

Social Networks, Asset Allocation and Portfolio Diversification

by

Qiutong Wang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Quantitative Finance

Waterloo, Ontario, Canada, 2015

© Qiutong Wang 2015

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In this thesis we consider the problem of choosing financial assets from the equity markets for portfolio construction purposes. We adapt various measures to model the dependence structure among financial assets, taking both the linear and the non-linear relationships into consideration. The dependence structure is reflected by the social networks. We apply the data clustering technique (Frey and Dueck, 2007) to the social networks and study the equity selections based on different dependence measures. The regime switching model (Perlin, 2014) is considered as well in order to identify the changes in the market phases. The performance of the equity selections is evaluated within the mean-variance framework. In addition, we present a diversification analysis of the equity selections with the methodology proposed by Meucci (2009). The numerical tests are applied on three major Chinese equity markets. Through changing the market environment, we acquire a good understanding of the influencing factors for choosing financial assets.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Tony Wirjanto for supporting me accomplishing this thesis. His guidance is of great benefit to my Master's study. I would also like to thank my second readers Prof. Adam Kolkiewicz and Prof. Ken Seng Tan for their valuable suggestions of my thesis. At last, I would like to thank my friends for their help in my research.

Table of Contents

List of Tables	vii
List of Figures	viii
1 Introduction	1
1.1 Overview	1
1.2 Selected Literature Review	2
1.3 Issues and Methods	6
2 Theoretical Backgrounds	8
2.1 Modeling Dependence Structure	8
2.1.1 Introduction of Copulas	9
2.1.2 Classes of copulas	14
2.1.3 Dependence	17
2.1.4 Fitting Copulas to Data	22
2.2 Social Network Clustering	23
2.2.1 Introduction of Social Network Analysis	24
2.2.2 Clustering by Affinity Propagation	27
2.2.3 Between-Within Proportion	29
2.3 Mean-variance Framework for Portfolio Selection	32
2.4 Diversification Technique	34

2.4.1	Non-additive Risk Sources	35
2.4.2	Principal Component Analysis	36
2.4.3	Diversification Distribution	37
2.4.4	Entropy as a Diversification Risk Measure	39
3	Social Network Analysis and Clustering	41
3.1	Data Collection and Cleaning	42
3.2	Modeling Dependence Structures via Copulas	44
3.3	Markov Regime Switching Analysis	59
3.3.1	The Switching Model	59
3.3.2	Regime Switching Results	63
3.4	Social Network Clustering Experiment	67
4	Portfolio Selection Evaluation	77
4.1	Mean-Variance Analysis	77
4.2	Portfolio Diversification Analysis	86
5	Conclusion	99
	APPENDICES	102
A	Additional Figures and Tables	103
A.1	Estimation of Tail Dependence and Mutual Information on Other Markets	103
A.2	Markov Regime Switching Results on Other Markets	106
	References	109

List of Tables

3.1	Data statistics of all three indexes after cleaning	43
3.2	Statistics of Estimated Linear Correlations and Rank Correlations	44
3.3	Statistics of Estimated Tail Dependence and Mutual Information	57
3.4	Statistics of Markov Switching Results	66
3.5	Statistics of Social Network Clustering Results	75
3.6	Statistics of Social Network Clustering Results	75
3.7	Selection of Stocks by Social Network Clustering on Linear Correlations	76
4.1	Mean-Variance Analysis of the Equity Selection from Different Market Settings	84
4.2	Mean-Variance Analysis of the Equity Selection from Different Market Settings	85
4.3	Mean-Diversification Analysis of the Equity Selection from Different Market Settings	97
4.4	Mean-Diversification Analysis of the Equity Selection from Different Market Settings	98

List of Figures

2.1	Sample graphs of stock network after clustering based on linear correlation.	25
3.1	Distributions of Estimated Linear Correlations and Rank Correlations	45
3.2	Kernel Estimation on Empirical Distribution of SHASHR Index.	49
3.3	Fitting Copulas to Returns	52
3.4	Distributions of Estimated Tail Dependence and Mutual Information	56
3.5	Markov Regime Switching Results on SHASHR Index.	64
3.6	Markov Switching Autoregressive Model with Two Market Indexes.	65
3.7	Social Network Clustering on Linear Correlations.	72
3.8	Social Network Clustering on Multiple Dependence Measures with Illustration of 16 Clusters.	74
4.1	MV Efficient Frontiers Consist of Small or Large Number of Equities based on Various Dependence Measures.	79
4.2	MV Efficient Frontiers Consist of Large Number of Equities based on Various Dependence Measures.	80
4.3	Mean-Diversification Analysis of the Equity Selection (21 Equities) based on the Linear Correlation.	88
4.4	Mean-Diversification Analysis of the Equity Selection (21 Equities) based on the Lower Tail Dependence.	91
4.5	Mean-Diversification Analysis of the Equity Selection (21 Equities) based on the Tail Dependence Indicated by t Copula.	92

4.6	Mean-Diversification Analysis of the Equity Selection (22 Equities) based on the Linear Correlation in the Bear Market Phase.	93
A.1	Distributions of Estimated Tail Dependence and Mutual Information	104
A.2	Distributions of Estimated Tail Dependence and Mutual Information	105
A.3	Markov Regime Switching on SHASHR Index with 4-states Mixture Distributions.	106
A.4	Markov Regime Switching Results on SHASHR Index.	107
A.5	Markov Regime Switching Results on SHASHR Index.	108

Chapter 1

Introduction

1.1 Overview

Over the past few decades, asset allocation theories have been actively studied by researchers and widely used by practitioners. To fulfill the need of seeking profits while avoiding potential risk of loss during the investment process, a robust portfolio strategy plays a key role.

Markowitz's mean-variance model was introduced more than 60 years ago and it is still considered as one of the most popular approaches in portfolio optimization. Under a given set of constraints, the investor could either seek minimizing its portfolio's variance for a given expected return or maximizing the expected return with a specified variance.

A major flaw of the classical mean-variance model is the assumption that returns admit multivariate normality with constant correlation. Nowadays, there is compelling evidence that the assets' return distributions are asymmetrical with extreme correlations. The traditional linear correlation alone is not sufficient to reveal the dependence structure among returns. As a result, new dependence measures such as tail dependence have been introduced to financial literatures. The emergence of these dependence measures has energized the study of portfolio theories as they are capable of capturing the observed non-linear dependence structure.

In this thesis, we will consider a portfolio construction problem focus on stock selection. More explicitly, we assume that an investor is participating in a large equity market, e.g. Shanghai Stock Exchange A which contains over 800 trading stocks. The task would be how to select stocks from the market. This is highly intuitive because the risky assets

composing a portfolio should be chosen based on some principles. Also, a portfolio which is either too small or too large will impose difficulties on the investor's management. We are interested in studying this topic from a new perspective called *social network analysis* (**SNE**). The good feature about using this technique is that the **SNE** offers a possibility of dimension reduction based on various dependence structures. We can view the stock selection process as a dimension reduction process. Stocks chosen by such a technique are usually less correlated.

An investor will also like to evaluate risk from suffering a heavy loss in his portfolio. In mitigating such risk, the utilization of a diversification principle in a portfolio shall not be ignored. To tackle this problem, we introduce a proper measure of diversification for portfolios. Many popular methodologies have been proposed to act on diversification analysis. In this thesis, we are interested in using a mean-diversification framework to evaluate our portfolios.

In summary, the main goal of this thesis is to reasonably select stocks for portfolio construction from large equity markets. Such a selection methodology should be built upon consideration of various dependence structures. Then, we evaluate the portfolio selections in a mean-variance framework and measure their diversification in a mean-diversification framework. Before dwelling into our work, let us introduce some related works from the literature.

1.2 Selected Literature Review

Since the appropriateness of linear correlation for financial time series came under close scrutiny, a number of study has been performed to study nonlinear correlations. [Longin and Solnik \(2001\)](#) study the conditional correlation structure of stock returns on five largest international equity markets (US, UK, France, Germany and Japan). A method for deriving a distribution of conditional tail correlation using extreme value theory is developed in their paper. Their results suggest that equities are more correlated when US market is moving downside than upside. [Ang and Chen \(2002\)](#) perform tests on US equity market, comparing the correlations implied by a normal distribution with those conditional on bear markets. They find that the latter are higher on average. Their study also shows stocks exhibiting stronger asymmetric correlations. [Patton \(2004\)](#) explores influences on portfolio decisions by asymmetric dependence between stocks. They use a copula approach to model dependence structure and find that models capturing asymmetric dependence yield better portfolio performance. [Hong et al. \(2007\)](#) extend Ang and Chen(2002)'s work by providing a test on symmetry which does not require specifying a statistical model on data. Their

empirical facts show that portfolios with strong asymmetric betas and covariances tend to have better performance. [Chordia et al. \(2011\)](#) search for reasons that account for asymmetric correlations between return distributions. They conjecture that trading activity in small stocks performed by retail investors leads to asymmetric correlations. [Zhao and Lin \(2011\)](#) measure the dependence structure in stock markets from the perspective of a copula entropy. The copula entropy approach they apply under a non-Gaussian distribution assumption results in a superior numerical analysis. They show that the copula entropy could be used as an alternative approach to compute another dependence measure named mutual information. As so many approaches have been proposed in the literature to measure the dependence structure of financial data more accurately, we are motivated in putting various measures of dependence structures into our portfolio construction framework.

On the other hand, equity market usually consists of thousands of securities. From the computational point of view, we are dealing with high dimensional financial data. Hence, one has to find a suitable low-dimensional representation of the data that can be used for portfolio problems, i.e. dimension reduction ([Putzig et al., 2010](#)). In various financial applications, dimension reduction techniques have proved to be quite helpful in tackling the aforementioned problem. [Boyle et al. \(2008\)](#) show that using a dimension reduction technique leads to a dramatic improvement in the efficiency of simulation-based computation of optimal portfolios. [Resta \(2011\)](#) proposes a model to uncover the underlying assets in a portfolio. He suggests that assets' dynamics can be characterized by a number of factors, but only a few of them play as dynamics' natural drivers. By utilizing dimension reduction techniques, an asset drivers framework is developed such that the driving factors are extracted, while those less significant ones are excluded from consideration. [Bai and Shi \(2011\)](#) examine different dimension reduction techniques in estimating high dimensional covariance matrices of financial data, including factor analysis and principal components analysis. [Lai et al. \(2011\)](#) demonstrate that dimension reduction can facilitate parameter estimation with the issue of large number of assets relative to trading periods. Factor models are implemented in the article to reduce the number of parameters to be estimated in an empirical Bayes approach. [Papanicolaou \(2013\)](#) considers a market with partial information. An approximate dynamic programming algorithm is proposed because a typical dynamic programming problem in an optimal portfolio is non-Markovian and it is difficult to compute. The approximate dynamic program in the paper initializes a dimension reduction process from the high-dimensionality of the non-Markovian problem. [Takano and Gotoh \(2014\)](#) study an optimal portfolio problem in multi-period using a kernel-based framework. A kernel principal component analysis is introduced to reduce the size of the optimization problem. The optimization model with a dimension reduction technique employed has several advantages, e.g. higher computation efficiency and better investment

performance. At this point, we have already understood the importance of dimension reduction in studying high dimensional problems. Indeed, the initial task of choosing a certain number of securities from a huge market in our case is a high dimensional problem.

