031
6	0.000	0.006	-0.235	-0.009	-0.013	0.036
7	0.000	0.008	0.042	-0.012	-0.017	0.032
Naive MVP	0.000	0.005	-0.153	-0.008	-0.011	0.015
EURO STOXX 50	-0.000	0.007	0.064	-0.011	-0.017	-0.013

Table 6.16. Basic statistics related to the log–returns of the portfolios lying on the convex frontier (Figure 6.7) by λ_L and by fuzzy clustering. Period: August 1, 2011 – September 9, 2011. The order from top to bottom in the table is the order from left to right in the figure.

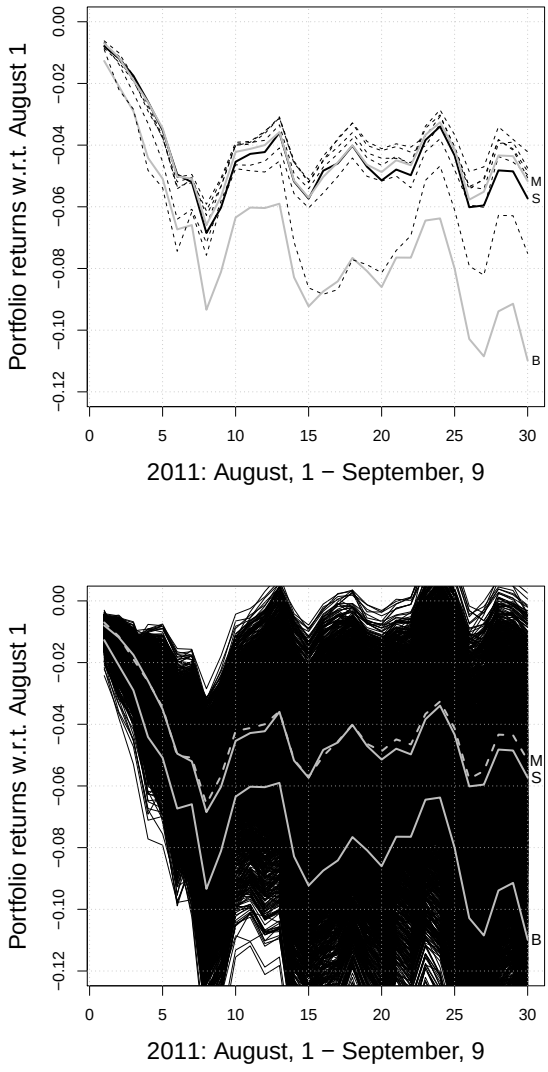


Figure 6.11. 7 minimum variance portfolios on the efficient frontier of all the possible 5-asset portfolios obtained by $\rho_{0.10}$ and by fuzzy clustering. The returns of the minimum variance portfolio on the frontier denoted by '*S*') and other 6 portfolios (*black dotted lines*) are compared (*upper*) with the returns of EURO STOXX 50 denoted by '*B*') and with the returns of the naive minimum variance portfolio denoted by '*M*') composed by all 50 assets. They are also compared (*lower*) with the returns of 15000 minimum variance portfolios built with 5 randomly selected assets, from August 1, 2011 to September 9, 2011.

The composition of the minimum variance portfolio in the convex frontier of Figures 6.8 to Figure 6.11 are shown in Table 6.17. Notice that all gave significant weight to one single asset, while the other assets are chosen and weighted in different manners. Anyway, as can be read from the basic statistics of the selected portfolios (Table 6.18), the returns of the portfolio obtained by hierarchical clustering (HC Portfolio) and the returns of portfolio obtained by fuzzy clustering (FC Portfolio) are quite similar. Thus, the clustering method does not have a strong influence on the overall results.

λ_L	**Hierarchical Clustering**	Assets	F.EI	F.FTEL	F.BSN	D.SAPX	E.IND
		Weights	0.41	0.25	0.15	0.14	0.06
	Fuzzy Clustering	Assets	F.EI	H.UNIL	E.TEF	D.RWEX	I.G
		Weights	0.43	0.20	0.15	0.13	0.09
$\rho_{0.10}$	**Hierarchical Clustering**	Assets	F.EI	F.FTEL	E.TEF	I.ENEL	D.VO3X
		Weights	0.47	0.23	0.13	0.09	0.07
	Fuzzy Clustering	Assets	F.EI	F.FTEL	F.BSN	D.RWEX	I.G
		Weights	0.42	0.22	0.16	0.12	0.09

Table 6.17. Composition of the minimum variance portfolio selected by hierarchical cluster procedure and fuzzy cluster procedure using TDC and $\rho_{0.10}$ measure.

		Mean	S.D.	Skewness	5% VaR	5% E.S.	Sharpe Ratio
λ_L	HC Portfolio	0.000	0.005	-0.043	-0.008	-0.012	0.012
	FC Portfolio	0.000	0.005	0.030	-0.008	-0.012	0.011
$\rho_{0.10}$	HC Portfolio	0.000	0.005	-0.064	-0.009	-0.012	0.007
	FC Portfolio	-0.000	0.005	0.062	-0.008	-0.012	-0.000
	Naive MVP	0.000	0.005	-0.153	-0.008	-0.011	0.015
	EURO STOXX 50	-0.000	0.007	0.064	-0.011	-0.017	-0.013

Table 6.18. Basic statistics related to the log–returns of selected minimum variance portfolios in the convex frontier by hierarchical clustering and fuzzy clustering. Period: August 1, 2011 – September 9, 2011.

6.4 Conclusion

We have introduced a procedure aiming at selecting a portfolio from a group of assets in such a way that the assets are diversified in their tail behavior. It is grounded on previous idea by [De Luca and Zuccolotto, 2011]. While, they proposed a procedure for portfolio diversification with assuming a parametric model. In our work, we changed the tail dependence measures, checked the feasibilities, and implemented two clustering algorithms: fuzzy and hierarchical.

It has been shown that the method works under more general assumptions about the involved tail dependence measure. It may provide a tool (mainly graphical) to show the possible evolution of our portfolios in crisis periods, and motivate us to more sound, mathematical constructions about tail dependence. It should not be thought as a tool for obtaining big gains, but, it may serve to have a more detailed view about how we may diversify a portfolio in a non-linear way.

But we also need to keep in mind the possible limitations of the methodology. One is that we need to calculate all the possibilities of the K-asset combinations, if we want to have a complete picture.

However, this is clearly a computational unfeasible for large portfolios. Moreover, the frontier chosen by us is also critical. To detect whether this method works or not, we also need to compare it with other portfolio construction methods, we also need to carry out this method under different markets conditions.

Chapter 7

A Spatial Financial Contagion Measure for Financial Time Series

7.1 Introduction

Measuring co-movements among financial time series is a widely debated issue since the seminal work of [Grubel, 1968] underlined the benefits from international portfolio diversification. In fact, it has been recognized that investors could reduce the risk of their portfolios when the comovements among the assets is taken into account (See also Chapter 6).

Therefore, the portfolio diversification issue naturally poses the question of investigating the relationship between financial time series and of checking whether they can be grouped together in a way that may be helpful to portfolio selection. As such, there has been a growing interest in exploring clustering methods to financial time series, mainly motivated by the need to reduce the complexity of a portfolio by grouping similar assets. A class of such methods is based on the pairwise dependence among assets. For instance, a popular methodology is grounded on the use of the Pearson correlation; e.g., see [Mantegna, 1999, Bonanno et al., 2004] and the references therein. Most recently, instead, several investigations have focused on the extreme linkage among financial markets; i.e., on those dependence properties that arise when one or more markets are exhibiting large losses. In this respect, the dissimilarity measures among assets have been derived by the tail dependence coefficient, as done by [De Luca and Zuccolotto, 2011, Durante et al., 2013d], or by conditional correlation coefficients, like Spearman's correlation, as suggested by [Durante et al., 2013c]. Other methods include, for instance, [Basalto et al., 2007, Corduas and Piccolo, 2008, Otranto, 2008, Brida and Risso, 2010, Bastos and Caiado, 2013, D'Urso et al., 2013, Aghabozorgi and Teh, 2013].

However, especially during the last global financial crisis (i.e., 1997 Asian crisis, the so–called 2008 Subprime Mortgage, and 2011 Sovereign Debt crisis) a clear need arises among academic and financial engineering communities; namely, how to distinguish between markets interdependence, i.e. the presence of comovements among the markets, and contagion, i.e. the presence of a mechanism that allows the propagation of financial difficulties from one economy to the others (see, e.g., [Kolb, 2011] and the references therein). In fact, although these two aspects are closely related, they underline two different features of the market behavior. In particular, recognizing and managing the presence of contagion provides useful benefits when dealing with financial risks. For instance, diversification strategies for building portfolios may fail when the dependence among assets changes in crisis period with respect to untroubled periods.

The notion of contagion is basically a new concept in the economic literature. "A search of the

EconLit data set for 1969–February 2000, yielded 147 entries that had the word 'contagion' in either the title or in the abstract. Of these, only 17 corresponded to works published before 1990"([Edwards, 2000, P.1]). In the late nineties, the economists began to focus more on contagion. This interest originated in the spread of the financial difficulties between international financial markets, especially in emerging markets. The Asian financial crisis of 97-98 exacerbated concerns about the financial contagion ([Hunter et al., 1999]). Researchers found surprisingly that global finance is not stable anymore, a local-national financial crisis would affect the other nations with "apparently healthy fundamentals and whose policies, 'only' a few months earlier, has been praised by market analysts and the multilateral institutions"([Edwards, 2000, P.1]). Moreover, the 2007-2009 financial crisis showed that contagion is no longer limited to the emerging markets([Kolb, 2011]).

Now, many scholars have studied the channels of contagion or the spread mechanisms. When researchers talk about this concept, one key related question is how to identify the occurrence of a contagion. As noted in [Kolb, 2011, P.4], "there is some danger of conflating contagion with the evidence of contagion. That is, there is a risk of taking evidence of contagion as the malady itself". For example, the highlighted correlation may be caused by the increased volatilities ([Forbes and Rigobon, 2002]).

One difficulty for dealing with financial contagion is that, there is no agreement in the literature on the meaning for it. For some scholars, contagion is merely the diffusion of financial stress without connotations of disease: ([Caramazza et al., 2004, P.51])

> "the spread of financial difficulties from one economy to others in the same region and beyond in a process that has come to be referred to as 'contagion'. "

In [Pericoli and Sbracia, 2003, PP. 574–575], five definitions of contagion that reflect the wide variety of meanings ascribed to this term are considered:

- Contagion is a significant increase in the probability of a crisis in one country, conditional on a crisis occurring in another country;
- Contagion occurs when volatility of asset prices spills over from the crisis country to other countries;
- Contagion occurs when cross-country comovements of asset prices cannot be explained by fundamentals;
- Contagion is a significant increase in comovements of prices and quantities across markets, conditional on a crisis occurring in one market or group of markets;
- (Shift-) contagion occurs when the transmission channel intensifies or, more generally, changes after a shock in one market".

As concluded in [Kolb, 2011], these five definitions exhibit substantial conceptual differences: "the first is defined as a change in probabilities of a crisis, while the second focuses on a change in volatilities, and the third speaks of markets or groups of markets".

However, defining contagion based on how fast financial difficulties spread is debatable. For [Kaminsky et al., 2003, P. 6], contagion is

> "... an episode in which there are significant immediate effects in a number of countries following an event– that is, when the consequences are fast and furious and evolve over a matter of hours or days. "

Here, we follow on one of these definitions that refers to contagion as "a significant increase in comovements of prices and quantities across markets, conditional on a crisis occurring in one market or group of markets". In the literature, several theoretical and empirical works have been devoted to the search for the change in the correlation/dependence structure of the underlying distribution governing the behavior of historical financial time series as an indicator of the presence of contagion.

Referred to this, some tests for contagion based on correlation coefficients were developed (see. e.g. the discussion in [Corsetti et al., 2011]). Specifically, if there is a significant increase in the correlation coefficient in financial returns between two markets after a shock, with respect to the cross-country correlation during a stable period, then it can be argued that the transmission mechanism between the two markets strengthened after the shock and contagion occurred. As noticed, for instance, by [Kaminsky et al., 2003], correlation-based contagion appears "only if there is excess comovement in financial and economic variables across countries in response to a common shock". Namely, if two markets are strongly correlated at any time (but the link is not changing), they exhibit no contagion.

But these tests for contagion based on correlation coefficients may be problematic. As noticed in [Ronn, 1998], changes in market volatility can bias the estimate of correlation coefficients and, hence, the related detection of contagion. In the later years, a series of papers began to investigate how the bias affects cross-market correlations or the measurement of contagion in more detail. [Boyer et al., 1999] proposed an adjustment to the correlation coefficient

$$\rho_A = \frac{\rho}{\sqrt{\rho^2 + (1-\rho^2)\frac{\mathrm{Var}(x)}{\mathrm{Var}(x|x\in A)}}}, \tag{7.1}$$

where ρ_A in (7.1) is the conditional correlation coefficient between x and y, conditional on the event $x \in A$. If the conditional correlation during the turmoil (high volatility) period is greater than the unconditional correlation during the full period, they say that contagion happens between these two markets. In [Bradley and Taqqu, 2004], this phenomenon was referred to as *correlation breakdown*. But this conditional correlation coefficient is still biased by heteroskedasticity. As concluded in [Bradley and Taqqu, 2004]:"during the crisis periods A, the volatility $\mathrm{Var}(x|x \in A)$ increases leading to $\mathrm{Var}(x|x \in A) > \mathrm{Var}(x)$. This makes $\rho_A > \rho$ automatically, even though the transmission mechanism from one market to another market remains constant. Moreover, the choice of conditioning event may lead to spurious conclusions".

In [Forbes and Rigobon, 2002], this difficulty was pointed out and a heteroscedasticity corrected version was proposed. An unconditional correlation coefficient during the turmoil period was then defined

$$\rho = \frac{\rho^*}{\sqrt{1+\delta(1-(\rho^*)^2)}}, \tag{7.2}$$

where ρ^* is the conditional correlation coefficient during the turmoil period, ρ is the unconditional correlation coefficient during the turmoil period, and δ is the relative increase in the variance of the variable considered x:

$$\delta \equiv \frac{\sigma_{xx}^h}{\sigma_{xx}^l} - 1, \tag{7.3}$$

where σ_{xx}^l and σ_{xx}^h are the variance of the period of relative market stability and of market turmoil directly after the shock or crisis, respectively. Under this definition, if ρ, the unconditional correlation during the full period, is less than or equal to ρ_t^h, that is ρ in 7.2, then contagion is presented. They

called "*interdependence*" when the conditional correlation efficient $(\rho_t^*)^h$ is greater than ρ, but the unconditional correlation efficient ρ_t^h is less than or equal to ρ. The definition of Forbes and Rigobon was an asymmetric contagion, meaning that the contagion starts from x.

But one problem of this test was pointed out by Bradley and Taqqu, namely that "the power of this test is very low due to the short crisis period". Meanwhile, they proposed a different approach to attempt to overcome this problem in [Bradley and Taqqu, 2004, 2005a,b]. In these papers, they considered a definition of contagion based on a local correlation coefficient and provided a test for contagion. The local correlation coefficient is a nonparametirc and nonlinear dependence measure. It was proposed by [Bjerve and Doksum, 1993, Doksum et al., 1994].

The local correlation of Bradley and Taqqu between Y and X at $X = x$ is given by

$$\rho(x) = \frac{\sigma_X \beta(x)}{\sqrt{\sigma_X^2 \beta^2(x) + \sigma^2(x)}} \tag{7.4}$$

where σ_X denotes the standard deviation of X, $\beta(x) = m'(x)$ is the slope of the regression function $m(x) = \mathbb{E}(Y|X = x)$ and $\sigma_x^2 = \mathrm{Var}(Y|X = x)$ is the nonparametric residual variance. Notice that the measure of local correlation does not require the specification of a crisis and non-crisis period or the use of a heteroscedasticity correction as in (7.2).

