Analysis of the Dependence Structure for Time Series of Hedge Funds Returns

by

He Zhu(Emma)

Bachelor of Mathematics, University of Waterloo, 2009

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Science

in the Program of

Applied Mathematics

Toronto, Ontario, Canada, 2011

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

_____

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

_____

Analysis of the Dependence Structure for Time Series of Hedge Funds Returns

Master of Science 2011

He Zhu(Emma)

Applied Mathematics

Ryerson University

The aim of the thesis is to emphasize the different dependence measures beyond the well known Pearson correlation. The study is developed in the setting of a fund of funds that deals with multiple strategies hedge funds under risk constraints. The relevance of our analysis is made clear by noticing that the Pearson correlation is sensitive only to linear relationships and it does not capture tail co-movements. Specifically, the dependence measures we focus on are Kendall's tau, Spearman's rho and tail dependence. This thesis attempts to suggest some other solutions to an effective optimization that combines various hedge fund strategies by using the aforementioned dependence measures.

Acknowledgements

From the formative stages of this thesis, to the final draft, I owe an immense debt of gratitude to my supervisor Dr. Marcos Escobar and co-supervisor Dr. Sebastian Ferrando. I would like to express my gratitude to them, who guided me along the way. Without their help and advice it would not have been possible to complete this work.

Last but not least, I would like to thank my family for their love and support.

Dedication

I would like to dedicate my thesis to my beloved grandmother Guiqin Huang, my father Chunyu Zhu and mother Jing Yu.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The hedge fund industry has encountered enormous growth in the last couple of decades. The number of active hedge funds has grown from as few as 300 funds in 1990 to approximately 10,000 today (Seco 2006). They have become highly visible, and their capital have reached a peak of almost \$1.9 trillion worldwide in 2007. What makes hedge funds such a popular investment? There are several reasonable points of view that can answer this question: (see Seco 2006) the growing number of new capital markets, the creation of opportunities for smaller firms that can take advantage of faster and better investment tools than larger financial firms, or most likely, the high profits attained by several hedge funds managers. On top of all this, it is the appeal of absolute returns, low volatility and risk-controlled strategies that hedge funds could offer.

Since the performance of a hedge fund can be closely approximated with a portfolio of as few as 20 hedge funds (see Peskin, Urias, Anjilvel and Boudreau 2000), this suggests a pooled fund-of funds approach as a viable alternative investment strategy. This can be explained in part by the relatively significant Pearson's correlation among individual hedge funds. A paper by Suppal (2004) provides a development of a systematic allocation methodology to combine hedge funds with multiple strategies as a "fund of funds". He finds that organizing distinct strategies into clusters or rational strategy groups such that they represent cohesiveness within a cluster yet maintain low Pearson's correlation with other clusters is one suggested approach to create a well diversified portfolio. More research is being conducted in this area since overlooking the dependence structure could give misleading allocation results detrimental for risk management. Suppal's paper suggests using, for optimization, a risk measure that looks beyond variance through Pearson correlation and considers higher moments, specifically skewness and kurtosis. In my thesis, we provide some alternative solutions to the multiple strategies hedge funds optimization by looking at other dependence measures beyond the Pearson correlation.

The most commons alternative choices of dependence measures are tail dependence, Kendall's tau and Spearman's rho. As the traditional concept of Pearson's correlation applied basically to linear relationships, it follows that it is not able to correctly reflect the behavior of assets in extreme situations,

which generally do not show a linear behavior. Sibuya (1959) was the first to introduce the concept of tail dependence. Thus, a coefficient of tail dependence has been developed and defined in a financial context as the probability that one asset undergoes a large loss or gain simultaneously with another asset. The other best two alternatives to the Pearson correlation as a measure of dependence are Kendall's tau and Spearman's rho. Spearman's rho applies to a ranking of the data making it a nonparametric measure of dependence, so no knowledge of the joint distribution of the data is required to obtain properties of this measure. It provides a measure of a monotonic relationship between two random variables as opposed to Pearson which only measures the linear relationship. It is most useful with ordinal data and, more importantly, it is robust to outliers while Pearson's correlation is not. Spearman's correlation is also a more widely used measure of rank correlation because it is easier to compute than Kendall's tau (Newson 2002). In order to provide contrast, notice that main advantages for using Kendall's tau, which is also a nonparametric measure of dependence, are that the distribution of this statistic has slightly better properties than that of Spearman's rho and there is a direct interpretation of Kendall's tau in terms of probabilities of observing concordant and discordant pairs (Conover, 1980). The literature on hedge funds has focused largely on Pearson's correlation measure disregarding the fact that hedge funds are complex instruments with not only nonstandard marginal behaviours (presence of Kurtosis and Skewness) but also nonstandard joint behaviours (dependence structure) leading to the analysis of the measures described in this thesis.

Additionally, we have also concentrated our efforts on defining the above dependence measures in the context of copulas. Copulas are well known for expressing and modelling the full dependence structure of a vector of random variables and, therefore, useful in multivariate financial and quantitative analysis. Copulas were first defined by Sklar (1959), it is an ideal tool for stress test extreme correlation movements and more general measures of dependence for a wide variety of financial portfolios and products in insurance and finance. As such, they are gradually becoming a standard element included in the best practice of integrated risk management. Therefore, in this work, copulas are going to be introduced and related to the previously mentioned dependence measures. Note that much empirical and theoretical work has been done in developing copulas for financial analysis (Jun Yan 2006) still, it is a difficult issue to find the copula that fits the data best. This is why we do not focus on finding the best copula but rather on estimating parameters and therefore the measures of dependence (Pearson, Kendall's tau, Spearman's rho and tail dependence) for some examples of copulas.

Here is a description of the thesis' contents. Chapter 2 provides the concept of copula together with some important families of copulas and the associated estimation methods. Chapter 3 introduces the four dependence measures which are discussed from two perspectives: parametric estimation using examples of copulas and nonparametric estimation. Chapter 4 provides a detailed analysis of the estimation results for the dependence measures within a hedge fund strategy and between hedge fund strategies using a population of hedge funds available from CISDM. Chapter 5 concludes with the key findings of this study. Last but not least, the code used in estimation and data manipulating for this thesis, implemented in Matlab, is displayed in Appendix for the interested reader.

1.1 Hedge Funds

There is no unique, precise definition for hedge funds because of its variety of investment styles, tools and strategies. A loosely speaking definition is that a hedge fund is an investment fund open to a limited range of investors that undertakes a wider range of investment and trading activities than traditional long-only investment funds, and that, in general, pays a performance fee to its investment manager.

Fung and Hsieh (1997) classify a hedge funds strategy according to both style and location. Refer to Capital Beason (2011), there are four major types of investment strategies using these hedge fund tools: global macro, directional, event driven, and relative value. The first element is style, which concerns how the investor will be investing the money in general. Your style of investment could be something like event-driven or global macro. Each style dictates a different way of trading and investing. Next, one has to consider the market: different markets have different rules and regulations. The instrument is also important, or the method. This could be futures, options, and so on. Table 1.1 provides us with an example of categorizing hedge funds. We also provides several definitions of the most popular financial markets and 21 strategies' names that are used in the thesis.

Table 1.1: Categorizing Hedge Fund Strategies: grouping under four themes

STRATEGY CLASS	SPECIFIC STRATEGY
Event Driven	
	Event Driven Multi-Startegy
	Merger/Risk Arbitrage
	Distressed
Relative Value	
	Arbitrage
	Statistical Arbitrage
	Specialist Credit
	Convertible Arbitrage
	Fixed Income Arbitrage
	Relative value Arbitrage
Long/Short	
	long/Short equity
	Dedicated Short-sellers
	Equity Market Neutral
Tactical	
	Global Macro
	Emerging markets
	Managed Futures

Equity long/short Equity long/short is the same as equity market neutral except without any explicit promise to maintain market neutrality. This increases the flexibility of the manager to choose net-long

or net-short (positive beta or negative beta) market exposure, while still focussing primarily on stock-selection opportunities. (see Connor and Lasarte.)

Equity market neutral Equity Market Neutral Strategy is the classic (and original) hedge fund strategy. This was the strategy proposed by Alfred Winslow Jones, who started the first hedge fund in 1949. A market neutral strategy uses the combination of longs and short-sales to offset any correlation between the portfolio return and the overall market return (See Connor and Lasarte.)

Fixed income arbitrage This strategy relies heavily on mathematical models of the term structure of interest rates to identify mis-pricing and manage positions. (See Connor and Lasarte) "Fixed income arbitrage has its historical roots in the fixed income trading desks of brokerage houses and investment banks", according to Connor and Lasarte.

Emerging markets An emerging market hedge fund is a hedge fund that specializes its investments in the securities of emerging market countries (Barclay Hedge.) Loosely speaking, emerging market countries are in the process of developing. Emerging market countries include a very large range of countries (See Connor and Lasarte).

FOF A "fund of funds" (FOF) is an investment strategy of holding a portfolio of other funds rather than investing directly in shares, bonds or other securities. This type of investing is often referred to as multi-manager investment. A fund of funds may be mutual fund FOF or hedge fund FOF. In our database, we faced several different FOF of hedge funds. FOF-Multi Strategy, FOF- Single Strategy, FOF- Invest Funds in Parent Company and FOF- Market Neutral. The diversified FOF strategies are characterized by different hedge fund investment strategies within a single hedge fund.

Table 1.2: Strategy Name and Size

Strategy Name Short Form	Strategy Name	Number of Companies
C.A.	Convertible Arbitrage	169
CTA-D	CTA- Discretionary	121
CTA-S/T-F	CTA- Systematic/ Trend- Following	558
D.S	Distressed Securities	112
E.M	Emerging markets	271
E.L.O.	Equity Long Only	76
E. L/S	Equity Long/Short	1276
E.M.N.	Equity Market Neutral	215
E.D. M.S.	Event Driven Multi Strategy	152
F I.	Fixed Income	108
F I.MB	Fixed Income-MBS	72
F. I. A.	Fixed Income Arbitrage	104
FOF-I F. P. C.	FOF- Invest Funds in Parent Company	69
FOF-M. N.	FOF- Market Neutral	65
FOF-M. S.	FOF- Multi Strategy	1205
FOF-S. S.	FOF- Single Strategy	218
G. M.	Global Macro	144
M. A.	Merger Arbitrage	93
M. S.	Merger Arbitrage	72
R. V. M. S.	Relative Value Multi Strategy	73
S	Sector	305

Chapter 2

Copulas

According to Embrechts, Lindskog and McMeil (2001), the standard "operational" definition of a copula is a multivariate distribution function defined on the unit cube $[0,1]^n$, with uniformly distributed marginals. This definition is a natural one considering how a copula is derived from a continuous multivariate distribution function. Loosely speaking, the copula is simply the original multivariate distribution function with transformed univariate margins. The exact general definition is provided next.

2.1 Copula

Definition A copula is a multivariate joint distribution defined on the n-dimensional unit cube $[0,1]^n$ such that every marginal distribution is uniform on the interval $[0,1]$. Specifically, $C : [0,1]^n \to [0,1]$ is an n-dimensional copula if:

1. $C(u) = 0$ whenever $u \in [0,1]^n$ has at least one component equal to 0;

2. $C(u) = u_i$ whenever $u \in [0,1]^n$ has all the components equal to 1 except the ith 1, which is equal to u_i;

3. C is n-increasing.

The following theorem is known as Sklar's theorem (1996). It states that given a joint distribution function, there exists a copula C such that the copula binds the marginals to provide the joint distribution. Moreover, as previously mentioned, the joint distribution can be used to define the dependence measures.

Theorem 2.1.1 *Let H be an n-dimensinal distribution function with marginals $F_1,...,F_n$. Then there exists an n-copula C such that for all $\mathbf{X}$ in $\mathbf{R}^n$,*

$$H(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n)). \tag{2.1}$$

If F_1, ... , F_n are all continuous, then C is unique; otherwise C is uniquely determined on Ran $F_1 \times \cdots \times$ Ran F_n. conversely, if C is an n-copula and $F_1, ..., F_n$ are distribution functions, then the function H defined above is an n-dimensional distribution function with marginals $F_1, ..., F_n$.

From Theorem 2.1.1 we know that the copula function C "separates" an n-dimensinal distribution function from its univariate margins. Further, let F be a univariate distribution function. We define the generalized inverse of F (Sklar 1996) as $F^{-1}(t) = inf\{x \in \mathbf{R}|F(x) \geq t\}$ for all t in [0, 1], using the convention inf $\emptyset = -\infty$

Corollary 2.1.2 *Let H be an n-dimensional distribution function with continuous marginals $F_1, ..., F_n$ and copula C (where C satisfies Equation (2.1)). Then for any $\mathbf{u}$ in $[0, 1]^n$,*

$$C(u_1, ..., u_n) = H(F_1^{-1}(u_1), ..., F_n^{-1}(u_n)). \tag{2.2}$$

Copulas and Random Variables . Let $X_1, ..., X_n$ be random variables with continuous distribution functions $F_1, ..., F_n$, respectively, and joint distribution function H. Then $(X_1, ..., X_n)^T$ has a unique copula C, where C is given in equation (2.1). The standard copula representation of the distribution of the random vector $(X_1, ..., X_n)^T$ then becomes:

$$H(x_1, ..., x_n) = \Pr\{X_1 \leq x_1, ..., X_n \leq x_n\} = C(F_1(x_1), ..., F_n(x_n)). \tag{2.3}$$

2.2   Survival Copula and Joint Survival Function

Let us consider the probability:

$$\bar{H}(\mathbf{x}) = \Pr(X_1 > x_1, X_2 > x_2, ..., X_n > x_n).$$

As in the bi-dimensional case, this probability is called the joint survival probability or survival function of the n agents or components X_i, while the marginal survival probabilities or survival functions are:

$$\bar{F}_i(x_i) = \Pr(X_i > x_i).$$

The copula that represents the joint survival probability in terms of the survival probabilities of the n components X_i separately is named the survival copula (Cherubini, Luciano and Vecchiato 2004). The existence of the latter is guaranteed by the survival version of Sklar's theorem, which guarantees that there is a copula $\bar{C}$, unique on Rank $\bar{F}_1 \times \cdots \times$ Rank $\bar{F}_n$. We then introduce the following:

Definition The survival copula of the random variables $X_1, X_2, ..., X_n$ is the copula $\bar{C}$, unique on Rank $\bar{F}_1 \times \cdots \times$ Rank $\bar{F}_n$, such that

$$\bar{H}(\mathbf{x}) = \bar{C}(\bar{F}_1(x_1), \bar{F}_2(x_2), ..., \bar{F}_n(x_n)). \tag{2.4}$$

Applying equations (2.3) and (2.4), for the bi-variate case, we have the following relationship between copula and survival copula:

$$\bar{C}(v,z) = v + z - 1 + C(1-v, 1-z). \tag{2.5}$$

It is also possible to express, via the survival copula, the conditional probability

$$\Pr(U_1 > v | \ U_2 > z) = \frac{1 - v - z + C(v,z)}{1-z} = \frac{\bar{C}(1-v, 1-z)}{1-z}$$

and, therefore,

$$\Pr(X > x | Y > y) = \frac{\bar{C}(\bar{F}_1(x), \bar{F}_2(y))}{\bar{F}_2(y)}. \tag{2.6}$$

We will develop equation (2.6) in a latter chapter in connection with the topic of "tail dependence coefficient".

2.3 Examples

This section introduces four examples of copulas; the first two are a Gaussian copula and a t-copula, both belong to the class of elliptical copulas. The advantage of elliptical copulas is that they can be simulated from elliptical distributions easily as a consequence of Sklar's Theorem. For further details on elliptical distributions we refer to Fang, Kotz, and NG(1987) and Cambanis, Huang, and Simons (1981). The other two examples are Archimedean copula and Marshall-Olkin copula. The class of Archimedean copulas allows for a great variety of different dependence measures, and some commonly encountered Archimedean copulas have closed form expressions. We will discuss them in the next section.