Therefore, we are in need of a scheme which takes both various dependence structures and dimension reduction into consideration. In this sense, *social network analysis* (**SNE**), may act as a powerful tool. The idea underlying the **SNE** is intuitive. Firstly, social network is used to study different kinds of relationships in social science, which is also applicable to dependence structures in assets' returns. Secondly, a dimension reduction process can be perceived as a clustering process in **SNE**. Few social network techniques have been applied to stock selection process in the literature so far. However, networks have already been introduced into different financial aspects. Mantegna (1999) seems to be the first who utilizes network methods to study financial markets. Mantegna computes the correlation coefficients between the logarithm of stock price and uses them to measure the distances between stocks trading in the market. A graph representing the hierarchical structure of the financial market is obtained based on those distances. This graph is known as the *minimal spanning tree* (**MST**). Extending Mantegna (1999)'s work, Bonanno et al. (2003) study stock network's properties of minimal spanning tree based on correlations. Their results point out that the spanning tree obtained by empirical data is a complex network, which can not be accurately reproduced or approximated by ordinary models. Onnela et al. (2003) also study minimal spanning tree of correlations between stocks. Their work focuses more on the time dependence and shows that during the episode of a "bear market", the whole network shrinks topologically due to correlations between stocks. This is related to the fact that the network graph is built upon some highly connected nodes. The tree network developed in their paper can also facilitate a characterization of the market taxonomy and proves to be robust. Eom et al. (2006) focus their study on finding factors that affect a specific stock's relations to other stocks in the network. Eom et al. (2009)'s study combine property of stock networks and random matrix theory. They compare the stock network developed from actual returns and the one from correlation matrix created by random matrix theory. Eom et al. (2006) suggest that the consistency between two networks is positively related to the number of eigenvalues considered. In the following year, Mantegna and other researchers dedicate more work to correlation based networks. In Tumminello et al. (2010), they propose some ways to construct different network models such as hierarchical trees from the correlation matrix. Partial correlation network is investigated in Kenett et al. (2010) to unveil the underlying backbone of the correlation structure of the market. The idea in the paper is to detect the influence on the correlation between two stocks caused by a third one. Tabaka et al. (2010) provide a network analysis on different sectors in a stock market. Štefan Lyócsa et al. (2012)

demonstrate a comparison of properties of minimal spanning trees obtained by dynamic conditional correlations and by rolling window correlations. Apart from the literature which only focuses on properties of minimal spanning tree or other patterns of stock networks, some researchers view the term 'network' from wider aspects. Schweitzer et al. (2009) emphasize the need for a good understanding of economic networks in the background of financial crisis. Ferrara and Fiumara (2012) particularly examine different models' ability to describe social network's structure. Nettleton (2013) reviews the popular concepts and tools in social networks focusing on their graph structure and discuss their applications in some topics. Hereby, we have introduced the development of study on stock networks originated from Mantegna (1999) and some applications of social network analysis. Social network is another branch of network theory and has not drawn much attention in financial application yet. Recall that our main goal is to select a pool of stocks from the market. In the context of SNE, we want to initialize a stock clustering process. Social network graph's property of forming clusters around its vertices suitably provides a path to solve this question. Meanwhile, the "distances" between all nodes on the graph work as a measure of dependence structure among assets' returns. Different dependence structures will lead to different graphs. If we are able to locate precise vertices on some well constructed stock clusters, these "nodes" stocks would be our choice of elements to be inserted into the portfolio.

Imagine now that we had already selected some stocks to compose our initial portfolio. A natural step to follow is to reasonably allocate our wealth to assets in order to construct the portfolio. Here, we briefly review the development of portfolio theory. The modern portfolio theory is built upon Markowitz (1952)'s work "Portfolio Selection". A framework of describing portfolios is established in the article with the insight that assets' risks and returns are viewed together in terms of variances and means. For this reason, the framework is also known as a mean-variance model. The framework states that a portfolio is considered to be optimal if the expected return of the portfolio is maximized for a given level of risk (proxied by the standard deviation of the portfolio's expected returns) or has a minimized risk level for a given level of the portfolio's expected return. In both views, portfolio optimizations are achieved but they differ in the fact that in the former the objective function is linear with the constraints being quadratic whereas the latter is a quadratic optimization with linear constraints. In the following few decades, various new approaches are introduced to extend this famous framework. The *Capital Asset Pricing Model* (CAPM) proposed by Sharpe (1964) considers a very important application of mean-variance analysis. CAPM takes into consideration the equilibrium asset pricing consequences of investors' individually rational actions and provides a foundation for an asset pricing model (Pennacchi, 2007). It introduces a capital market line and sug-

gests an efficient portfolio is actually a linear combination of the market portfolio and the risk-free asset. [Ross \(1976\)](#) proposes *Arbitrage Pricing Theory (APT)*, which is a generalization of **CAPM**. Instead of considering a single risk factor, assets' returns are driven by multiple risk factors. [Merton \(1972\)](#) derives an analytical solution of portfolio weights in mean-variance framework when assets' returns and their covariance matrix are given. Nowadays, portfolio theory is being enriched as new techniques are introduced. The new methods include but are not limited to evaluating portfolios in new risk measures (Value at Risk, Expected Shortfall), robust portfolio construction approaches (Black-Litterman approaches, shrinkage approaches, resampled approaches), regime switching techniques etc.

Evaluating portfolios plays an active role in risk management. At this stage we assume that a portfolio consisting of the chosen stocks is constructed. We have efficiently allocated our wealth. An adequate diversification will guarantee a less risky portfolio against fat-tailed underlying distributions or other risk source. Put differently, a well diversified portfolio is not exposed to any risk factor that might evidently drive returns. [Meucci \(2005\)](#) introduces different backgrounds of diversification in the context of asset allocation. An effective methodology to perform diversification analysis on portfolios is then proposed by [Meucci \(2009\)](#). In the paper Meucci utilizes principal component analysis to decompose risk sources and introduces [Shannon \(1948\)](#)'s entropy as the representation of a diversification measure. [Xiong \(2009\)](#) and [d-fine GmbH \(2011\)](#) put Meucci's diversification technique into practice separately. Other applications of Shannon's entropy is reviewed in [Zhou et al. \(2013\)](#).

1.3 Issues and Methods

As we mention above, researchers study the dependence structure of assets' returns from a new perspective. Also, various techniques of dimension reduction have been implemented to process large financial data. A network method is capable of taking both aspects into account. However, the minimal spanning tree method suffers from certain limitations. Firstly, the stock networks built from this methodology can only reveal linear relationships whereas other nonlinear patterns of dependence structure are not considered. Secondly, such networks are mainly constructed in hierarchical forms (trees) other than the clustering forms by k-means or fuzzy-C method. The expansion of a hierarchical tree can reflect the overall relationship among its nodes but it is still difficult to capture some nodes (or stocks) as its "representatives". In other words, it is not efficient when we are looking for some cluster centers.

In the thesis, we tackle the stock selection problem within a social network framework to overcome those issues. We study the association structure among all stocks through various dependence measures, i.e. linear correlation, rank correlations, tail dependence and mutual information. The nonlinear measures are constructed via copulas. Then, we construct stock networks through an efficient clustering technique proposed by [Frey and Dueck \(2007\)](#) so that the networks' structures are revealed by various dependence measures. We choose the cluster centers of a network as the stocks to be included in a portfolio. Such a scheme also considers regime changes in the market. We use a *Markov Regime Switching Model* to identify different market phases.

Then, we evaluate the portfolio selection based on different dependence measures and market phases. Such a task is achieved within a mean-variance framework. In order to mitigate the risk of suffering heavy loss, we also measure the portfolio diversification against different risk sources with a diversification framework proposed by [Meucci \(2009\)](#).

The thesis structure is designed as follows:

In Chapter 2, we introduce the foundation for the theoretical framework used in this thesis.

In Chapter 3, we perform the social network clustering analysis to achieve the goal of stock selection. Various dependence measures and the regime switching effects are studied within the social network clustering framework.

In Chapter 4, we make a comparative analysis to evaluate the portfolio selection through the mean-variance and the mean-diversification framework. The analysis suggests the suitable stock selection under various market environment. The influence of the dependence measures, portfolio sizes and regime switching effect are examined in the analysis.

In Chapter 5, we conclude our findings.

Chapter 2

Theoretical Backgrounds

In this chapter, we will introduce some background studies to lay the theoretical foundation for methods used in this thesis. We firstly discuss how to construct various dependence measures via copulas. Then, we introduce the framework of social network analysis and the mechanism of *Affinity Propagation Clustering* with an improvement criterion. Finally, we introduce the mean-variance framework and the diversification technique as evaluation tools for the portfolio selections.

2.1 Modeling Dependence Structure

The statistical features of financial time series can be described by multi-variate distributions of random vectors. Both univariate and multivariate distributions of financial time series are observed with some stylized facts such as heavy-tailed return series, extreme returns clustering, coincidence of extreme returns between some series, etc. Hence, to obtain a good understanding of the dynamics of our underlying securities, we need a model which is able to describe probabilistic properties of distributions of time series and the dependence structure among them (McNeil et al., 2005). The concept of copula, since 1990s, is often used in the financial literature to tackle the above question due to its various advantages in the flexibility and completeness of characterizing multivariate distributions' properties. If we say a joint distribution implicitly consists of two parts: individual properties of its marginal distributions and a pattern of dependence structure associated with them, copula method will make it possible to decompose the two parts and explain the latter with copula's special features. Copula also leads to the derivation of some important dependence

measures, such as rank correlations and coefficients of tail dependence. In this sense, copula is chosen to be the base point of our introduction of this section. There are many good textbooks discussing copulas, such as [Nelsen \(2006\)](#), [Cherubini et al. \(2004\)](#), [McNeil et al. \(2005\)](#), [Kemp \(2010\)](#), [Cherubini et al. \(2011\)](#), [Rzschendorf \(2013\)](#). We will only introduce the basics of copulas and some dependence measures that serve the purposes of this thesis. We mainly refer to [McNeil et al. \(2005\)](#) in the definition of relevant concepts.

2.1.1 Introduction of Copulas

In this subsection, we discuss the concept of copulas.

Consider a general d -dimensional random vector $\mathbf{X} = (X_1, \dots, X_d)'$. The *joint distribution function* of \mathbf{X} can be written as

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}(x_1, \dots, x_d) = \mathbb{P}[\mathbf{X} \leq \mathbf{x}] = \mathbb{P}[X_1 \leq x_1, \dots, X_d \leq x_d]. \quad (2.1)$$

For simplicity, we write F instead of $F_{\mathbf{X}}$. Then, the *marginal* distributions of \mathbf{X} can be characterized by the *marginal distribution function* F_{X_i} , or simply F_i . For all i we have

$$F_i(x_i) = \mathbb{P}[X_i \leq x_i] = F(\infty, \dots, \infty, x_i, \infty, \dots, \infty). \quad (2.2)$$

We refer to f_i , the i th partial derivative of F , as the i th *marginal density function* of \mathbf{X} , if the marginal distribution function F_i is absolutely continuous. Inversely, if there exists some non-negative function f so that

$$F(x_1, \dots, x_d) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f(u_1, \dots, u_d) du_1 \dots du_d, \quad (2.3)$$

we say that the distribution function F is absolutely continuous with f being the corresponding *joint density function* of \mathbf{X} . We note here that the existence of marginal densities for all k -dimensional marginals can be implied by the existence of a joint density. Nonetheless, the existence of marginal densities does not necessarily imply the existence of a joint density ([McNeil et al., 2005](#)).

For simplicity, we introduce the following concepts in terms of bivariate cases. They can also be generalized to a high dimensional case.

Let \mathbb{R} denote the ordinary real line $(-\infty, \infty)$, $\overline{\mathbb{R}}$ the extended real line $[-\infty, \infty]$. The extended real plane $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$ is denoted by $\overline{\mathbb{R}}^2$. Let the Cartesian product $B = [x_1, x_2] \times [y_1, y_2]$ denote a *rectangle* in $\overline{\mathbb{R}}^2$. Then, the points (x_1, y_1) , (x_1, y_2) , (x_2, y_1) and (x_2, y_2) are all the *vertices* of rectangle B . A *2-place real function* H is a function whose domain, denoted by

$\text{Dom}H$, is a subset of $\overline{\mathbb{R}}^2$ and whose range, denoted by $\text{Ran}H$, is a subset of \mathbb{R} (Nelsen, 2006).

Definition 2.1.1. Let A_1 and A_2 be two non-empty subsets of $\overline{\mathbb{R}}$. Let a_i denote the least element of A_i , $i = 1, 2$. A 2-place real function $H : A_1 \times A_2 \rightarrow \mathbb{R}$ is called *grounded* if, for all (x, y) in $A_1 \times A_2$,

$$H(a_1, y) = 0 = H(x, a_2).$$

Definition 2.1.2. Consider the same function $H : A_1 \times A_2 \rightarrow \mathbb{R}$. Let $B = [x_1, x_2] \times [y_1, y_2]$ denote a rectangle all of whose vertices lie in $A_1 \times A_2$, such that $x_1 \leq x_2, y_1 \leq y_2$. Then the *H-volume* of B is defined as

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1). \quad (2.4)$$

Definition 2.1.3. A 2-place real function $H : A_1 \times A_2 \rightarrow \mathbb{R}$ is called *2-increasing* if $V_H(B) \geq 0$ for all rectangles B whose vertices are in $A_1 \times A_2$.

The *H-volume* of B can actually be regarded as a measure of mass of rectangle B in the domain of 2-place real function H (Cherubini et al., 2004).

Lemma 2.1.4. Let a 2-place real function $H : A_1 \times A_2 \rightarrow \mathbb{R}$ be grounded and 2-increasing. Then H is nondecreasing in each argument. Proof. see Nelsen (2006).