Under the definition of local correlation coefficient, there is contagion from market X to market Y if

$$\rho(x_L) > \rho(x_M) \tag{7.5}$$

where $x_M = F_X^{-1}(0.5)$ is the median of the distribution of X, and $x_L = F_X^{-1}(0.025)$ is a low quantile of the distribution. Following their idea, there is contagion from market X to market Y if there is more dependence between X and Y when X is doing badly than when X exhibits typical performance, that is, if there is more dependence at the loss distribution of X than at its center ([Durante and Jaworski, 2010]).

In the same spirit of [Bradley and Taqqu, 2004], [Durante and Jaworski, 2010] defined and investigated a related notion of *spatial contagion* between two financial markets X and Y, by describing their dependence not by means of the local correlation coefficient, but by using the information contained in the copula of (X, Y). Notice that copulas have been used several times for checking financial contagion: see, for instance, [Rodriguez, 2007, Kenourgios et al., 2011, Peng and Ng, 2012, Ye et al., 2012]. However, most of these approaches require the specification of a parametric (copula-based models) for asset returns. Such models that have several advantages as illustrated in [Patton, 2012], may suffer from the fact that most statistical goodness-of-fit tests cannot identify the correct copula (see, e.g., [Grundke and Polle, 2012]).

Instead, as discussed by [Durante and Jaworski, 2010] (see also [Durante et al., 2013b, Durante and Foscolo, 2013]), a notion of spatial contagion can be introduced by using the dependence (i.e., the copula) associated with the random variables X and Y representing the returns of two financial markets. Specifically, a practical way to check contagion may be implemented by comparing Spearman's correlation among the time series when X and Y are experiencing severe losses with Spearman's correlation when X and Y are in an untroubled (i.e., tranquil) scenario. Such an idea is grounded on some previous investigations about exceedance correlations [Longin and Solnik, 2001]. However, Spearman's correlation is used instead of Pearson's correlation in order to detect nonlinear (yet monotone) dependence among the data (see, e.g., [McNeil et al., 2005]).

The aim of the present work is to analyze spatial financial contagion by providing a new measure that goes beyond the use of linear correlation and does not require, in addition, the a priori specification

of crisis/non-crisis periods via suitable thresholds (as done for instance in [Durante and Jaworski, 2010]). The new measure is completely data-driven and can be calculated via non-parametric methods. As such, it avoids possible misspecification in the dependence structure. Spatial contagion measure can be calculated empirically via simple procedures and could be also used to identify sub-groups of assets that have similar behavior in crisis periods. In fact, compared with related approaches in the literature, the spatial contagion measure detects the changes in the dependence structure among financial markets and not the presence of a persistent extreme dependence, which is due to interdependence, but not to contagion (for such viewpoint, compare with the discussion in [Forbes and Rigobon, 2002]). Moreover, the measure of symmetric spatial contagion could be used to implement clustering procedures for financial time series, aiming at finding sub-groups of assets that have similar behavior in periods of market distress.

This chapter is organized as follows. In Sections 7.2 and 7.3, a novel contagion measure is introduced and a detailed description on how to compute it is given. Section 7.4 also reports a simulation study where we show our method at work according to different conditions of (extreme) dependence. Section 7.5 illustrates the methodology in a practical analysis of a portfolio of financial markets. In particular, a clustering procedure is illustrated in order to provide some insights for portfolio selection. Finally, Section 7.6 concludes.

7.2 The Spatial Contagion Measure: Theoretical Aspects

We recall here some basic aspects of this procedure that will be useful in the following.

Let X and Y be two random variables on a suitable probability space representing the returns (or log–returns) of financial markets whose dependence is described by means of a copula C. Consider the following Borel sets of $\mathbb{R}^2$:

- the *tail set* T_{α_1,α_2} given by

$$T_{\alpha_1,\alpha_2} = [-\infty, q_X(\alpha_1)] \times [-\infty, q_Y(\alpha_2)],$$

  where $\alpha_1, \alpha_2 \in [0,1]$ and q_X and q_Y are the quantile functions associated with X and Y, respectively.

- the *central set* (or *mediocre set*) M_{β_1,β_2} given by

$$M_{\beta_1,\beta_2} = [q_X(\beta_1), q_X(1-\beta_1)] \times [q_Y(\beta_2), q_Y(1-\beta_2)]$$

  where $\beta_1, \beta_2 \in [0,1]$.

Intuitively, T_{α_1,α_2} represents the "risky scenario" for the pair (X,Y), since it includes the bivariate observations that are less than a given threshold; while M_{β_1,β_2} represents the so-called "untroubled scenario", since it is related to all the observations that are in the central region of the joint distribution (the extreme values are excluded). For sake of being completed, we recall that the quantile function associated with a continuous random variable X with d.f. F is given by $q_X(\alpha) = \inf\{x\colon F(x) \geq \alpha\}$ for every $\alpha \in [0,1]$. Moreover, we set $q_X(1) = +\infty$ and $q_X(0) = -\infty$ for any random variable X.

Following [Durante and Jaworski, 2010] (see also [Durante et al., 2013b, Durante and Foscolo, 2013, Jaworski and Pitera, 2013]), the notion of spatial contagion may be given in terms of (conditional) Spearman's correlation ρ in the following way:

- There is *symmetric contagion* of (X, Y) at level $\alpha \in (0, 0.5)$ when

$$\rho(X, Y \mid (X, Y) \in T_{\alpha,\alpha}) > \rho(X, Y \mid (X, Y) \in M_{\alpha,\alpha})$$

- There is *contagion from X to Y* at level $\alpha \in (0, 0.5)$ when

$$\rho(X, Y \mid (X, Y) \in T_{\alpha,1}) > \rho(X, Y \mid (X, Y) \in M_{\alpha,0})$$

- There is *contagion from Y to X* at level $\alpha \in (0, 0.5)$ when

$$\rho(X, Y \mid (X, Y) \in T_{1,\alpha}) > \rho(X, Y \mid (X, Y) \in M_{0,\alpha})$$

Thus, loosely speaking, there is contagion from X to Y when the dependence between X and Y is stronger given that X is having severe losses (as defined by a suitable level α) than given that X is experiencing typical performances. Analogously, there is symmetric contagion when the dependence between X and Y is stronger given that both X and Y are experiencing extreme losses than given that X and Y are experiencing typical performances. In particular, if two markets are perfectly comonotone (i.e., $Y = f(X)$ for a strictly increasing f), then there is no contagion between them. In fact, they remain comonotone with respect to any conditioning and no change of dependence can be detected.

For sake of simplicity, in the following we adopt the convention $\rho(T_{\alpha,\alpha})$ and $\rho(M_{\alpha,\alpha})$ (and similar modifications), when the random variables we are referring to can be identified from the context.

In practical applications, the presence of spatial contagion can be easily checked and tested. However, typically, the choice of the threshold α strongly influences the final result. In fact, it is common that for some $\alpha \in (0, 0.5)$ contagion appears (i.e. $\rho(T_{\alpha,\alpha}) - \rho(M_{\alpha,\alpha})$ is significantly greater than 0), while for other α there is no such evidence. An example is given by Figure 7.1.

Now, while it is sometimes convenient to leave to the decision maker (i.e., the risk manager) the choice of α according to her/his risk level or taking into account regulatory constraints, in other cases the arbitrariness is neither appropriate nor useful. In [Jaworski and Pitera, 2013], for instance, an automatic method is presented for the choice of the optimal α, which is determined in comparison with the Gaussian model.

Another possibility is considered here. Loosely speaking, instead of selecting a single α, we determine a set L of all possible α's to be considered, and calculate how often there is contagion at the level α, for α belonging to L. Then we may do a kind of average of all the α's that induce contagion with respect to the whole set (compare also with Eq.'s (6) and (7) in [Jaworski and Pitera, 2013]). The formal definition of this procedure is stated below.

Definition 7.2.1. Let X and Y be two random variables representing the (log–) returns of two financial markets. The *symmetric contagion measure* between X and Y is defined by the formula

$$\gamma(X, Y) = \frac{1}{\lambda(L)} \lambda\left(\{\alpha \in L \mid \rho(T_{\alpha,\alpha}) - \rho(M_{\alpha,\alpha}) > 0\}\right), \tag{7.6}$$

where λ is the Lebesgue measure on $[0, 1]$ and $L \subseteq [0, 0.5]$ is a (connected) set of possible values taken by α.

In other words, $\gamma(X, Y)$ is defined as the (normalized) measure of the set of all α's such that there is symmetric contagion at α. Analogously, the following definitions can be given.

Definition 7.2.2. Let X and Y be two random variables representing the (log–) returns of two financial markets. The *contagion measure from X to Y* is defined by the formula

$$\gamma(X \to Y) = \frac{1}{\lambda(L)} \lambda\left(\{\alpha \in L \mid \rho(T_{\alpha,1}) - \rho(M_{\alpha,0}) > 0\}\right). \tag{7.7}$$

Analogously, the *contagion measure from Y to X* is defined by the formula

$$\gamma(Y \to X) = \frac{1}{\lambda(L)} \lambda\left(\{\alpha \in L \mid \rho(T_{1,\alpha}) - \rho(M_{0,\alpha}) > 0\}\right). \tag{7.8}$$

The contagion measures so defined take the following values in $[0, 1]$: the value 0 corresponds to the absence of contagion at any level $\alpha \in L$, while the value 1 corresponds to the case when, at any $\alpha \in L$ a contagion is detected. Notice that, in absence of any further specification, L should be taken as large as possible, apart from some numerical difficulties in calculating the conditional Spearman's ρ (see below).

As it can be seen, the contagion measures depend on the quantity $\rho(T) - \rho(M)$, for some tail set T and central set M, that, in its turn, is related to the copula of (X, Y) as stressed by [Durante and Jaworski, 2010]. We recall that copulas are exactly the objects that allow the capture of the dependence between random variables and, as such, they have been employed in a number of studies about financial time series (see, for instance, [Jaworski et al., 2010, Jaworski and Pitera, 2013] and the references therein). Specifically, for given tail set T and central set M, $\rho(T)$ depends on the copula of the conditional distribution function of $[X, Y \mid (X, Y) \in T]$, while $\rho(M)$ depends on the copula of the conditional distribution function of $[X, Y \mid (X, Y) \in M]$. Copulas related to conditional distributions are also called *threshold copulas* [McNeil et al., 2005, Durante et al., 2008].

The analytical computation of the contagion measure introduced above is quite cumbersome. In fact, in general, the process of selecting appropriate copula model for the data at hand may be subjected to estimation error and misspecification; two sources of possible errors may amplify the calculation of the contagion measures. Moreover, even if the associated copula of (X, Y) was known, it is quite difficult to obtain explicit formulas for the associated threshold copulas. Therefore, also in order to avoid any parametric assumption, we rely on numerical approximations, as we illustrate below.

7.3 The Spatial Contagion Measure: Computational Aspects

After introducing the contagion measure, it is useful to discuss some computational aspects that are related to its practical implementation.

Suppose that the symmetric contagion measure between two financial markets X and Y has to be calculated (the case of asymmetric contagion can be done analogously). Usually, we have at disposal time series, $(X_t)_{t=1,\dots,n}$ and $(Y_t)_{t=1,\dots,n}$, representing the log-returns of the given stock indices in a given period of time.

The first step of the procedure is to calculate the conditional Spearman's correlation associated with the observations. These procedures run as follows. For more details about their practical implementation, see [Dobric et al., 2014, Durante and Jaworski, 2010].

Algorithm 7.3.1 (Calculation of $\rho(T_{\alpha,\alpha})$ and $\rho(M_{\alpha,\alpha})$)**.**

1. Set $\alpha \in (0, 0.5)$.
2. Calculate the empirical cumulative distribution functions $\widehat{F}_X$ and $\widehat{F}_Y$ associated with $(X_t)_{t=1,\dots,n}$ and $(Y_t)_{t=1,\dots,n}$, respectively.

3. For any $t = 1, \ldots, n$, let $(R_t, S_t) = (\widehat{F}_X(X_t), \widehat{F}_Y(Y_t))$.

4. Select all the observations in the sets

$$\widehat{T} = \{(R_t, S_t) \mid R_t \leq \alpha, S_t \leq \alpha\}$$

and

$$\widehat{M} = \{(R_t, S_t) \mid \alpha < R_t < 1 - \alpha, \alpha < S_t < 1 - \alpha\}.$$

5. Denote by $\mathscr{I}_T$ (respectively, $\mathscr{I}_M$) the set of all indices t's such that $(R_t, S_t) \in \widehat{T}$ (respectively, $(R_t, S_t) \in \widehat{M}$)

6. Calculate the univariate empirical cumulative distribution functions $\widehat{F}_{T,X}$ and $\widehat{F}_{T,Y}$ associated with all the observations $(X_t, Y_t)_{t \in \mathscr{I}_T}$.

7. Calculate the univariate empirical cumulative distribution functions $\widehat{F}_{M,X}$ and $\widehat{F}_{M,Y}$ associated with all the observations $(X_t, Y_t)_{t \in \mathscr{I}_M}$.

8. For any index $t \in \mathscr{I}_T$ (respectively, $\mathscr{I}_M$), let $(R'_t, S'_t) = (\widehat{F}_{T,X}(X_t), \widehat{F}_{T,Y}(Y_t))$ (respectively, $(R''_t, S''_t) = (\widehat{F}_{M,X}(X_t), \widehat{F}_{M,Y}(Y_t))$).

9. Calculate Spearman's correlation $\widehat{\rho}(T_{\alpha,\alpha})$ and $\widehat{\rho}(M_{\alpha,\alpha})$ given by

$$\widehat{\rho}(T_{\alpha,\alpha}) = \frac{12}{n_T} \sum_{t \in \mathscr{I}_T} R'_t S'_t - 3 \qquad \widehat{\rho}(M_{\alpha,\alpha}) = \frac{12}{n_M} \sum_{t \in \mathscr{I}_M} R''_t S''_t - 3$$

where n_T is the cardinality of $\widehat{T}$ and n_M is the cardinality of $\widehat{M}$.

Notice that, by using similar arguments as in [Schmid and Schmidt, 2007], it can be proved that the conditional versions of Spearman's rho described above are consistent and asymptotically normally distributed; i.e.,

$$\sqrt{n_T}(\widehat{\rho}(T_{\alpha,\alpha}) - \rho(T_{\alpha,\alpha})) \xrightarrow{d} N(0, \sigma_T^2) \qquad \sqrt{n_M}(\widehat{\rho}(M_{\alpha,\alpha}) - \rho(M_{\alpha,\alpha})) \xrightarrow{d} N(0, \sigma_M^2)$$

as n_T and n_M tend to ∞, provided that the threshold tail copulas and central copulas exist and satisfy some regularity assumptions. Here σ_T^2 and σ_M^2 depend on the threshold copulas associated to $T_{\alpha,\alpha}$ and $M_{\alpha,\alpha}$, respectively.

Furthermore, in the previous algorithm, the calculation of Spearman's correlation depends on the number of points n_T and n_M in tail and central regions, respectively. If such number is small, the estimated correlations would be affected by this small sample size [Dobric et al., 2014]. Typically, a convenient sample size may be reached by selecting an appropriate threshold α (in this chapter, if not otherwise specified, we consider the point 0.05 as lower bound for L).

Now, given the way to approximate the conditional Spearman's correlation, the calculation of the contagion measure $\gamma(X, Y)$ proceeds by a discrete approximation of Equation (7.6). Notice that the input value is p, which is the number of division of the interval L, i.e. the accuracy of the discrete approximation (the larger p the finer the partition). The algorithm goes as it follows.

Algorithm 7.3.2 (Symmetric contagion - naive version)**.**

1. Set $\varepsilon \geq 0$.

2. Fix $L = [a, b] \subseteq [0, 0.5]$ and divide it into a finite number of equidistant points $L_1 = a, \ldots, L_p = b$ such that $|L_i - L_{i+1}| = \ell$ for each i.