2.3.1 Gaussian Copula

Definition Multivariate Gaussian Copula [U. Cherubini, E. Luciano, W. Vecchiato (2004)]. Let R be a symmetric, positive definite matrix with $diag(R) = (1, 1, ...1)^T$ and Φ_R the standardized multivariate normal distribution with correlation matrix R. The multivariate Gaussian copula is defined as follows:

$$C_R^{Ga}(\mathbf{u}) = \Phi_R(\Phi^{-1}(u_1), \Phi^{-1}(u_2), ..., \Phi^{-1}(u_n)) \tag{2.7}$$

where Φ^{-1} , is the inverse of the standard univariate normal distribution function Φ.

Proposition 2.3.1 *The Gaussian copula generates the standard joint normal distribution function (via Sklar's theorem) if and only if the margins are standard normal.*

Since it is the standard joint normal distribution function, we can find the corresponding density by taking the derivatives. We have:

$$\frac{1}{(2\pi)^{n/2} \mid \mathbf{R} \mid^{1/2}} \exp(-\frac{1}{2}\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}) = c_R^{Ga}(\Phi(x_1), \Phi(x_2), ..., \Phi(x_n)) \times \prod_{j=1}^{n}(\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x_j^2)) \tag{2.8}$$

where $| \mathbf{R} |$ is the determinant of $\mathbf{R}$.

$$c_R^{Ga}(\Phi(x_1), \Phi(x_2), ..., \Phi(x_n)) = \frac{\frac{1}{(2\pi)^{n/2}|\mathbf{R}|^{1/2}} \exp(-\frac{1}{2}\mathbf{x}^T \mathbf{R}^{-1} \mathbf{x})}{\prod_{j=1}^{n}(\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x_j^2))} \tag{2.9}$$

let $u_j = \Phi(x_j)$, so that $x_j = \Phi^{-1}(u_j)$. We have the density function:

$$c_R^{Ga}(u_1, u_2, ..., u_n) = \frac{1}{| \mathbf{R} |^{1/2}} \exp(-\frac{1}{2}\varsigma^T(\mathbf{R}^{-1} - I)\varsigma) \tag{2.10}$$

where $\varsigma = (\Phi^{-1}(u_1), \Phi^{-1}(u_2), ..., \Phi^{-1}(u_n))^T$

2.3.2 Student t copula

Definition Multivariate Student's t copula[U. Cherubini, E. Luciano, W. Vecchiato (2004)]. Let R be a symmetric, positive definite matrix with $diag(R) = (1, 1, ...1)^T$ and $t_{R,v}$ the standardized multivariate Student's t distribution with correlation matrix R and v degrees of freedom, i.e.

$$t_{R,v}(x_1, x_2, ..., x_n) = \int_{-\infty}^{x_1} ... \int_{-\infty}^{x_n} \frac{\Gamma(\frac{v+n}{2}) | R |^{-1/2}}{\Gamma(\frac{v}{2})(v\pi)^{n/2}} (1 + \frac{1}{v}\mathbf{s}^T \mathbf{R}^{-1} \mathbf{x})^{-\frac{v+n}{2}} dx_1 dx_2 ... dx_n \tag{2.11}$$

The multivariate t copula is defined as follows:

$$T_{R,v}(u_1, u_2, ..., u_n) = t_{R,v}(t_v^{-1}(u_1), t_v^{-1}(u_2), ..., t_v^{-1}(u_n)) \tag{2.12}$$

$$T_{R,v}(u_1, u_2, ..., u_n) = \int_{-\infty}^{t_v^{-1}(u_1)} ... \int_{-\infty}^{t_v^{-1}(u_n)} \frac{\Gamma(\frac{v+n}{2}) | R |^{-1/2}}{\Gamma(\frac{v}{2})(v\pi)^{n/2}} (1 + \frac{1}{v}\mathbf{s}^T \mathbf{R}^{-1} \mathbf{x})^{-\frac{v+n}{2}} dx_1 dx_2 ... dx_n \tag{2.13}$$

$$\tag{2.14}$$

where t_v^{-1} is the inverse of the univariate c.d.f. of Student's t with v degrees of freedom. Using the canonical representation, it turns out that the copula density is:

$$c_{R,v}(u_1, u_2, ..., u_n) = | R |^{-\frac{1}{2}} \frac{\Gamma(\frac{v+n}{2})}{\Gamma(\frac{v}{2})} (\frac{\Gamma(\frac{v}{2})}{\Gamma(\frac{v+1}{2})})^n \frac{(1 + \frac{1}{v}\varsigma^T \mathbf{R}^{-1} \Gamma)^{-\frac{v+n}{2}}}{\prod_{j=1}^{n}(1 + \frac{\varsigma_j^2}{v})^{-\frac{v+1}{2}}} \tag{2.15}$$

where $\varsigma_j = t_v^{-1}(u_j)$.

2.3.3 Archimedean copulas

The class of Archimedean copulas has been named by Ling(1965), but it was recognized by Schweizer and Sklar(1961). Before being introduced in Finance, Archimedean copulas have been used in the Actuarial field: the idea arose indirectly in Clayton(1978) and was developed in Oakes(1982), Cook and Johnson(1981). Among Archimedean copulas, we are going to consider in particular the one-parameter ones, which are constructed using a generator $\psi_\alpha(t)$, indexed by the (real) parameter α. By choosing

the generator, one obtains a family of Archimidean copulas.

Definition Let ψ be a strict generator, with $\psi-1$ completely monotonic on $[0, \infty)$, Then an n-variate Archimedean copula is the function

$$C(u_1, u_2, ..., u_n) = \psi^{-1}(\psi(u_1) + \psi(u_2) + ... + \psi(u_n)). \tag{2.16}$$

As in the bi-dimensional case, an important source of generators for Archimedean n-colulas consists of the inverse of the Laplace transforms of c.d.f.

The remaining of the section provides three examples from a one-parameter Archimedean copula family. For more details see Genest and MAcKay (1986), Genest Rivest (1993), Nelsen(1999) and Joe (1997).

Gumbel n-copula

The generator is given by $\psi(u) = (-\ln(u))^{\alpha}$, hence $\psi^{-1}(t) = \exp(-t^{\frac{1}{\alpha}})$; it is completely monotonic if $\alpha > 1$. The Gumbel n-copula is therefore:

$$C(u_1, u_2, ..., u_n) = \exp\{-[\sum_{i=1}^{n}(-\ln u_i)^{\alpha}]^{\frac{1}{\alpha}}\}, \quad \alpha > 1. \tag{2.17}$$

Clayton n-copula

The generator is given by $\psi(u) = u^{-\alpha} - 1$, hence $\psi^{-1}(t) = (t+1)^{\frac{-1}{\alpha}}$; It is completely monotonic if $\alpha > 0$. The Clayton n-copula is therefore:

$$C(u_1, u_2, ..., u_n) = [\sum_{i=1}^{n} u_i^{-\alpha} - n + 1]^{\frac{-1}{\alpha}}, \quad \alpha > 0. \tag{2.18}$$

Frank n-copula

The generator is given by

$$\psi(u) = \ln(\frac{\exp(-\alpha u) - 1}{\exp(-\alpha) - 1}) \tag{2.19}$$

therefore,

$$\psi^{-1}(u) = -\frac{1}{\alpha}\ln(1 + e^t(e^{-\alpha} - 1)), \tag{2.20}$$

it is completely monotonic if $\alpha > 0$. The Frank n-copula is given by:

$$C(u_1, u_2, ..., u_n) = -\frac{1}{\alpha}\ln[1 + \frac{\prod_{i=1}^{n}(e^{-\alpha u_i} - 1)}{(e^{-\alpha} - 1)^{n-1}}]. \tag{2.21}$$

2.3.4 Marshall-Olkin copulas

Embrechts, Lindskog and McNeil (2001) consider a two- component system where the components are subject to shocks, which are fatal to one or both components. Let X_1 and X_2 denote the lifetimes of the

two components. Assume that the shocks follow three independent Poisson processes with parameters λ_1, λ_2, $\lambda_{12} \leq 0$, where the index indicates whether the shocks effect only component 1, component 2 or both. Then the times Z_1, Z_2 and Z_{12} of occurrence of these shocks are independent exponential random variables with parameters λ_1, λ_2 and λ_{12} respectively. Hence

$$\bar{H}(x_1, x_2) = \Pr\{X_1 > x_1, X_2 > x_2\} = \Pr\{Z_1 > x_1\} \Pr\{Z_2 > x_2\} \Pr\{Z_{12} > max(x_1, x_2)\}, \qquad (2.22)$$

with the survival functions for X_1 and X_2:

$$\bar{F}_1(x_1) = exp(-(\lambda_1 + \lambda_{12})x_1) \qquad (2.23)$$

$$\bar{F}_2(x_2) = exp(-(\lambda_2 + \lambda_{12})x_2). \qquad (2.24)$$

$$(2.25)$$

Since $\max(x_1, x_2) = x_1 + x_2 - \min(x_1, x_2)$,

$$\bar{H}(x_1, x_2) = \exp(-(\lambda_1 + \lambda_{12})x_1 - (\lambda_2 + \lambda_{12})x_2 + \lambda_{12}\min(x_1, x_2)) \qquad (2.26)$$

$$\bar{H}(x_1, x_2) = \bar{F}_1(x_1)\bar{F}_2(x_2)\min(\exp(\lambda_{12}x_1), \exp(\lambda_{12}x_2)). \qquad (2.27)$$

Let $\alpha_1 = \lambda_{12}/(\lambda_1 + \lambda_{12})$ and $\alpha_2 = \lambda_{12}/(\lambda_2 + \lambda_{12})$. This construction leads to a copula family given by

$$C_{\alpha_1,\alpha_2}(u_1, u_2) = min(u_1^{1-\alpha_1}u_2, u_1 u_2^{1-\alpha_2}) = \begin{cases} u_1^{1-\alpha_1}u_2, & u_1^{\alpha_1} \geq u_2^{\alpha_2} \\ u_1 u_2^{1-\alpha_2}, & u_1^{\alpha_1} \leq u_2^{\alpha_2}. \end{cases} \qquad (2.28)$$

2.4 Estimation

This section provides an introduction to estimation procedures for copulas. Depending on the assumptions made on a copula model, we will discuss the estimation procedures in two parts, parametric estimation in Section 2.4.1 and non-parametric estimation in Section 2.4.2.

2.4.1 Parametric Estimation

Exact Maximum Likelihood Method

Let $\mathbf{X} = (x_{1t}, x_{2t}, ..., x_{nt})_{t=1}^T$ be the sample data matrix. Thus, the expression for the log-likelihood function is

$$l(\theta) = \sum_{t=1}^T \ln c(F_1(x_{1t}), F_2(x_{2t}), ..., F_n(x_{nt})) + \sum_{t=1}^T \sum_{j=1}^n \ln f_j(x_{jt}), \qquad (2.29)$$

where θ is the set of all parameters for both, the marginals and the copula. At this point, the reader is referred to Serfling (1980) and Shao (1999). Hence, given a set of marginal p.d.f. $f_j(x_{jt})$ and a copula, the previous log-likelihood may be evaluated, and, by maximization we obtain the maximum likelihood

estimator:

$$\widehat{\theta}_{MLE} = ArgMax\, l(\theta). \tag{2.30}$$

IFM Method

The maximum likelihood method, previously shown, could be very computationally intensive, especially in the case of a high dimension. Joe and Xu (1996) proposed that these set of parameters should be estimated in two steps:

1. First of all, they estimate the margins' parameters θ_1 by performing the estimation of the univariate marginal distributions:

$$\widehat{\theta_1} = ArgMax_{\theta_1} \sum_{t=1}^{T} \sum_{j=1}^{n} \ln f_j(x_{jt}; \theta_1). \tag{2.31}$$

2. Secondly, given $\hat{\theta}_1$, they perform the estimation of the copula parameter θ_2:

$$\hat{\theta}_2 = ArgMax_{\theta_2} \sum_{t=1}^{T} \ln c(F_1(x_{1t}), F_2(x_{2t}), ..., F_n(x_{nt}); \theta_2, \hat{\theta}_1). \tag{2.32}$$

This method is called inference for the margins or IFM. The IFM estimator is defined as the vector:

$$\hat{\theta}_{IFM} = (\hat{\theta}_1, \hat{\theta}_2)'. \tag{2.33}$$

We call l the entire log-likelihood function, l_j the log-likelihood of the j_{th} marginal. Hence, the IFM estimator is the solution of:

$$(\frac{\partial l_1}{\partial \theta_{11}}, \frac{\partial l_2}{\partial \theta_{12}}, ..., \frac{\partial l_n}{\partial \theta_{1n}}, \frac{\partial l}{\partial \theta_2}) = 0' \tag{2.34}$$

while the MLE comes from solving

$$(\frac{\partial l}{\partial \theta_{11}}, \frac{\partial l}{\partial \theta_{12}}, ..., \frac{\partial l}{\partial \theta_{1n}}, \frac{\partial l}{\partial \theta_2}) = 0'. \tag{2.35}$$

From the above two equations, one can notice that these two estimators are not equivalent, and it is easy to see that the IFM estimator provides a good starting point for obtaining an exact MLE. (See Cherubini, Luciano and Vecchiato 2004). According to Cherubini, Luciano and Vecchiato, since it is computationally easier to obtain the IFM estimator than the MLE, it is worth addressing a question about the IFM asymptotic efficiency compared with the MLE. Thus, one has to compare the asymptotic covariance matrix of the two estimators.

2.4.2   Non-Parametric Estimation

We provide the notion of the empirical copula introduced by Deheuvels (1979,1981). Let $X_t = (X_{1t}, X_{2t}, ..., X_{nt}) \in \mathbf{R}_n$ be an i.i.d. sequence with continuous joint c.d.f. F and continuous margins F_j. Let $\{x_1^{(t)}, x_2^{(t)}, ..., x_n^{(t)}\}$

be the order statistic and let $\{r_1^{(t)}, ..., r_n^{(t)}\}$ be the rank statistic of the sample, which are linked by the relationship $x_n^{(r_n^t)} = x_{nt}, t = 1, 2, ..., T$.

Deheuvels' empirical copula The empirical copula defined on the lattice

$$l = \{(\frac{t_1}{T}, \frac{t_2}{T}, ..., \frac{t_n}{T}) : 1 \leq j \leq n, t_i = 0, 1, ..., T\} \tag{2.36}$$

is the following function:

$$\hat{C}(\frac{t_1}{T}, ..., \frac{t_n}{T}) = \frac{1}{T} \sum_{t=1}^{T} \prod_{j=1}^{n} \mathbf{I}(r_j^t \leq t_j) \tag{2.37}$$

where **I** is the indicator function that takes value equal to 1 when its argument condition $r_j^{(t)} \leq t_j$ is satisfied.

The analog of the Radon-Nikodym density for the empirical copula is the following empirical copula frequency, as defined by Nelsen (1999):

$$\hat{c}(\frac{t_1}{T}, ..., \frac{t_n}{T}) = \sum_{i_1=1}^{2} \sum_{i_2=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{\sum_{j=1}^{n} i_j} \times \hat{C}(\frac{t_1 - i_1 + 1}{T}, \frac{t_2 - i_2 + 1}{T}, ..., \frac{t_n - i_n + 1}{T}). \tag{2.38}$$

Nelson (1999) notes that the concept of empirical copula permits us to define the sample version of many dependence measure and, also, the sample version of other concepts expressed in terms of copulas (See Cherubini, Luciano and Vecchiato 2004). Besides that, empirical copulas may also be used to construct non-parametric tests for independence (Deheuvel, 1981). The so-called empirical copulas resemble usual multivariate empirical cumulative distribution functions. The drawbacks of this method is that these empirical copula are highly discontinuous and cannot be exploited as a graphical device. (Fermanian and Scaillet 2003)

Chapter 3

Measures of Dependence

The Pearson correlation coefficient is one of the most frequently-used dependence measures but it is sensitive only to a linear relationship between two variables. Other correlation coefficients have been developed to be more robust than the Pearson correlation, or sensitive to nonlinear relationships and extreme events. In this Chapter, Pearson's correlation, Kendall's rank correlation coefficient τ, Spearman's rank correlation coefficient ρ and the coefficients of tail dependence are provided with their parametric and nonparametric estimators.