Lemma 2.1.5. Let a 2-place real function $H : A_1 \times A_2 \rightarrow \mathbb{R}$ be grounded 2-increasing and with margins. Let (x_1, y_1) and (x_2, y_2) be any points in $A_1 \times A_2$. Then

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|.$$

Proof. see Nelsen (2006).

With the above results, we can proceed to the definition of copulas.

Definition 2.1.6. (Nelsen, 2006) A *two-dimensional subcopula* \mathcal{C} is a real function with the following properties:

1. \mathcal{C} is defined on $S_1 \times S_2$, where S_1 and S_2 are non-empty subsets of $\mathbf{I} = [0, 1]$ containing 0 and 1:

$$\mathcal{C} : S_1 \times S_2 \rightarrow \mathbf{I};$$

2. \mathcal{C} is grounded and 2-increasing;

3. For every u in S_1 and v in S_2 , $\mathcal{C}(u, 1) = u$ and $\mathcal{C}(1, v) = v$.

As $0 \leq \mathcal{C}(u, v) \leq 1$ for every (x, y) in the domain of \mathcal{C} , the range of \mathcal{C} is a subset of \mathbf{I} .

Definition 2.1.7. (Nelsen, 2006) A *two-dimensional copula* C is two-dimensional subcopula whose domain is \mathbf{I}^2 :

$$C : \mathbf{I}^2 \rightarrow \mathbf{I}$$

with the following properties:

1. For every u, v in \mathbf{I} ,

$$C(u, 0) = 0 = C(0, v), \tag{2.5}$$

and

$$C(u, 1) = u, C(1, v) = v; \tag{2.6}$$

2. For every u_1, u_2, v_1, v_2 in \mathbf{I} such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0. \tag{2.7}$$

Note here since $C(u, v) = V_C([0, u] \times [0, v])$, $C(u, v)$ can be viewed as a number in \mathbf{I} , which measures rectangle $[0, u] \times [0, v]$.

Theorem 2.1.8. \mathcal{C} is a subcopula, then for every (u, v) in the domain of \mathcal{C} ,

$$\max(u + v - 1, 0) \leq \mathcal{C}(u, v) \leq \min(u, v). \tag{2.8}$$

Proof. See Nelsen (2006).

Let $M(u, v) = \min(u, v)$ and $W(u, v) = \max(u + v - 1, 0)$. Then for every copula C and every (u, v) in \mathbf{I}^2 ,

$$W(u, v) \leq C(u, v) \leq M(u, v). \tag{2.9}$$

Inequality 2.9 is the *Fréchet-Hoeffding bounds* inequality for copula, with M being the *Fréchet-Hoeffding upper bound* and W being the *Fréchet-Hoeffding lower bound*. $M(u, v)$ and $W(u, v)$ are called *maximum copula* and *minimum copula* respectively, another important copula is the *product copula* (or *independence copula*) $\Pi(u, v) = uv$.

Theorem 2.1.9. If \mathcal{C} is a subcopula, then for every $(u_1, v_1), (u_2, v_2)$ in the domain of \mathcal{C} ,

$$|\mathcal{C}(u_2, v_2) - \mathcal{C}(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|. \tag{2.10}$$

Proof. Let \mathcal{C} denote H and set $F(x) = x, G(y) = y$ in Lemma 2.1.5. □

Theorem 2.1.9 reveals that a subcopula \mathcal{C} is uniformly continuous on its domain. Since \mathcal{C} is grounded 2-increasing, it is also nondecreasing on its domain. These are two very important properties of subcopula. Definition 2.1.6, 2.1.7 and Theorem 2.1.9 tell us that a copula is actually a joint distribution function of standard uniform distributions:

$$C(u, v) = \mathbb{P}[U \leq u, V \leq v]. \quad (2.11)$$

Definition 2.1.10. (Nelsen, 2006) The *diagonal section* of a copula C is the function δ_C from \mathbf{I} to \mathbf{I} defined by $\delta_C(t) = C(t, t)$.

Theorem 2.1.11. (Nelsen, 2006) Let C be a copula. For any v in \mathbf{I} , the partial derivative $\partial C(u, v)/\partial u$ exists for almost all u , and for such v and u ,

$$0 \leq \frac{\partial}{\partial u} C(u, v) \leq 1. \quad (2.12)$$

Likewise, for any u in \mathbf{I} , the partial derivative $\partial C(u, v)/\partial v$ exists for almost all v , and for such v and u ,

$$0 \leq \frac{\partial}{\partial v} C(u, v) \leq 1. \quad (2.13)$$

Furthermore, the functions $u \mapsto \partial C(u, v)/\partial v$ and $v \mapsto \partial C(u, v)/\partial u$ are defined and non-decreasing almost everywhere on \mathbf{I} . *Proof.* See Nelsen (2006)

Definition 2.1.12. (Nelsen, 2006) A *quasi-inverse* of distribution function F is any function $F^{(-1)}$ with domain \mathbf{I} such that

1. if u is in $\text{Ran}F$, then $F^{(-1)}(u)$ is any number x in $\overline{\mathbb{R}}$ such that $F(x) = u$, i.e., for all u in $\text{Ran}F$,

$$F(F^{(-1)}(u)) = u;$$

2. if u is not in $\text{Ran}F$, then

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

If F is strictly increasing, then it has a unique quasi-inverse, which is known as the ordinary inverse F^{-1} . In the following context, we mainly refer to F^{-1} .

Next, we introduce the core in copula theory: *Sklar's theorem*, which links multivariate distribution functions and their margins together.

Theorem 2.1.13. *Sklar (1959)'s theorem.* Let H be a joint distribution function with margins F and G . Then there exists a copula C such that for all x, y in $\overline{\mathbb{R}}$,

$$H(x, y) = C(F(x), G(y)). \quad (2.14)$$

If the margins F and G are continuous, then copula C is unique; otherwise C is uniquely determined on $\text{Ran}F \times \text{Ran}G$. Conversely, if C is a copula and F and G are distribution functions, then the function H defined in 2.14 is a joint distribution function with margins F and G .

Proof. For full proof, please see Nelsen (2006). Here we give a very intuitive proof. By Equation 2.11 and Definition 2.1.12 we have:

$$\begin{aligned} H(x, y) &= \mathbb{P}[X \leq x, Y \leq y] \\ &= \mathbb{P}[F^{-1}(U) \leq x, G^{-1}(V) \leq y] \\ &= \mathbb{P}[U \leq F(x), V \leq G(y)] \\ &= C(F(x), G(y)). \end{aligned} \quad (2.15)$$

□

If we substitute $F(x)$ and $G(y)$ with u and v respectively in Equation 2.14 and reverse the proof of 2.15, then we can obtain the following corollary.

Corollary 3.3.13. *Let H be a joint distribution function with margins F and G , and let $F^{(-1)}$ and $G^{(-1)}$ be their quasi-inverses respectively. Then there exists a subcopula \mathcal{C} , for any (u, v) in the domain of \mathcal{C} ,*

$$\mathcal{C}(u, v) = H(F^{(-1)}(u), G^{(-1)}(v)). \quad (2.16)$$

If F and G are continuous, Corollary 2.1.1 applies to copulas and Equation 2.16 can be written as:

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)). \quad (2.17)$$

Equation 2.17 offers us a way of constructing copulas from joint distribution functions (Nelsen, 2006). On the other hand, in practice, we can use multiple statistical tools to accurately estimate marginal distributions $F(u), G(v)$, nonetheless the joint distribution

is usually difficult to describe due to the complex dependence structure. In this sense, Equation 2.14 simply explains how copulas can handle this task.

Before we head into the next subsection that deals with different classes of copulas, we will introduce one more concept, the *copula density*. Theorem 2.1.11 tells us about the property of partial derivative of copula. Although the *joint density* of copula may not always exist, in our context we can say the density exists almost everywhere in the interior of \mathbf{I}^2 and is non-negative.

Definition 2.1.14. Let C be a copula, the density c associated to copula C is described by

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}. \quad (2.18)$$

Suppose that in Equation 2.14, the joint density of $H(x, y)$ is denoted by $h(x, y)$ while the densities of $H(x, y)$'s marginal distributions $F(x), G(y)$ are denoted by $f(x), g(y)$ respectively, and also suppose they are all continuous. Then, by Sklar's Theorem 2.1.13, we can see the relationship between the copula density and the density of distribution H :

$$\begin{aligned} H(x, y) &= C(F(x), G(y)) \\ \frac{\partial^2 H(x, y)}{\partial x \partial y} &= \frac{\partial^2 C(F(x), G(y))}{\partial x \partial y} \\ h(x, y) &= c(F(x), G(y))f(x)g(y). \end{aligned} \quad (2.19)$$

Hence, Equation 2.19 indicates that the joint density of distribution H equals the product of its copula density and densities of marginal distributions.

2.1.2 Classes of copulas

In this subsection, we briefly present some frequently used copulas in financial applications.

There are various kinds of copulas in their big family and each of them possesses some unique features. For those most popular ones, McNeil et al. (2005) divided them into three categories: *fundamental* copulas, e.g. *maximum* copula, *minimum* copula and *independence* copula we introduced in Theorem 2.1.8, which represent some extreme dependence structures; *implicit* copulas which are built from Sklar's Theorem 2.1.13, e.g. *Gaussian* copula and *t-copula*, which nonetheless are difficult to be expressed in closed-form in high dimensions; *explicit* copulas, e.g. *Gumble* copula, *Clayton* copula and *Frank* copula, which have

simple closed-forms and can be easily constructed from Sklar's Theorem 2.1.13. Cherubini et al. (2011) also call implicit copulas and explicit copulas as *elliptical* copulas and *Archimedean* copulas, respectively, based on the way they are constructed. We will first represent the *Archimedean* families and then discuss *Gaussian* copula and *t*-copula.

For simplicity, we still stick to bivariate cases. To step into the *Archimedean* copulas, consider two heuristic concepts: *pseudo-inverse* and *generator*.

Definition 2.1.15. (Nelsen, 2006) Let φ be a continuous, strictly decreasing function $\varphi : \mathbf{I} \rightarrow \mathbb{R}_+$ such that $\varphi(1) = 0$. The *pseudo-inverse* of φ is the function $\varphi^{[-1]}$ with $\text{Dom}\varphi^{[-1]} = \mathbb{R}_+$ and $\text{Ran}\varphi^{[-1]} = \mathbf{I}$ given by

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & \text{if } 0 \leq t \leq \varphi(0), \\ 0, & \varphi(0) \leq t \leq +\infty. \end{cases} \quad (2.20)$$

Note here $\varphi^{[-1]}$ is continuous and non-increasing on \mathbb{R}_+ , and strictly decreasing on $[0, \varphi(0)]$. In addition,

$$\varphi^{[-1]}(\varphi(u)) = u, \quad \text{for every } u \in \mathbf{I}.$$

Finally,

$$\varphi^{[-1]} = \varphi^{-1}, \quad \text{if } \varphi(0) = +\infty.$$

Definition 2.1.16. Let φ be a continuous, strictly decreasing and *convex* function $\varphi : \mathbf{I} \rightarrow \mathbb{R}_+$ such that $\varphi(1) = 0$, then φ is called a *generator*. φ is called a *strict generator* whenever $\varphi(0) = +\infty$.

Definition 2.1.17. Let φ be a generator and $\varphi^{[-1]}$ be its pseudo-inverse given by 2.20. An *Archimedean copula* $C : \mathbf{I}^2 \rightarrow \mathbf{I}$ is generated as follows:

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)). \quad (2.21)$$

Given Equation 2.21, copulas can be constructed with proper generators. Next we will introduce three most popular *one-parameter Archimedean copulas* whose generators $\varphi_\theta(t)$ are indexed by one parameter θ .

Definition 2.1.18. Let generator $\varphi_\theta(t) = (-\ln t)^\theta$, with $\theta \in [1, +\infty]$. The *Gumbel* copula is given by:

$$C_\theta(u, v) = \exp\left(-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right). \quad (2.22)$$

Gumbel copula equals product copula $\Pi(u, v)$ if $\theta = 1$, and Fréchet-Hoeffding upper bound $M(u, v)$ if $\theta \rightarrow +\infty$.