3. For $L_1, L_2, \ldots, L_p$, calculate

$$\widehat{\Delta}_i = \widehat{\rho}(T_{L_i,L_i}) - \widehat{\rho}(M_{L_i,L_i})$$

4. Return

$$\hat{\gamma}\,(X, Y) = \frac{\#\{i\colon \widehat{\Delta}_i > \varepsilon\}}{p}$$

Algorithm 7.3.2 is a direct consequence of the definition of contagion measure as soon as $\varepsilon = 0$. However, in this case, it performs poorly, especially when the dependence between the time series is close to the independent case. In fact, consider that the data generating process is equal to the independence copula. Then the contagion measure should be equal to 0 (there is no difference in the dependence strength in the tail or in the center of the distribution); however, due to sample variability, it turns out to have much bigger values, as we noticed in a series of simulations (see Tables 7.1 and 7.2). In fact, $\widehat{\rho}(T_{\alpha,\alpha}) - \widehat{\rho}(M_{\alpha,\alpha})$ is approximately Gaussian distributed with 0 mean [Durante and Jaworski, 2010]. Thus, due to sample variability, $\widehat{\rho}(T_{\alpha,\alpha}) - \widehat{\rho}(M_{\alpha,\alpha})$ is strictly positive with probability 0.5.

A remedy to this fact is to count only the i's where the $\widehat{\Delta}_i := \widehat{\rho}(T_{L_i,L_i}) - \widehat{\rho}(M_{L_i,L_i})$ is "significantly" different from 0, that is to fix $\varepsilon > 0$ according to a given criterion.

Here we propose a heuristic procedure for selecting $\varepsilon > 0$ when the financial time series of interest has length equal to n. The idea is that the optimal ε should be chosen in such a way that, under a sampling of n independent and uniform random variables, the average contagion measure should not exceed $1/p$ (the accuracy of the discrete approximation). This could be done by supposing that ε is a power function of the sample size n, i.e. $\varepsilon = 1/n^k$ where $k \in \{k_1, \ldots, k_d\}$ (where $k_1 > k_2 > \cdots > k_d > 0$), and then find the optimal k^* by simulations. The procedure goes as follows.

Algorithm 7.3.3 (Choice of optimal ε)**.**

1. Set $i = 1$

2. Set $k = k_i$.

3. Set $\varepsilon = 1/n^k$.

4. For $b = 1, \ldots, B$, simulate n variates from two independent and uniform random variables (U_b, V_b) and calculate

$$m = \frac{1}{B}\sum_{b=1}^{B} \hat{\gamma}(U_b, V_b)$$

where $\widehat{\gamma}$ is obtained by using Algorithm 7.3.2 with ε defined in point 3.

5. If $m < 1/p$, then return k.

6. Else, $i = i + 1$ and return to point 2.

Loosely speaking, the value ε determined by Algorithm 7.3.3 represents a kind of uniform noise level that we may have due to sample variability. Such noise could be calculated in the easy situation when independence occurs and, as said, the contagion measure is equal to 0. Notice that another situation where the contagion measure is equal to 0 is in the comonotonic case, i.e. one return is an increasing monotone function of the other one. However, in such a case, there is no possibility of bias in the detection of contagion since the scatter plot of the data is represented by a line segment.

Figure 7.2 represents a typical graph for the choice of the optimal k when a sample of 5000 points is considered, together with $B = 100$ replications, and the naive version of the symmetric contagion measure is calculated with $L = [0.05, 0.45]$ and $\ell = 0.01$ $(1/p = 0.0244)$. The optimal k is given by the value $k^* = 0.1675$, which corresponds to an optimal $\varepsilon^* = 0.2398$. The performance of such procedure in a simulation study is shown in Section 7.4.

An alternative to the procedure of Algorithm 7.3.2 may be considered by using the asymptotic properties of the population version of the process $\Delta_\alpha = \rho(T_{\alpha,\alpha}) - \rho(M_{\alpha,\alpha})$. Specifically, by repeating similar arguments as in [Dobric et al., 2014], we have

$$\sqrt{n}(\widehat{\Delta}_\alpha - \Delta_\alpha) \xrightarrow{d} N\left(0, \frac{\sigma_T^2}{V_C(T_{\alpha,\alpha})} + \frac{\sigma_M^2}{V_C(M_{\alpha,\alpha})}\right)$$

where C is the copula linking the variables of interest, while $V_C(M_{\alpha,\alpha})$ is the C–volume of the central set $M_{\alpha,\alpha}$ and $V_C(T_{\alpha,\alpha})$ is the C–volume of the tail set $T_{\alpha,\alpha}$. The variance of the previous process depends on the underlying copula and it is, in general, difficult to compute. However, we may rely on bootstrap procedures to approximate it. In particular, if we fix a confidence level κ (usually, $\kappa = \{0.75, 0.90, 0.95, 0.99\}$)), we may obtain a lower bound for the confidence interval for the estimation of Δ_α. Now, the idea behind the new algorithm for calculating the contagion measure is to check whether such a lower bound is strictly positive. To this end, the previous naive algorithm is modified in the following way.

Algorithm 7.3.4 (Symmetric contagion - robust version)**.**

1. Fix $\varepsilon = 0$.
2. Fix $L = [a, b] \subseteq [0, 0.5]$ and divide it into a finite number p of equidistant points $L_1 = a, \ldots, L_p = b$ such that $|L_i - L_{i+1}| = \ell$ for each i.
3. For $L_1, L_2, \ldots, L_p$:
   (a) Calculate
   $$\hat{\Delta}_i = \widehat{\rho}(T_{L_i,L_i}) - \widehat{\rho}(M_{L_i,L_i}).$$
   (b) Generates B bootstrap samples from the original sample and for each replication determine $\hat{\Delta}_i^B$.
   (c) Set $\underline{\hat{\Delta}}_i^B$, the empirical quantile of order $(1 - \kappa)$ for the bootstrapped values $\hat{\Delta}_i^B$'s.
4. Return
   $$\hat{\gamma}(X, Y) = \frac{\#\{i : \underline{\hat{\Delta}}_i > \varepsilon\}}{p}$$

This alternative algorithm has been tested in the case of independent random variables (see Section 7.4) and has proved to give quite stable results in terms of the calculated contagion measure. However, we have to notice that it is quite demanding in terms of computational time since it requires a series of bootstrap replications for each division of the interval L.

7.4 Contagion Measures and Dependence: a Simulation Study

In this section, we calculate the contagion measure between two financial assets X and Y supposing that their dependence is described by a specific copula belonging to different known families. Here the main goal is to show whether the contagion measure just introduced is able to recover some stylized facts that are already known in copula theory. Notice that all the calculations have been performed by using the symmetric contagion measure, but they are very similar to the case of asymmetric contagion measure (these results are not reported for the sake of simplicity).

Specifically, we consider data generated from different dependence structures, namely Gaussian, Student-t, Frank, Clayton and Survival Gumbel with different values of Spearman's correlation (for formal definitions of these families, see [McNeil et al., 2005, Durante and Sempi, 2010]. Moreover, in order to check the performance of our contagion measure, we also include the independence copula Π; i.e., the copula associated with independent random variables, which does not describe contagion.

We start our numerical simulation in order to detect symmetric contagion and we suppose that the financial markets (X, Y) are linked by a copula from one of the previous classes. The sampling algorithm for each copula family are provided by the R package `copula`. The simulation study has been performed by using `R` 3.0.1 (64-bit). Moreover, we have exploited `doSNOW` for using multicore parallel processing to speed up Monte Carlo (MC) simulation.

For each copula we proceed as follows.

1. We choose a copula C belonging to one of the considered families and with a prescribed (unconditional) Spearman's correlation ρ;
2. We generate a 2-dimensional sample of size N from the copula C;
3. We estimate $\gamma(X, Y)$ following the procedures described in section 7.3.
4. Return to step 2 and redo it S times;
5. Finally, we compute the mean over all S-simulations of $\gamma(X, Y)$.

A comment is needed here in order to stress the role of the involved setting parameters.

- The set L has been chosen in the following way. The lower bound is equal to 0.05, because a smaller value would not guarantee an adequate number of observations in the tail set and may give biased results. The upper bound is selected to be equal to 0.18 or 0.45. The value 0.18 according to the optimal threshold calculated by [Jaworski and Pitera, 2013]. Under this assumption, Gaussian copulas with a positive correlation have a symmetric contagion measure equal to 0, so that this bound may serve in order to detect the possible presence of non–Gaussianity in the data. The value 0.45 is somehow a standard choice if one wants to consider the whole range of possible losses.
- The partition of L has been obtained by a step $\ell = 0.01$, which has been empirically tested as the most convenient alternative. By making it finer, not only do we not improve the results, but also add some noise to the computations.
- The number of bootstrap replications of Algorithm 7.3.4 is fixed to 100 in order to avoid large time consuming procedures. As can be seen from the results, this number already allows good overall performances.

The results related to Algorithm 7.3.2 with $\varepsilon = 0$ are reported in Tables 7.1 and 7.2. As it can be seen, the values of the contagion measure are quite large, especially when the dependence structure is close to the independent case or when no contagion should be detected, as in the case of Gaussian copula in Table 7.2. Overall, as also stressed above, it would be necessary to fix a suitable $\varepsilon > 0$. This procedure is performed in Tables 7.3 and 7.4. Here, on the contrary, the detection of contagion is much more precise in the independent case and in the case of Gaussian copulas with $L = [0.05, 0.18]$. Moreover, generally, all the values are much lower than in the naive algorithm with $\varepsilon = 0$.

	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 0$
Gaussian	0.6028	0.6647	0.6514	0.4987
Student-t	0.9992	0.9906	0.9031	
Frank	0.3268	0.3359	0.3593	
Clayton	0.9734	0.9955	0.9968	
Survival Gumbel	0.9897	0.9868	0.9367	

Table 7.1. Calculation of symmetric contagion measure via Algorithm 7.3.2 with $\varepsilon = 0$ from different copula families. Sample size $N = 5000$, $L = [0.05, 0.45]$. $S = 250$.

	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 0$
Gaussian	0.3081	0.1293	0.0429	0.4950
Student-t	0.9989	0.9756	0.7284	
Frank	0.0891	0.0001	0.0000	
Clayton	0.9317	0.9849	0.9900	
Survival Gumbel	0.9717	0.9621	0.8097	

Table 7.2. Calculation of symmetric contagion measure via Algorithm 7.3.2 with $\varepsilon = 0$ from different copula families. Sample size $N = 5000$, $L = [0.05, 0.18]$. $S = 250$.

	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 0$
Gaussian	0.0140	0.0974	0.3994	0.0222
Student-t	0.8708	0.6762	0.5268	
Frank	0.0083	0.0242	0.1358	
Clayton	0.2917	0.7866	0.7264	
Survival Gumbel	0.3721	0.6868	0.6343	

Table 7.3. Calculation of symmetric contagion measure via Algorithm 7.3.2 with $\varepsilon = 0.2335$ (calculated using Algorithm 7.3.3) from different copula families. Sample size $N = 5000$, $L = [0.05, 0.45]$. $S = 250$.

	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 0$
Gaussian	0.0043	0.0000	0.0000	0.0691
Student-t	0.7823	0.1949	0.0000	
Frank	0.0006	0.0000	0.0000	
Clayton	0.1440	0.4403	0.2411	
Survival Gumbel	0.2706	0.2383	0.0311	

Table 7.4. Calculation of symmetric contagion measure with Algorithm 7.3.2 with $\varepsilon = 0.2225$ (calculated using Algorithm 7.3.3) from different copula families. Sample size $N = 5000$, $L = [0.05, 0.18]$. $S = 250$.

$\kappa = 0.90$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 0$
Gaussian	0.2910	0.5395	0.6132	0.1607
Student-t	0.9898	0.9515	0.8378	
Frank	0.1112	0.1888	0.3180	
Clayton	0.8927	0.9756	0.9800	
Survival Gumbel	0.9356	0.9493	0.8939	
$\kappa = 0.95$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 0$
Gaussian	0.2368	0.5120	0.6046	0.1188
Student-t	0.9844	0.9368	0.8237	
Frank	0.0788	0.1632	0.3071	
Clayton	0.8641	0.9671	0.9749	
Survival Gumbel	0.9112	0.9383	0.8841	
$\kappa = 0.99$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 0$
Gaussian	0.1793	0.4673	0.5917	0.0783
Student-t	0.9744	0.9188	0.8049	
Frank	0.0532	0.1263	0.2900	
Clayton	0.8173	0.9571	0.9666	
Survival Gumbel	0.8756	0.9212	0.8729	

Table 7.5. Calculation of average symmetric contagion measure from different copula families. Sample size $N = 5000$, $L = [0.05, 0.45]$, $\kappa = \{0.90, 0.95, 0.99\}$, $S = 100$. Algorithm 7.3.4.

The results related to Algorithm 7.3.4 are instead reported in Tables 7.5 and 7.6. Here the results seem to be closer to the expected (theoretical) values. The procedure, in some sense, seems better to implement, although the computational effort due to bootstrap replications is larger than in the other cases.

Overall, the main indications that we obtain from the simulation study are the following. Copulas with a non-zero lower tail dependence coefficient (like Student-t, Clayton and survival Gumbel) exhibit, as expected, a larger measure of contagion. Whereas, copulas with zero lower tail dependence coefficients (Gaussian and Frank) have a contagion measure close to 0 (apart from the naive algorithm with $\varepsilon = 0$). Moreover, the larger the Spearman's correlation is, the smaller the contagion measure seems to be. This is due to the fact that the comonotone copula (which has Spearman's correlation equal to 1) has a contagion measure equal to 0, because the strength of dependence does not change in the tail or in the central region of the distribution, where the variables are coupled always in a completely monotonic way.

The results, overall, are compatible with the general intuition that we could infer from the copula families. For instance, it is not surprising that the greater the dependence, the clearer the distinction among the families with and without non-zero tail dependence coefficients is. Moreover, it seems that Gaussian copulas exhibit larger values than Frank copulas, although the two families have both zero tail dependence coefficient. This may be due to the different tail expansion of these copulas near the corner $(0, 0)$ (compare with [Jaworski et al., 2010]).

About the other parameters used in Algorithm 7.3.4, we notice that the role of the confidence level κ is not so evident.

As a by-product, the simulation study has underlined that the choice of the copula may influence largely the presence of the contagion among markets and, as a consequence, may under- (or over-) estimate the risk of joint downturns. For instance, these numerical studies indicate which copulas can be employed in the construction of copula-based multivariate models of financial time series, where it

$\kappa = 0.90$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 0$
Gaussian	0.0536	0.0321	0.0129	0.1693
Student-t	0.9736	0.8879	0.5464	
Frank	0.0136	0.0000	0.0000	
Clayton	0.7129	0.9421	0.9486	
Survival Gumbel	0.8529	0.8614	0.6550	
$\kappa = 0.95$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 0$
Gaussian	0.0407	0.0186	0.0107	0.1250
Student-t	0.9657	0.8571	0.5036	
Frank	0.0093	0.0000	0.0000	
Clayton	0.6600	0.9264	0.9407	
Survival Gumbel	0.8221	0.8393	0.6200	
$\kappa = 0.99$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 0$
Gaussian	0.0229	0.0107	0.0086	0.0929
Student-t	0.9457	0.8093	0.4507	
Frank	0.0043	0.0000	0.0000	
Clayton	0.5829	0.9050	0.9243	
Survival Gumbel	0.7750	0.8029	0.5886	

Table 7.6. Calculation of average symmetric contagion measure from different copula families. Sample size $N = 5000$, $L = [0.05, 0.18]$, $\kappa = \{0.90, 0.95, 0.99\}$, $S = 100$. Algorithm 7.3.4

is allowed that all time series have time-varying conditional mean and variance, and the standardized residuals are joined by a suitable copula (for more details, see [Patton, 2012, 2013]).