3.1 Examples

3.1.1 Coefficients of Tail Dependence

The concept of tail dependence arises from the need to characterize the asymptotic dependence structure of bivariate (or multivariate) data for the purpose of modelling their extreme values. For example, it is known that a portfolio's Value at Risk (VaR) is determined by the risk behaviour of each single asset in the portfolio (See Cherubini, Luciano and Vecchiato 2004). In this work, tail dependence is characterized by the so-called coefficient of tail dependence and embedded into the general framework of copula. In two dimensions, tail dependence describes the limiting proportion that one marginal exceeds a certain threshold given that the other marginals has already exceeded that certain threshold. There are various ways to define the tail dependence in multivariate dimensions, and this variety depends on the number of given marginals that have already exceeded the certain extreme value. Let X and Y be continuous random variables with marginal distributions G_x and G_y, with G_y^{-1} and G_x^{-1} denoting the inverse distribution functions of Y and X, respectively, and with joint function copula C defined in $u \in [0, 1]$. Then the coefficients of lower and upper tail dependence are given in the following displays.

Lower Tail Dependence Coefficient (Lower Tail Index):

$$\lambda_L = lim_{u \to 0} Pr[Y < G_y^{-1}(u)|X < G_x^{-1}(u)], \tag{3.1}$$

in case the limit exists (the same remark applies to several use of limits below). We say (X,Y) is lower tail-independent if $\lambda_L = 0$. Alternatively, (X,Y) is said to have lower tail dependence iff $\lambda_L \in (0,1]$.

Upper Tail Dependence Coefficient (Upper Tail Index):

$$\lambda_U = lim_{u \to 1} Pr[Y > G_y^{-1}(u) | X > G_x^{-1}(u)] \tag{3.2}$$

$$\lambda_U = lim_{u \to 1} \frac{Pr[Y > G_y^{-1}(u), X > G_x^{-1}(u)]}{Pr[X > G_x^{-1}(u)]}. \tag{3.3}$$

Analogously to the lower tail dependence coefficient, we say that (X,Y) is upper tail-independent if $\lambda_u = 0$. Alternatively, (X,Y) is said to have upper tail dependence if and only if $\lambda_u \in (0,1]$. The lower tail dependence coefficient describe the "left corner associations" among the data set (X,Y), and the upper tail dependence coefficient gives the "right corner associations" among the data set(X,Y).

The concept of tail dependence can be also described using copula theory,

$$\lambda_L = lim_{u \to 0} \frac{C(u,u)}{u}, \tag{3.4}$$

$$\lambda_U = lim_{u \to 1} \frac{1 - 2u + C(u,u)}{u}. \tag{3.5}$$

Let $X = (X_1, X_2, ..., X_n)$ be a n-dimensional vector, an extension to the multivariate case is (see Schmidt and Stadtmuller 2003):

$$\lambda_L = lim_{u \to 0} \frac{C(u, ..., u)}{u} \tag{3.6}$$

$$\lambda_U = lim_{u \to 1} \frac{\overline{C}(u, ..., u)}{u}. \tag{3.7}$$

3.1.2   Pearson's correlation

Pearson's correlation ρ_p or the Pearson product-moment correlation coefficient is the most familiar measure of dependence between two variables. It is defined as:

$$\rho_p = corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{3.8}$$

where the μ_X and μ_Y are the expected values of random variables X and Y with standard deviations σ_X and σ_Y.

The Pearson correlation is +1 indicating the case of perfect positive linear relationship, -1 indicating a perfect negative linear relationship. The closer the coefficient is to either -1 or 1, the stronger the linear correlation between the variables. As we mentioned, the Pearson's correlation is only sensitive to linear relationship moreover in the special case when the two variables X and Y are jointly normal, $\rho_s = 0$ is equivalent to independence.

3.1.3 Kendall's Tau

Kendall' s tau is a commonly used alternative measure of dependence to the benchmark Pearson's correlation. For many joint distributions these two measures have different values, as they measure different aspects of the dependence structure.

To explain this difference, the concept of concordance will be introduced first here. Refer to Cherubini, Luciano and Vecchiato (2004), the concordance concepts aim at capturing the fact that the probability of having "large (or small)" values of both X and Y is high, while the probability of having the probability of having "large" values of X together with "small" values of Y, or vice versa, is low.

For instance, refer to Fredricks and Nelsen (2006), if X and Y are random variables with marginal distribution functions F and G, respectively, then Kendall's τ is the difference between the probability of concordance $\Pr[(X_1 - X_2)(Y_1 - Y_2) > 0]$ and the probability of discordance $\Pr[(X_1 - X_2)(Y_1 - Y_2) < 0]$ for two independent pairs (X_1, Y_1) and (X_2, Y_2) of observations drawn from the distribution.

Definition (Nelsen 2001) Kendall's tau. Let (x_1, y_1), (x_2, y_2), ... , (x_n, y_n) be a set of joint observations from two random variables X and Y respectively. Any pair of observations (x_i, y_i) and (x_j, y_j) are said to be concordant if the ranks for both elements agree: that is, if both $x_i > x_j$ and $y_i > y_j$ (i.e., if $(x_i - x_j)(y_i - y_j) > 0$); and discordant if $x_i < x_j$ and $y_i > y_j$ or if $x_i > x_j$ and $y_i < y_j$; If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant nor discordant. Assume N_c be the number of concordant pairs, and N_d be the number of discordant pairs. Then the Kendall' τ coefficient is defined as:

$$\tau(X, Y) = \frac{N_c - N_d}{(n-1)/2} \tag{3.9}$$

For example, Kendall's tau for the given (X_1, Y_1) and (X_2, Y_2) is:

$$\tau = \Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - \Pr[(X_1 - X_2)(Y_1 - Y_2) < 0]. \tag{3.10}$$

Theorem 3.1.1 *Follows Nelsen (1991), Kendall's τ can be computed using a copula C as:*

$$\tau(X, Y) = 1 - 4 \iint_{[0,1]^2} \frac{\partial C(v, z)}{\partial v} \frac{\partial C(v, z)}{\partial z} \partial v \partial z \tag{3.11}$$

Kendall's tau takes values between -1 and +1. The coefficient has value 1 indicating the agreement between the two rankings is perfect. The coefficient has values -1 indicating the disagreement between the two ranking is perfect. If X and Y are independent, then we would expect the coefficient to be approximately zero.

3.1.4 Spearman's Rho

Spearman's rank correlation coefficient or Spearman's rho, denoted by ρ or r_s, was proposed in 1904 (Nelsen 2001). It is a normalized expected value and represents rank correlation. In other words, Spearman's ρ is the Pearson correlation coefficient of the transformed random variables $F(X)$ and $G(Y)$ where F and G represent the distributions functions of X and Y respectively.

Definition Spearman's ρ for random variable X and Y with copula C is:

$$\rho = 12 \iint_{[0,1]^2} C(v,z) dvdz - 3 = 12 \iint_{[0,1]^2} vz dC(v,z) - 3 \tag{3.12}$$

We can also use integral transform, $U_1 = F_1(X)$ and $U_2 = G_2(Y)$, which are standard uniform, with joint distribution function C. The formula for Spearman's ρ can be written as:

$$\rho = 12E[U_1 U_2] - 3 = \frac{E[U_1 U_2] - 1/4}{1/12} \tag{3.13}$$

Knowing the property of standard uniforms distribution, one can transform the equation by using its mean (1/2) and variance (1/12), the equation becomes (See Cherubini, Luciano and Vecchiato 2004):

$$\rho = \frac{cov(F(X), G(X))}{\sqrt{var(F(X)) var(G(Y))}} \tag{3.14}$$

Therefore, we will define such a ratio as the linear correlation coefficient between $F(X)$ and $G(Y)$: Spearman's ρ is the rank correlation, in the sense of correlation of the integral transforms, of X and Y. The meanings of the range value for both Kendall's tau and Spearman's rho are similar. They both called the "rank correlation coefficient",but the difference is the ways of ranking. According to Durbin and Stuart (1951), there exists a relationship between Kendall's tau and Spearman's rho. For a given copula, i.e. for a given dependence structure between X and Y we have (See Cherubini, Luciano and Vecchiato 2004):

$$\begin{cases} \frac{3\tau}{2} - \frac{1}{2} \le \rho \le \frac{1}{2} + \tau - \frac{\tau^2}{2} & \tau \ge 0 \\ \frac{-1}{2} + \tau + \frac{\tau^2}{2} \le \rho \le \frac{3\tau}{2} + \frac{1}{2} & \tau < 0 \end{cases} \tag{3.15}$$

3.2 Estimations

One of our objectives is to provide estimators for the various dependence measures. In this Section, the estimation procedure will be discussed in two parts, parametric estimation and non-parametric estimation. In Section 3.2.1, we list a few parametric estimators based on the choice of the copula model. In this work, we select the Gaussian copula, Archimedean copula family and Marshall-Olkin copula in bivariate case. According to the definitions of the dependence measures, see equations (3.4) (3.11) and (3.12), we can generate the estimators for each dependence measure by signing different copula models. In Section 3.2.2, we have the estimators for the dependence measures by not making any assumptions on the parametric copula model.

3.2.1   Parametric Estimation

Gaussian Copula

Recalling the the definitions of Gaussian Copula, see equation (2.7), let ρ be linear correlation coefficient. We can use the definition of Kendall's τ and Spearman's ρ_s, and have:

$$\tau = \frac{2\arcsin\rho}{\pi} \tag{3.16}$$

and

$$\rho_s = \frac{6\arcsin\frac{\rho}{2}}{\pi} \tag{3.17}$$

For tail dependency, the Gaussian copulas have neither upper nor lower tail dependence unless $\rho = 1$,see U. Cherubini, E. Luciano, W. Vecchiato (2004):

$$\lambda_U = \lambda_L = \begin{cases} 0 & \rho < 1 \\ 1 & \rho = 1 \end{cases} \tag{3.18}$$

Archimedean Copulas

Archimedean copulas are easily related to measures of association. (See Cherubini, Luciano and Vecchiato 2004)

Theorem 3.2.1 *See Nelson (1991). Let ψ be a strict generator such that ψ^{-1} belongs to the class of Laplace transforms of a.s. strictly positive r.v.s. If $\psi'(0)$ is finite and different from zero (See Cherubini, Luciano and Vecchiato 2004), then*

$$C(u,v) = \psi^{-1}(\phi(v) + \phi(z)) \tag{3.19}$$

does not have upper tail dependency. If instead C has upper tail dependency, then $1/\psi'(0) = -\infty$ and the coefficient of upper tail dependency is (See Cherubini, Luciano and Vecchiato 2004)

$$\lambda_U = 2 - 2\lim_{s\to 0}\frac{\varphi'(s)}{\varphi'(2s)} \tag{3.20}$$

The coefficient of lower tail dependency is (See Cherubini, Luciano and Vecchiato 2004)

$$\lambda_U = 2\lim_{s\to\infty}\frac{\varphi'(s)}{\varphi'(2s)} \tag{3.21}$$

Furthermore, pick one-parameter copula from Archimedean family. We have Gumbel family introduced by Gumbel (1960), Clayton family (1978) and Frank family (1979). The generators of these copula families are listed in Table 3.2.1. By applying the formula of the dependence measures, we can have its estimator, see Table 3.2.

Gumbel	
$\phi_\alpha(t)$	$(-\ln t)^\alpha$
C(v, z)	$exp\{-[(-lnv)^\alpha] + (-lnz)^\alpha]^{\frac{1}{\alpha}}\}$
Clayton	
$\phi_\alpha(t)$	$\frac{1}{\alpha}(t^{-\alpha} - 1)$
C(v, z)	$\max[(v^{-\alpha} + z^{-\alpha} - 1)^{-1/\alpha}, 0]$
Frank	
$\phi_\alpha(t)$	$-ln\frac{exp(-\alpha t)-1}{exp(-\alpha)-1}$
C(v, z)	$-\frac{1}{\alpha}ln(1 + \frac{(exp(-\alpha v)-1)(exp(-\alpha z)-1)}{exp(-\alpha)-1})$

Table 3.1: Archimedean Copulas with parameter α, where ϕ_α is the Archimedean copula generators

Family	kendall's τ	Spearman's ρ	Tail Dependence
Gumbel(1960)	1-α^{-1}	no closed form	$2 - 2^{1/\alpha}$
Clayton(1978)	$\alpha/(\alpha+2)$	complicated expression	$2^{-1/\alpha}$
Frank (1979)	$1 + 4[D_1(\alpha) - 1]/\alpha$	$1 - 12[D_2(-\alpha) - D_1(-\alpha)]/\alpha$	0

Table 3.2: Dependence measures for one-parameter Archimedean Copula family (See Cherubini, Luciano and Vecchiato 2004)

The concordance measures of the Frank copula require the computation of the so-called "Debye" functions, defined as (See Cherubini, Luciano and Vecchiato 2004)

$$D_k(\alpha) = \frac{k}{\alpha^k} \int_0^\alpha \frac{t^k}{\exp(t) - 1}dt, k = 1, 2 \tag{3.22}$$

The Marshall-Olkin Copula

Recall that Marshall-Olkin family is characterized by two parameters, m and n, belonging to $[0, 1]$. It is defined as follows (See Cherubini, Luciano and Vecchiato 2004):

$$C^{MO}(v, z) = \min(v^{1-m}z, vz^{1-n}) = \begin{cases} v^{1-m}z, & v^m \geq z^n \\ vz^{1-n}, & v^m < z^n \end{cases} \tag{3.23}$$

The family is positively ordered w.r.t. each parameter. As for its relationship with measures of concordance, we have(Nelsen, 1999):

$$\tau = \frac{mn}{m - mn + n}, \rho_s = \frac{3mn}{2m + 2n_m n} \tag{3.24}$$

The Marshall-Olkin copulas have upper tail dependence: (Embrechts, Lindskog and McNeil 2001)

$$\lambda_U = \min(m, n) \tag{3.25}$$

3.2.2 Non-Parametric Estimation

Tail Dependence Coefficient Estimators (TDC)

For a given bivariate random sample of length n $(X_1, Y_1), ..., (X_n, Y_n)$ from (X, Y), let $X_{(1)} \equiv \min\{X_1, ..., X_n\} \leq X_{(2)} \leq ... \leq X_{(n)} \equiv \max\{X_1, ..., X_n\}$ denote the corresponding order statistics. All of the relative non-parametric TDC-estimators $\hat{\lambda}_U$ of λ_U come from the non-parametric copula estimator (see Fisher and Dorflinger 2005):

$$C_n(j/n, j/n) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}(X_l \leq X(i), (Y_l \leq Y(j)) \tag{3.26}$$

Substituting the above into equation (3.2), will lead to the first two $\hat{\lambda}_U$ as follows:

$$\hat{\lambda}_U^{(1)} \equiv \frac{C_n((1 - k/n, 1] \times (2 - k/n, 1])}{1 - (1 - k/n)}$$

and

$$\hat{\lambda}_U^{(2)} = 2 - \frac{\log C_n(1 - k/n, 1 - k/n)}{\log(1 - k/n)}$$

where $k \approx \sqrt{n}$ (Dobric and Schmid, 2005). Then Dobric and Schmid (2005) interpret equations (3.2), (3.5) after suitable re-formulations as regression equations $C_n((1-i/n, 1] \times (1-i/n, 1]) = \lambda_U \cdot \frac{i}{n} + \varepsilon_i, i = 1, ..., k$, and drives $\hat{\lambda}_U^{(3)}$ as OLS-estimator, where ε_i is the error term in regression.

$$\hat{\lambda}_U^{(3)} = \frac{\lambda_U \cdot \frac{i}{n} + \varepsilon_i}{1 - (1 - k/n)},$$

In accordance to Dobric and Schmid (2005), another two estimators for λ_U are given by:

$$\hat{\lambda}_U^{(4)} = ArgMin_{\lambda \in [0,1]} \sum_{i=1}^{k} (C_n(1 - i/n, 1 - i/n) - (1 - i/n)^{2-\lambda})^2$$

and

$$\hat{\lambda}_U^{(5)} = ArgMin_{\lambda \in [0,1]} \sum_{i=1}^{k} (\ln C_n(1 - \frac{i}{n}, 1 - \frac{i}{n}) - (2 - \lambda) \cdot \ln(1 - \frac{i}{n}))^2 \tag{3.27}$$

One can prove that the above two estimators are equivalent by using the relationship $\ln(y^d) \approx 1 - y^d$ for $y \approx 1$ and $d \in [0, 1]$ We apply equation (3.27) in our later empirical analysis chapter as this seems to be the most robust and precise of them all.