Definition 2.1.19. Let generator $\varphi_\theta(t) = \frac{1}{\theta}(t^{-\theta} - 1)$, with $\theta \in [-1, 0) \cup (0, +\infty)$. The *Clayton* copula is given by:

$$C_\theta(u, v) = \max \left[\left(u^{-\theta} + v^{-\theta} - 1 \right)^{-1/\theta}, 0 \right]. \quad (2.23)$$

Clayton copula equals Fréchet-Hoeffding lower bound $W(u, v)$ if $\theta = -1$, and Fréchet-Hoeffding upper bound $M(u, v)$ if $\theta \rightarrow +\infty$. Furthermore, Clayton copula becomes product copula if $\theta = 0$.

Definition 2.1.20. Define a generator $\varphi_\theta(t) = -\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$, with $e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$ and $\theta \in (-\infty, 0) \cup (0, +\infty)$. The *Frank* copula is given by:

$$C_\theta(u, v) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right). \quad (2.24)$$

Frank copula attains a Fréchet-Hoeffding lower bound $W(u, v)$ if $\theta \rightarrow -\infty$, and a Fréchet-Hoeffding upper bound $M(u, v)$ if $\theta \rightarrow +\infty$. Furthermore, Frank copula becomes product copula if $\theta = 0$.

Next, we will present the Gaussian copula and t -copula. Recall that Equation 2.17 could be used to construct copulas from marginal distributions. This illustrates how Gaussian copula and t -copula can be defined.

Definition 2.1.21. Let Φ_ρ denote the joint distribution function of a bivariate standard normal vectors, with a linear correlation coefficient ρ . Let Φ denote the marginal standard normal distribution function. The *Gaussian* copula is given by:

$$\begin{aligned} C^{Ga}(u, v) &= \Phi_\rho \left(\Phi^{-1}(u), \Phi^{-1}(v) \right) \\ &= \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)} \right) ds dt. \end{aligned} \quad (2.25)$$

The density of Gaussian copula is then given by:

$$c^{Ga}(u, v) = \frac{1}{\sqrt{1-\rho^2}} \exp \left(\frac{s^2 + t^2}{2} - \frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)} \right). \quad (2.26)$$

Gaussian copula has a wide range of financial applications, especially in credit market (Cherubini et al., 2011).

Definition 2.1.22. Let t_ν denote the univariate Student's t distribution function with ν degrees of freedom. Let $t_{\nu,\rho}$ denote the bivariate distribution corresponding to t_ν with ρ in **I**. The Student's t -copula is given by:

$$\begin{aligned} C_{\nu,\rho}^t(u, v) &= t_{\nu,\rho}(t_\nu^{-1}(u), t_\nu^{-1}(v)) \\ &= \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(1 + \frac{s^2 - 2\rho st + t^2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} ds dt. \end{aligned} \quad (2.27)$$

t -copula will converge to Gaussian copula as the degree of freedom ν diverges (Cherubini et al., 2004). However, the t -copula captures more observations in the tails comparing to the Gaussian copula, making it more suitable in capturing tail dependence. In the next subsection, we will discuss the concept of dependence structure and how the copulas relate to it.

2.1.3 Dependence

This subsection plays a key role in this thesis as the social network clustering methodology for stock selection totally relies on some descriptions of dependence structures. There is a variety of forms to measure the dependence, namely:

1. the traditional *linear correlation* which captures the linear relationship between random vectors;
2. the *rank correlations* which aim at capturing *concordance* (roughly speaking, concordance represents the fact that extreme values tend to appear together with respect to a pair of random vectors, while non extreme values of random vectors are less associated);
3. the *tail dependence* measures, which flexibly describe tails of the joint distribution function through different copulas.

Most of the above measures of dependence structures are related to copulas in terms of the ways they are constructed. We will start from the basic linear correlation.

Suppose we have two continuous univariate random vector X, Y . Let $\mathbb{E}[\cdot]$ denote the *mean* of a random vector, and $\text{var}(\cdot)$ denote the *variance* of a random vector. Then the *covariance* is denoted by:

$$\sigma_{X,Y} = \text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \quad (2.28)$$

Definition 2.1.23. Given two continuous univariate random vector X, Y , the *Pearson linear correlation coefficient* is given by:

$$\rho_{X,Y}^L = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}. \quad (2.29)$$

The linear correlation coefficient is designed to effectively capture the linear relationship between random vectors, however it is not very suited to measure dependence structures that are nonlinear. In other words, it can cause issues if we are studying some non normal distributions (Cherubini et al., 2011).

Different from the linear correlation which depends both the joint distribution and marginal distributions, rank correlations, as measures of dependence, can be constructed independently of the marginal distribution (McNeil et al., 2005). Normally, the rank correlations refer to *Kendall's τ* and *Spearman's ρ_S* .

Kendall's τ is a measure of concordance between random vectors. Suppose that we have bivariate random vectors (X, Y) . If Y tends to increase with X , then we say that the probability of concordance is relatively high; if Y tends to decrease with increasing X , then we say the opposite. To understand this measure, we use the following definition of Kendall's τ .

Definition 2.1.24. (Nelsen, 2006) Let (X_1, Y_1) and (X_2, Y_2) denote two independent and identically distributed bivariate random vectors with the same joint distribution function H . Then the *Kendall's τ* is defined as the probability of concordance minus the probability of discordance:

$$\tau_{X,Y} = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]. \quad (2.30)$$

Kendall's τ can also be presented in terms of copula (for details please see Cherubini et al. (2004) and Nelsen (2006)). The following definition presents the relationship between them.

Definition 2.1.25. Let X, Y be continuous random variables whose copula is C . Then the *Kendall's τ* is given by:

$$\tau_{X,Y} = \tau_C = 4 \iint_{\mathbf{I}^2} C(u, v) dC(u, v) - 1. \quad (2.31)$$

In addition to Kendall's τ , another rank correlation measuring concordance is Spearman's $\rho_{X,Y}^S$.

Definition 2.1.26. (Nelsen, 2006) Let $(X_1, Y_1), (X_2, Y_2)$ and (X_3, Y_3) denote three independent random vectors with the same joint distribution function H . Then the *Spearman's* ρ is defined as follows:

$$\rho_{X,Y}^S = 3 (\mathbb{P} [(X_1 - X_2) (Y_1 - Y_3) > 0] - \mathbb{P} [(X_1 - X_2) (Y_1 - Y_3) < 0]). \quad (2.32)$$

Similar to Kendall's τ , Spearman's ρ can be interpreted in terms of a copula as well.

Definition 2.1.27. Let X, Y be continuous random variables whose copula is C . Then the *Spearman's* ρ is given by:

$$\rho_{X,Y}^S = \rho_C = 12 \iint_{\mathbf{I}^2} C(u, v) du dv - 3. \quad (2.33)$$

Spearman's ρ is somehow linked to the linear correlation. In Equation 2.29, we present a linear correlation by random vectors X, Y . Suppose X, Y are characterized by marginal distribution function $F(X), G(Y)$ and joint distribution $H(X, Y)$ with copula $C(u, v)$. If we substitute random vectors X, Y with their corresponding marginal probability measures $F(X), G(Y)$, we obtain another interpretation of Spearman's ρ :

$$\begin{aligned} \rho_{X,Y}^S &= \frac{\text{cov}(F(X), G(Y))}{\sqrt{\text{var}(F(X))\text{var}(G(Y))}} \\ &= \frac{\text{cov}(U, V)}{\sqrt{\text{var}(U)\text{var}(V)}} \\ &= \frac{\mathbb{E}[UV] - \mathbb{E}[U]\mathbb{E}[V]}{\sqrt{\text{var}(U)\text{var}(V)}}. \end{aligned} \quad (2.34)$$

Since in copula $C(u, v)$, U, V are uniformly distributed, $\mathbb{E}[U] = \mathbb{E}[V] = 1/2$, $\text{var}[U] = \text{var}[V] = 1/12$, we can rewrite Equation 2.34 in the form of Equation 2.33. In this sense, Spearman's ρ can be perceived as the linear correlation of probability-transformed random vectors (Cherubini et al., 2011):

$$\rho^S(X, Y) = \rho^L(F(X), G(Y)).$$

Pearson's linear correlation, Kendall's τ and Spearman's ρ are all symmetric dependence measures ranging from -1 to 1 . If the random vectors are mutually independent, they all take the value 0 , but not visa versa. As we mentioned above, rank correlations do not depend on marginal distributions. Hence, they can be estimated by the *ranks* of empirical

data alone. In addition, due to the connection to copulas, rank correlations can capture nonlinear dependence which linear correlation can not offer. In practical applications, rank correlations are quite useful in calibrating copulas to data (McNeil et al., 2005).

Previously we mentioned financial time series are often not distributed as Normal with extreme correlations, presenting fat tails and other extreme correlations. The association between extreme values can be described by tail dependence. In other words, tail dependence measures the dependence between tail distributions of random vectors. The coefficients of tail dependence can be defined using the concept of "limiting conditional probabilities of *quantile exceedances*" (McNeil et al., 2005). A detailed definition is given by Nelsen (2006):

Definition 2.1.28. Let X, Y denote two continuous random variables with distribution functions F and G , respectively. The coefficient of *upper tail dependence* λ_U is defined as the limit (if it exists) of the conditional probability that Y exceeds the t -th quantile of G given that X exceeds the t -th quantile of F as t approaches 1, i.e.

$$\lambda_U = \lim_{t \rightarrow 1^-} \mathbb{P} \left[Y > G^{(-1)}(t) | X > F^{(-1)}(t) \right]. \quad (2.35)$$

Likewise, the coefficient of *lower tail dependence* λ_L is defined as the limit (if it exists) of the conditional probability that Y is less than or equal to the t -th quantile of G given that X is less than or equal to the t -th quantile of F as t approaches 0, i.e.

$$\lambda_L = \lim_{t \rightarrow 0^+} \mathbb{P} \left[Y \leq G^{(-1)}(t) | X \leq F^{(-1)}(t) \right]. \quad (2.36)$$

Similar to rank correlations, tail dependence depends only on the copula of random vectors. Equation 2.35 and 2.36 can be rewritten in terms of copulas. Let C be the copula of X, Y in Definition 2.1.28, then for lower tail dependence λ_L we have:

$$\begin{aligned} \lambda_L &= \lim_{t \rightarrow 0^+} \mathbb{P} \left[Y \leq G^{(-1)}(t) | X \leq F^{(-1)}(t) \right] \\ &= \lim_{t \rightarrow 0^+} \mathbb{P} [G(Y) \leq t | F(X) \leq t] \\ &= \lim_{t \rightarrow 0^+} \frac{\mathbb{P} [G(Y) \leq t, F(X) \leq t]}{\mathbb{P} [F(X) \leq t]} \\ &= \lim_{t \rightarrow 0^+} \frac{C(t, t)}{t} \end{aligned} \quad (2.37)$$

$$= \delta'_C(0^+), \quad (2.38)$$

where $\delta_C(\cdot)$ is the diagonal section of copula (see Definition 2.1.10).

Accordingly, the upper tail dependence λ_U can be rewritten as:

$$\begin{aligned}\lambda_U &= \lim_{t \rightarrow 1^-} \mathbb{P} \left[Y > G^{(-1)}(t) | X > F^{(-1)}(t) \right] \\ &= \lim_{t \rightarrow 1^-} \frac{1 - 2t + C(t, t)}{1 - t}\end{aligned}\tag{2.39}$$

$$= 2 - \delta'_C(1^-).\tag{2.40}$$

By looking at Equation 2.37 and Equation 2.39, we understand the way a copula interprets tail dependences. Copula C admits upper (or lower) tail dependence if λ_U (or λ_L) takes value in $(0, 1]$, and presents no upper (or lower) tail dependence if $\lambda_U = 0$ (or $\lambda_L = 0$). In the last part of this subsection, we will present the expressions of tail dependence of some copulas discussed before.

The bivariate Gumbel and Clayton family of Archimedean copulas admits upper and lower tail dependence respectively. Suppose that the bivariate Archimedean copulas are indexed by one parameter θ , then the coefficient of upper tail dependence of Gumbel copula is given by:

$$\lambda_U^{Gu} = \lim_{t \rightarrow 1^-} \frac{1 - 2t + C_\theta^{Gu}(t, t)}{1 - t} = 2 - \lim_{t \rightarrow 1^-} \frac{C_\theta^{Gu}(t, t) - 1}{t - 1} = 2 - 2^{1/\theta}.\tag{2.41}$$

Accordingly, the coefficient of lower tail dependence of Clayton copula is given by:

$$\lambda_L^{Cl} = \lim_{t \rightarrow 0^+} \frac{C_\theta^{Cl}(t, t)}{t} = 2^{-1/\theta}.\tag{2.42}$$

Neither Gumbel copula nor Clayton copula admits tail dependence on the other end. Frank copula shows no tail dependence on both ends. Due to the property of uniquely presenting tail dependence on one side, Gumbel copula and Clayton copula are frequently implemented in studying asymmetric distributions. For example, Gumbel copula enables us to model the association of extreme gains between two assets while Clayton copula enables us to handle the case of extreme losses.