Thanks to previous considerations, we may outline a practical application of spatial contagion measure to the description of tail dependence in bivariate copula models. Specifically, the problem of the detection of tail dependence in bivariate models is usually a difficult task that has a relevant impact in the calculation of risk measure associated with a stochastic model, like Value-at-Risk in portfolio management [McNeil et al., 2005], or return period in hydrology (Salvadori et al., 2007), etc. Therefore, several methods have been presented in the literature for a meaningful description and approximation of the tail: see, for instance, [Jaworski, 2010, 2013]; [Li, 2013] and references therein. Here, we illustrate how spatial contagion measure may help in this respect.

Example 7.4.1. Consider the dataset collected at the Copenhagen Reinsurance Company and formed by 2167 fire losses over the period 1980–1990. The data have been described in detail in [Haug et al., 2011] and are available at `www.ma.hw.ac.uk/ mcneil/`. These losses have been adjusted for inflation to reflect 1985 values and are expressed in millions of Danish Kroner. Moreover, we consider the data related to loss to building and loss to contents, provided that both components are greater than or equal to 1 million DK (this avoids the presence of ties). As shown in [Haug et al., 2011], this selection reduces the dimension to 301 pairs of observations. A suitable copula for such data seems to be the Gumbel copula that is an extreme-value copula and admits a non-negative upper tail dependence coefficient (i.e. large losses tend to occur together). Now, let us calculate the symmetric contagion measure of the dataset via Algorithm 7.3.4. Notice that, since we are interested in the behavior of extreme losses, we consider the opposite of our observations (i.e. the gains) in order to be coherent with our setting (thus, extreme losses correspond to lower tail of the distributions of gains). Specifically, we consider all the observations, that are larger than a given threshold, as tail set. The calculations are performed by setting a step $\ell = 0.05$ and the number of bootstrap replications equals to 100. When the

set L is given by $[0.05, 0.18]$, according to confidence level of 95%, the symmetric contagion measure is equal to 0.370, which is greater than the value of 0 that is related to the Gaussian assumption. So, contagion measure indicates evidence of non-Gaussianity in the dependence structure of the data, coherently to previous approaches in the literature. As a further remark, notice that the symmetric contagion measure for the interval $L = [0.05, 0.45]$ at the same confidence level is equal to 0.160. Thus, we may say that there is more contagion in extreme losses than in moderate losses, and, in both cases, Gaussianity assumptions seem to be violated.

7.5 Empirical Analysis

In order to illustrate our procedure we analyze daily log-returns of Morgan Stanley Capital International (MSCI) Developed Markets indices, available at Datastream. The return indices are related to the following markets: Austria (AT), Belgium (BE), Switzerland (CH), Germany (DE), Spain (ES), FInland (FI), France (FR), Greece (GR), Ireland (IE), Italy (IT), Netherlands (NL), Portugal (PT), United Kingdom (UK), and the United States (US). Summary statistics of indices returns are reported in Table 7.7. We find some evidence of skewness and the strong presence of kurtosis. The dataset refers to the time interval from of September 12, 2008 (coinciding with the Lehman Brothers's bankruptcy) to December 31, 2012.

	Mean	Median	Stdev	Skewness	Kurtosis
AT	-0.0006	0.0002	0.03	-0.04	3.22
BE	-0.0001	0.0005	0.02	-0.62	6.60
CH	0.0002	0.0003	0.02	0.05	4.32
DE	-0.0000	0.0004	0.02	0.05	3.40
ES	-0.0004	0.0000	0.02	0.20	3.72
FI	-0.0004	-0.0003	0.02	-0.03	2.23
FR	-0.0001	0.0001	0.02	0.10	3.92
GR	-0.0018	-0.0010	0.03	0.22	1.74
IE	-0.0006	0.0000	0.03	-0.62	5.68
IT	-0.0005	0.0001	0.02	0.06	2.77
NL	-0.0001	0.0001	0.02	-0.06	4.43
PT	-0.0004	0.0000	0.02	-0.00	5.08
UK	0.0000	0.0007	0.02	-0.06	6.56
US	0.0001	0.0004	0.02	-0.30	7.60

Table 7.7. Basic statistics of the considered European stock indices in the period from September 12, 2008 to December 31, 2012.

Two applications of the introduced contagion measure are illustrated here. The first one is related to the determination of asymmetric contagion, namely to determine whether large losses in one market may increase the dependence with another market. The second one is related to the use of symmetric contagion measure for discovering sub-groups of markets that have stronger dependence in crisis periods than in untroubled periods.

7.5.1 Detecting spatial contagion among financial markets

The recent financial crisis has created renewed interest in the systemic risk, namely the risk of a collection of institutions and markets that are interconnected and may have losses that quickly propagate during periods of financial distress [Billio et al., 2012]. Such studies also require the definition of

suitable measure of interconnectedness between different markets. To this end, the asymmetric spatial contagion measure may give some information about the one-directional link from one market to another one during financial distress compared with the link that exists in non-crisis periods. Roughly speaking, it measures the relative increment of dependency between two markets in financial distress.

Thus, in order to check such linkages, we can calculate the contagion measures on the original time series data. Namely, for each pairs of indices (i, j), we calculate the asymmetric contagion measure $\hat{\gamma}(i \to j)$ where $L = [0.05, 0.45]$ or $L = [0.05, 0.21]$ (this upper bound has been given by [Jaworski and Pitera, 2013] in order to ensure that Gaussian distribution does not contain any contagion effect). The results are reported in Figures 7.3 and 7.4.

When we look at extreme scenarios providing by $L = [0.05, 0.21]$, the only country that assumes large values of the measure is the US. For all the remaining countries contagious effects seem to be not remarkable. Moreover, we experienced a slightly weaker than expected performance for those countries (such as Greece, Italy, and Portugal) recently considered as source of contagion in Europe. We argue that it may represent the so-called flight–to–quality investor behavior; i.e., investors have already diversified their portfolios in such a way that they have hedged against contagion from countries that are in trouble.

However, when we also consider moderate losses by assuming $L = [0.05, 0.45]$, the output is different and shows the presence of several one-directional linkages among markets, being again the US market as the big player that drives most of the other countries.

As a second approach, we applied a univariate Student-t ARMA(1,1)-GARCH(1,1) model to each time series of log-returns to remove autocorrelation and heteroscedasticity from the data and we computed the standardized residuals. In fact, as discussed, for instance, by [Durante and Foscolo, 2013], the calculation of the conditional correlations could be biased when heteroscedastic effects are not removed from the original time series. Moreover, thanks to a copula approach, we may always separate the fit of univariate time series from the global information about the dependence (see [Patton, 2012, 2013] for more details). The adequacy of this univariate filtering is checked via the Box-Pierce and Ljung-Box tests for residual autocorrelation, and the ARCH test for residual volatility at lag 1 and 4. The output about contagion measures are reported in Figures 7.5 and 7.6.

As can be seen, filtered data produces, on average, smaller values of the contagion contagion in both tail sets choices. However, the general consideration about the market linkages remain the same, and US market maintains its role of big player in this group of countries.

7.5.2 Clusters of financial markets via spatial contagion

Here we present a further application of the introduced contagion measure to portfolio cluster analysis. We recall that cluster analysis is concerned with discovering group structure among a set of different observations. Such methods are usually based on the notion of dissimilarity; given a set S of objects, a dissimilarity measure d is a mapping $d\colon S \times S \to \mathbb{R}_+$ that is symmetric and such that $d(x, x) = 0$ for all $x \in S$. Starting with the dissimilarity matrix obtained by n objects, a number of different clustering techniques can be applied, like hierarchical clustering. See, for instance, [Kaufman and Rousseeuw, 1990, Hastie et al., 2009] and the references therein.

We consider the measure of symmetric contagion in order to define a suitable dissimilarity index. In view of the discussion above, we focus on unfiltered data and we calculate the symmetric contagion measure with the robust algorithm when $L = [0.05, 0.45]$ and the confidence level $\kappa = 0.95$. First of all, we calculate it for all considered markets. The obtained values $\hat{\gamma}\,(X, Y) = \widehat{\gamma}_{ij}$ are reported in Figure 7.7.

Using these values, we may construct a dissimilarity matrix $D = (d_{ij})$, where $d_{ij} = \sqrt{1 - \widehat{\gamma}_{ij}}$.

Such a matrix can be used to obtain a hierarchical clustering (as, for instance, in [Brida and Risso, 2010, Durante et al., 2013c]). Alternatively, we could use this matrix to visualize the different financial time series in a Cartesian plane via multidimensional scaling (see, e.g., [De Luca and Zuccolotto, 2011]), or we may perform the minimum spanning tree as done in [Mantegna, 1999]. All these different features are exploited here and the results can be visualized in Figures 7.8, 7.9 and 7.10. The output of a clustering procedure via spatial contagion should be intended as a tool that could assist the decision maker (e.g. the trader) in the selection of portfolio of assets when diversification procedures in crisis periods need to be considered. Moreover, interestingly, an automatic selection procedure for choosing a diversified portfolio with a fixed clustering structure can be implemented following the ideas of [De Luca and Zuccolotto, 2011] (see also [Durante et al., 2013a]). It could be applied here as well.

Finally, notice that clusters obtained via symmetric spatial contagion detect aspects of financial time series that are not covered by other (dependence-based) clustering methods (and, obviously, they have different outputs).

In fact, consider, for example, the clustering method based on Pearson's correlation coefficient, as done by [Mantegna, 1999]. As known, these methods provide information about the linear dependence among the different time series, since Pearson's correlation cannot detect the non-linear behavior of financial times series [McNeil et al., 2005]. Moreover, they may fail in the identification of diversified portfolios when the market is under distress (see, e.g., [Bradley and Taqqu, 2004]). Contrarily, spatial contagion clusters are essentially non-linear (since they are copula-based) and take into account possible effects of underlying financial distress (through the choice of the set L). Secondly, compared with clusters based on tail-dependence measures like conditional Spearman's correlation [Durante et al., 2013c] or tail dependence coefficients [De Luca and Zuccolotto, 2011], spatial contagion measure has two advantages: it is not grounded on asymptotic measures (as tail dependence coefficients), giving a more realistic (i.e. at a finite scale) tail description and, at the same time, it does not require the specification of a single threshold related to the definition of extreme losses.

7.6 Conclusion

The main contribution of this paper has been the introduction of a novel measure of spatial contagion. It aims at quantifying the difference between the dependence among financial time series in normal periods and in periods of market distress.

The proposed measure is based on the calculation of suitable conditional Spearman's correlations extracted from the financial time series of interest. Moreover, compared with previous approaches from the literature, it does not require the specification of a suitable threshold associated with the definition of extreme losses.

The whole procedure is illustrated in details and algorithms are provided, showing several numerical issues that may arise in practice. Moreover, as a major advantage, it is non-parametric and does not require the specification of any fully parametric models that drive all the time series of interest. As a matter of future application, the procedure is illustrated in practice, by showing also how it can be used to detect the presence of contagion effects from one market to the other ones and, hence, identify networks in the financial system. Furthermore, by using its symmetric version, it can create clusters of financial time series according to their dependence in adverse events. Such clusters underline different aspects of time series that cannot be detected by classical correlation-based clustering methods.

The presented methodology is expected to be useful to portfolio risk managers in order to provide some practical (and graphical) tools that may help in the selection of a diversified portfolio in crisis periods.

Possible drawbacks that are related to the proposed methodology are basically numerical. First, the contagion measure needs to work with an adequate number of data, since it also requires the approximation of Spearman's correlation for extreme losses. Second, the significance of the proposed contagion measure needs to be evaluated by bootstrapping, a tool that usually increases the computation burden especially when the contagion in a large portfolio needs to be detected. Both these aspects need to be further investigated in the future, also in order to enhance the application of this spatial contagion measure. In particular, a profound analysis of different estimators of the spatial contagion measure (and their asymptotic behavior) may help to fix these practical inconveniences.

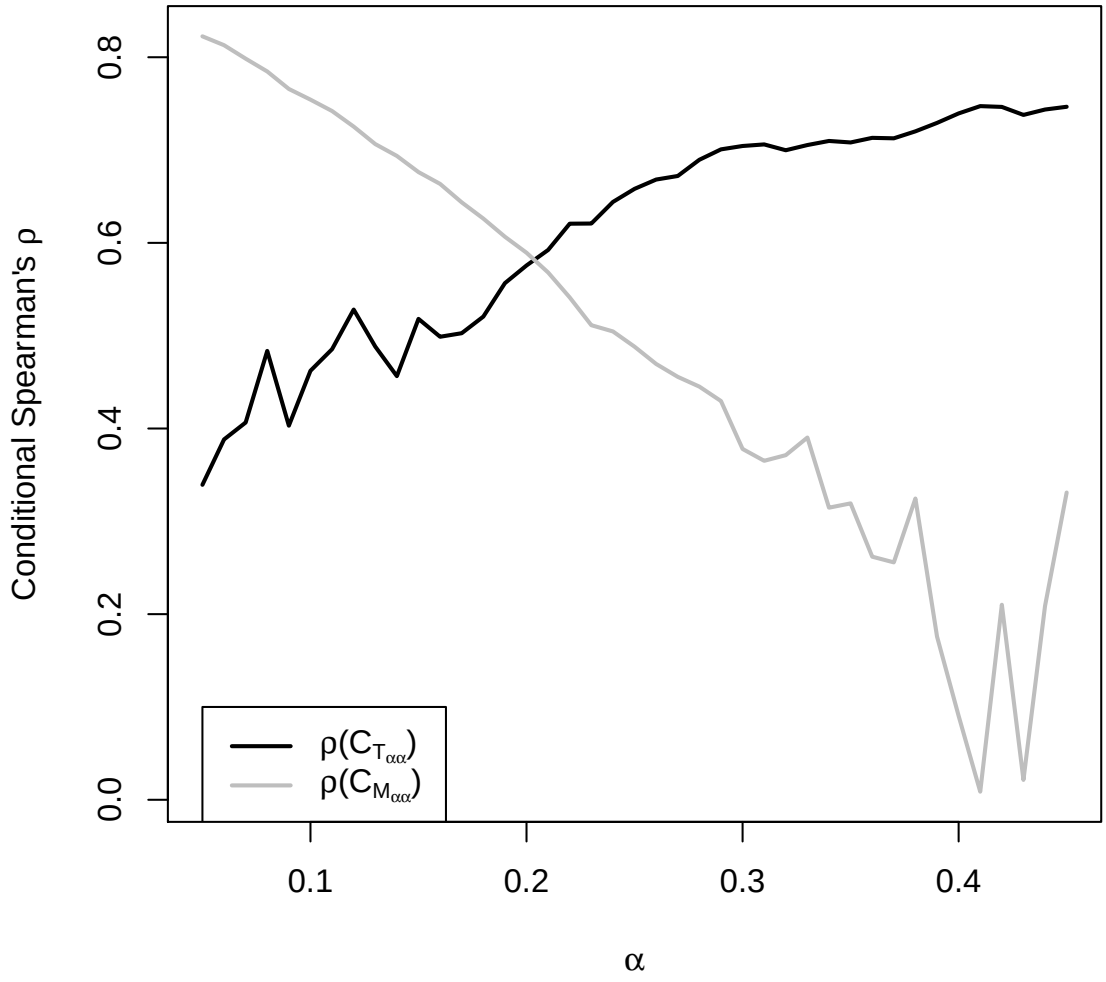


Figure 7.1. Evolution of conditional Spearman's $\rho(T_{\alpha,\alpha})$ and $\rho(M_{\alpha,\alpha})$ for different values of α. The data are related to the MSCI indices of Germany and UK, as illustrated in Section 7.5