Pearson's correlation

Suppose there are n measurements of X and Y, x_i and y_i, where $i = 1, ..., n$, then the sample correlation coefficient can be used to estimate the Pearson correlation r between X and Y as (See Cherubini, Luciano and Vecchiato 2004):

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \tag{3.28}$$

where $\bar{x}$ and $\bar{y}$ are the sample means of X and Y, s_x and s_y are the sample standard deviations of X and Y.

Kendall's Tau

Here is another equation to define τ. In order to estimate τ from a random sample of n pairs (X_i, Y_i), $i = 1, 2, 3..., n$ having defined the indicator variables

$$A_{ij} = sgn(X_i - X_j)(Y_i - Y_j) \tag{3.29}$$

as in Gibbons (1992), one can notice that

$$E(A_{ij}) = \Pr\left((X_i - X_j)(Y_i - Y_j) > 0\right) - \Pr\left((X_i - X_j)(Y_i - Y_j) < 0\right) = \tau$$

It follows that an unbiased estimator of Kendall's coefficient is the so-called Kendall's Sample τ (See Cherubini, Luciano and Vecchiato 2004):

$$\tau = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} A_{ij}$$

Spearman's Rho

Given a random sample of n pairs (X_i, Y_i), $i = 1, ..., n$, and recalling that ρ is the rank correlation coefficient, according to equation (3.14), one can switch to the ranks of the sample variates: $R_i \equiv rank(X_i)$ and $S_i \equiv rank(Y_i)$. The Spearman's sample ρ_s can be written as (See Cherubini, Luciano and Vecchiato 2004):

$$\rho_s = \frac{\sum_i^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_i^n (R_i - \bar{R}^2) \sum_i^n (S_i - \bar{S})^2}} \tag{3.30}$$

We will apply the formulas mentioned above in the Chapter 4.

Chapter 4

Empirical Analysis

In this chapter we will estimate the dependence measures non-parametrically and analyze the estimation results for a large variety of hedge fund strategies. As a complement, we will also have a look at one example of parametric estimation.

In the non-parametric estimation part, the empirical analysis will be discussed from two perspectives, the dependence measures between hedge fund strategies and the dependence measures within hedge fund strategies. The database used in this chapter lists 33 strategy types and 8977 companies. It measures the monthly returns of hedge funds starting in January 2001 until to December 2005. Moreover, 21 out of 33 strategies have been chosen in our empirical analysis due to the large number of companies using these strategies in the hedge funds industry (see Table 4.1). These 21 strategies represent roughly 85% of the market. The standard line to define "popularity" in this Chapter is whether or not the strategy is employed by more than 60 companies; it means that the strategy is chosen if more than 60 companies have adopted this strategy. Note that the strategies Equity Long/Short and the family of FOF strategies are by far the most popular accounting for almost 30% of the industry. For each paired time series, only those longer than 36 months (3 years) are selected as valid data, this is done to avoid small sample sizes and, therefore, to achieve better estimators of the chosen dependence measures.

More statistical information for the chosen hedge funds returns is given in Table 4.2, from which we have the expected value of hedge fund monthly returns for each chosen strategy and its standard deviation of the monthly returns. We can realize that strategies "Distressed Securities" and "Emerging Markets" lead the best performances overall in terms of mean returns with values of 0.0121 and 0.0166, while strategies "Relative Value Multi Strategy" and "FOF-Invest Funds in Parent Company" lead to the least variance and therefore financial risk of all with values 0.0003. We can also notice that the strategies "Merger Arbitrage" and "Equity Market Neutral" could have companies with very poor performances based on the minimum mean return among the companies in that strategy. Curiously, the best mean returns among strategies "Emerging Markets" are still very low while the worst mean returns in strategies "Merger Arbitrage" and "Equity Market Neutral" are not so negative (values around 0.10). Overall a useful measure of risk-return trade off is the Sharp Ratio, defined as the ratio of the mean

return and the standard deviation, in this regards the best performance was obtained from strategy CTA-Systematic/Trend-Following.

Later in this chapter, we are going to use the abbreviation for each hedge fund strategy listed in Table 4.1. Together with the short form names of hedge fund strategies, the number of companies that adopt the certain strategy are also given.

Table 4.1: Strategy Name and Size

Strategy Name Short Form	Strategy Name	Number of Companies
C.A.	Convertible Arbitrage	169
CTA-D	CTA- Discretionary	121
CTA-S/T-F	CTA- Systematic/ Trend- Following	558
D.S	Distressed Securities	112
E.M	Emerging markets	271
E.L.O.	Equity Long Only	76
E. L/S	Equity Long/Short	1276
E.M.N.	Equity Market Neutral	215
E.D. M.S.	Event Driven Multi Strategy	152
F I.	Fixed Income	108
F I.MB	Fixed Income-MBS	72
F. I. A.	Fixed Income Arbitrage	104
FOF-I F. P. C.	FOF- Invest Funds in Parent Company	69
FOF-M. N.	FOF- Market Neutral	65
FOF-M. S.	FOF- Multi Strategy	1205
FOF-S. S.	FOF- Single Strategy	218
G. M.	Global Macro	144
M. A.	Merger Arbitrage	93
M. S.	Merger Arbitrage	72
R. V. M. S.	Relative Value Multi Strategy	73
S	Sector	305

4.1 Non-Parametric Estimation Analysis

This section is divided into three parts, the analysis of companies within a given strategy, the analysis of companies between two given strategies and an application to portfolio theory of these findings. We use modern portfolio theory to compute and analyze the portfolio allocations resulting from employing the four different dependence measures.

4.1.1 Empirical Analysis within Hedge Fund Strategies

The empirical analysis within a hedge fund strategy studies the dependence structures between companies that invest in hedge fund by using the same strategy. In this part of empirical analysis, the measures of

	Mean	Variance	Minimum	Maximum
C A.	0.0066	0.0006	-0.2056	0.5200
CTA-D	0.0091	0.0040	-0.5748	0.5735
CTA-S/T-F	0.0070	0.0041	-0.7530	0.9798
D.S	0.0121	0.0011	-0.2620	0.4740
E.M	0.0166	0.0033	-0.5313	0.8176
E.L.O.	0.0086	0.0036	-0.4000	0.4747
E. L/S	0.0073	0.0024	-0.9300	1.2246
E.M.N.	0.0043	0.0007	-0.8200	0.2896
E.D. M.S.	0.0093	0.0012	-0.2065	0.8847
F I.	0.0066	0.0004	-0.1200	0.1210
F I.MB	0.0084	0.0005	-0.2702	0.3221
F. I. A.	0.0058	0.0004	-0.2500	0.2724
FOF-I F. P. C.	0.0068	0.0004	-0.0947	0.1049
FOF-M. N.	0.0052	0.0003	-0.1528	0.1148
FOF-M. S.	0.0061	0.0004	-0.8900	0.7900
FOF-S. S.	0.0065	0.0004	-0.3928	0.2620
G. M.	0.0081	0.0025	-0.5181	0.6227
M. A.	0.0026	0.0003	-0.1608	0.2517
M. S.	0.0077	0.0011	-0.2150	0.4080
R. V. M. S.	0.0072	0.0003	-0.2200	0.2443
S	0.0070	0.0034	-0.4106	0.6081

Table 4.2: Summary Table of Hedge Fund Monthly Returns: listed the mean, variance, minimum and maximum value of hedge fund returns for each strategy.

dependence between any two different companies within one strategy are computed and the inferential statistics results are shown and compared among the different strategies.

Table 4.3: Non-Parametric estimation within the Hedge Fund Strategies

Non-Para.	Tail Dep. Coef.	Spearman's	Kendall's	Pearson
C A.	0.2849	0.0897	0.3422	0.4533
CTA-D	0.1138	0.0519	0.0583	0.0704
CTA-S/T-F	0.1593	0.0208	0.1701	0.2334
D.S	0.2195	0.0733	0.3106	0.4007
E.M	0.1620	0.0346	0.2625	0.3506
E.L.O.	0.2855	0.1151	0.3345	0.4154
E. L/S	0.1387	0.0128	0.1692	0.2306
E.M.N.	0.1307	0.0470	0.0633	0.0750
E.D. M.S.	0.2132	0.0369	0.2661	0.3610
F I.	0.2458	0.1131	0.2231	0.2896
F I.MB	0.1911	0.1926	0.1719	0.1999
F. I. A.	0.1502	0.0799	0.1097	0.1406
FOF-I F. P. C.	0.3052	0.1066	0.3860	0.4723
FOF-M. N.	0.3314	0.0799	0.3906	0.4948
FOF-M. S.	0.2739	0.0428	0.3921	0.5250
FOF-S. S.	0.2013	0.0458	0.3181	0.4154
G. M.	0.1476	0.0478	0.1110	0.1493
M. A.	0.2573	0.0498	0.3046	0.4391
M. S.	0.1665	0.0601	0.1849	0.2305
R. V. M. S.	0.1959	0.0793	0.2189	0.2722
S	0.1670	0.0421	0.2315	0.2976

We estimate the four dependence measures within each selected strategy, as presented in Table (4.3). The numbers here indicate the mean values of each estimated dependence measures. This table shows that the Pearson correlation leads to the largest range and variability of values among strategies. These mean values are basically population values due to the fact that we are using the whole population of hedgefunds within each strategy in the given period. Still we also provide 95% confidence intervals for each coefficient based on assumptions of independence between pairs of companies ((X_1, X_2) independent of (X_3, X_4)) and therefore the use of the Central Limit Theorem. Even though the assumption of independence is questionable, the CI gives an idea of the range for the true mean and this is the purpose of building such a CI. Note the size of its C.I. is less than $1/10$ of the corresponding coefficient. As observed, the values go from 0.0704 for strategy CTA- Discretionary to 0.5250 for strategy FOF- Multi Strategy. This means those strategies show the lowest and highest Pearson correlations respectively among its constituents. Based on the confidence intervals, the mean Pearson correlation for CTA-Discretionary is greater than zero but still the companies in this strategy are in average curiously linearly independent from each other. On the other hand, the Spearman's has the lowest mean value and the smallest range and variety among strategies. It has a minimum of 0.0208 for strategy "CTA- Systematic/ Trend-Following" and a maximum of 0.1926 for strategy "Event Driven Multi Strategy". 80% of the strategies

give us significant low rank correlations (i.e. less than 0.1). This shows that according to Spearman's rho the companies using the same strategy are close to independent from each other too. Furthermore, like Pearson correlation, strategy "CTA-Discretinary" and "FOF-Multi Strategy" of Kendall's tau also give us the minimum and maximum value of 0.0519 and 0.3921 respectively. The highest coefficient of 0.392 for strategy "FOF-Multi Strategy" shows a difference between the probability of concordance and the probability of discordance for two companies is 39.2%. On the other hand, the lowest coefficient of 0.0519 of strategy "CTA-Discretinary" show the difference between the probabilities of concordance and discordance is 5.19%. The range of 95% the confidence interval is around 0.06 which is significantly small.

Taking a closer look at the estimation result for tail dependence. FOF- Market Neutral gives the highest tail dependence coefficient as 0.3314 meaning intuitively that 33.14% of companies in this strategy could default all together. While a relative low tail coefficient was obtained for CTA-discretionary 0.1138, but still shows a worrisome statistic that 11.38% of companies could default all together. Recall that among the four dependence measures, Spearman's rho, Kendall's tau and Pearson correlation all have the range of $[-1, 1]$; but tail dependence coefficient is in the range of $[0, 1]$, and because of this range, we compare the movements of Spearman's, Kendall's and Pearson separately from Tail dependence. In summary. Kendall's correlation gives the highest coefficient of 0.3921 and the lowest of 0.0583. However, Spearman's rho gives us the highest of 0.1926 and the lowest of 0.0128 while Pearson is between 0.0704 and 0.5250. There are some strategies indicating a consistency of these three dependence measures, those are CTA- Discretionary and Fixed Income-MBS. All other strategies show quite different behaviors among dependence measures. For example, companies within strategies "Equity Long/Short", "FOF-Market Neutral" and "FOF-Multi Strategy" are almost rank-independent based on Spearman while showing a strong linear dependence (Pearson); and the strategies of "FOF-Multi Strategy" and "FOF-Market Neutral" show a relative high dependency at the tail, as well. The strategy "Emerging Markets" have a low dependence shown at the tail but have a relatively high positive linear dependence among the companies.

Furthermore, among the 21 selected hedge fund strategies, four of them (FOF- Invest Funds in Parent Company, FOF- Market Neutral, FOF- Multi Strategy and FOF- Single Strategy) have a relatively high correlation coefficients and tail dependence coefficient (i.e. meaning high probability of large co-movements among their companies) indicating the commonality of their interest and handling of investments.

4.1.2 Empirical Analysis between Hedge Fund Strategies

In this section, we compute the four dependence measures between hedge fund strategies. We call the dependence coefficients critical if it is negative or if the 95% confidence interval for the mean of the population of pairs includes zero. By using the central limit theorem for sampling distribution , we can compute the 95% confidence interval for the means of the dependence measures (by assuming the independent distribution). Still note the CI is just as a proxy but in general the values obtained can be considered as those of the population. Next an analysis per dependence measure is provided.

Pearson Correlation Coefficients between hedge fund strategies

The empirical results of Pearson correlation coefficients are shown in Figure 4.1. We can observe that there are very few zero (i.e. 6 pairs of companies) or negative coefficients (i.e. 3 pairs of companies) in general as well as compared to other dependence measures. The majority of critical coefficient concentrate in Fixed Income-MBS, particularly on the pairs Fixed Income-MBS with CTA-Discretinary, with Emerging Markets, with Equity Long Only, Equity Long/Short, Merger Arbitrage and Sector . This implies hedge funds are positively linearly correlated with few valuable exceptions. We can also highlight that Pearson correlation between hedge funds is lower than within hedge funds (see Table 4.3), but in particular some pairs of strategies related linearly stronger than within strategies, for example, strategy Equity Market Neutral shows a linear-independent relationship among companies (i.e. small Pearson correlation =0.07) while the correlation coefficient between strategy "Equity Market Neutral" and "FOF-invest Funds in Parent Company" shows a relative stronger and positive linear dependency of 0.157. This phenomenon could be attributed to the similarity in the objectives and the adaptability of hedge funds which lead them to change strategies sometimes without reporting it. Another finding is that FOF- Multi Strategy gives significant high correlation coefficients between strategies, with a maximum value of 0.4376 between "FOF- Multi Strategy" and "FOF-Invest Funds in Parent Company", and coefficients greater than 0.3 between this strategy and 10 other strategies. Additionally, it shows the high correlation among "fund of funds" investment styles. It should be noticed that, when analyzing the population of pairs of companies from two given strategies, the 95% confidence intervals for the mean value of the Pearson correlation is very small (i.e. 0.008). This implies that the mean value for Pearson correlation between two strategies is statistically different from the mean value of any other pairs of strategies. There are also many cases of low but nonzero correlations that could be useful to investors, for example the correlation between FOF-Single Strategy and Fixed Income-MBS is positive low but non-zero of 0.0113.

Kendall's Rank Correlation Coefficient between Hedge Fund Strategies

The empirical results of Kendall's tau are shown in Figure (4.2). The majority of Kendall's tau correlation between strategies are positive similarly to Pearson. Only three pairs of hedge fund strategies give negative coefficients : strategy Fixed Income-MBS with Equity Long Only, strategy Fixed Income-MBS with Equity Long/Short, and strategy Fixed Income-MBS with Sector. Two pairs of hedge fund strategies give zero for Kendall's rank correlation coefficient, the only ones are: strategy Equity Long Only with CTA- Discretionary and strategy Fixed Income-MBS with Emerging markets. Similar with Pearson correlation, investors will not expect many negative rank correlations between between hedge fund strategies.