Gaussian copulas show no tail dependence on two ends unless the correlation coefficient ρ equals 1. In this extreme case, $\lambda_U = \lambda_L = 1$. This is called the asymptotic independence of the Gaussian copula (McNeil et al., 2005).

In contrast, Student t -copula presents both upper and lower tail dependence of the same magnitude. Follow the settings in Definition 2.1.22, the symmetrical tail dependence of t -copula is given by:

$$\lambda_U^t = \lambda_L^t = 2t_{\nu+1} \left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}} \right).\tag{2.43}$$

Given $\rho > -1$, bivariate t -copula shows asymptotic dependence in both tails (also called radial symmetry (McNeil et al., 2005)).

Up to now, we have introduced three main dependence measures: linear correlation, rank correlations and tail dependence. Linear correlation mainly captures the linear relationship between random vectors while rank correlations and tail dependence focus more on the non-linear association between random vectors. The rank correlations are designed to describe the concordance while tail dependence measures the strength of association in the tails of a bivariate distribution. Dependence measures are closely linked to copulas in different aspects and their coefficients can be estimated from the data. In the next subsection, we will quickly introduce the basic technique of calibrating copula from data.

2.1.4 Fitting Copulas to Data

As we presented above, copulas are good expressions for multivariate distributions. If we can obtain an accurate estimation of the copula parameters from raw data, we say the copula is properly fitted to the data. Currently, the most popular methodology that can be applied is the *maximum likelihood estimation* (**MLE**). The exhaustive theory of MLE is too long (usually due to the needs of solving complex numerical optimizations and mixed derivatives involved in likelihood (Cherubini et al., 2004)) for this thesis and out of its purpose. Here we provide the basic idea of implementing the method.

Recall that the multivariate distribution functions given in Equations 2.1, 2.2 and the copula density given in Equation 2.19. In a d -dimensional random vector, we have:

$$f(x_1, x_2, \dots, x_d) = c(F_1(x_1), F_2(x_2), \dots, F_d(x_d)) \cdot \prod_{i=1}^d f_i(x_i), \quad (2.44)$$

where c is the copula density of $F(\mathbf{X})$, f is the joint density of \mathbf{X} and f_i is the marginal density of X_i .

Based on Equation 2.44, a scheme for estimating copula could generally be split into two steps:

1. Estimation of the marginal distributions of raw data;
2. Estimation of proper copula parameters via MLE.

In practice, we rarely observe copula data directly. Hence, to estimate copulas, we have to model their marginal distributions first, i.e. obtaining estimation of F_i and f_i for $i =$

$1, \dots, d$. Then, we transform the sample marginal distribution $(\hat{X}_1, \dots, \hat{X}_d)$ into standard uniform distribution $(\hat{U}_1, \dots, \hat{U}_d)$ and estimate copula parameters through MLE.

Suppose that we have N observations for each marginal random variable. Thus, the log-likelihood function is given by:

$$l(\boldsymbol{\theta}) = \sum_{j=1}^N \ln c(F_1(x_{1j}), \dots, F_d(x_{dj})) + \sum_{j=1}^N \sum_{i=1}^d \ln f_i(x_{ij}), \quad (2.45)$$

where $\boldsymbol{\theta}$ denotes all of the parameters (including those of both copula and marginal p.d.f.s) to be estimated. Given the forms of sample marginal p.d.f.s f_i and a target copula function, the maximum likelihood estimator is then given by:

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \max_{\boldsymbol{\theta} \in \Theta} l(\boldsymbol{\theta}). \quad (2.46)$$

Moreover, in a so-called *Canonical Maximum Likelihood (CML)* method (Cherubini et al., 2004) where only copula parameters are to be estimated, the log-likelihood 2.45 could simply be written as:

$$\ln L(\boldsymbol{\theta}; \hat{U}_1, \dots, \hat{U}_d) = \sum_{i=1}^d \ln c_{\boldsymbol{\theta}}(\hat{U}_i), \quad (2.47)$$

where $\boldsymbol{\theta}$ is the set of copula parameters to be estimated in C , and \hat{U}_i denotes the observations of copula transformed from marginals \hat{X}_i . By maximizing CML 2.47, we obtain the MLE $\boldsymbol{\theta}$.

There are other methods that can be used to estimate the copula parameters, e.g non-parametric estimation, Method-of-Moment using rank correlation, eigenvalue method and so on (McNeil et al., 2005). As to MLE, the task of estimation presents difficulties in high dimensions as it carries a heavy computational load. Decomposing the whole task into two steps of estimating marginals and copulas separately allow us to pick the statistical model that best fits marginal distribution and plug in the copula function.

2.2 Social Network Clustering

This section is devoted to the question on how to view the dependence structure between assets' returns under a social network framework. We first look at the basic concepts of social network analysis and discuss the essential tools related to this thesis. Then, we introduce a data clustering technique proposed by Frey and Dueck (2007), in order to separate assets into different groups.

2.2.1 Introduction of Social Network Analysis

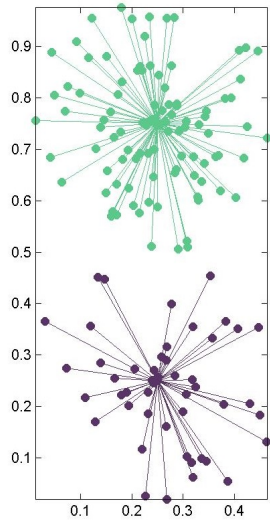
As often used in studying social science in the past few decades, *Social Network Analysis* (SNA) has various definitions. In a way, SNA is a branch of Network Science, whose objective is to understand networks emerging in nature, technology and society using a unified set of tools and principles (Du, 2014). Wasserman and Faust (1994) argue that SNA focuses on the importance of relationships among interacting units and contains theories, models, and applications which involves interpretations of relational concepts or processes. They also mention that the units in SNA is not the individual, but an entity consisting of a collection of individuals and the linkages among them. Briefly speaking, SNA could be seen as the mapping and measuring of relationships and flows between entities. The term **entities** could represent different concepts and in our case, they are the stocks trading in market. Intuitively, we know there are no independent stocks and we want to study the patterns of ties among them.

Social network analysis can be performed by mathematical models or statistical methods, and in this thesis, we mainly need two concepts from them: *Graph Theory* and *Network Structure*. There are many common terminologies used in both concepts such as *centrality* and *clustering*.

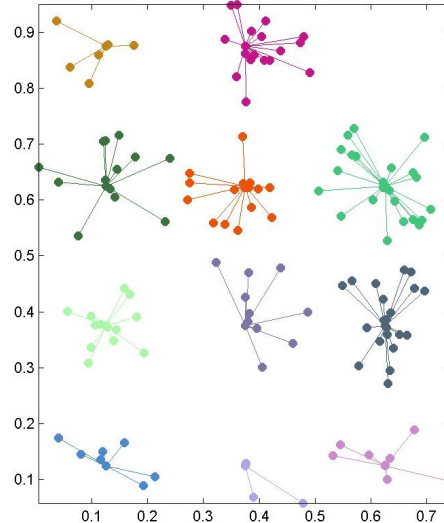
Graph theory generally speaks of the study of *graphs*. We could view a graph as a mathematical representation of a network modeling pairwise relations between entities. A graph basically consists of a set of:

1. *Nodes* or *Vertices* which stand for the individuals in the whole network;
2. *Edges* or *Arcs* that connect certain pairs of nodes.

In this thesis, we set each node as an individual security. In this sense, the edges connecting the nodes (securities) could be seen as a certain pattern of the dependence structure among them, e.g. correlations or tail dependence. Note that a graph may be directed, meaning that there is a direction on the edge associated with a certain pair of nodes. But we can skip this concern since we perceive dependence structure as a mutually equivalent relationship among two financial assets. For instance, we show in Figure 2.1 two sample graphs of stock networks. Each node on both graphs represents a certain stock trading in the market. The edges illustrate linear relationships among them. Note that there should be such ties between all of the nodes, the above figures are designed to facilitate the understanding of clustering, which reflects a relatively strong relationships inside one group.



(a) 2 clusters of stocks



(b) 11 clusters of stocks

Figure 2.1: Sample graphs of stock network after clustering based on linear correlation.

Alternatively, graphs can be represented in a quantified way. One of the basic method for mathematically describing a social network is the *Adjacency Matrix*:

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{pmatrix}.$$

In the above matrix, each “1” means there exists a certain pattern of relationship (or edge) between the corresponding two entities (or nodes). Likewise, “0” means no ties among them. Since we are studying such ties between nodes but nodes themselves, all diagonal cells are set to “0”. One drawback of this kind of adjacency matrix is that there could exists too many “0” entries, thus making the density of social network quite low. To deal with

this issue, we alter the above matrix so that it could be interpreted as a “valued graph”:

$$A = \begin{pmatrix} 0 & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & 0 & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & 0 \end{pmatrix}. \quad (2.48)$$

In this new matrix, each cell $a_{i,j}$ stands for a quantified relationship between nodes i and j , e.g, a linear correlation ranging from -1 to 1 among returns of two securities. With the knowledge that all stocks are correlated with each other at certain levels (we will show it later in Chapter 3), almost all entries except diagonals are filled with nonzero values.

Alternatively, each $a_{i,j}$ in the adjacency matrix A could be transformed and then viewed as a measure of *distance* between nodes on a graph. In other words, we are looking at how close or how far the nodes are to each other instead of directly saying whether there exists ties among them. We mainly refer to the *Euclidean distance* which measures nodes’ *similarity*. The similarity can be obtained by a certain transformation of each cell in the above adjacency matrix A . For instance, suppose that each $a_{i,j}$ in A is originally given by $\rho_{i,j}^L$ (2.29), the linear correlation between certain entities i and j as an indicator of similarity. Then we assign $a'_{i,j} = \sqrt{2(1 - a_{i,j})}$. We can easily tell that the more positively correlated a pair of nodes is, the shorter their distance is and vice versa. In this sense, the new matrix A' of $a'_{i,j}$ is treated as a numerical expression of similarity among all nodes. In addition, when we study this pattern of relationship, it is always symmetrical, thus making $a'_{i,j} = a'_{j,i}$. This means that we only need to learn the network from a highly dense triangular matrix. The intriguing feature of this transformation is that all types of parameters measuring dependence introduced in 2.1.3 can be inserted into the adjacency matrix, yielding new similarity matrices which capture the specific dependence structure in place of distance. We will introduce afterwards how it contributes to the process of clustering.

Now we describe the basic features of graphs for social network analysis. The building block of **SNA** is the network structure. Imagine a case in which we can numerically explain entities’ relationships by means of a similarity matrix. We might be interested in the question: “which is(are) the most important or central entity(entities) in a network?” This question may have multiple answers, depending on what we mean by importance. It naturally leads to the term *centrality* in network structure. There are various *centrality measures*, e.g. degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, page rank centrality etc. In this thesis, we are particularly interested in closeness centrality since it analogously follows the content of similarity matrix. Generally, closeness

centrality in a network states that the more central a node is, the lower its total distance to all other nodes.

To be able to locate those “central nodes” on a social network, we will need assistance from the clustering technique. Strictly speaking, we will divide all nodes into certain groups so that within each group all nodes are geometrically close while the nodes from different groups are relatively distant. Such groups are called *clusters*. Then we will find a centrality within each cluster. Such centrality is treated as a representative locally since it is close or strongly tied to all of the remaining nodes inside its cluster.

The ordinary clustering methods include *hierarchical clustering*, *k-means clustering*, *fuzzy C-means clustering* and etc. Each of them possesses certain advantages and limitations. In applications, we often face the clustering problem with enormous data and the number of potential clusters is unknown. Due to these reasons, Frey and Dueck (2007) devised *clustering by affinity propagation* which efficiently provides good solutions. In the next subsection, we briefly discuss this technique.

2.2.2 Clustering by Affinity Propagation

A typical task in data clustering is to identify certain “central nodes” in the whole network. A traditional way to accomplish this is the *K-means Clustering*. Briefly speaking, we first assign a presumed number of clusters and randomly pick the same amount of nodes from all data points which are treated as initial centrality. Then we construct initial clusters based on these nodes and look for a set of better centrality within each cluster. We recursively refine the results by this step until we reach a stable solution. However, we could obtain some very distinguished solutions from the same experiment when choosing initial choices randomly¹. To get a convincing result requires picking the nodes closest to the “real centrality”. In addition, the number of clusters we presume in advance might not be coherent with the real data structure. The above two issues are the major drawbacks of K-means Clustering. The *Clustering by Affinity Propagation (AP)* (Frey and Dueck, 2007) effectively resolves those issues since it does not assume the amount of clusters and treats all points as potential centrality (which are given the name *exemplars* in their paper).