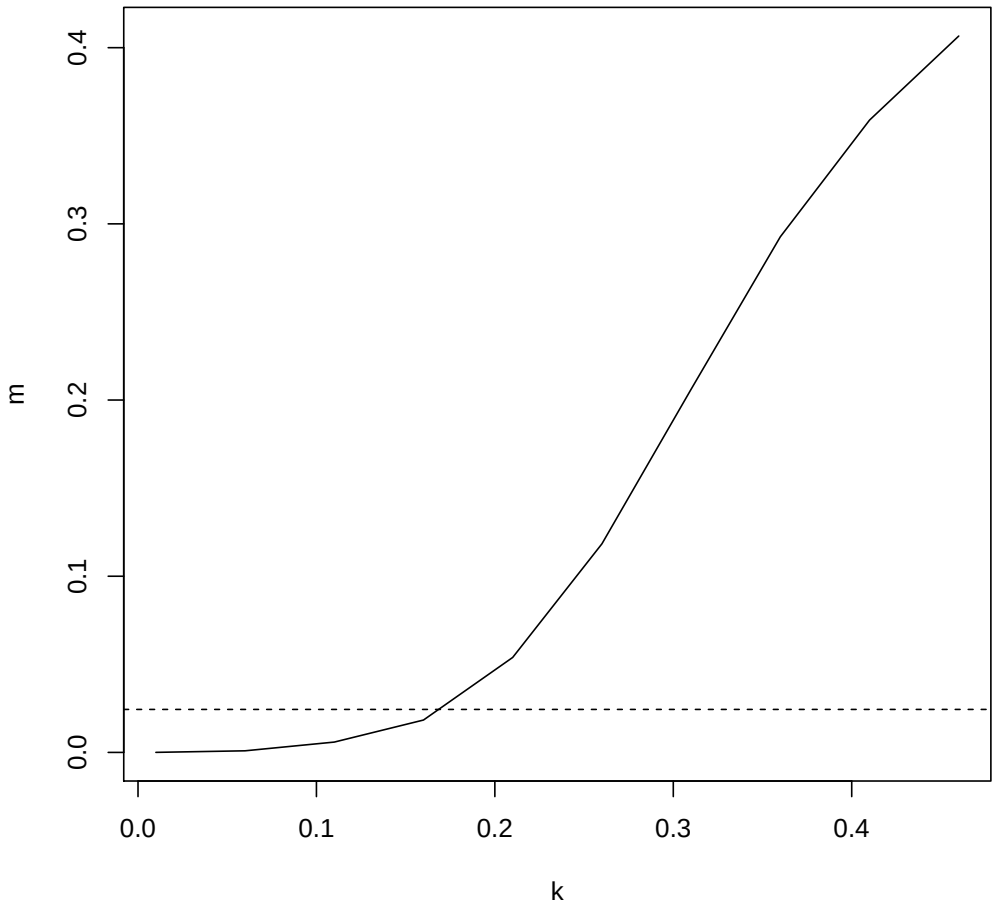


Figure 7.2. Choice of optimal k using Algorithm 7.3.3.

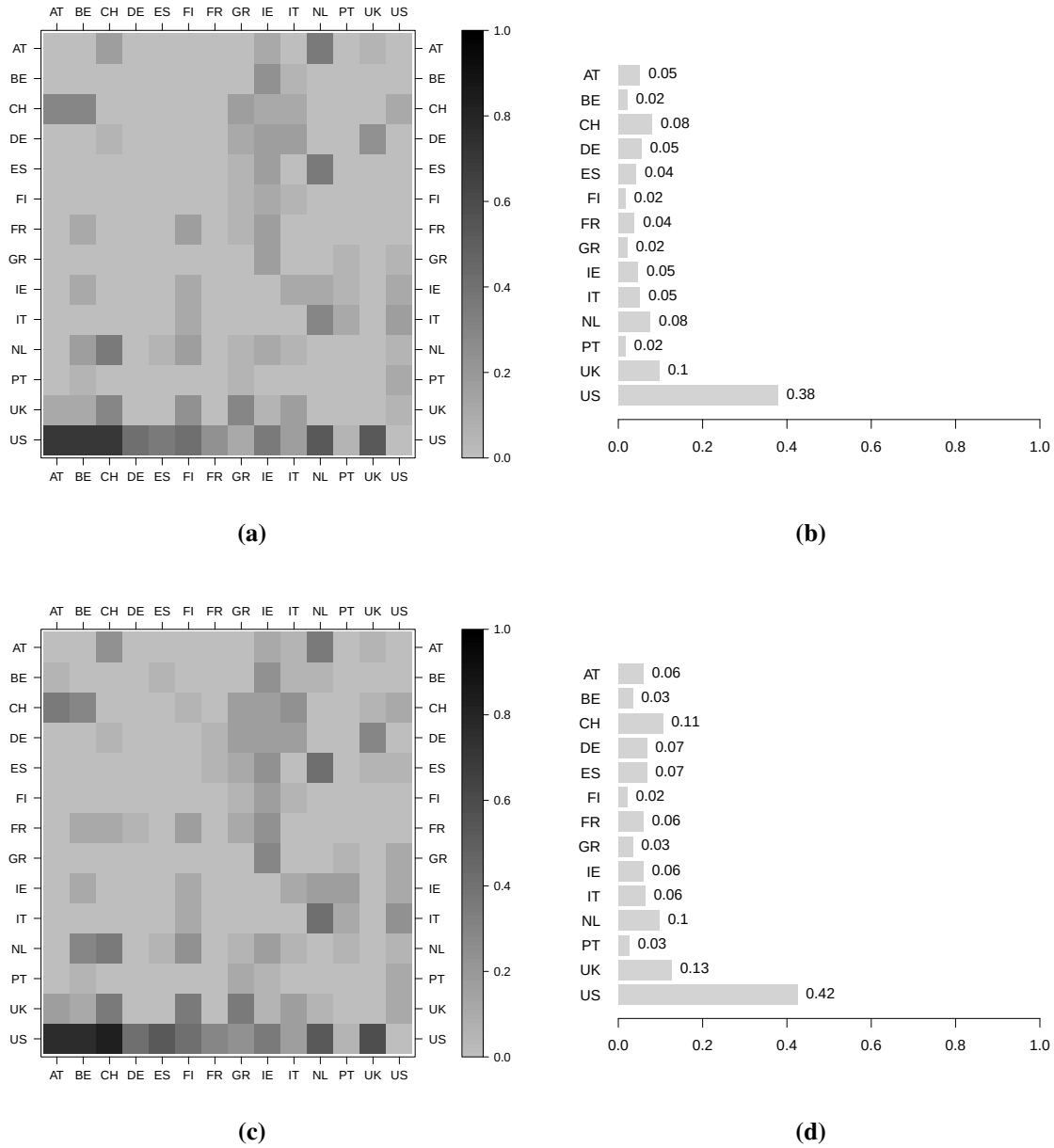
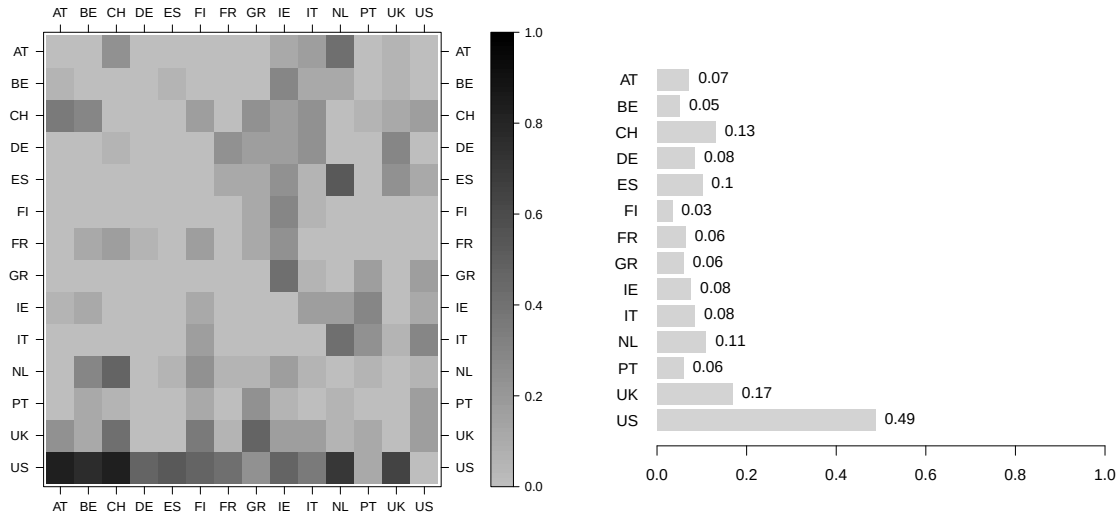
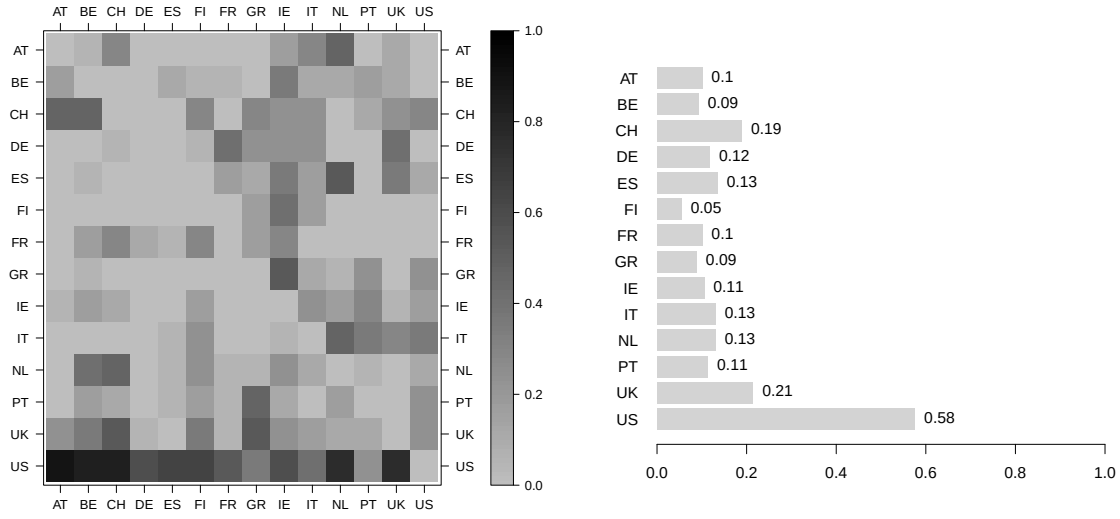


Figure 7.3. Unfiltered data. Asymmetric contagion measure in MSCI data with $L = [0.05, 0.21]$ and $\kappa = 0.99$ (a and b), 0.95 (c and d), 0.90 (e and f), 0.75 (g and h). (Left) Each cell (i, j) corresponds to the asymmetric contagion measure from market i to market j. (Right) Average asymmetric contagion measure from market i to all other markets $j \neq i$.



(e) **(f)**



(g) **(h)**

Figure 7.3. Cont. Unfiltered data. Asymmetric contagion measure in MSCI data with $L = [0.05, 0.21]$ and $\kappa = 0.99$ (a and b), 0.95 (c and d), 0.90 (e and f), 0.75 (g and h). (Left) Each cell (i, j) corresponds to the asymmetric contagion measure from market i to market j. (Right) Average asymmetric contagion measure from market i to all other markets $j \neq i$.

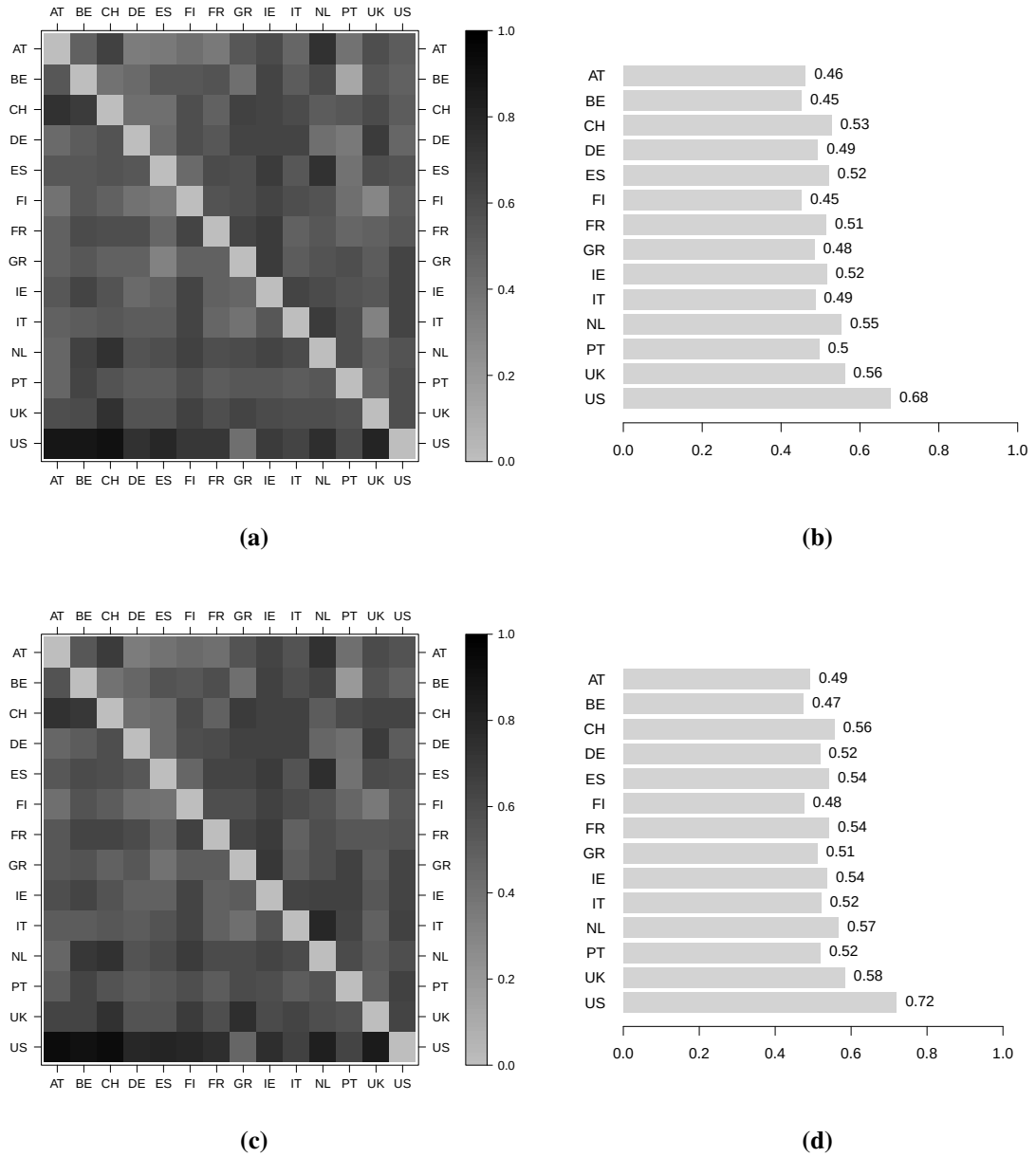
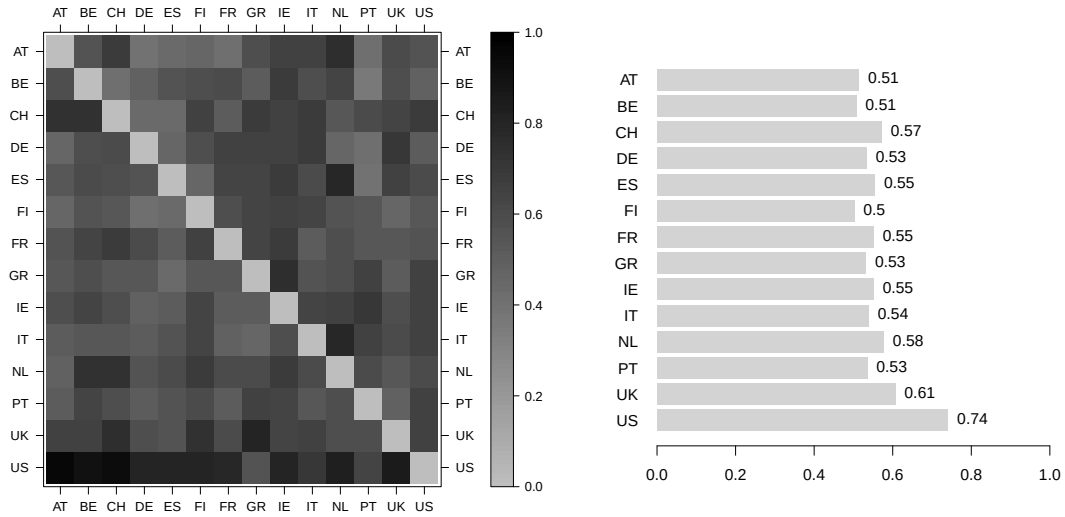
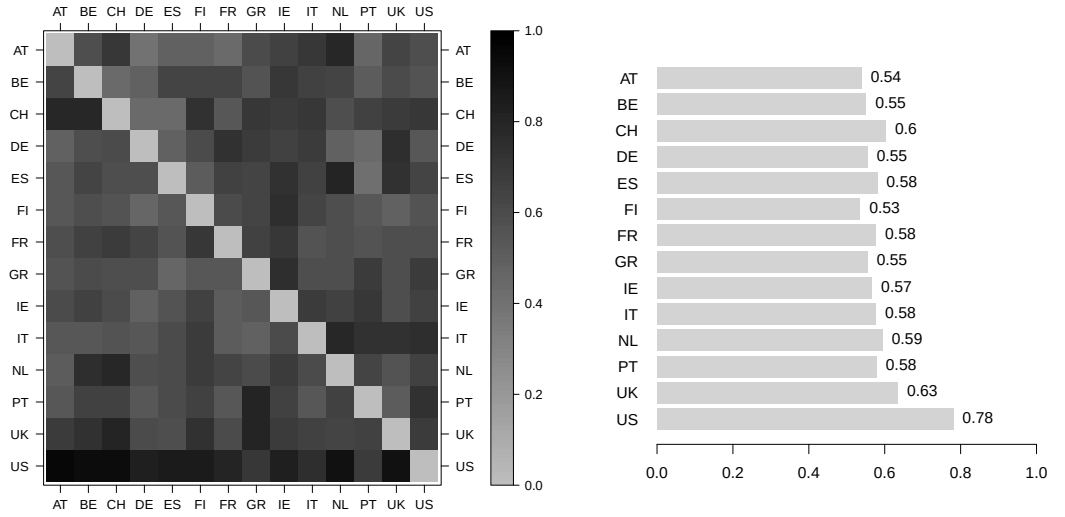