We can also state that Kendall's rank correlation between hedge funds is lower than within hedge funds in general (see Table 4.3). In particular some pairs of strategies related way stronger than within strategies, for example, strategy "FOF- Market Neutral" shows a relative significant concordant rela-

Pearson Corr	C.A.	CTA-D	CTA-S/T-F	D.S	E.M	E.L.O.	E.L/S	E.M.N.	E.D.M.S.	F.I.	F.I.MB	F.I.A.	FOF-I.F.P.C.	FOF-M.N.	FOF-M.S.	FOF-S.S.	G.M.	M.A.	M.S.	R.V.M.S.	S
C.A.	1																				
CTA-D	0.0457	1																			
CTA-S/T-F	0.0679	0.0565	1																		
D.S	0.2336	0.0208	0.0516	1																	
E.M	0.1393	0.0254	0.0179	0.2183	1																
E.L.O.	0.1535	**0.0009**	0.0351	0.2842	0.3147	1															
E.L/S	0.1160	0.0303	0.0124	0.2046	0.2363	0.2664	1														
E.M.N.	0.0563	0.0173	0.0400	0.0495	0.0378	0.0472	0.0534	1													
E.D.M.S.	0.2381	0.0344	0.0486	0.3357	0.2973	0.3218	0.2981	0.0983	1												
F.I.	0.2362	0.0332	0.0254	0.2422	0.2950	0.2286	0.2091	0.0595	0.2460	1											
F.I.MB	0.0455	**0.0087**	0.0432	0.0247	**0.0014**	-0.0689	-0.0439	0.0186	**0.0026**	0.0375	1										
F.I.A.	0.1631	0.0119	**0.0066**	0.1502	0.0879	0.0490	0.0756	0.0489	0.1122	0.0863	0.0801	1									
FOF-I.F.P.C.	0.2043	0.0324	0.1030	0.2939	0.2453	0.2033	0.2360	0.1557	0.2486	0.1912	0.0317	0.0592	1								
FOF-M.N.	0.3342	0.0555	0.0963	0.3562	0.2740	0.2052	0.2090	0.1125	0.2506	0.2057	0.0749	0.1167	0.3784	1							
FOF-M.S.	0.2773	0.0571	0.1021	0.3512	0.3358	0.3439	0.2940	0.1179	0.3817	0.2664	0.0130	0.1059	0.4376	0.4373	1						
FOF-S.S.	0.2485	0.0439	0.0794	0.2942	0.2720	0.3040	0.2854	0.1213	0.3152	0.2235	0.0113	0.0797	0.3773	0.3638	0.4226	1					
G.M.	0.1225	0.0418	0.0900	0.1456	0.1558	0.1452	0.1374	0.0632	0.1363	0.1212	0.0225	0.0445	0.1990	0.1730	0.1917	0.1764	1				
M.A.	0.1558	0.0287	0.0766	0.2166	0.2346	0.2748	0.2481	0.1115	0.2685	0.1898	**0.0065**	0.0566	0.2975	0.3113	0.3581	0.3346	0.1171	1			
M.S.	0.1595	0.0146	0.0747	0.2029	0.2021	0.1902	0.2021	0.1023	0.1858	0.1395	0.0323	0.0620	0.2761	0.2753	0.2534	0.2570	0.1351	0.1743	1		
R.V.M.S.	0.3132	0.0354	0.0233	0.2411	0.2216	0.1828	0.1703	0.0653	0.1843	0.1884	0.0677	0.1269	0.2155	0.2420	0.2586	0.3243	0.1151	0.1799	0.1989	1	
S	0.1343	0.0335	0.0292	0.2219	0.2508	0.3036	0.2473	0.0527	0.2742	0.1842	**-0.0210**	0.0482	0.2409	0.1992	0.2763	0.2791	0.1174	0.2321	0.1811	0.1428	1

Figure 4.1: Mean value of non-parametrically estimated Pearson Correlation between strategies. The bolded number without underlined shows the negative dependence coefficient and the underlined number shows its 95 % interval including zero. This table is generated by applying the formula equation 3.28

Kendall's Tau	C.A.	CTA-D	CTA-S/ D.S	E.M	E.L.O.	E. L/S	E.M.N.	E.D. M/F.I.	F.I.MB	F. I. A.	FOF-I.F	FOF-M	FOF-M	FOF-S	G.M.	M.A.	M.S.	R.V.MS
C.A.	1																	
CTA-D	0.0285	1																
CTA-S/T-F	0.0548	0.0361	1															
D.S	0.1602	0.0122	0.0383	1														
E.M	0.0906	0.0177	0.0166	0.1606	1													
E.L.O.	0.1119	**0.0081**	0.0304	0.2062	0.2098	1												
E.L/S	0.0682	0.0193	0.0116	0.1507	0.1671	0.1972	1											
E.M.N.	0.0396	0.0118	0.0289	0.0410	0.0303	0.0396	0.0442	1										
E.D. M.S.	0.1593	0.0227	0.0392	0.2437	0.2026	0.2328	0.2115	0.0714	1									
F.I.	0.1586	0.0196	0.0173	0.1801	0.1971	0.1568	0.1498	0.0512	0.1741	1								
F.I.MB	0.0514	0.0124	0.0388	0.0381	**0.0006**	-0.0479	-0.0280	0.0119	0.0088	0.0364	1							
F.I.A.	0.1048	0.0101	0.0101	0.1125	0.0613	0.0352	0.0428	0.0380	0.0751	0.0629	0.0606	1						
FOF-I.F.P.C.	0.1376	0.0270	0.0767	0.2076	0.1667	0.1441	0.1659	0.1115	0.1869	0.1220	0.0267	0.0422	1					
FOF-M.N.	0.2420	0.0372	0.0726	0.2577	0.1799	0.1322	0.1380	0.0922	0.1885	0.1336	0.0710	0.0755	0.2853	1				
FOF-M.S.	0.1841	0.0366	0.0830	0.2647	0.2360	0.2497	0.2066	0.0900	0.2374	0.1881	0.0103	0.0710	0.3287	0.2835	1			
FOF-S. S.	0.1636	0.0298	0.0613	0.2254	0.1915	0.2289	0.2105	0.0920	0.2294	0.1544	0.0121	0.0554	0.2835	0.1349	0.3108	1		
G. M.	0.0822	0.0265	0.0621	0.0988	0.1048	0.0957	0.0878	0.0441	0.0926	0.0823	0.0208	0.0218	0.2179	0.1184	0.1284	0.1193	1	
M.A.	0.1062	0.0201	0.0611	0.1608	0.1536	0.1800	0.1658	0.0813	0.1866	0.1231	0.0208	0.0442	0.2092	0.2128	0.2404	0.2338	0.1276	1
M.S.	0.1151	0.0163	0.0609	0.1606	0.1377	0.1508	0.1401	0.0756	0.1480	0.0948	0.0208	0.0437	0.1685	0.1975	0.1959	0.2045	0.0911	0.1276	1
R.V.M.S.	0.2215	0.0235	0.0223	0.1836	0.1522	0.1252	0.1176	0.0535	0.1308	0.1247	0.0208	0.0828	0.2291	0.1797	0.1713	0.2045	0.0859	0.1249	0.1438 1
S	0.0872	0.0209	0.0189	0.1613	0.1784	0.2306	0.1649	0.0399	0.1984	0.1241	0.0208	0.0089	0.1723	0.1361	0.1986	0.2086	0.0756	0.1527	0.1311 0.1029 1

Figure 4.2: Mean value of non-parametrically estimated Kendall's tau between strategies. The bolded number without underlined shows the negative dependence coefficient and the underlined number shows its 95 % interval including zero. This table is generated by applying the formula equation 3.29

tionship among companies (τ =0.3906) while the correlation coefficient between strategy "FOF-Market Neutral" and "CTA-Discretionary" shows a relative strong independency ($\tau = 0.0372$). strategy "FOF-Invest Funds in Parent Company" shows a relative significant concordant relationship among companies (τ =0.3860) while the correlation coefficient between strategy "FOF- Invest Funds in Parent Company" and "CTA-Discretionary" shows a relative strong independency ($\tau = 0.0372$). We did not find any pair of strategies where the dependence between them was stronger/larger than the dependence within the strategy.

Another finding is that, like Pearson correlation, FOF- Multi Strategy gives significant high correlation coefficients between strategies, with a maximum value of 0.3287 between "FOF- Multi Strategy" and "FOF-Invest Funds in Parent Company", and among all strategies only two coefficients are greater than 0.3 and both of them are associated with FOF- Multi Strategy. They are strategy "FOF-Invest Funds in Parent Company" and "FOF-Market Neutral". The rest pairs of Kendall's rank correlations are less than 0.3. It shows us the correlation between strategies generated by using Kendall's are less than Pearson's overall. It also should be noticed that, when analyzing the population of pairs of companies from two given strategies, the 95% confidence intervals for the mean value of the Kendall's tau is very small (less than 1/10 of the mean value). This implies that the difference of the mean values for Kendall's between any two strategies is statistically significant.

Spearman's Rank Correlation Coefficient between Hedge Fund Strategies

The empirical results of Pearson correlation coefficients are shown in Figure (4.3). It also gives the feature that the Spearman's rho between strategies is still smaller than within strategies in general. Here, we did find some pair of strategies where the Spearman's correlation between them was stronger/larger than the dependence within the strategy. (See figure 4.3). Again this happens to the strategy of Equity Long/Short and Strategy Distressed Securities. Also, unlike Pearson correlation and Kendall's tau, Spearman's rho gives a lot more negative (i.e. 27 pairs of companies) and zero values (i.e. 73 pairs of companies) with smaller variation. Particularly, strategy FOF- Invest Funds in Parent Company gives the most critical values (negative or zero) of correlation coefficient. It has positive rank correlations with only 2 strategies, FOF-Multi Strategy and FOF- Single Strategy with values of 0.0124 and 0.0132 respectively. However, comparing with Figure (4.3), the Spearman's rho of value 0.1066 within strategy for FOF-Invest Funds in Parent Company shows a certain level of positive rank dependency but a rank-independency with respect to other strategies. Additionally, strategies "FOF-Multi Strategy" and "FOF-Market Neutral" do not show us neither significant nor apparent rank-dependency between strategies like Pearson's or Kendall's do.

Moreover, among the non-zero Spearman's correlation coefficients, their absolute values are still very close to zero. In other words, the correlation coefficients skew to zero in general and show us that there is no significant association between most of the strategies in terms of their ranks, and we can trust these resulted mean values of Spearman's rank correlation coefficients within its 95% confidence interval with size of 0.03. This gives us an alternative view to the dependence structure as comparing to Pearson and Kendall's correlation, both giving us relatively high and positive coefficients.

Going back to the definition and assumptions involving these two dependence measures. Spearman's rho is the Pearson correlation coefficient of the ranking so of the transformed random variables $F(X)$ and $G(Y)$ where F and G represent the distributions functions of X and Y respectively. It is robust to outliers while Pearson's correlation is not. Pearson correlation holds under the assumption of linearity in the relationship, therefore, we should consider the Spearman's rank correlation a better proxy for nonlinear relationships between these strategies.

Tail Dependence Coefficient between Hedge Fund Strategies

One can observe the following statistical results about the tail dependence coefficients between strategies from Figure 4.4. Similarly with other dependence measures, the tail dependence coefficient between strategies is smaller than within strategies (see figure 4.3). This is reasonable because the probability of default companies adopting the same investment strategy should be higher than the probability of default of companies adopting different strategies. In general the numbers are quite stable showing and average tail dependence among all strategies of roughly 0.13. In particularly, strategy CTA-Discretionary has tail dependence coefficients less than 0.1 with all other strategy, while it is 0.1138 among its companies. "Relative value Multi Strategy" and "CTA-Systematic/Trend-Following" gives the minimum value of 0.0461, which means it has only 4.61% chance for these two strategies to default together, while they have the default probabilities of 15% and 19.59% respectively among their companies. The highest number in this coefficient matrix is 0.2341, generated between strategy FOF-Market Neutral and FOF-Multi Strategy, which is around the mean value of the tail dependence coefficients within those strategies. As illustrated in the previous section, CTA-Discretionary also has the lowest tail dependence coefficient within strategy, we may conclude that CTA-Discretionary strategy has feature of the lowest tail dependence.

Referring to tail dependence within strategies, one observes two strategies, namely FOF-Invest Funds in Parent Company and FOF-Market Neutral having relative high co-default probabilities among its companies. For the tail dependence coefficients between strategies, the relative high values are given among the strategies FOF-Invest Funds in Parent Company, FOF-Market Neutral and FOF- Multi Strategy. 5 pairs of strategies give probabilities of co-default greater than 20% among all selected strategies, and 3 of them are the combinations from these three strategies. Even though the tail dependence coefficient is lower between strategies than within strategy, these 3 strategies still indicate the co-movement feature beyond the tail. We also find some pairs of strategies where the dependence between them was stronger/larger than the dependence within the strategy. For example, the Strategy Equity Long/ Short and Distressed Securities shows the relatively stronger correlation again.

Spearman's rho	C.A.	CTA-D	CTA-S/T-F	D.S	E.M	E.L.O.	E.L/S	E.M.N.	E.D. M.S.	F.I.	F.I.MB	F.I.A.	FOF-I.F.P.C.	FOF-M.N.	FOF-M.S.	FOF-S.S.	G.M.	M.A.	M.S.	R.V. MS
C.A.	1																			
CTA-D	0.0143	1																		
CTA-S/T-F	0.0172	0.0148	1																	
D.S	0.0230	0.0160	0.0021	1																
E.M	0.0075	0.0149	0.0109	0.0170	1															
E.L.O.	0.1119	0.0200	0.0110	0.0158	0.0164	1														
E.L/S	-0.0034	0.0075	0.0041	0.0140	0.0118	0.0311	1													
E.M.N.	0.0014	0.0186	0.0167	0.0104	0.0067	0.0311	0.0107	1												
E.D. M.S.	0.0056	0.0095	0.0004	0.0179	0.0121	0.0159	0.0057	0.0034	1											
F.I.	0.0215	0.0231	0.0133	0.0258	0.0130	0.0203	0.0042	0.0184	0.0168	1										
F.I.MB	0.0818	0.0520	0.0297	0.0291	0.0082	0.0047	-0.0048	0.0212	0.0012	0.0173	1									
F.I.A.	0.0433	0.0261	0.0133	0.0219	0.0178	0.0000	-0.0057	0.0126	0.0038	0.0189	0.0741	1								
FOF-I.F.P.C.	-0.0100	0.0166	0.0181	0.0118	0.0118	0.0053	-0.0100	-0.0060	-0.0122	-0.0011	0.0134	-0.0129	1							
FOF-M.N.	0.0274	0.0152	0.0071	0.0179	0.0019	0.0086	0.0047	0.0115	-0.0028	0.0187	0.0220	0.0097	0.0126	1						
FOF-M.S.	-0.0278	-0.0215	-0.0101	-0.0076	-0.0041	0.0000	-0.0166	0.0042	-0.0083	-0.0609	-0.0274	0.0124	0.0057	0.0023	1					
FOF-S.S.	-0.0098	-0.0111	-0.0147	-0.0022	0.0016	0.0081	-0.0031	0.0077	0.0042	-0.0417	-0.0099	0.0132	0.0023	0.0209	0.0132	1				
G.M.	0.0197	0.0223	0.0137	0.0202	0.0075	0.0117	0.0078	0.0023	0.0052	0.0249	0.0133	-0.0039	0.0093	-0.0124	-0.0089	0.0209	1			
M.A.	-0.0038	-0.0021	-0.0081	0.0050	0.0013	0.0073	-0.0060	0.0052	-0.0031	-0.0202	-0.0007	0.0014	-0.0039	0.0262	0.0133	0.0053	0.0138	1		
M.S.	0.0055	-0.0041	-0.0111	0.0195	0.0098	0.0000	-0.0044	0.0014	0.0176	-0.0067	0.0111	-0.0286	0.0112	0.01	0.0046	0.0097	0.0106	0.0109	1	
R.V.M.S.	0.0442	0.0281	0.0178	0.0209	0.0036	-0.0109	-0.0112	0.0106	-0.0067	0.0145	0.0469	-0.0189	0.0126	-0.0056	-0.0171	-0.0079	-0.0016	0.0132	-0.0079	1
S	-0.0033	0.0190	0.0080	0.0170	0.0186	0.0287	0.0082	0.0173	0.0095	0.0241	0.0102	0.0082	0.0028	0.0104	0.0043	0.0152	0.0104	0.0049	0.0104	0.0000