AP considers a measure of similarities between pairs of nodes in a network. Suppose that we have n nodes in total with a $n \times n$ similarity matrix S taking the form as in 2.48 (note that the diagonal entries will no longer be zeros). Then each entry $s(i, k)$ in the matrix is a real value indicating to what degree is node k suitable as the exemplar for node

¹In K-means Clustering, distinguished solutions could exist due to the structure of raw data.

i. Such $s(i, k)$ could be obtained by different applications. One typical example is the *Euclidean distance* introduced in Section 2.2.1, in which $s(i, k)$ takes value as $\sqrt{2(1 - \rho_{i,k}^L)}$. Meanwhile, the feature of not prescribing the number of clusters in **AP** is realized by the diagonal elements $s(k, k)$. They are not taking the value of 0. Instead, they take real values referred to as “*preferences*” with notion $p(k)$ ¹ for node k . Certain node k ($k = 1, 2, \dots, n$) associated with a greater input value $s(k, k)$ (greater “*preference*” $p(k)$) possesses a higher probability of being chosen as an exemplar. Alternatively, given that the data structure is known at the beginning, we can assign a common value to all $s(k, k)$ ($k = 1, 2, \dots, n$), meaning all nodes share the same probability of being identified as exemplars initially. In this case, assigning a great value of preferences² initially will result in a large number of clusters at last.

Then, the mechanism of **AP** is built upon two kinds of messages transferred between all nodes. Frey and Dueck (2007) define two terms to interpret them, the “*responsibility*” with notion $r(i, k)$ and “*availability*” with notion $a(i, k)$. The “*responsibility*” $r(i, k)$ is the message sent from node i to a potential exemplar k . It works as a collection of evidence and indicates how good is node k at serving as the exemplar for node i , considering all other potential exemplars for i . On the other hand, the “*availability*” $a(i, k)$ is the message sent from a potential exemplar k to node i . It is also a collection of evidence telling the fact to what extent it will be appropriate for node i to choose node k as its exemplar, considering all other nodes choosing node k as their exemplar. $r(i, k)$ shares an accumulative relationship with $a(i, k)$. It can be displayed by iterating the following rules:

$$\begin{cases} r(i, k) \leftarrow s(i, k) - \max \{a(i, k') + s(i, k')\}, & k' \text{ s.t. } k' \neq k, \\ a(i, k) \leftarrow \min \{0, r(k, k) + \sum \max \{0, r(i', k)\}\}, & i' \text{ s.t. } i' \neq i, k. \end{cases} \quad (2.49)$$

Initially, we will set all availabilities $a(i, k)$ at 0. Hence, during the initial iteration of $r(i, k)$ in 2.49, it does not take into consideration that the rest of the nodes will choose other exemplars. After a few iterations, some nodes may find it better to be assigned to other exemplars rather than node k . This can be shown in the second rule of 2.49: their availabilities $a(i, k)$ become negative. In return, certain negative values $a(i, k')$ will weaken the influences of their pairwise similarities $s(i, k')$ in the first rule, kicking the corresponding potential exemplars out of the game. In order to control the positive responsibilities in the second rule, a minimum threshold is added to assure that $a(i, k)$ is not greater than 0. When it comes to a special case such as $i = k$, Frey and Dueck (2007) name $r(k, k)$ and

¹The “*preference*” $p(k)$ means the initial value assigned to the diagonal element $s(k, k)$ in the similarity matrix S . The input value of $p(k)$ will have an influence on the potential exemplar node k .

² $p(k)$ usually takes value from the interval of similarities.

$a(k, k)$ as “self-responsibility” and “self-availability” respectively. They are given by:

$$\begin{cases} r(k, k) \leftarrow s(k, k) - \max \{s(k, k')\}, & k' \text{ s.t. } k' \neq k, \\ a(k, k) \leftarrow \sum \max \{0, r(i', k)\}, & i' \text{ s.t. } i' \neq k. \end{cases} \quad (2.50)$$

We implement Equations 2.49 and 2.50 iteratively to update responsibilities and availabilities. At any stage of **AP**, responsibilities and availabilities are put together to choose exemplars for the whole network. Frey and Dueck (2007) claim that for node i , the node k associated with the greatest value of $a(i, k) + r(i, k)$ will be its suited exemplar and **AP** procedure could be terminated by different options, e.g. returning stable solution after some iterations, obtaining a desired result based on some conditions and etc. When updating $r(i, k)$ and $a(i, k)$ during iterations, we might encounter numerical oscillations. To avoid such situations, a damping factor λ is introduced so that in each iteration step j , $r(i, k)$ and $a(i, k)$ will be updated with their previous values, i.e.:

$$\begin{cases} r(i, k)_{(j)} = (1 - \lambda) \times r(i, k)_{(j)} + \lambda \times r(i, k)_{(j-1)}, \\ a(i, k)_{(j)} = (1 - \lambda) \times a(i, k)_{(j)} + \lambda \times a(i, k)_{(j-1)}, \end{cases} \quad (2.51)$$

where λ usually takes value in $[0, 1]$ and takes 0.5 as a default value.

In conclusion, each iteration of **AP** consists of three steps:

1. updating all $r(i, k)$ based on availabilities and similarities;
2. updating all $a(i, k)$ based on responsibilities;
3. computing $a(i, k) + r(i, k)$ to identify possible exemplars.

After each step the solution will be monitored to decide whether **AP** could be terminated.

2.2.3 Between-Within Proportion

In applications we usually use a criterion to evaluate the clustering result, such as the *Silhouette Value* which measures how similar a node is to other nodes in its own cluster compared to nodes in other clusters. The silhouette value ranges in $[-1, 1]$ and indicates a well-suited result when taking high value. When the majority of the nodes indicate good silhouette values, the clustering solution is deemed as acceptable (Kaufman and Rousseeuw, 2005). The silhouette value has been widely proved to be useful and reliable when associated

with K-means clustering, but maybe not so seamless with **AP** (Zhou et al., 2010b) (Zhou et al., 2011). Zhou et al. (2010a) devise a new criterion called *Between-Within Proportion* (**BWP**) for clustering assessment. Zhou et al. (2011) revise **BWP** to make it particularly fit for working with **AP** clustering algorithm. Here we briefly interpret the definitions of the criterion.

Definition 2.2.1. Let $K = \{X, R\}$ denote the clustering space, where $X = (x_1, \dots, x_n)$ stand for all the nodes in a network. Suppose that n nodes are grouped into c clusters, then we define the minimal average distance between the i th node in the j th cluster and all of the rest nodes in its cluster as *Between Distance*, given by $bd(j, i)$:

$$bd(j, i) = \min_{1 \leq k \leq c, k \neq j} \left(\frac{1}{n_k} \sum_{p=1}^{n_k} \|x_p^{(k)} - x_i^{(j)}\| \right), \quad (2.52)$$

where k, j are indexes for clusters. $x_i^{(j)}$ stands for the i th node in the j th cluster and $x_p^{(k)}$ stands for the p th node in the k th cluster. n_k indicates the total number of nodes in cluster k and $\|\cdot\|$ represents a *Euclidean distance*.

Definition 2.2.2. Given same preliminaries as in Definition 2.2.1, we define the average distance between the i th node in cluster j and all the other nodes in cluster j as the *Within Distance*, denoted by $wd(j, i)$:

$$wd(j, i) = \frac{1}{n_j - 1} \sum_{q=1, q \neq i}^{n_j} \|x_q^{(j)} - x_i^{(j)}\|, \quad (2.53)$$

where the settings of notations are identical to those used in Definition 2.2.1.

Definition 2.2.3. Given same preliminaries as Definition 2.2.1, we define the clustering distance for the i th node in cluster j as sum of $bd(j, i)$ and $wd(j, i)$, namely *Between-and-Within Distance* and denoted by $bawd(j, i)$:

$$\begin{aligned} bawd(j, i) &= bd(j, i) + wd(j, i) \\ &= \min_{1 \leq k \leq c, k \neq j} \left(\frac{1}{n_k} \sum_{p=1}^{n_k} \|x_p^{(k)} - x_i^{(j)}\| \right) + \frac{1}{n_j - 1} \sum_{q=1, q \neq i}^{n_j} \|x_q^{(j)} - x_i^{(j)}\|, \end{aligned} \quad (2.54)$$

where the settings of notations are identical to those used in Definition 2.2.1.

Definition 2.2.4. Given same preliminaries as those in Definition 2.2.1, we define the clustering deviation distance for the i th node in cluster j as difference between $bd(j, i)$ and

$wd(j, i)$, denoted by $bswd(j, i)$:

$$\begin{aligned} bswd(j, i) &= bd(j, i) - wd(j, i) \\ &= \min_{1 \leq k \leq c, k \neq j} \left(\frac{1}{n_k} \sum_{p=1}^{n_k} \|x_p^{(k)} - x_i^{(j)}\| \right) - \frac{1}{n_j - 1} \sum_{q=1, q \neq i}^{n_j} \|x_q^{(j)} - x_i^{(j)}\|, \end{aligned} \quad (2.55)$$

where the settings of notations are identical to those used in Definition 2.2.1.

Definition 2.2.5. Given same preliminaries as those in Definition 2.2.1, we define the *Between-Within Proportion* for the i th node in cluster j as ratio of $bswd(j, i)$ to $bawd(j, i)$, denoted by $BWP_d(j, i)$:

$$\begin{aligned} BWP_d(j, i) &= \frac{bswd(j, i)}{bawd(j, i)} \\ &= \frac{bd(j, i) - wd(j, i)}{bd(j, i) + wd(j, i)} \\ &= \frac{\min_{1 \leq k \leq c, k \neq j} \left(\frac{1}{n_k} \sum_{p=1}^{n_k} \|x_p^{(k)} - x_i^{(j)}\| \right) - \frac{1}{n_j - 1} \sum_{q=1, q \neq i}^{n_j} \|x_q^{(j)} - x_i^{(j)}\|}{\min_{1 \leq k \leq c, k \neq j} \left(\frac{1}{n_k} \sum_{p=1}^{n_k} \|x_p^{(k)} - x_i^{(j)}\| \right) + \frac{1}{n_j - 1} \sum_{q=1, q \neq i}^{n_j} \|x_q^{(j)} - x_i^{(j)}\|}, \end{aligned} \quad (2.56)$$

where the settings of notations are identical to those used in Definition 2.2.1.

The previous definitions are built upon the concept of distance. In our thesis, since we aim at clustering securities based on the dependence structure among them, the similarities measured by linear correlations or tail dependence will be preferred. In fact, distance corresponds to non-similarity. Zhou et al. (2011) also offer the option of measuring **BWP** based on non-similarity. It simply alters Definition 2.2.1 and Definition 2.2.2.

Definition 2.2.6. Let $K = \{X, R\}$ denote the clustering space, where $X = (x_1, \dots, x_n)$ stand for all the nodes in a network. Suppose n nodes are grouped into c clusters, then we define the minimal average non-similarity between the i th node in the j th cluster and all of the rest nodes in its cluster as $bd(j, i)$:

$$bd(j, i) = \min_{1 \leq k \leq c, k \neq j} \left(\frac{1}{n_k} H(x_p^{(k)}, x_i^{(j)}) \right), \quad (2.57)$$

where k, j are indexes for clusters. $x_i^{(j)}$ stands for the i th node in the j th cluster and $x_p^{(k)}$ stands for the p th node in the k th cluster. n_k indicates the total number of nodes in cluster k and $H(\cdot)$ represents *non-similarity*.

Definition 2.2.7. Given same preliminaries as those in Definition 2.2.6, we define the average distance between the i th node in cluster j and all the other nodes in cluster j as the *Within Distance*, denoted by $wd(j, i)$:

$$wd(j, i) = \frac{1}{n_j - 1} \sum_{q=1, q \neq i}^{n_j} H(x_p^{(k)}, x_i^{(j)}), \quad (2.58)$$

where the settings of notations are same as those used in Definition 2.2.1.