Figure 7.4. Unfiltered data. Asymmetric contagion measure in MSCI data with $L = [0.05, 0.45]$ and $\kappa = 0.99$ (a and b), 0.95 (c and d), 0.90 (e and f), 0.75 (g and h). (Left) Each cell (i, j) corresponds to the asymmetric contagion measure from market i to market j. (Right) Average asymmetric contagion measure from market i to all other markets $j \neq i$.



(e) **(f)**



(g) **(h)**

Figure 7.4. Cont. Unfiltered data. Asymmetric contagion measure in MSCI data with $L = [0.05, 0.45]$ and $\kappa = 0.99$ (a and b), 0.95 (c and d), 0.90 (e and f), 0.75 (g and h). (Left) Each cell (i, j) corresponds to the asymmetric contagion measure from market i to market j. (Right) Average asymmetric contagion measure from market i to all other markets $j \neq i$.

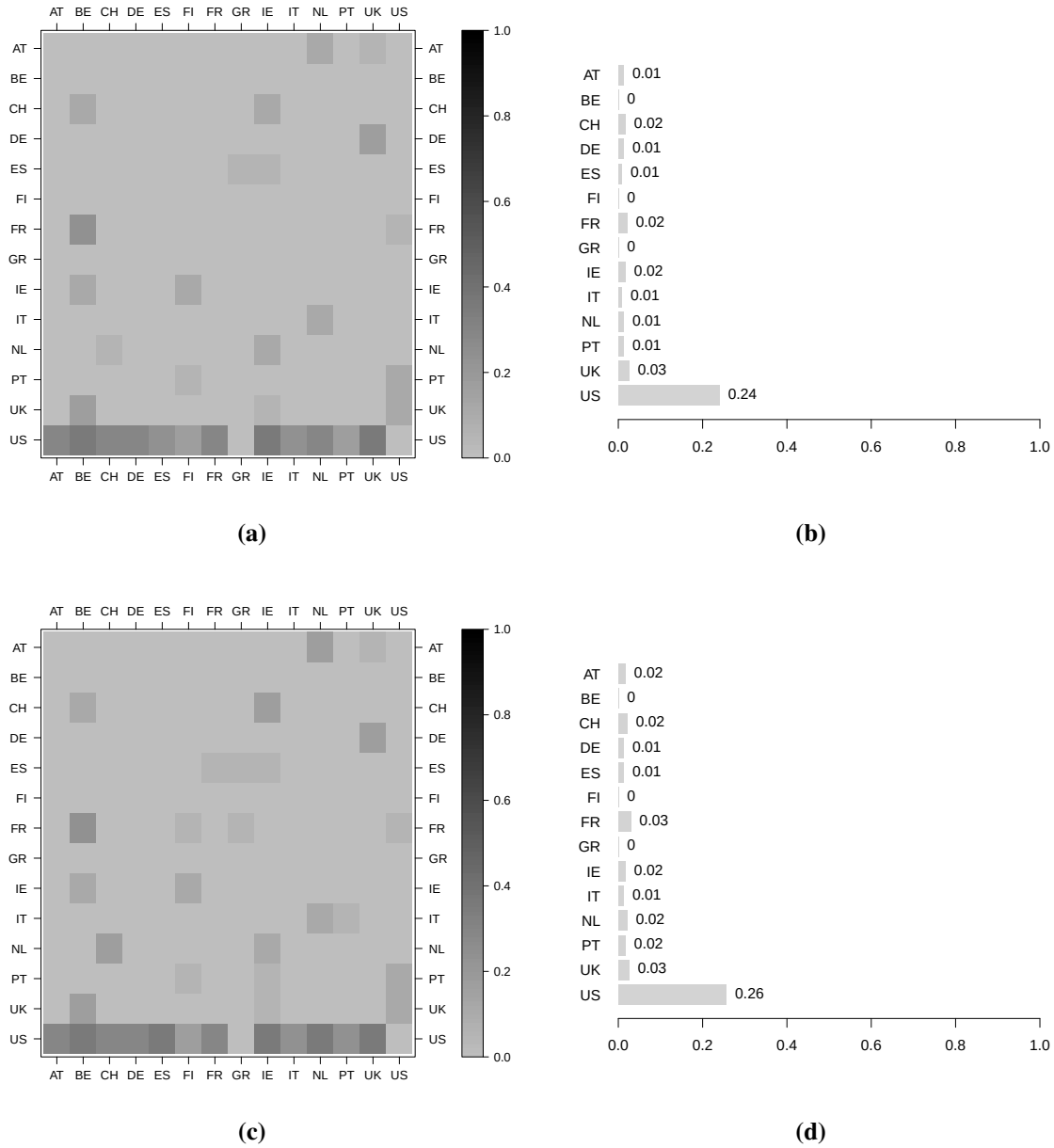
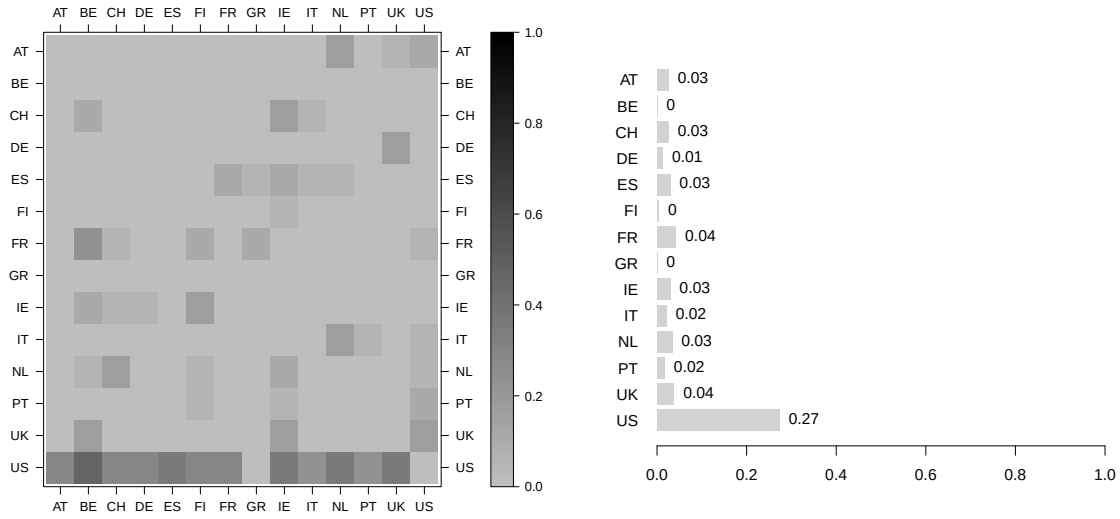
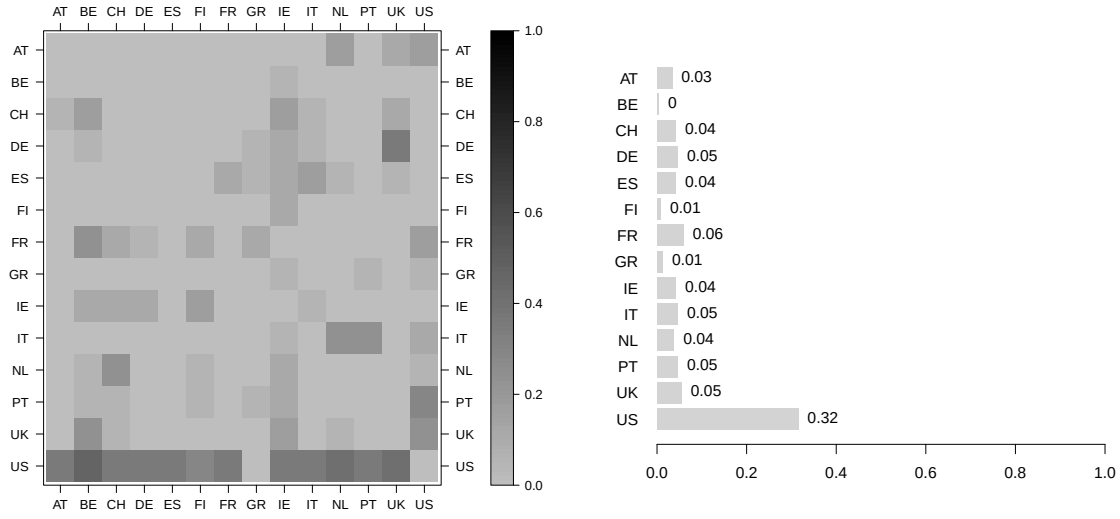


Figure 7.5. Filtered data. Asymmetric contagion measure in MSCI data with $L = [0.05, 0.21]$ and $\kappa = 0.99$ (a and b), 0.95 (c and d), 0.90 (e and f), 0.75 (g and h). (Left) Each cell (i, j) corresponds to the asymmetric contagion measure from market i to market j. (Right) Average asymmetric contagion measure from market i to all other markets $j \neq i$.



(e) **(f)**



(g) **(h)**

Figure 7.5. Cont. Filtered data. Asymmetric contagion measure in MSCI data with $L = [0.05, 0.21]$ and $\kappa = 0.99$ (a and b), 0.95 (c and d), 0.90 (e and f), 0.75 (g and h). (Left) Each cell (i, j) corresponds to the asymmetric contagion measure from market i to market j. (Right) Average asymmetric contagion measure from market i to all other markets $j \neq i$.

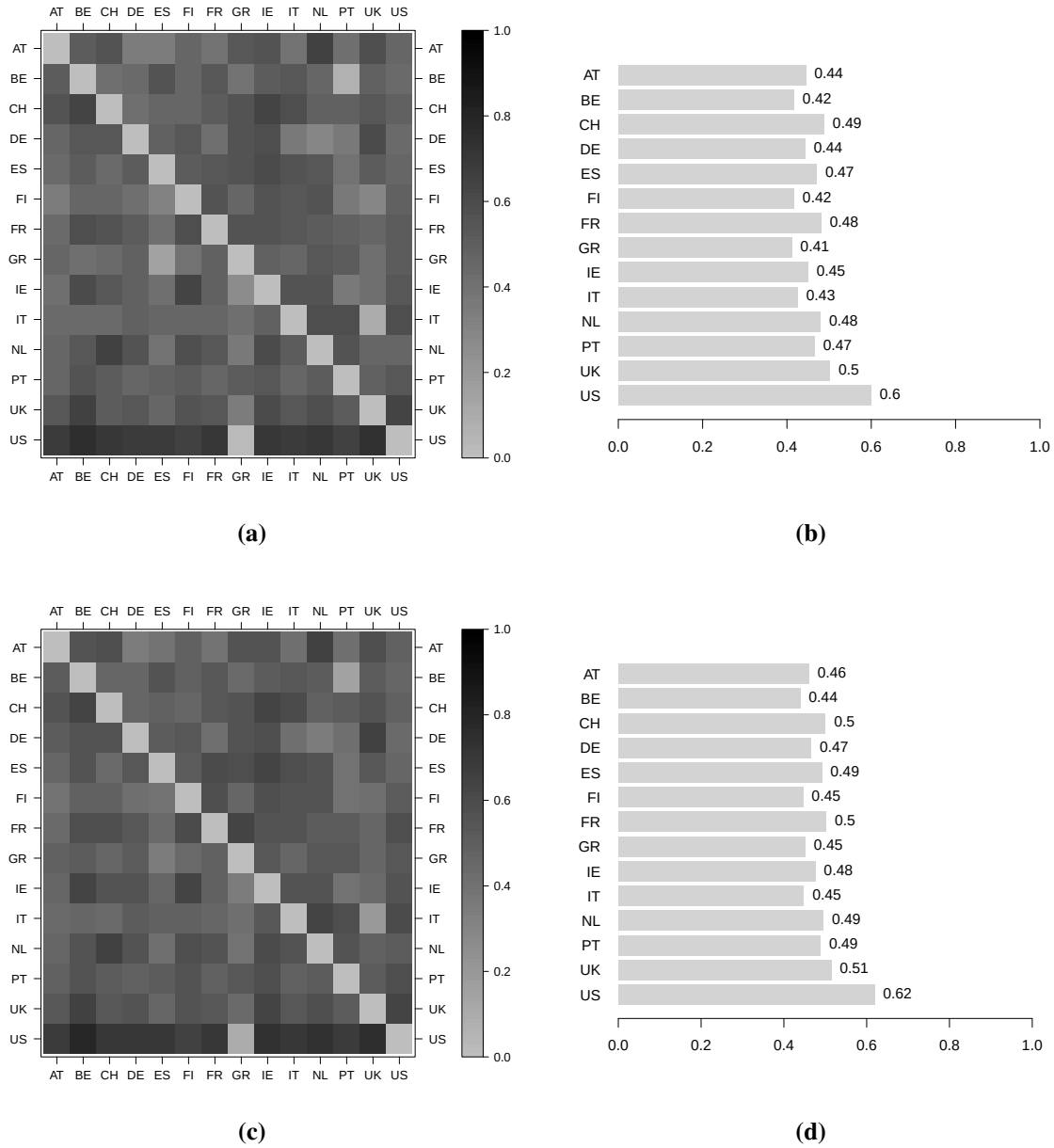
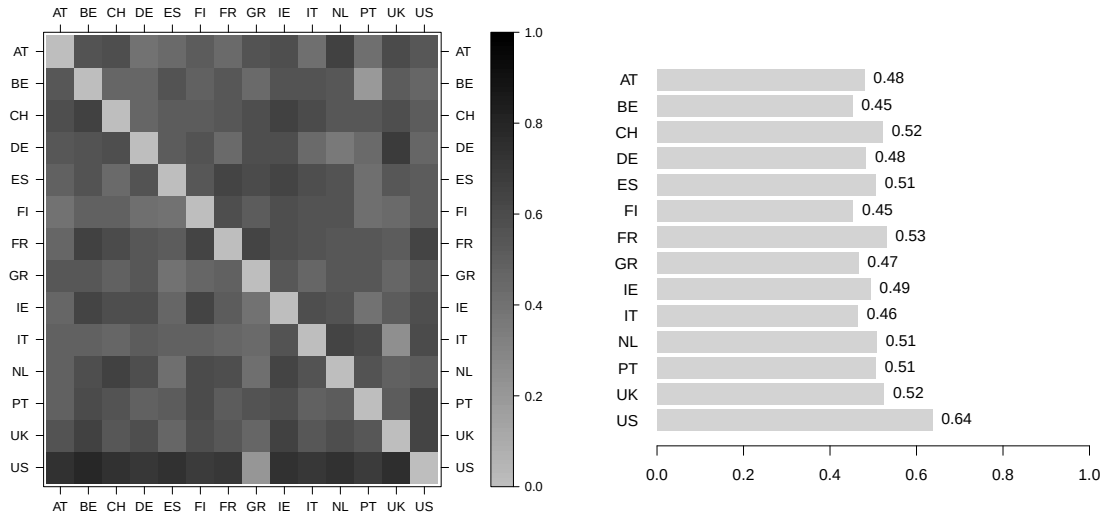
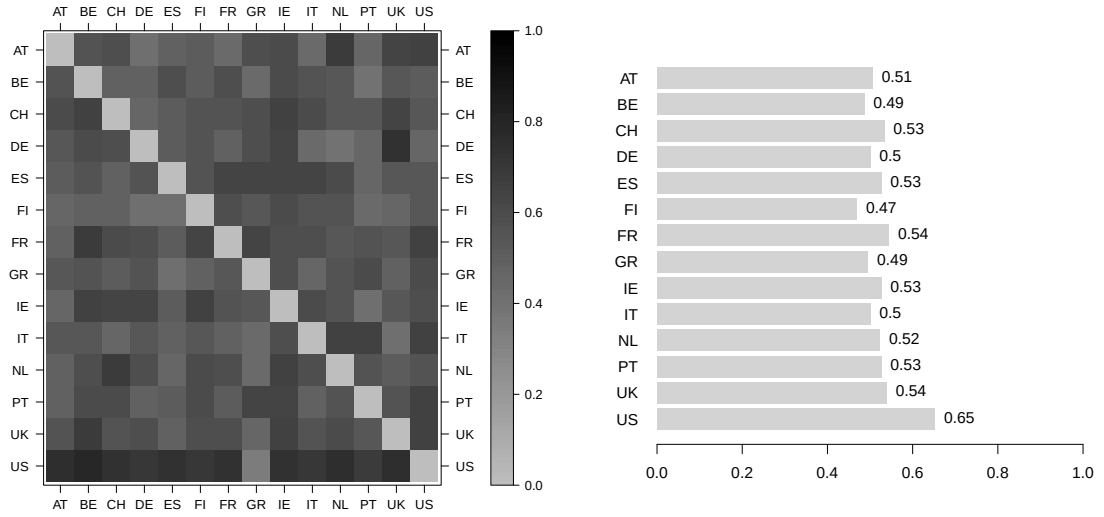