Figure 4.3: Mean value of non-parametrically estimated Spearman's rho between strategies. The bolded number without underlined shows the negative dependence coefficient and the underlined number shows its 95 % interval including zero. This table is generated by applying the formula equation (3.30)

Tail Dep. Coef.	C.A.	CTA-D	CTA-S/t-F	D.S	E.M	E.L.O.	E. L/S	E.M.N.	E.D. M.S.	F.I.	F.I.MB	F.I. A.	FOF-I.F.P.C.	FOF-M. N.	FOF-M. S.	FOF-S. S.	G. M.	M. A.	M. S.	R. V. MS	S
C.A.	1																				
CTA-D	0.0779	1																			
CTA-S/t-F	0.0830	0.0849	1																		
D.S	0.1608	0.0608	0.0747	1																	
E.M	0.0901	0.0673	0.0682	0.0965	1																
E.L.O.	0.1300	0.0692	0.1026	0.1490	0.1519	1															
E. L/S	0.1077	0.0697	0.0717	0.1109	0.1145	0.1549	1														
E.M.N.	0.1175	0.0804	0.0766	0.0944	0.0802	0.1070	0.0956	1													
E.D. M.S.	0.1519	0.0658	0.0898	0.1808	0.1363	0.1704	0.1612	0.1226	1												
F.I.	0.1669	0.0817	0.0834	0.1430	0.1619	0.1802	0.1521	0.1166	0.1571	1											
F.I.MB	0.0975	0.0871	0.0756	0.0824	0.0651	0.0766	0.0716	0.0714	0.0819	0.0824	1										
F.I. A.	0.1311	0.0825	0.0615	0.1057	0.0863	0.0900	0.0963	0.1056	0.1052	0.1190	0.1216	1									
FOF-I.F.P.C.	0.1434	0.0807	0.0892	0.1719	0.1053	0.1086	0.1360	0.1252	0.1397	0.1190	0.0902	0.0794	1								
FOF-M. N.	0.2281	0.0928	0.0820	**0.2291**	0.1051	0.1376	0.1383	0.1314	0.1439	0.1593	0.0942	0.1088	0.2008	1							
FOF-M. S.	0.1381	0.0713	0.0929	0.1558	0.1285	0.1674	0.1497	0.1105	0.1901	0.1418	0.0636	0.0899	0.2191	0.2341	1						
FOF-S. S.	0.1404	0.0727	0.1371	0.1160	0.1362	0.1393	0.1130	0.1560	0.1299	0.0768	0.0909	0.1553	0.1722	0.1898	0.1975	1					
G. M.	0.0847	0.0855	0.1082	0.0940	0.0992	0.1156	0.1082	0.0950	0.1560	0.1067	0.0827	0.0844	0.1313	0.1212	0.1553	0.1084	1				
M. A.	0.0963	0.0708	0.0634	0.1157	0.0893	0.1092	0.1125	0.0922	0.1560	0.0943	0.0751	0.0822	0.1535	0.1669	0.1880	0.1723	0.0945	1			
M. S.	0.0935	0.0621	0.0817	0.1205	0.0851	0.1238	0.1027	0.1078	0.1901	0.0983	0.0773	0.0823	0.1393	0.1277	0.1181	0.1325	0.1019	0.1221	1		
R. V. M. S.	0.1915	0.0651	**0.0461**	0.1448	0.1143	0.1223	0.1105	0.1135	0.1208	0.1525	0.0985	0.1189	0.0956	0.1688	0.1302	0.1356	0.0858	0.1212	0.1077	1	
S	0.1128	0.0706	0.0757	0.1236	0.1178	0.1654	0.1368	0.0959	0.1539	0.1452	0.0874	0.0964	0.1367	0.1317	0.1460	0.1377	0.0977	0.1229	0.1088	0.1071	1

Figure 4.4: Mean value of non-parametrically estimated tail dependence coefficient between strategies. The bolded number shows the highest and smallest value of the tail dependence coefficient. This table is generated by applying the formula in equation (3.27)

4.1.3   Modern Portfolio Theory

Harry Markowitz introduced Modern Portfolio Theory in 1952, he aimed to maximize the portfolio expected return with a given portfolio risk. One variant of this approach, which is very robust when dealing with real data, is working with the minimum variance portfolio, therefore without targeting any portfolio mean return. This approach has shown to perform better on real data than a mean-variance approach due to relying on less parameters (only variances and correlations and no mean returns involved) and therefore avoiding some estimation risk (DeMiguel, Garlappi and Uppal 2006 and 2007). Other approaches avoid working with Pearson's correlations and uses more robust estimators instead (DeMiguel, Garlappi and Uppal 2009). In this section, we apply modern portfolio theory by minimizing our portfolio risk and using Kendall's, Spearman's and Tail dependence as the measures of dependence instead of Pearson. This would give a different perspective of portfolio allocation when considering nonlinear and extremal dependence measures.

First of all, suppose we have a portfolio P, which is composed by 21 different strategies of hedge funds. The portfolio returns would be given in the following way:

$$R_p = \sum_i^{21} w_i \cdot R_i \tag{4.1}$$

where R_i is the return of hedge fund i and w_i is the weighting of each type of hedge fund strategy i (that is , the share of hedge fund i in the portfolio). Recall $E(R_p)$ is the expected return of the portfolio, $E(R_i)$ is the expected return of hedge fund i:

$$E(R_p) = \sum_{i=1}^{21} w_i \cdot E(R_i) \tag{4.2}$$

Then, we attempt to minimize the portfolio risk, which means minimizing the portfolio variance σ_p^2, here σ_i and σ_j are the standard deviations of each hedge fund strategy. Note the expected return $E(R_i)$, σ_i and σ_j can be found in Figure 4.2

$$\sigma_{p,r}^2 = \sum_{i=1}^{21} \sum_{j=1}^{21} w_i w_j \sigma_i \sigma_j r_{ij}, \tag{4.3}$$

$$\sigma_{p,\tau}^2 = \sum_{i=1}^{21} \sum_{j=1}^{21} w_i w_j \sigma_i \sigma_j \tau_{ij}, \tag{4.4}$$

$$\sigma_{p,\rho}^2 = \sum_{i=1}^{21} \sum_{j=1}^{21} w_i w_j \sigma_i \sigma_j \rho_{ij}, \tag{4.5}$$

$$\sigma_{p,\lambda}^2 = \sum_{i=1}^{21} \sum_{j=1}^{21} w_i w_j \sigma_i \sigma_j \lambda_{ij}. \tag{4.6}$$

Here r_{ij} is the Pearson's correlation coefficient, τ_{ij} is the Kendall's rank correlation coefficient, ρ_{ij} is the Spearman's rho and λ_{ij} is tail dependence coefficient between the returns of hedge fund strategies i and j; and $r_{ij} = \tau_{ij} = \rho_{ij} = \lambda_{ij} = 1$ for $i = j$. Therefore, the coefficient matrix we used here are the four matrices that are listed in the previous section, you may find them in Figures 4.2, 4.3 and 4.4. It should be mentioned that all these matrices are valid correlations matrices as they are positive definite matrices (a fact that can be checked numerically).

Applying the above formula and using Matlab build in code frontcon(), we can generate the minimized portfolio standard deviation and its expect returns as follows:

- $\sigma_p = 0.005675$, $E(R_p) = 0.008373$ for Pearson Correlation Coefficient.

- $\sigma_p = 0.005755$, $E(R_p) = 0.007883$ for Kendall's Rank Correlation Coefficient.

- $\sigma_p = 0.006317$, $E(R_p) = 0.005846$ for Spearman's Rank Correlation Coefficient.

- $\sigma_p = 0.005592$, $E(R_p) = 0.008064$ for Tail Dependence Coefficient.

One can observe that Spearman's coefficient gives the highest risk and lowest return (so lowest Sharpe Ratio). For more information, we generate portfolio weighting table displayed as bars in Figure (4.4) and its graph as lines in Figure (4.5). Spearman's rank correlation coefficient scattered the weights into all 21 hedge fund strategies more evenly as it can be seen in the variances of the weights provided next (smallest variance of the weights). On the contrast, Pearson correlation coefficient set the weights concentrated merely in a minority of hedge funds strategies, making it also the least diversified of all 4 allocation strategies. In this case only 4 hedge fund strategies take the majority of the portfolio weights when consider Pearson correlation. In particular the strategy Merger Arbitrage takes the largest proportion with 37.7% of the portfolio; and strategy Fixed Income-MBS have the second largest weights with 18.7%. 11 out of 21 strategies have zero weights or less than 1%.

On the other hand there is no zero weights generated by Spearman's coefficients. Strategy Merger Arbitrage still take the largest proportion in the portfolio (i.e. but it shrinks to 20.6%) for Spearman's rho. The weights variation for Kendall's and tail dependence are in between those of Pearson and Spearman. There are seven zero weights generated by using tail dependence correlation, and 6 zeros generated by Kendall's tau. The largest proportions for using both of these two methods is located at Merger Arbitrage Strategy as well with values of 32.9% (Kendall's) and 33.4% (Tail Dependence). The second largest weighting is concentrated at strategy Fixed Income-MBS for both Kendall's and Tail Dependence methods, which is the same as using Pearson's. However, comparing to Kendall's with tail dependence correlation (TDC), TDC allows the rest of the weights to be allocated much evenly in a few strategies (i.e. roughly 6 strategies), See Figure (4.5).

We can illustrate this phenomenon in terms of variation of the weighting shown as following and see the Figure (4.6). For all four cases, one can observe that strategies Merger Arbitrage take largest weighting, but Fixed Income-MBS varies the most as well as strategy FOF-Market Neutral. Fixed Income-MBS take the second largest proportion except for Spearman. It shrinks 60% of weights proportion from 18.7 % (Pearson) to 7.6% (Pearson). On contrast, strategies Fixed Income, FOF- Invest

Funds in Parent Company, Merger Arbitrage and Relative Value Multi Strategy are relatively stable. Summarizing, Pearson lead to the least diversified portfolio while Spearman aim at the most diversified.

- Variance of portfolio weights for Pearson is 0.0082.

- Variance of portfolio weights for Spearman's is 0.0023.

- Variance of portfolio weights for Kendall's is 0.0061.

- Variance of Portfolio weights for Tail is 0.0060.

- l^2-norm of portfolio weights for Pearson is 0.3507.

- l^2-norm of portfolio weights for Spearman's is 0.2763.

- l^2-norm of portfolio weights for Kendall's is 0.3282.

- l^2-norm of Portfolio weights for Tail is 0.3245.

Table 4.4: Portfolio Allocations, with minimized risk (standard deviation)

| Weights	Pearson	Kendall's	Spearman's	Tail Dep. Coef
C A.	0.018	0.028	0.042	0.024
CTA-D	0.011	0.010	0.003	0.000
CTA-S/T-F	0.003	0.003	0.005	0.000
D.S	0.000	0.000	0.027	0.003
E.M	0.000	0.000	0.006	0.000
E.L.O.	0.000	0.000	0.006	0.000
E. L/S	0.000	0.000	0.013	0.000
E.M.N.	0.048	0.046	0.032	0.023
E.D. M.S.	0.000	0.000	0.019	0.000
F I.	0.031	0.040	0.041	0.033
F I.MB	0.187	0.164	0.078	0.129
F. I. A.	0.101	0.095	0.054	0.071
FOF-I F. P. C.	0.078	0.075	0.095	0.100
FOF-M. N.	0.007	0.033	0.076	0.056
FOF-M. S.	0.000	0.022	0.085	0.050
FOF-S. S.	0.021	0.038	0.078	0.060
G. M.	0.003	0.008	0.013	0.000
M. A.	0.377	0.329	0.206	0.334
M. S.	0.000	0.000	0.022	0.000
R. V. M. S.	0.115	0.109	0.090	0.116
S	0.000	0.000	0.007	0.000
Weight SUM	1.000	1.000	1.000	1.000
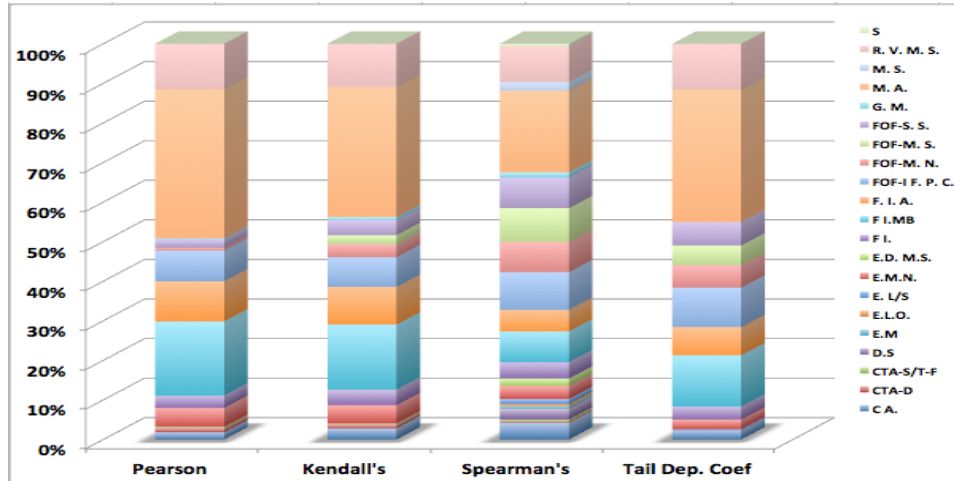
Figure 4.5: Portfolio Stacked Weighting Graph in 3D: colored stack indicates different hedge fund strategies.
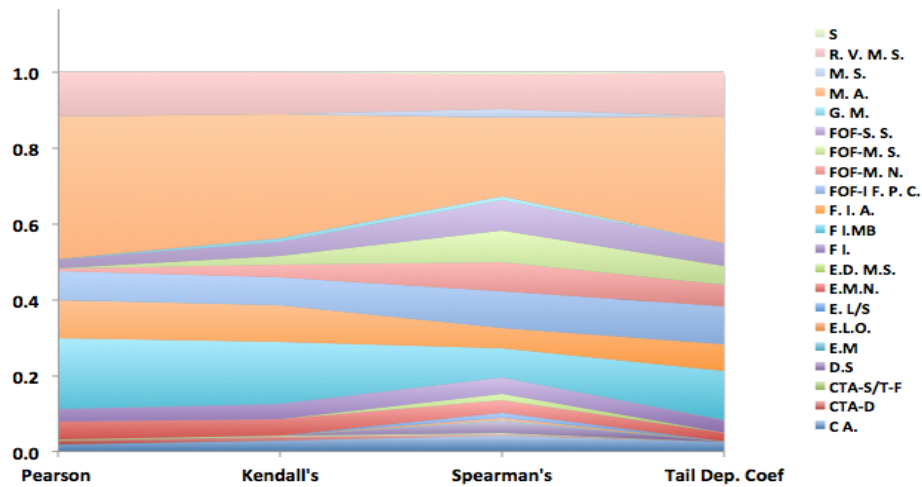

Figure 4.6: Variation of Portfolio Weighting for different correlation coefficient.

4.2   Parametric Estimation Analysis

In this section, we provide an example of parametric estimation results assuming t-student marginals and a Clayton copula. Recall the estimation method introduced in Chapters (2.4.1) and (3.2.1) with equations (4.7) and (4.8), we used the bivariate case of the IFM method.

1. First, we estimate the marginals' parameters θ_1 by performing the estimation of the univariate marginal student t distributions:

$$\widehat{\theta_1} = ArgMax_{\theta_1} \sum_{t=1}^{T} [\ln f_1(x_{1t}; \theta_{1_1}) + \ln f_2(x_{2t}; \theta_{1_2})] \tag{4.7}$$

where $f_1(x_{1t}; \theta_{1_1}$ and $f_2(x_{2t}; \theta_{1_2}$ are the pdf of student t distribution, θ_{1_1} and θ_{1_2} are the parameter vectors.