Definition 2.2.8. Given same preliminaries as those used in Definition 2.2.6, we define the *Between-Within Proportion* based on non-similarity, denoted by $BWP_s(j, i)$:

$$\begin{aligned} BWP_s(j, i) &= \frac{bd(j, i) - wd(j, i)}{bd(j, i) + wd(j, i)} \quad (2.59) \\ &= \frac{\min_{1 \leq k \leq c, k \neq j} \left(\frac{1}{n_k} \sum_{p=1}^{n_k} H(x_p^{(k)}, x_i^{(j)}) \right) - \frac{1}{n_j - 1} \sum_{q=1, q \neq i}^{n_j} H(x_p^{(k)}, x_i^{(j)})}{\min_{1 \leq k \leq c, k \neq j} \left(\frac{1}{n_k} \sum_{p=1}^{n_k} H(x_p^{(k)}, x_i^{(j)}) \right) + \frac{1}{n_j - 1} \sum_{q=1, q \neq i}^{n_j} H(x_p^{(k)}, x_i^{(j)})}, \end{aligned}$$

where the settings of notations are same as those used in Definition 2.2.6.

BWP is able to reflect closeness within clusters and dispersion between them based on $bd(j, i)$ and $wd(j, i)$ (Zhou et al., 2011). The value of **BWP** ranges in $[-1, 1]$. Individually, a large value of $BWP_s(j, i)$ indicates a well-suited solution for a single node. As to the entire network, the greater the average value of all nodes' BWP is, the better solution the clustering algorithm indicates. This also helps to decide the best solution. For instance, if we were using **AP** towards a network and obtained several results with different number of clusters, then we would choose the optimal solution which maximize the average BWP value.

2.3 Mean-variance Framework for Portfolio Selection

A complete investment strategy requires a portfolio optimization. The primary purpose of this thesis is to select stocks based on their dependence structure but we also need to make evaluations on the portfolio selections. We firstly choose the basic mean-variance framework to accomplish the task. In fact, to emphasize the influence of clustering method, a naive strategy which equally allocates wealth among all assets may express the performance more clearly. However, applying mean-variance framework is more practical.

Assume our investment universe consists of N risky assets with T price periods. Let $S_i^{(p)}$ denote the price of asset i at time p , where $i = 1, 2, \dots, N$ and $p = 1, 2, \dots, T$. Then the *logarithmic return* of asset i at time p is denoted by:

$$r_i^{(p)} = \ln \left(\frac{S_i^{(p)}}{S_i^{(p-1)}} \right), \quad \text{where } p \geq 2. \quad (2.60)$$

We use a vector $\mathbf{r}_i = (r_i^{(1)}, r_i^{(2)}, \dots, r_i^{(T)})'$ to denote the logarithmic returns of asset i and let $R = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$ denote return matrix of all assets over the entire investment horizon. Then the i th asset's *mean return* and *variance of return* is respectively given by:

$$\mu_i = \mathbb{E}[\mathbf{r}_i], \quad (2.61)$$

$$\text{Var}(\mathbf{r}_i) = \sigma_i^2 = \mathbb{E}[(\mathbf{r}_i - \mathbb{E}[\mathbf{r}_i])(\mathbf{r}_i - \mathbb{E}[\mathbf{r}_i])']. \quad (2.62)$$

In a vectorized form for all assets, we have $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_N)'$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_N)'$. Following Equation 2.28, the covariance of returns between assets i and j is given by:

$$\text{Cov}(\mathbf{r}_i, \mathbf{r}_j) = \mathbb{E}[(\mathbf{r}_i - \mathbb{E}[\mathbf{r}_i])(\mathbf{r}_j - \mathbb{E}[\mathbf{r}_j])'] = \rho_{ij}\sigma_i\sigma_j, \quad (2.63)$$

where ρ_{ij} is the linear correlation between returns i and j . Let $\boldsymbol{\Sigma}$ denote the covariance matrix corresponding to R . Next, suppose that $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_N)'$ is an $N \times 1$ vector of portfolio proportions, such that ω_1 is the proportion of total portfolio wealth invested in asset i . It follows that the expected return of the portfolio is given by:

$$\mu_{\text{port}}(\boldsymbol{\omega}) = \boldsymbol{\omega}'\boldsymbol{\mu} = \sum_{i=1}^N \omega_i \mu_i, \quad (2.64)$$

and the total variance of the portfolio returns is given by:

$$\sigma_{\text{port}}^2(\boldsymbol{\omega}) = \boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} = \sum_{j=1}^N \sum_{i=1}^N \omega_i \omega_j \sigma_i \sigma_j \rho_{ij}. \quad (2.65)$$

Given all assets returns and the corresponding covariance matrix, the mean-variance framework (Markowitz, 1952) considered a portfolio to be optimal if it is risk minimal for a given level of return μ_{obj} ,

$$\begin{aligned} \min_{\boldsymbol{\omega}} \quad & \boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} \\ \text{s.t.} \quad & \boldsymbol{\omega}'\boldsymbol{\mu} = \mu_{\text{obj}} \\ & A\boldsymbol{\omega} = \mathbf{b}, \end{aligned} \quad (2.66)$$

where A and b define budget constraints¹. If we do not allow for short selling, there will be additional constraint $0 \leq \omega_i \leq 1, \forall i$. The inequality constraints can be integrated into A .

Alternatively, the optimal portfolio in the mean-variance framework could be constructed by maximizing its return for a given level of risk σ_{obj} ,

$$\begin{aligned} \max_{\omega} \quad & \omega' \mu \\ \text{s.t.} \quad & \omega' \Sigma \omega = \sigma_{obj} \\ & A\omega = b . \end{aligned} \tag{2.67}$$

The optimal portfolio can be obtained by solving either one of the problems, depending on the investment goal. Mathematically, if we do not take the inequality constraint into consideration, 2.66 works as a quadratic optimization problem with linear constraints. In contrast, 2.67 has a linear objective function with quadratic constraints.

If we choose to solve 2.66 with respect to different levels of μ_{obj} , then in a σ, μ space the optimal portfolios will form a curve which is called *efficient frontier*. The optimal portfolios are thus called *frontier portfolios* as well. Suppose that we have found optimal portfolio weights ω^* for a given μ_{obj} , then the variance of the frontier (optimal) portfolio is given by (Pennacchi (2007) pp.53-54):

$$\begin{aligned} \sigma_{port}^2(\omega^*) &= \omega^{*'} \Sigma \omega^* \\ &= \frac{\delta \mu_{obj}^2 - 2\alpha \mu_{obj} + \varsigma}{\varsigma \delta - \alpha^2} \\ &= \frac{1}{\delta} + \frac{\delta \left(\mu_{obj} - \frac{\alpha}{\delta} \right)^2}{\varsigma \delta - \alpha^2} . \end{aligned} \tag{2.68}$$

where $\alpha \equiv \mu' \Sigma^{-1} \mathbf{1}_N = \mathbf{1}'_N \Sigma^{-1} \mu$, $\varsigma \equiv \mu' \Sigma^{-1} \mu$, and $\delta \equiv \mathbf{1}'_N \Sigma^{-1} \mathbf{1}_N$ are scalars. Equation 2.68 stands for a parabola in a $\sigma_{port}^2, \mu_{port}$ space. In practice we often substitute the variance σ_{port}^2 with volatility σ_{port} and the curve becomes the efficient frontier.

2.4 Diversification Technique

Diversification techniques play an important role in reducing the chance of suffering great investment loss. The loss could be caused by the investment project being heavily exposed

¹Usually we set $A = \mathbf{1}'_N$ and $b = 1$ as default choice, where $\mathbf{1}'_N = (1, \dots, 1)$ is a N -dimensional vector of ones.

to some significant risk factors. In financial applications, diversification could be utilized to construct portfolios or to evaluate the risk of investment by some definitions. Since we would accomplish our major work of choosing stocks based on dependence structure in Chapter 3 and then build up corresponding portfolios by means of mean-variance framework in Chapter 4, the remaining task would be to evaluate the diversification of this investment. Note that if we say “measuring how well certain risks are diversified”, the “risks” are not restricted to volatilities of financial assets only. Instead, the term could refer to any abstract risk factors that exist within the investment universe.

In this thesis we choose to implement a diversification technique proposed by Meucci (2009). The methodology aims at describing portfolios in terms of uncorrelated risk source by applying *Principal Component Analysis*. It makes diversification easier to handle due to several reasons. In what follows, we briefly introduce how to obtain uncorrelated risk factors and what are the advantages of this technique.

2.4.1 Non-additive Risk Sources

Consider a portfolio $P(\boldsymbol{\omega})$ ¹ which consists of N risky assets with investor’s wealth allocation characterized by $\boldsymbol{\omega}$. The total return of the portfolio $P(\boldsymbol{\omega})$ is given by Equation 2.64, which is a sum of weight-adjusted returns of all risky assets. Then, the portfolio risk is characterized by the variance of total return, reads: Equation 2.65. Given an additional condition that all assets are perfectly uncorrelated in our portfolio, 2.65 can be rewritten in a form which consists of additive sources of risk:

$$\sigma_{port}^2(\boldsymbol{\omega}) = \boldsymbol{\omega}' \boldsymbol{\Sigma} \boldsymbol{\omega} = \sum_{i=1}^N \text{Var}(\omega_i \mu_i) = \sum_{i=1}^N \omega_i^2 \sigma_i^2. \quad (2.69)$$

In this case, the maximum diversification of $P(\boldsymbol{\omega})$ is easily achieved by equal volatility-adjusted weights. However, such a portfolio formed by mutually uncorrelated risky assets is not feasible in the real world. From the statistical aspect, the correlation structure of the assets can not be expressed by a diagonal matrix (d-fine GmbH, 2011). However, we are still attracted by the above amazing feature. If we are not able to find uncorrelated assets, maybe we can decompose the risk sources of them into uncorrelated parts which are therefore additive. In Meucci (2009), *Principal Component Analysis* is used to perform such decomposition.

¹ $\boldsymbol{\omega}$ follows the setting in subsection 2.3.

2.4.2 Principal Component Analysis

The aim of *Principal Component Analysis* (**PCA**) is to reduce the dimensionality of highly correlated data (McNeil et al., 2005). Given a covariance matrix Σ of return matrix R , **PCA** is concerned with using a few uncorrelated linear combinations to explain most of the variations in the structure of Σ . It could be also perceived as a data-rotation technique. Suppose that we have certain data points scattered on *Cartesian Coordinate System*. We rotate axes so that in a new coordinate system the data points will have the largest variance in their first coordinate. The primary idea of **PCA** is to find the principal components with maximal variances and are mutually uncorrelated.

The $N \times N$ covariance matrix Σ in 2.63 is symmetric. Hence, Σ is *orthogonally diagonalizable* and has an *orthonormal* set of N *eigenvectors*:

$$E' \Sigma E = \Lambda, \quad (2.70)$$

where $E = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$ is an $N \times N$ matrix whose columns form an orthonormal set of *eigenvectors* of Σ and $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_N\}$ is a diagonal matrix containing all the corresponding eigenvalues (normally $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$). Note that since E is orthogonal, $E' = E^{-1}$ or $E' E = I$. Then, by applying a spectral decomposition, we have:

$$\Sigma = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_N \mathbf{e}_N \mathbf{e}_N'. \quad (2.71)$$

With the above knowledge, the idea of **PCA** is that the i th principal component of return matrix $R = \{\mathbf{r}_1, \mathbf{r}_1, \dots, \mathbf{r}_N\}$ in 2.60 is the linear combination (Tsay, 2010):

$$\tilde{\mathbf{r}}_i = \mathbf{e}_i' R, \quad i = 1, \dots, N, \quad (2.72)$$

such that,

$$\text{Var}(\tilde{\mathbf{r}}_i) = \mathbf{e}_i' \Sigma \mathbf{e}_i = \lambda_i, \quad (2.73)$$

while between different principal components i and j ,

$$\text{Cov}(\tilde{\mathbf{r}}_i, \tilde{\mathbf{r}}_j) = \mathbf{e}_i' \Sigma \mathbf{e}_j = 0. \quad (2.74)$$

It indicates that the principal component $\tilde{\mathbf{r}}_i$ and $\tilde{\mathbf{r}}_j$ of R are mutually uncorrelated. **PCA** makes the variances of each principal component as large as possible with the constraint that $\mathbf{e}_i' \mathbf{e}_i = 1$. It follows that the first principal component $\tilde{\mathbf{r}}_1 = \mathbf{e}_1' R$ accounts for the largest variance and it is denoted as λ_1 . λ_2 accounts for the second largest variance among them, and the N th principal component has minimum variance among them. The variances of principal components are exactly the eigenvalues of the corresponding eigenvector. Let

$\tilde{R} = \{\tilde{\mathbf{r}}_1, \tilde{\mathbf{r}}_2, \dots, \tilde{\mathbf{r}}_N\}$ denote a new matrix consisting of principal components. Then it can be proved that

$$\text{Var}(R) = \text{Var}(\tilde{R}) = \text{tr}(\Sigma) = \sum_1^N \lambda_i = \sum_1^N \text{Var}(\tilde{\mathbf{r}}_i), \quad (2.75)$$

and

$$\frac{\text{Var}(\tilde{\mathbf{r}}_i)}{\text{Var}(R)} = \frac{\lambda_i}{\lambda_1 + \dots + \lambda_N}. \quad (2.76)$$

It means that the total variance of R and \tilde{R} are the same while the variances are additive in terms of uncorrelated principal component. Moreover, by Equation 2.76, we can compute how much portion of total variance each component contributes to.