Figure 7.6. Filtered data. Asymmetric contagion measure in MSCI data with $L = [0.05, 0.45]$ and $\kappa = 0.99$ (a and b), 0.95 (c and d), 0.90 (e and f), 0.75 (g and h). (Left) Each cell (i, j) corresponds to the asymmetric contagion measure from market i to market j. (Right) Average asymmetric contagion measure from market i to all other markets $j \neq i$.



(e) **(f)**



(g) **(h)**

Figure 7.6. Cont. Filtered data. Asymmetric contagion measure in MSCI data with $L = [0.05, 0.45]$ and $\kappa = 0.99$ (a and b), 0.95 (c and d), 0.90 (e and f), 0.75 (g and h). (Left) Each cell (i, j) corresponds to the asymmetric contagion measure from market i to market j. (Right) Average asymmetric contagion measure from market i to all other markets $j \neq i$.

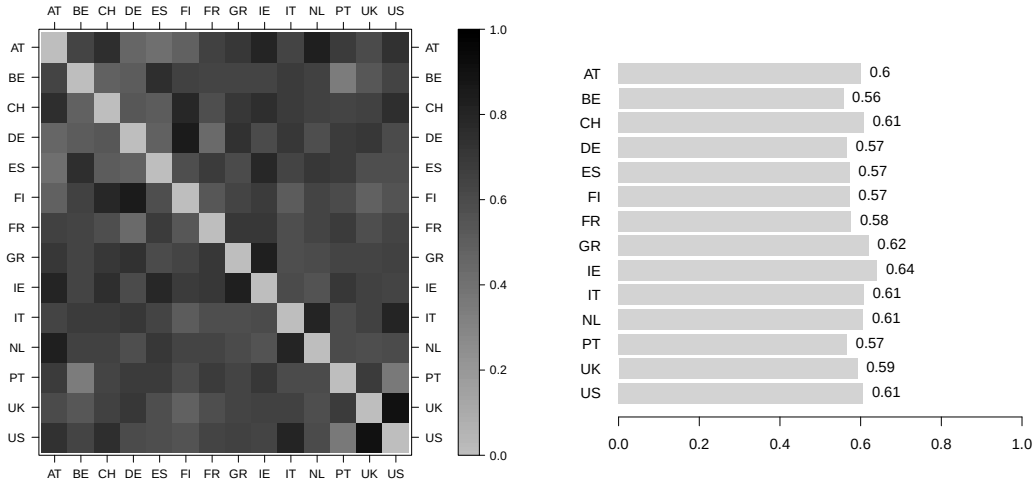


Figure 7.7. Unfiltered data. Symmetric contagion measure in MSCI data with $L = [0.05, 0.45]$ and $\kappa = 0.95$. (Left) Each cell (i, j) corresponds to the symmetric contagion measure from market i to market j. (Right) Average symmetric contagion measure from market i to all other markets $j \neq i$.

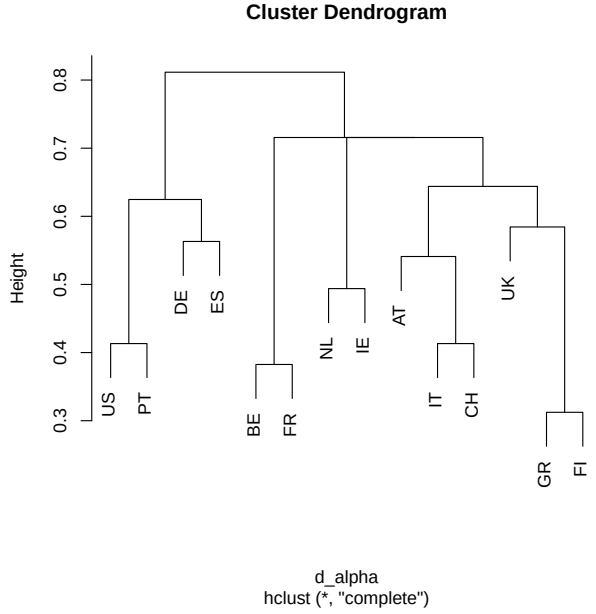


Figure 7.8. Unfiltered data. Hierarchical clustering in MSCI data.

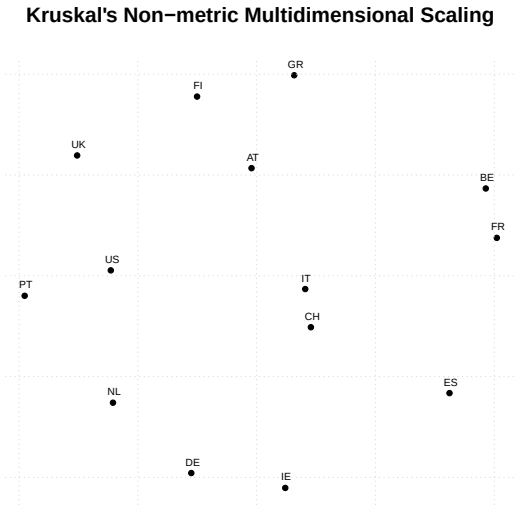


Figure 7.9. Unfiltered data. Multidimensional scaling in MSCI data.

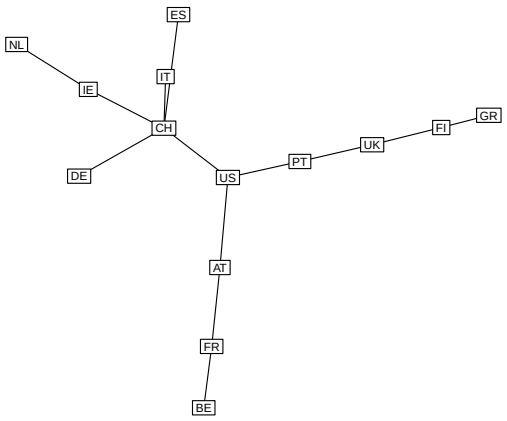


Figure 7.10. Unfiltered data. Minimum spanning tree in MSCI data.

List of Figures

List of Tables

Bibliography

Kjersti Aas, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.

Saeed Aghabozorgi and Ying Wah Teh. Stock market co-movement assessment using a three-phase clustering method. *Expert Systems with Applications*, 2013.

Amir AghaKouchak, András Bárdossy, and Emad Habib. Conditional simulation of remotely sensed rainfall data using a non-gaussian v-transformed copula. *Advances in Water Resources*, 33(6): 624–634, 2010.

Nicolas Basalto, Roberto Bellotti, Francesco De Carlo, Paolo Facchi, Ester Pantaleo, and Saverio Pascazio. Hausdorff clustering of financial time series. *Physica A: Statistical Mechanics and its Applications*, 379(2):635–644, 2007.

João A Bastos and Jorge Caiado. Clustering financial time series with variance ratio statistics. *Quantitative Finance, In press*, 2013.

Basel Committee on Banking Supervision BCBS. Basel II: International convergence of capital measurement and capital standards: a revised framework. Technical report, Bank for International Settlements (BIS), 2006. URL http://www.bis.org/publ/bcbs128.htm.

Basel Committee on Banking Supervision BCBS. Basel III: a global regulatory framework for more resilient banks and banking systems. Technical report, Bank for International Settlements (BIS), 2011. URL http://www.bis.org/publ/bcbs189.htm.

Viktor Benes and Josef Stepan. *Distributions with given marginals and moment problems*. Kluwer Academic Publishers, 1997.

Daniel Bienstock. Computational study of a family of mixed-integer quadratic programming problems. *Mathematical programming*, 74(2):121–140, 1996.

Monica Billio, Mila Getmansky, Andrew W Lo, and Loriana Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559, 2012.

Steinar Bjerve and Kjell Doksum. Correlation curves: measures of association as functions of covariate values. *The Annals of Statistics*, pages 890–902, 1993.

Giovanni Bonanno, Guido Caldarelli, Fabrizio Lillo, Salvatore Miccichè, Nicolas Vandewalle, and Rosario Nunzio Mantegna. Networks of equities in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):363–371, 2004.

Brian H Boyer, Michael S Gibson, and Mico Loretan. Pitfalls in tests for changes in correlations. In *Federal Reserve Boars, IFS Discussion Paper No. 597R*. Citeseer, 1999.

Brendan Bradley and Murad Taqqu. Empirical evidence on spatial contagion between financial markets. *Finance Letters*, 3(1):77–86, 2005a.

Brendan Bradley and Murad Taqqu. How to estimate spatial contagion between financial markets. *Finance Letters*, 3(1):64–76, 2005b.

Brendan O Bradley and Murad S Taqqu. Financial risk and heavy tails. *Handbook of heavy tailed distributions in finance*, pages 35–103, 2003.

Brendan O Bradley and Murad S Taqqu. Framework for analyzing spatial contagion between financial markets. *Finance Letters*, 2(6):8–15, 2004.

Juan Gabriel Brida and Wiston Adrián Risso. Hierarchical structure of the german stock market. *Expert Systems with Applications*, 37(5):3846–3852, 2010.

Francesco Caramazza, Luca Ricci, and Ranil Salgado. International financial contagion in currency crises. *Journal of International Money and Finance*, 23(1):51–70, 2004.

Francesco Cesarone, Andrea Scozzari, and Fabio Tardella. A new method for mean-variance portfolio optimization with cardinality constraints. *Annals of Operations Research*, pages 1–22, 2013.

T-J Chang, Nigel Meade, John E Beasley, and Yazid M Sharaiha. Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, 27(13):1271–1302, 2000.

Umberto Cherubini, Elisa Luciano, and Walter Vecchiato. *Copula methods in finance*. Wiley. com, 2004.

Umberto Cherubini, Sabrina Mulinacci, Fabio Gobbi, and Silvia Romagnoli. *Dynamic Copula methods in finance*, volume 625. John Wiley & Sons, 2012.

Stuart Coles, Janet Heffernan, and Jonathan Tawn. Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365, 1999.

Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *QUANTITATIVE FINANCE*, 1:223–236, 2001.

Delores Conway. Plackett family of distributions. *Encyclopedia of Statistical Sciences*, 1986.

Marcella Corduas and Domenico Piccolo. Time series clustering and classification by the autoregressive metric. *Computational Statistics & Data Analysis*, 52(4):1860–1872, 2008.

Giancarlo Corsetti, Marcello Pericoli, and Massimo Sbracia. Correlation analysis of financial contagion. *In: Kolb, R. (Ed.), Financial Contagion: The Viral Threat to the Wealth of Nations*, pages 11–20, 2011.

Carles M Cuadras, Josep Fortiana, et al. *Distributions with given marginals and statistical modelling*. Springer, 2002.

Giorgio Dall'Aglio, Samuel Kotz, Gabriella Salinetti, and Universitá degli Studi di Roma. *Advances in probability distributions with given marginals: beyond the copulas*. Kluwer Academic Publishers, 1991.

Giovanni De Luca and Paola Zuccolotto. A tail dependence-based dissimilarity measure for financial time series clustering. *Advances in Data Analysis and Classification*, 5(4):323–340, 2011.

Giovanni De Luca, Giorgia Rivieccio, and Paola Zuccolotto. Combining random forest and copula functions: a heuristic approach for selecting assets from a financial crisis perspective. *Intelligent Systems in Accounting, Finance and Management*, 17(2):91–109, 2010.

Casper G de Vries and Chen Zhou. Discussion of "copulas: Tales and facts", by thomas mikosch. *Extremes*, 9(1):23–25, 2006.

Paul Deheuvels. La fonction de dépendance empirique et ses propriétés. un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci.(5)*, 65(6):274–292, 1979.

Jadran Dobric, Gabriel Frahm, and Friedrich Schmid. Dependence of stock returns in bull and bear markets. *Dependence Modeling*, 1:94–110, 2014.

Kjell Doksum, Stephen Blyth, Eric Bradlow, Xiao-Li Meng, and Hongyu Zhao. Correlation curves as local measures of variance explained by regression. *Journal of the American Statistical Association*, 89(426):571–582, 1994.

Fabrizio Durante and Enrico Foscolo. An analysis of the dependence among financial markets by spatial contagion. *International Journal of Intelligent Systems*, 2013.

Fabrizio Durante and Piotr Jaworski. Spatial contagion between financial markets: a copula-based approach. *Applied Stochastic Models in Business and Industry*, 26(5):551–564, 2010.

Fabrizio Durante and Carlo Sempi. Copula theory: an introduction. In *Copula Theory and Its Applications*, pages 3–31. Springer, 2010.