2. Secondly, given the student's t parameters $\widehat{\theta_1}$, we perform the estimation of the Clayton copula (Archimedean Copula Family) parameter θ_2 using Matlab:

$$\alpha = \hat{\theta}_2 = ArgMax_{\theta_2} \sum_{t=1}^{T} \ln c(F_1(x_{1t}), F_2(x_{2t}); \theta_2, \hat{\theta}_1) \tag{4.8}$$

This combination of marginals and copula are popular due to its stylized features (for the marginals) and thee simplicity (for the copula) but in general there is no consensus on this joint model, particularly on the best copula for financial data. Moreover, unfortunately the Clayton copula didn't give results consistent enough with the non-parametric estimations. In part because most copulas underestimate the tail dependence and Kendall's correlations (see Escobar, Frielingsdorf and Zagst 2011) and also the length of the paired time series is not long enough to allow for good estimations.

Table (3.2.1) shows the within strategies estimation results from the nonparametric and parametric approaches respectively. We can observe several estimation consistency, such as strategies CA, .M. A. Sector for Tail Dependence; and FOF-M. S. FOF-S. S. for Kendall's tau. The rest of the strategies are all underestimated. Kendall's tau gives the largest difference between parametric estimation and non-parametric estimation. The difference for Kendall's varies from 0.0009 to 0.3515, while it varies from 0.0023 to 0.2744 for tail dependence coefficient. Because of the underestimation issue we did not adopt the parametric estimation results in portfolio optimization section.

Table 4.5: Parametric Estimation VS. Non-Parametric Estimation for TDC and Kendall's

non-para/para	Tail Dep. Coef.		Kendall's	
C A.	0.2849	0.2826	0.3422	0.1971
CTA-D	0.1138	0.0327	0.0583	0.0360
CTA-S/T-F	0.1593	0.0912	0.1701	0.2321
D.S	0.2195	0.0896	0.3106	0.0761
E.M	0.1620	0.0613	0.2625	0.0641
E.L.O.	0.2855	0.2232	0.3345	0.1559
E. L/S	0.1387	0.0724	0.1692	0.0881
E.M.N.	0.1307	0.0281	0.0633	0.0343
E.D. M.S.	0.2132	0.1587	0.2661	0.1232
F I.	0.2458	0.0810	0.2231	0.062
F I.MB	0.1911	0.0478	0.1719	0.0542
F. I. A.	0.1502	0.0267	0.1097	0.0330
FOF-I F. P. C.	0.3052	0.0308	0.3860	0.0345
FOF-M. N.	0.3314	0.2616	0.3906	0.1883
FOF-M. S.	0.2739	0.2191	0.3921	0.3902
FOF-S. S.	0.2013	0.1859	0.3181	0.3052
G. M.	0.1476	0.1194	0.1110	0.0817
M. A.	0.2573	0.2301	0.3046	0.2793
M. S.	0.1665	0.1232	0.1849	0.1428
R. V. M. S.	0.1959	0.1617	0.2189	0.1858
S	0.1670	0.1542	0.2315	0.2198

Chapter 5

Conclusions

The main objective of this work is the analysis of the dependence structure of the returns of hedge funds' strategies with respect to three other dependence measures of Kendall's tau, Spearman's rho and tail dependence coefficient beyond the commonly used Pearson correlation. We also suggest alternative approaches to an effective optimization when combining various hedge fund strategies based on using different dependence measures. This involves the estimation of these three dependence measures parametrically through the concept of copula and no parametrically without constraining for any particular model. Ideally, the parametric results should be consistent with the non-parametric estimation. However, due to at least two reasons the parametric estimators perform poorly, one being the size of our paired time series sample which is not significantly large enough (most of the sample size are less than 50) affecting the accuracy of the estimation results. Secondly and more importantly, any copula chosen implies a rigid structure on its dependence measures constraining their real values. Therefore some copulas allow for specific measures to be captured better than others, this together with the lack of a clear favourite copula in the hedge fund's literature limits the usefulness of parametric estimations. As a result of these two points, we only focus on the non-parametric estimation results.

For the non-parametric estimation within hedge fund strategies, we found that the strategies "FOF-Invest Funds in Parent Company" and "FOF- Market Neutral" have the highest tail dependence and therefore higher probabilities of companies defaulting simultaneously than other strategies. The Pearson correlation coefficients show the largest variation and higher values among the strategies for all four measures of dependence considered. Additionally, in the analysis between strategies, the correlations coefficients for the four dependence measures are smaller than within strategies as expected. Spearman's rho shows the smallest variation and the weakest association between strategies, since all its coefficients are skewed to zero and within the value -0.1 0.1. Also, in general the numbers are quite stable showing and average tail dependence among all strategies.Furthermore, the cases of FOF- Invest Funds in Parent Company , FOF- Market Neutral and FOF- Multi Strategy show a significant large and positive association, greater than any other groups for all measures, but except for Spearman. Curiously, Spearman's correlation gives approximately zero values among these strategies. This point out that Pearson corre-

41

lation coefficient is not the only measure of dependence that should be considered and therefore makes questionable any analysis based solely on Pearson.

Relying on the empirical covariance matrixes obtained with non-parametric estimation between strategies, we provide four alternative solutions to the multiple strategies hedge funds portfolio optimization. Spearman's rho generates the smallest, least varied correlations, in addition the weights of the portfolio were spread out among all chosen 21 hedge fund strategies more evenly leading to higher diversification. However, for Pearson's correlation, the portfolio weights were concentrated on 5 out of 21 hedge fund strategies leading to the least diversified portfolio. Poorly diversified portfolios are known in the "estimation risk" literature to perform worst out of sample. Although Pearson still gives the highest expected return and smallest variance, the reality is that the Pearson's correlation coefficient do not fully reflect the dependence between hedge funds strategies. The investors should be aware of this fact and make adjustment referring to other alternative dependence measures. Furthermore, among four of the dependence measures, Kendall's rank correlation give the most moderate solution to optimization of the portfolio, since it has the same variation trend with Pearson's correlation, but not as most diversified as Pearson's. Another point that should be brought to our attention is that the tail dependence coefficient also provide a relatively high expected return and small variance for the portfolio so when an investor encounters extreme market conditions, optimization of portfolio for tail dependence should be considered.

Putting the results into a broader perspective, dependence measures in the world of hedge funds lead to diverse values as diverse as hedge funds itself so there is almost no homogeneous behaviour within or between strategies. There are many ways and perspectives to interpret dependence measures and each one gives a feature of the full copula structure. In real life, investors should consider different dependence measures, when investing in the large existing variety of strategies and styles, in order to achieve an optimal way to create a funds of hedge funds.

References

[1] Abdi, H. (2007) *" The Kendall rank correlation coefficient."* The University of Texas at Dallas, USA,pp. 2-7

[2] Alvarez, M. and Levinson, M. (2007) *" Hedge Fund Risk Modeling"* MSCI Barra, MODEL IN-SIGHTS, pp. 2-27

[3] Brown, S. J. and Goetzmann, W. N.(2001) *"Hedge Funds with Styles"* Working paper,National Bureau of Economic Research. pp. 2-33

[4] Cherubini, U. and Luciano, E. and Vecchiato, W. (2004). *"Copula methods in finance"*. The Wiley Finance Series. England. pp. 24- 164

[5] Cizek,P., Hardle, W. and Weron, R. (2005). *"Statistical tools for finance and insurance"*. Springer.

[6] Conover, W. J. (1980). *"Practical nonparametric stratistics"*.Wiley(New York)

[7] Connor, G and lasarte, T. (2010) *"An overview of Hedge Fund Strategies"*. London School of Economics and Lasarte at Standard Life Investment Management

[8] DeMiguel, V., Garlappi, L., Uppal, R. (2006) *"Implementation Details and Robustness Checks"*. London Business School, University of Texas at Austin and London Business School, working paper.

[9] DeMiguel, V., Garlappi, L., Uppal, R. (2007). *"Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy"*. London Business School, University of Texas at Austin and London Business School, Working paper.

[10] DeMiguel, V., Garlappi, L., Nogales, J. R., Uppal, R. (2009) *"Implementation Details and Robustness Checks"*. Anppendix to *"Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?"* London Business School, University of Texas at Austin and London Business School, Working paper.

[11] Dorflinger, M. and Fischer, M. *"Hedge A note on a Non-Parametric tail dependence estimator"* Working paper,University of Erlangen-Nurnberg, Germany.

[12] Easterling, Ed. (2007) *"Hedge Funds: Myths and Facts"* CRESTMONT RESEARCH

[13] Escobar, M., Frielingsdorf, T., Zagst, R. (2010). *"Impact of factor models on portfolio risk measures. A structural approach"*. Ryerson Applied Mathematics Laboratory. Technical Report.

[14] Fung, W. and Hsieh, D. A. (1997) *"Emperical Characteristics of Dynamic Trading Strategies: Case of Hedge Funds"* Oxford Jornals.

[15] Nelsen, R. B. (1999) *"An Introduction to Copulas"*, New York: Springer, ISBN 0387986235.,

[16] Salkind, N. (2007) *"Encyclopedia of measurement and Statistics"* The University of Texas at Dallas, USA.

[17] Seco, L.A. *"Hedge Funds: Truths and Myths"* Department of Mathematics University of Toronto. (3-4)

[18] Sklar, A. (1959) *"Fonctions de repartition a n dimensions et leurs marges"* , Publ. Inst. Statist. Univ. Paris 8: 229231,

[19] Suppal, K. CFA (2004) *"Constructing multi-strategy fund of hedge funds"* Simon Fraser University.

[20] Trudel, David (2008) *"Tail Dependence of Hedge Funds"* . Swiss Federal Institute of Technology Zurich. (98-99) (4-6)

[21] Peskin, M. W. , Urias M. S., Anjilvel, S. I. and Boudreau B. E. (2000) *"Why Hedge Funds Make Sense"* Quantitative Strategies, Morgan Stanley Dean Writer

[22] Prokhorov, A.V. (2001) *"Kendall coefficient of rank correlation"*, in Hazewinkel, Michiel, Encyclopaedia of Mathematics, Springer, .

[23] Yan, J. *"Enjoy the Joy of Copulas"*,Department of Statistics and Actuarial Science, University of Iowa.

Appendix A

Comments on Code

In this chapter, I summarize the coding work created in MATLAB for my thesis. This was the most time consuming part of this work and took me around 6-8 months from manipulating the data to outputting the empirical results. Roughly, the coding contains 2 main components corresponding to chapter 4 (Empirical Analysis), the non-parametric estimation and the parametric estimation. There were around 80 functions created in total from cleaning, organizing the data to performing within and between strategies analysis for the four measures of dependence via parametric and non-parametric. Before coding the estimation parts, the first challenge I encountered was to manipulate and organize the raw data.

The original data was written in one spread sheet of Excel document. It had 234,309 monthly returns (Jan. 2001- Dec. 2005) characterized by 33 different hedge fund strategies and 8,977 financial companies listed in one column. As we known for MATLAB, it is easy to import data from excel using the build-in function called "importdata()". However, after importing the data, we have two separated data sets. One is constructed by "strings (text)", what we had in there was the matrix illustrating only the strategy names with the number part as "NA". The other one was constructed by "data (numbers)", and what we have in here is the matrix illustrating only the returns, the date time and the company IDs with the text part as "NA". We needed the time series of the returns with the same strategy and same company in a matrix in order to perform most of the calculations. As a result, the core idea was to let MATLAB read and pick the time series of returns with the same strategy from the "data" set, but the "starting and ending points" were recognized from the "text" set. For the convenience of future use, I create a big cell of matrices of time series of returns, where its rows indicating the strategy types and its columns indicating the company ID. In other words, every matrix in the same row of the cell corresponds to the same type of hedge fund strategy with other matrices in the same row. For example, the cell{1,1} indicates the time series of returns with strategy "Capital Structure Arbitrage" and company ID 1; the cell{1,2} indicates the time series of returns with strategy "Capital Structure Arbitrage" and company ID 2.

After having the matrices of time series of hedge fund returns (for each strategy and company) ready, it was not difficult to pair and match the time series to another one. Next was to implement the equations