Fabrizio Durante, Rachele Foschi, and Fabio Spizzichino. Threshold copulas and positive dependence. *Statistics & Probability Letters*, 78(17):2902–2909, 2008.

Fabrizio Durante, Enrico Foscolo, Roberta Pappadà, and Hao Wang. A clustering procedure for selecting a portfolio of financial assets in a risky scenario. Manuscript submitted, 2013a.

Fabrizio Durante, Enrico Foscolo, and Miroslav Sabo. A spatial contagion test for financial markets. In *Synergies of Soft Computing and Statistics for Intelligent Data Analysis*, pages 313–320. Springer, 2013b.

Fabrizio Durante, Roberta Pappadà, and Nicola Torelli. Clustering of financial time series in risky scenarios. *Advances in Data Analysis and Classification*, pages 1–18, 2013c.

Fabrizio Durante, Roberta Pappadà, and Nicola Torelli. Clustering of extreme observations via nonparametric tail dependence estimation. Manuscript submitted, 2013d.

Fabrizio Durante, Enrico Foscolo, Piotr Jaworski, and Hao Wang. A spatial contagion measure for financial time series. *Expert Systems with Applications*, 41(8):4023 – 4034, 2014.

Pierpaolo D'Urso, Carmela Cappelli, Dario Di Lallo, and Riccardo Massari. Clustering of financial time series. *Physica A Statistical Mechanics and its Applications*, 392:2114–2129, 2013.

Sebastian Edwards. Contagion. *The World Economy*, 23(7):873–900, 2000.

Paul Embrechts. *Modelling extremal events: for insurance and finance*, volume 33. Springer, 1997.

Paul Embrechts. Extreme value theory: Potential and limitations as an integrated risk management tool. *Derivatives Use, Trading & Regulation*, 6(1):449–456, 2000.

Paul Embrechts. Discussion of "copulas: Tales and facts", by thomas mikosch. *Extremes*, 9(1):45–47, 2006.

Paul Embrechts. Copulas: A personal view. *Journal of Risk and Insurance*, 76(3):639–650, 2009.

Paul Embrechts, Alexander McNeil, and Daniel Straumann. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, pages 176–223, 2002.

Paul Embrechts, Filip Lindskog, and Alexander McNeil. Modelling dependence with copulas and applications to risk management. *Handbook of heavy tailed distributions in finance*, 8(1):329–384, 2003.

Hong-Bin Fang, Kai-Tai Fang, and Samuel Kotz. The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, 82(1):1–16, 2002.

Kai-Tai Fang, Samuel Kotz, and Kai Wang Ng. *Symmetric Multivariate and Related Distributions Monographs on Statistics and Applied Probability*. London: Chapman and Hall Ltd. MR1071174, 1990.

Nicholas I Fisher. Copulas. *Encyclopedia of Statistical Sciences, Update*, 1:159–163, 1997.

Kristin J Forbes and Roberto Rigobon. No contagion, only interdependence: measuring stock market comovements. *The Journal of Finance*, 57(5):2223–2261, 2002.

Gabriel Frahm, Markus Junker, and Alexander Szimayer. Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*, 63(3):275–286, 2003.

Gabriel Frahm, Markus Junker, and Rafael Schmidt. Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics*, 37(1):80–100, 2005.

Maurice Fréchet. Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon Sect. A*, 9:53–77, 1951.

Christian Genest. Frank's family of bivariate distributions. *Biometrika*, 74(3):549–555, 1987.

Christian Genest. Dependence modelling: Statistical theory and applications in finance and insurance (demostafi), Quebec city, 20-22 May 2004-preface, 2005a.

Christian Genest. Special issue: Papers presented at the demostafi conference, Quebec, 20-22 May 2004-preface, 2005b.

Christian Genest and Anne-Catherine Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347–368, 2007.

Christian Genest and Jock Mackay. The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, 40(4):280–283, 1986.

Christian Genest and Bruno Rémillard. Discussion of "copulas: tales and facts", by thomas mikosch. *Extremes*, 9(1):27–36, 2006.

Christian Genest and Johan Segers. Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, 37(5B):2990–3022, 2009.

Christian Genest, Anne-Catherine Favre, Julie Béliveau, and C Jacques. Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research*, 43(9): W09401, 2007.

Christian Genest, Michel Gendron, and Michaël Bourdeau-Brien. The advent of copulas in finance. *The European Journal of Finance*, 15(7-8):609–618, 2009.

Herbert G Grubel. Internationally diversified portfolios: welfare gains and capital flows. *The American Economic Review*, 58(5):1299–1314, 1968.

Peter Grundke and Simone Polle. Crisis and risk dependencies. *European Journal of Operational Research*, 223(2):518–528, 2012.

Gordon Gudendorf and Johan Segers. Extreme-value copulas. In *Copula Theory and Its Applications*, pages 127–145. Springer, 2010.

Emil Julius Gumbel and Neil Goldstein. Analysis of empirical bivariate extremal distributions. *Journal of the American Statistical Association*, 59(307):794–816, 1964.

Jiawei Han, Micheline Kamber, and Jian Pei. *Data mining: concepts and techniques*. Morgan kaufmann, 2006.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference and prediction, 2nd Edition.* Springer New York, 2009.

Stephan Haug, Claudia Klüppelberg, and Liang Peng. Statistical models and methods for dependence in insurance data. *Journal of the Korean Statistical Society*, 40(2):125–139, 2011.

Wassily Hoeffding. Masstabinvariante korrelationstheorie. *Schriften des Matematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin 5 Heft 3*, pages 179–233, 1940.

Wassily Hoeffding. Masstabinvariante korrelationsmasse für diskontinuierliche verteilungen. *Arkiv für matematischen Wirtschaften und Sozialforschung 7*, pages 49–70, 1941.

Marius Hofert. Construction and sampling of nested archimedean copulas. In *Copula Theory and Its Applications*, pages 147–160. Springer, 2010.

Marius Hofert, Ivan Kojadinovic, Martin Maechler, and Jun Yan. *copula: Multivariate Dependence with Copulas*, 2013. URL http://CRAN.R-project.org/package=copula. R package version 0.999-7.

Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.

Tak-Kee Hui. Portfolio diversification: a factor analysis approach. *Applied Financial Economics*, 15 (12):821–834, 2005.

Henrik Hult and Filip Lindskog. Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied probability*, 34(3):587–608, 2002.

William Curt Hunter, George G Kaufman, and Thomas H Krueger. *The Asian Financial Crisis: Origins, Implications, and Solutions;[Papers Presented at a Conference Held on Oct. 8-10, 1998]*. Springer, 1999.

Timothy P Hutchinson and Chin Diew Lai. *Continuous bivariate distributions, emphasising applications*. Rumsby Scientific Publishing Adelaide, 1990.

Piotr Jaworski. Tail behaviour of copulas. In *Copula Theory and Its Applications*, pages 161–186. Springer, 2010.

Piotr Jaworski. The limiting properties of copulas under univariate conditioning. In *Copulae in Mathematical and Quantitative Finance*, pages 129–163. Springer, 2013.

Piotr Jaworski and Marcin Pitera. On spatial contagion and multivariate garch models. *Applied Stochastic Models in Business and Industry*, 2013.

Piotr Jaworski, Fabrizio Durante, Wolfgang Karl Härdle, and Tomasz Rychlik. *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*, volume 198. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

Norbert J Jobst, Michael D Horniman, Cormac A Lucas, and Gautam Mitra. Computational aspects of alternative portfolio selection models in the presence of discrete asset choice constraints. *Quantitative Finance*, 1(5):489–501, 2001.

Harry Joe. Parametric families of multivariate distributions with given margins. *Journal of multivariate analysis*, 46(2):262–282, 1993.

Harry Joe. *Multivariate models and dependence concepts*, volume 73. CRC Press, 1997.

Harry Joe. Discussion of "copulas: Tales and facts", by thomas mikosch. *Extremes*, 9(1):37–41, 2006.

Kumar Jogdeo. Concepts of dependence. *Encyclopedia of statistical sciences*, 1:324–334, 1982.

Alessandro Juri and Mario V Wüthrich. Copula convergence theorems for tail events. *Insurance: Mathematics and Economics*, 30(3):405–420, 2002.

Graciela L Kaminsky, Carmen Reinhart, and Carlos A Vegh. The unholy trinity of financial contagion. Technical report, National Bureau of Economic Research, 2003.

Leonard Kaufman and Peter Rousseeuw. *Finding Groups in Data: an introduction to cluster analysis*. Wiley & Sons, Inc., New York, 1990.

Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

Dimitris Kenourgios, Aristeidis Samitas, and Nikos Paltalidis. Financial crises and stock market contagion in a multivariate time-varying asymmetric framework. *Journal of International Financial Markets, Institutions and Money*, 21(1):92–106, 2011.

Roger T. Kilgore and David B. Thompson. Estimating joint flow probabilities at stream confluences by using copulas. *Transportation Research Record: Journal of the Transportation Research Board*, 2262:200–206, 2011.

Robert W Kolb. *Financial contagion: The viral threat to the wealth of nations*, volume 604. Wiley. com, 2011.

Tõnu Kollo. Preface. *Journal of Statistical Planning and Inference*, 139(11):3740, 2009.

Samuel Kotz and Saralees Nadarajah. *Multivariate t-distributions and their applications*. Cambridge University Press, 2004.

Haim Levy and Marshall Sarnat. International diversification of investment portfolios. *The American Economic Review*, 60(4):668–675, 1970.

David Xianglin Li. On default correlation: a copula function approach. *Available at SSRN 187289*, 1999.

Haijun Li. Toward a copula theory for multivariate regular variation. In *Copulae in Mathematical and Quantitative Finance: Proceedings of the Workshop Held in Cracow, 10-11 July 2012*, page 177. Springer Science & Business, 2013.

Francois Longin and Bruno Solnik. Extreme correlation of international equity markets. *The Journal of Finance*, 56(2):649–676, 2001.

Martin Maechler, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. *cluster: Cluster Analysis Basics and Extensions*, 2013. R package version 1.14.4 — For new features, see the 'Changelog' file (in the package source).

Rosario N Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11(1):193–197, 1999.

Harry Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.

Alexander J McNeil and Johanna Nešlehová. Multivariate archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. *The Annals of Statistics*, pages 3059–3097, 2009.

Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: concepts, techniques, and tools*. Princeton university press, 2005.

Thomas Mikosch. Copulas: Tales and facts. *Extremes*, 9(1):3–20, 2006.

Boris Mirkin. *Clustering for Data Mining: A Data Recovery Approach*. Taylor & Francis, 2005.

S.R. Nanda, B. Mahanty, and M.K. Tiwari. Clustering indian stock market data for portfolio management. *Expert Systems with Applications*, 37(12):8793–8798, 2010.

Roger B Nelsen. Concordance and copulas: A survey. *Distributions with given marginals and statistical modelling. Kluwer, Dordrecht*, pages 169–178, 2002.

Roger B Nelsen. *An introduction to copulas*. Springer Verlag, 2006.

Arno Onken, Steffen Grünewälder, Matthias Munk, and Klaus Obermayer. Modeling short-term noise dependence of spike counts in macaque prefrontal cortex. *Advances in neural information processing systems*, 21:1233–1240, 2009a.

Arno Onken, Steffen Grünewälder, Matthias HJ Munk, and Klaus Obermayer. Analyzing short-term noise dependencies of spike-counts in macaque prefrontal cortex using copulas and the flashlight transformation. *PLoS computational biology*, 5(11):e1000577, 2009b.

Edoardo Otranto. Clustering heteroskedastic time series by model-based procedures. *Computational Statistics & Data Analysis*, 52(10):4685–4698, 2008.

Don B Panton, V Parker Lessig, and O Maurice Joy. Comovement of international equity markets: a taxonomic approach. *Journal of Financial and Quantitative Analysis*, 11(03):415–432, 1976.

Andrew J Patton. A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18, 2012.

Andrew J Patton. Copula methods for forecasting multivariate time series. *In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting*, 2:899–960, 2013.

Yue Peng and Wing Lon Ng. Analysing financial contagion and asymmetric market dependence with volatility indices via copulas. *Annals of Finance*, 8(1):49–74, 2012.

Marcello Pericoli and Massimo Sbracia. A primer on financial contagion. *Journal of Economic Surveys*, 17(4):571–608, 2003.

Robin L Plackett. A class of bivariate distributions. *Journal of the American Statistical Association*, 60 (310):516–522, 1965.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/.

William M Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.

Brian D Ripley. *Stochastic simulation*, volume 316. Wiley. com, 2009.

Juan Carlos Rodriguez. Measuring financial contagion: A copula approach. *Journal of Empirical Finance*, 14(3):401–423, 2007.

Ehud Ronn. The impact of large changes in asset prices on intra-market correlations in the stock and bond markets. *Manuscript, Department of Finance, University of Texas, Austin*, 1998.

Michael Roth. Linköping university electronic press report: On the multivariate t distribution. Technical report, Linköpings University Electronic Press, 2013.

Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

Ludger Rüschendorf, Berthold Schweizer, and Michael Dee Taylor. Distributions with fixed marginals and related topics. In *IMS Lecture Notes-Monograph Series Number 28*, pages 233–243. IMS, 1996.

Gianfausto Salvadori. *Extremes in Nature: An Approach to Using Copulas*, volume 56. Springer, 2007.

Marco Scarsini. On measures of concordance. *Stochastica: revista de matemática pura y aplicada*, 8 (3):201–218, 1984.

Ulf Schepsmeier, Jakob Stoeber, Eike Christian Brechmann, and Benedikt Graeler. *VineCopula: Statistical inference of vine copulas*, 2013. URL http://CRAN.R-project.org/package=VineCopula. R package version 1.1-2.

Friedrich Schmid and Rafael Schmidt. Multivariate extensions of spearman's rho and related statistics. *Statistics & Probability Letters*, 77(4):407–416, 2007.

Christian Schoelzel, Petra Friederichs, et al. Multivariate non-normally distributed random variables in climate research–introduction to the copula approach. *Nonlin. Processes Geophys.*, 15(5):761–772, 2008.

Philipp J Schönbucher. *Credit derivatives pricing models: models, pricing and implementation*. John Wiley & Sons, 2003.

Berthold Schweizer. Thirty years of copulas. In *Advances in probability distributions with given marginals*, pages 13–50. Springer, 1991.

Berthold Schweizer. Introduction to copulas. *Journal of Hydrologic Engineering*, 12(4):346–346, 2007.

Berthold Schweizer and Abe Sklar. Operations on distribution functions not derivable from operations on random variables. *Studia Mathematica*, 52(1):43–52, 1974.

Berthold Schweizer and Abe Sklar. *Probabilistic metric spaces*. DoverPublications. com, 1983.

Berthold Schweizer and Edward F Wolff. On nonparametric measures of dependence for random variables. *The annals of statistics*, pages 879–885, 1981.

George A.F Seber. Multivariate observations. *Wiley series in probability and mathematical statistics (Probability and mathematical statistics)*, 1984.

Abe Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8*, pages 229–231, 1959.

Abe Sklar. Random variables, distribution functions, and copulas: a personal look backward and forward. *Lecture notes-monograph series*, pages 1–14, 1996.

Charles Spearman. The proof and measurement of association between two things. *The American journal of psychology*, 15(1):72–101, 1904.

Ghasem Tarmast. Multivariate log–normal distribution. *ISI Proceedings: Seoul 53rd Session*, 2001. URL http://isi.cbs.nl/iamamember/CD2/pdf/329.PDF.

Vincenzo Tola, Fabrizio Lillo, Mauro Gallegati, and Rosario N Mantegna. Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control*, 32(1):235–258, 2008.

Malevergne Yannick and Didier Sornette. *Extreme financial risks*. Springer-Verlag, 2006.

Wuyi Ye, Xiaoquan Liu, and Baiqi Miao. Measuring the subprime crisis contagion: Evidence of change point analysis of copula functions. *European Journal of Operational Research*, 222(1):96–103, 2012.