and formulas accordingly. Additionally, the most time consuming function to run in MATLAB was the function of tail dependence coefficient estimation in general. Certainly, it depended on the sample size of corresponding paired time series and the longest one takes almost 12 hours even after upgrading and optimizing the code efficiency. Next the full code is provided:

```
%%paired all data for one stratege
function c_timeSeries=SortData_PairedAll(str,textdata,data)
i=2;
k=1;
j=1;
types=1;
data_nums_eachFundStr=[];
C_data={};
C_textdata={};
L=length(data(:,1));
%%sort the data by same fundStr, output the data size for each funStr and

for i=2:length(textdata(:,2))-1
TF = strcmp(textdata(i,2), textdata(i+1,2));
j=j+1;
data_nums_eachFundStr(types)=j-1;

    if TF==0
        C_textdata{types}=textdata(k+1:k+j-1,:);
        C_data{types}=data(k:k+j-2,:);
        types=types+1;
        k=i;
        j=1;

    end

i=i+1;
end
C_textdata{types}=textdata(k+1:k+j,:);
C_data{types}=data(k:k+j-1,:);

%%sort the data by funID within the same fundStr
i=1;
j=1;
ID={};
```

```
C_data_ID={};
C_textdata_ID={};
for i=1:types
m=0;
ID{i}=unique(C_data{i}(:,1));
    for j=1:length(ID{i}(:,1))
      k=count(C_data{i},ID{i}(j));
      C_data_ID{i,j}=C_data{i}(m+1:m+k,:);
      C_textdata_ID{i,j}=C_textdata{i}(m+1:m+k,:);
      m=m+k;

    end
end


%%sort into the same time series (for two different funds using the same
%%trading strategy. i.e., C_data_ID{i,j1} and C_data_ID{i,j2}
%%given ID1,ID2,str

function [timeSeries]=match2ID(ID1,ID2,str,cell_name)
timeSeries=[];
i=1;
j=1;
m=1;
l1=length(cell_name{str,ID1}(:,1));
l2=length(cell_name{str,ID2}(:,1));
for i=1:l1
    for j=1:l2
        if cell_name{str,ID1}(i,4)==cell_name{str,ID2}(j,4)
         timeSeries(m,1)=cell_name{str,ID1}(i,4);
         timeSeries(m,2)=cell_name{str,ID1}(i,5);
         timeSeries(m,3)=cell_name{str,ID2}(j,5);
         m=m+1;
        end
    end
end
    end
%%for all IDs within one strategy
%%count_cell(c_data_ID,str)=# of matrix in a row
function k= count_cell(cell_name,row_num)
```

```
i=1;k=0;
while  i<=length(cell_name(1,:)) && i~=0
    if  ~isempty(cell_name{row_num,i})
        k=i;
        i=i+1;
    else
        i=0;

    end
end
end
I=1;J=1;c_timeSeries={};
for  I=1:count_cell(C_data_ID,str)
    for  J=I:count_cell(C_data_ID,str)
        c_timeSeries{I,J}=match2ID(I,J,str,C_data_ID);
    end

end
end




function  [Tailmean,  Tailstd,Tailpercen]=StatTailEstimation(M,observed,Min)

lamda=[];
k=1;
for  i=1:observed
    for  j=i+1:observed
        if  ~isempty(M{i,j}) && length(M{i,j}(:,1))>=Min
    lamda(k)=fmincon(@(LAMDA)tailEstimation(M{i,j},LAMDA),0,0,1);
    %reduce the  negative  value  to  0
    if  lamda(k)<0
        lamda(k)=0;
    elseif  lamda(k)>1
        lamda(k)=1;
    end
    k=k+1;
```

```
            end
        end
end
% STEP 1 − rank the data
if length(lamda)>1
y = sort(lamda);
MIN=min(lamda)
MAX=max(lamda)
Tailmean=mean(lamda)
Tailstd=std(lamda)
SS=length(lamda)
hist(lamda);
end
end




function [SpearmanMean,SpearmanStd,SpearmanPercen]=StatSpearman(M,observed,Min)

rho=[];
k=1;
for i=1:observed
    for j=i+1:observed
        if ~isempty(M{i,j}) && length(M{i,j}(:,1))>=Min
            rho(k)=spearmanSample(M{i,j});
            k=k+1;
        end

    end
end
% STEP 1 − rank the data
y = sort(rho);
MIN=min(rho)
MAX=max(rho)
SpearmanMean=mean(rho)
SpearmanStd=std(rho)
SS=length(rho)
hist(rho)
```

```
end


function  [PearsonMean , PearsonStd , PearsonPercen]=StatPearson (M, observed , Min)


corr =[];
k=1;
for  i =1:observed
     for  j=i+1:observed
     if  ~isempty (M{i ,j}) && length (M{i ,j }(:,1)) >=Min
         corrMatrix=corrcoef (M{i ,j }(: ,2) ,M{i ,j }(: ,3));
         corr (k)=corrMatrix (1 ,2);
         k=k+1;
     end
     end
end
% STEP  1 − rank  the  data
y  =  sort (corr );
MIN=min( corr )
MAX=max( corr )
PearsonMean=mean( corr )
PearsonStd=std ( corr )
SS=length ( corr )
hist ( corr );
end


function  [KendalMean , KendalStd , KendalPercen]=StatKendals (M, observed , Min)


tau =[];
k=1;
for  i =1:observed
     for  j=i+1:observed
     if  ~isempty (M{i ,j}) && length (M{i ,j }(:,1)) >=Min
         tau (k)=kendalsample (M{i ,j });
         k=k+1;
     end
     end
end
```

```
% STEP 1 − rank the data
y = sort(tau);
MIN=min(tau)
MAX=max(tau)
KendalMean=mean(tau)
KendalStd=std(tau)
SS=length(tau)
hist(tau);
end


function f=tailEstimation(matrix,LAMDA)

l_matchedtime=length(matrix(:,1));
estF=[];
for j=1:round(sqrt(l_matchedtime))
%%copula:

n=0;
if l_matchedtime >1
sorted=sort(matrix(:,2:3));
index=l_matchedtime−j;
X_index=sorted(index,1);
Y_index=sorted(index,2);

 for i= 1: length(sorted)
   if matrix(i,2)<=X_index && matrix(i,3)<=Y_index
    n=n+1;
   end
 end
  copula(j)=n/l_matchedtime;
estF(j)=(log(copula(j))−(2−LAMDA)*log(1−j/l_matchedtime))^2;
end


end
f=sum(estF);
end
```

```
%%matrix is the timeSeries (1st col= time; 2nd col= ID1; 3rd col= ID2)
function tau=kendalsample(matrix)

l_time=length(matrix(:,1));
i=1;
A_ij=0;
function A12=sgn(x1,y1,x2,y2)

   if x1<x2 && y1<y2
     A12=1;
   elseif   x1>x2 && y1>y2
         A12=1;
   elseif x1==x2  ||  y1==y2
             A12=0;
   else
             A12=-1;
   end

end

for i=1:l_time −1
     for j=i+1:l_time
        a_ij= sgn(matrix(i,2),matrix(i,3),matrix(j,2),matrix(j,3));
        A_ij=A_ij+a_ij;
     end
end
tau=2*A_ij/(l_time*(l_time −1));



end


%C_data_ID=a cell such that {n,m}, where n=str number and modified ID
%number, In our example we have 32 different strategies.
i=2;
k=1;
j=1;
```

```matlab
types=1;
data_nums_eachFundStr=[];
C_data={};
C_textdata={};
L=length(data);
%%sort the data by same fundStr, output the data size for each funStr and
%%the number of the fundStr unique(A)
for  i=2:length(textdata(:,2))-1
TF = strcmp(textdata(i,2), textdata(i+1,2));
j=j+1;
data_nums_eachFundStr(types)=j-1;

    if TF==0
        C_textdata{types}=textdata(k+1:k+j-1,:);
        C_data{types}=data(k:k+j-2,:);
        types=types+1;
        k=i;
        j=1;

    end

i=i+1;
end
C_textdata{types}=textdata(k+1:k+j,:);
C_data{types}=data(k:k+j-1,:);

%%sort the data by funID within the same fundStr
i=1;
j=1;
ID={};
C_data_ID={};
C_textdata_ID={};
for  i=1:types
m=0;
ID{i}=unique(C_data{i}(:,1));
    for  j=1:length(ID{i})
      k=count(C_data{i},ID{i}(j));
      C_data_ID{i,j}=C_data{i}(m+1:m+k,:);
      C_textdata_ID{i,j}=C_textdata{i}(m+1:m+k,:);
```

```
    m=m+k;


    end
end


function n=count(array,number)
n=0;
i=1;
l=length(array);
for i=1:l
if number==array(i)
    n=n+1;
end

end
end


function k= count_cell(cell_name,row_num)
i=1;k=0;
while i~=0
    tf=isempty(cell_name{row_num,i});
    if tf==0

        k=i;
        i=i+1;
    else
        i=0;
    end
end
end




function c_timeSeries=btStr_SortData_PairedAll(str1,str2,textdata,data)
```

```matlab
%%paired all data for one stratege


i=2;
k=1;
j=1;
types=1;
data_nums_eachFundStr=[];
C_data={};
C_textdata={};
L=length(data(:,1));
%%sort the data by same fundStr, output the data size for each funStr and


for i=2:length(textdata(:,2))-1
TF = strcmp(textdata(i,2), textdata(i+1,2));
j=j+1;
data_nums_eachFundStr(types)=j-1;


    if TF==0
        C_textdata{types}=textdata(k+1:k+j-1,:);
        C_data{types}=data(k:k+j-2,:);
        types=types+1;
        k=i;
        j=1;


    end


i=i+1;
end
C_textdata{types}=textdata(k+1:k+j,:);
C_data{types}=data(k:k+j-1,:);


%%sort the data by funID within the same fundStr
i=1;
j=1;
ID={};
C_data_ID={};
C_textdata_ID={};
```

```
for i=1:types
m=0;
ID{i}=unique(C_data{i}(:,1));
    for j=1:length(ID{i}(:,1))
      k=count(C_data{i},ID{i}(j));
      C_data_ID{i,j}=C_data{i}(m+1:m+k,:);
      C_textdata_ID{i,j}=C_textdata{i}(m+1:m+k,:);
      m=m+k;


    end
end


%%sort into the same time series (for two different funds using the same
%%trading strategy. i.e.,C_data_ID{i,j1} and C_data_ID{i,j2}
%%given ID1,ID2,str

function [timeSeries]=match2ID(ID1,ID2,str1,str2,cell_name)
timeSeries=[];
i=1;
j=1;
m=1;
l1=length(cell_name{str1,ID1}(:,1));
l2=length(cell_name{str2,ID2}(:,1));
for i=1:l1
    for j=1:l2
        if cell_name{str1,ID1}(i,4)==cell_name{str2,ID2}(j,4)
          timeSeries(m,1)=cell_name{str1,ID1}(i,4);
          timeSeries(m,2)=cell_name{str1,ID1}(i,5);
          timeSeries(m,3)=cell_name{str2,ID2}(j,5);
         m=m+1;
        end
    end
end
    end
%%for all IDs within one strategy
%%count_cell(c_data_ID,str)=# of matrix in a row
function k= count_cell(cell_name,row_num)
i=1;k=0;
while i<=length(cell_name(1,:)) && i~=0
```

```
    if ~isempty(cell_name{row_num,i})
        k=i;
        i=i+1;
    else
        i=0;

    end
end
end
I=1;J=1;c_timeSeries={};a=1;
for  I=1:count_cell(C_data_ID,str1)
    for  J=I:count_cell(C_data_ID,str2)
        c_timeSeries{1,a}=match2ID(I,J,str1,str2,C_data_ID);
        a=a+1;
    end

end

end


%parametric estimation with student t marginal and Frank copula alpha

function [alpha]=BStF_alpha_para_estimation(M,observed,Min)
 M_new={};
 n=1;
 alpha=[];
for  j=1:observed

        para_1=[];
        para_2=[];
        if ~isempty(M{1,j}) && length(M{1,j}(:,1))>=Min
            para_1=mle(M{1,j}(:,2),'distribution','tlocationscale');
            para_2=mle(M{1,j}(:,3),'distribution','tlocationscale');
           M{1,j}(:,2)= (M{1,j}(:,2)-para_1(1))/para_1(2);
           M{1,j}(:,3)= (M{1,j}(:,3)-para_2(1))/para_2(2);
          F_2=cdf('t',M{1,j}(:,2),para_1(3));
          F_3=cdf('t',M{1,j}(:,3),para_2(3));
```

```
                M_new{1,j}(:,1)=F_2;
                M_new{1,j}(:,2)=F_3;
                Alpha = copulafit('Frank',M_new{1,j});
                if Alpha~=0
                alpha(n) = Alpha;
                n=n+1
                end
            end
end
end


function [alpha]=BStC_alpha_para_estimation(M,observed,Min)
 M_new={};
 n=1;
 alpha=[];
for j=1:observed
            para_1=[];
            para_2=[];
            if ~isempty(M{1,j}) && length(M{1,j}(:,1))>=Min
                 para_1=mle(M{1,j}(:,2),'distribution','tlocationscale');
                 para_2=mle(M{1,j}(:,3),'distribution','tlocationscale');
                M{1,j}(:,2)= (M{1,j}(:,2)-para_1(1))/para_1(2);
                M{1,j}(:,3)= (M{1,j}(:,3)-para_2(1))/para_2(2);
                F_2=cdf('t',M{1,j}(:,2),para_1(3));
                F_3=cdf('t',M{1,j}(:,3),para_2(3));
                M_new{1,j}(:,1)=F_2;
                M_new{1,j}(:,2)=F_3;
                alpha(n)=copulafit('Clayton',M_new{1,j});

            end
            n=n+1;
end

end


%%RETURN an array of the Spearman's rho of ID_i and ID_j, where i NOT equal
%%to j. Use this array we can calculate mean and std, histogram...
```

```matlab
function lamda=arrayTail(M, observed)

lamda=[];
k=1;
for i=1:observed
    for j=i+1:observed
        if ~isempty(M{i,j})
    lamda(k)=tail(M{i,j});
    k=k+1;
        end
    end
end

end


%%RETURN an array of the Spearman's rho of ID_i and ID_j, where i NOT equal
%%to j. Use this array we can calculate mean and std, histogram...
function rho=arraySpearman(M, observed, min)

rho=[];
k=1;
for i=1:observed
    for j=i+1:observed
        if ~isempty(M{i,j}) && length(M{i,j}(:,1))>=min
            rho(k)=spearmanSample(M{i,j});
            k=k+1;
        end

    end
end

end


%%RETURN an array of the Kendal's tau of ID_i and ID_j, where i NOT equal
%%to j. Use this array we can calculate mean and std, histogram...
function tau=arrayKendals(M, observed, min)
```

```
tau = [];
k=1;
for  i=1:observed
      for  j=i+1:observed
      if  ~isempty(M{i,j}) && length(M{i,j}(:,1))>=min
            tau(k)=kendalsample(M{i,j});
            k=k+1;
      end
      end
end

end


function  [SpearmanMean,SpearmanStd,SpearmanPercen]=BSStatSpearman(M,observed,Min)

rho = [];
k=1;
for  j=1:observed

        if  ~isempty(M{1,j}) && length(M{1,j}(:,1))>=Min
            rho(k)=spearmanSample(M{1,j});
            k=k+1;
        end

end
% STEP 1 - rank the data
y = sort(rho);
MIN=min(rho)
MAX=max(rho)
SpearmanMean=mean(rho)
SpearmanStd=std(rho)
SS=length(rho)
hist(rho)

end
```

```
function [PearsonMean, PearsonStd, PearsonPercen]=BSStatPearson(M, observed, Min)


corr = [];
k=1;
for j=1:observed

    if ~isempty(M{1,j}) && length(M{1,j}(:,1))>=Min
        corrMatrix=corrcoef(M{1,j}(:,2),M{1,j}(:,3));
        corr(k)=corrMatrix(1,2);
        k=k+1;
    end
    end
end
PearsonMean=mean(corr)
PearsonStd=std(corr)
SS=length(corr)
hist(corr);
end




function [KendalMean, KendalStd, KendalPercen]=BSStatKendals(M, observed, Min)


tau = [];
k=1;
for j=1:observed
    if ~isempty(M{1,j}) && length(M{1,j}(:,1))>=Min
        tau(k)=kendalsample(M{1,j});
        k=k+1;
    end
    end
end
% STEP 1 - rank the data
y = sort(tau);
MIN=min(tau)
MAX=max(tau)
KendalMean=mean(tau)
KendalStd=std(tau)
SS=length(tau)
hist(tau);
```

```
end


function [Tailmean, Tailstd, Tailpercen]=BSStatTailEstimation(M,observed,Min)

lamda=[];
k=1;
for j=1:observed
        if ~isempty(M{1,j}) && length(M{1,j}(:,1))>=Min
    lamda(k)=fmincon(@(LAMDA)tailEstimation(M{1,j},LAMDA),0,0,1);
    %reduce the negative value to 0
    if lamda(k)<0
        lamda(k)=0;
    elseif lamda(k)>1
        lamda(k)=1;
    end
    k=k+1;
        end
    end

% STEP 1 - rank the data
if length(lamda)>1
y = sort(lamda);
MIN=min(lamda)
MAX=max(lamda)
Tailmean=mean(lamda)
Tailstd=std(lamda)
SS=length(lamda)
hist(lamda);
end
end



function [alpha]=tC_alpha_para_estimation(M,observed,Min)
 M_new={};
 n=1;
 alpha=[];
for i=1:observed
```

```
    for j=i+1:observed
        para_1 =[];
        para_2 =[];
        if ~isempty(M{i,j}) && length(M{i,j}(:,1))>=Min
            para_1=mle(M{i,j}(:,2),'distribution','tlocationscale');
            para_2=mle(M{i,j}(:,3),'distribution','tlocationscale');
           M{i,j}(:,2)=  (M{i,j}(:,2)-para_1(1))/para_1(2);
           M{i,j}(:,3)=  (M{i,j}(:,3)-para_2(1))/para_2(2);
          F_2=cdf('t',M{i,j}(:,2),para_1(3));
          F_3=cdf('t',M{i,j}(:,3),para_2(3));
          M_new{i,j}(:,1)=F_2;
          M_new{i,j}(:,2)=F_3;
          alpha(n)=copulafit('Clayton',M_new{i,j});

        end
    end
    n=n+1
end
end


function tF_para_Spearman(alpha)

for i=1:length(alpha)
syms t
D1_intF(i)=double(int(t/(exp(t)-1), t,0, alpha(i)));
syms x
D2_intF(i)=double(int(x^2/(exp(x)-1), x,0, alpha(i)));
end
D1=D1_intF./alpha;
D2=2*(D2_intF./(alpha.^2));
rho=1-12*(D2+2*alpha/3-D1-alpha/2);
y = sort(rho);
MIN=min(rho)
MAX=max(rho)
SpearmanMean=mean(rho)
SpearmanStd=std(rho)
SS=length(rho)
hist(rho)
```

```
end


function tF_para_Kendals(alpha)

for i=1:length(alpha)
syms t
D1_intF(i)=double(int(t/(exp(t)-1), t,0, alpha(i)));

end
D1=D1_intF./alpha;
tau=1+4*(D1-1)./alpha;
% STEP 1 - rank the data
y = sort(tau);
MIN_tau=min(tau)
MAX_tau=max(tau)
KendalMean=mean(tau)
KendalStd=std(tau)
SS_tau=length(tau)
hist(tau);

end


function tC_para_LTD(alpha)


lambda=2.^(-1./alpha);
% STEP 1 - rank the data
y = sort(lambda);
MIN_lambda=min(lambda)
MAX_lambda=max(lambda)
TailMean=mean(lambda)
TailStd=std(lambda)
SS_lambda=length(lambda)
hist(lambda);
```

```
end


function tC_para_Kendals(alpha)


tau=alpha./(alpha+2);
% STEP 1 − rank the data
y = sort(tau);
MIN_tau=min(tau)
MAX_tau=max(tau)
KendalMean=mean(tau)
KendalStd=std(tau)
SS_tau=length(tau)
hist(tau);

end
```