In risk analysis, we perform **PCA** to find first few principal components (risk factors) that account for the major sources of uncertainty of a market. Then each of the major risk factors are extracted for further analysis.

2.4.3 Diversification Distribution

In the context of investment study, the orthonormal eigenvectors given by 2.70 form a set of N uncorrelated portfolios. Each eigenvector \mathbf{e}_i is given the name *principal portfolio* (Meucci, 2005). Therefore, the returns of principal portfolios read:

$$\tilde{R} = E^{-1}R. \quad (2.77)$$

According to the features of **PCA**, variances of the principal portfolios are reduced as their corresponding eigenvalues λ_i decrease. Meucci (2009) mentions that the such principal portfolios exist for any market with a well-defined covariance matrix.

Progressively, the portfolio $P(\boldsymbol{\omega})$ defined in Subsection 2.4.1 can be replicated as a linear combination of the uncorrelated principal portfolios, whose weights are given by (see also Partovi and Caputo (2004)):

$$\tilde{\boldsymbol{\omega}} = E^{-1}\boldsymbol{\omega}. \quad (2.78)$$

The “substituting weights” $\tilde{\boldsymbol{\omega}}$ is obtained by applying **PCA** to the original portfolio portions $\boldsymbol{\omega}$. $\tilde{\boldsymbol{\omega}}$ therefore represents the set of linear combination coefficients allocated to the principal portfolios.

Meucci (2009) follows this path and introduces the *variance concentration curve*:

$$v_i \equiv \tilde{\omega}_i^2 \lambda_i, \quad i = 1, 2, \dots, N, \quad (2.79)$$

where v_i represents the variance of the i th weighted principal portfolio. Once again, all weighted principal portfolios are uncorrelated, so that the total portfolio variance is:

$$\text{Var}(P(\boldsymbol{\omega})) = \boldsymbol{\omega}' \boldsymbol{\Sigma} \boldsymbol{\omega} = \boldsymbol{\omega}' E \boldsymbol{\Lambda} E' \boldsymbol{\omega} = \tilde{\boldsymbol{\omega}}' \boldsymbol{\Lambda} \tilde{\boldsymbol{\omega}} = \text{Var}(P(\tilde{\boldsymbol{\omega}})) = \sum_{i=1}^N \tilde{\omega}_i^2 \lambda_i = \sum_{i=1}^N v_i. \quad (2.80)$$

Next, let $\sigma(\tilde{\boldsymbol{\omega}})$ denote the standard deviation of $P(\tilde{\boldsymbol{\omega}})$. Then the *volatility concentration curve* is defined as:

$$s_i = \frac{v_i}{\sigma(\tilde{\boldsymbol{\omega}})} = \frac{\tilde{\omega}_i^2 \lambda_i}{\sqrt{\sum_{i=1}^N \tilde{\omega}_i^2 \lambda_i}}, \quad i = 1, 2, \dots, N. \quad (2.81)$$

The above expression actually implies the impact of changes in principal weights on variance contributions of the corresponding weighted portfolios. In fact, 2.81 is referred to as the decomposition of volatility or tracking error with respect to the contributions of each weighted principal portfolio in [Litterman \(1996\)](#).

Likewise, [Meucci \(2009\)](#) defines the *diversification distribution* as follows:

$$p_i = \frac{v_i}{\text{Var}(\tilde{\boldsymbol{\omega}})} = \frac{\tilde{\omega}_i^2 \lambda_i}{\sum_{i=1}^N \tilde{\omega}_i^2 \lambda_i}, \quad i = 1, 2, \dots, N. \quad (2.82)$$

In one way, the above expression could be regarded as a percentage measure of total variance each weighted principal portfolio contributes to. In another way, each p_i is equivalent to the “*r-square* from regression of total portfolio return on corresponding weighted principal portfolios” (see [Meucci \(2009\)](#)).

Given above definitions, we may have an intuitive idea of diversification that each weighted principal portfolio should have equal influence on the total portfolio risk. Since we know all components are mutually uncorrelated, well diversified investments would be some allocations of principal portfolios associated with uniform diversification distributions in terms of 2.82 (see also [Xiong \(2009\)](#) [d-fine GmbH \(2011\)](#)).

In real financial markets, we sometimes require portfolio management measured against a benchmark with weights \mathbf{b} . [Meucci \(2009\)](#)’s methodology is applicable to such a case by a simple modification:

$$\boldsymbol{\omega} \mapsto \boldsymbol{\omega} - \mathbf{b}, \quad (2.83)$$

where we replace original portfolio weights $\boldsymbol{\omega}$ with the vector of the *relative bets* $\boldsymbol{\omega} - \mathbf{b}$. 2.81 and 2.82 are renamed as *tracking error concentration curve* and *relative diversification distribution* respectively.

2.4.4 Entropy as a Diversification Risk Measure

It will be convenient to express the level of diversification in terms of a single value. Recall that from *diversification distribution* 2.82:

$$\sum_{i=1}^N p_i = 1, \quad \text{with } 0 \leq p_i \leq 1 \text{ for all } i, \quad (2.84)$$

and we state approximately equal probability masses p_i to indicate a well-diversified portfolio. The above settings of p_i naturally comply with the concept of *entropy*, which originated from thermodynamics but was adapted to information theory as a measure of uncertainty in a system (Shannon, 1948). Following the definition in Cover and Thomas (2006), the entropy of the entire portfolio characterized by *diversification distribution* 2.82 is given by:

$$H(P(\tilde{\omega})) = - \sum_{i=1}^N p_i \ln p_i. \quad (2.85)$$

The concept of entropy itself can work as a risk measure. Philippatos and Wilson (1972) regard entropy as a substitute of variance in looking for efficient portfolios. Subsequently, the *principal of maximum entropy* (MEP) becomes popular. Loosely speaking, MEP states that the probability density function with the largest entropy is the best fit to the situation with current state of knowledge and constraints. Such applications of entropy in finance are reviewed in Zhou et al. (2013).

Back to diversification, Meucci (2009) interprets it as the exponential of *diversification distribution's* entropy:

$$\mathcal{N}_{Ent} = \exp \left(- \sum_{i=K+1}^N p_i \ln p_i \right), \quad (2.86)$$

where K is the number of current existing constraints ($K = 0$ for the unconditional case). Such \mathcal{N}_{Ent} is coined as the “number of effective uncorrelated relative bets” by Meucci. We can treat the value of \mathcal{N}_{Ent} as an indicator of the current level of diversification. In a generic portfolio $P(\omega)$ consisting of N risky assets, $\mathcal{N}_{Ent} = 1$ means that the $P(\tilde{\omega})$ is entirely concentrated in one principal direction or the total risk is completely generated from the first principal portfolio alone. In such a case we see a sharp peak on diversification distribution and the portfolio is ill-diversified. On the contrary, $\mathcal{N}_{Ent} = N$ ($\mathcal{N}_{Ent} = N - K$ in conditional cases) means that the total risk of portfolio is equally spread among all N principal portfolios and the diversification distribution illustrates a perfect uniformity with $p_i = 1/N$. Indeed, the number of effective uncorrelated relative bets \mathcal{N}_{Ent} utilizes

principal of maximum entropy that the maximal degree of diversification is achieved when an entropy function of probability mass distribution p_i maximized its value.

Such a diversification technique built upon **MEP** is not unique. [Bera and Park \(2008\)](#) study an optimal portfolio with diversification based on **MEP**. They propose a cross-entropy measure and look for shrinkage estimation of portfolio weights. Using [Meucci \(2009\)](#) the words, “Bera and Park (2008) act on the portfolio weights, and thus do not account for the volatilities and the correlations in the market.” However, the portfolio optimization terminology *Mean-Diversification Efficient Frontier* is also applicable to [Meucci \(2009\)](#)’s work.

In the investment universe defined in 2.3, the mean-diversification efficient frontier can be described as:

$$\begin{aligned} \max_{\omega_\varphi} \quad & \mathcal{N}_{Ent}(\omega) \\ \text{s.t.} \quad & A\omega = b \\ & \boldsymbol{\mu}'\omega \geq \varphi, \end{aligned} \tag{2.87}$$

where $\omega, A, b, \boldsymbol{\mu}$ are defined as in 2.67, while parameter $\varphi \in [\underline{\varphi}, \bar{\varphi}]$ indicates our bias on objective function:

$$\begin{aligned} \underline{\varphi} &= \boldsymbol{\mu}' \arg \max_{A\omega = b} \mathcal{N}_{Ent}(\omega), \\ \bar{\varphi} &= \arg \max_{A\omega = b} \boldsymbol{\mu}'\omega. \end{aligned} \tag{2.88}$$

We are more concerned about diversification when φ approaches $\underline{\varphi}$. Accordingly, we are more concerned about expected returns when φ gets closer to $\bar{\varphi}$. In a simplified form, 2.87 and 2.88 are combined as ([Xiong, 2009](#)):

$$\omega_\varphi = \arg \max_{A\omega = b} \left\{ \varphi \boldsymbol{\mu}'\omega + (1 - \varphi) \mathcal{N}_{Ent}(\omega) \right\}, \tag{2.89}$$

where $\varphi \in [0, 1]$ and it adjusts the importance of diversification and expected return.

Analogous to the mean-variance frontier, we can draw the mean-diversification efficient frontier with the above expressions. But unlike the curve in the MV framework, the one with the mean-diversification could be non-smooth or even discontinuous. This is due to the difference between variance and entropy. In principle, a consistent relationship among volatility and expected return could ensure a smooth frontier. However, this does not occur in the actual markets. Although, the frontier is not perfectly smooth sometimes, it still reflects a convex curve. Thus, it suffices to evaluate the diversification of a portfolio based on its location comparing to the frontier and its number of effectively uncorrelated relative bets.

Chapter 3

Social Network Analysis and Clustering

This chapter is devoted to the selection of stocks based on their dependence structure. We first work with data from three large Chinese equity markets and perform essential data cleaning. Then, we model the dependence structure of market data via copula functions. We estimate the coefficients of dependence measures including linear correlation, rank correlations, tail dependence and mutual information. Such estimation is performed with respect to the entire length of investment horizon as well as different market phases based on Markov Regime Switching Result. Next, we conduct the social network clustering experiment. The coefficients of dependence measures are transitioned into similarity matrices for clustering algorithm. We improve the **AP** clustering method by reducing its computational intensity and integrating **BWP** value into it. The final solutions of data clustering are presented with the corresponding cluster centers denoting the stocks selected to be inserted into portfolios.

3.1 Data Collection and Cleaning

We collect daily adjusted closing prices¹ of all trading stocks from the following three Chinese markets²:

1. Shanghai Stock Exchange A Share Index (SHASHR), from 1/4/2006 to 11/28/2013;
2. Shenzhen Stock Exchange A Share Index (SZASHR), from 1/4/2006 to 11/28/2013;
3. Shanghai Shenzhen Csi 300 Index (SHSZ300), from 1/4/2006 to 11/28/2013.

The original data contains a lot of invalid values and thus can not be analyzed directly. It is due to the length of investment horizon and the sample size. Some stocks are not actively traded in the markets during the entire horizon and some do not have IPOs until later dates. As a result, we see a number of empty values in their historical prices. We also find some abnormal values in the form of sharp jumps in stocks' dynamics which go against the threshold set in the Chinese stock markets that the intra-day price fluctuation shall not be over 10%.

Since we organize the market data in a matrix with each column displaying a certain stock's entire price dynamics and each row showing the date, the data cleaning algorithm is designed as follows:

The threshold η in *Steps* 2 is set depending on the actual needs and a referenced value in our experiment is 0.20. In *Steps* 5 we consider three different interpolation methods: linear interpolation, polynomial interpolation and spline interpolation. It turns out that the linear interpolation presents fast convergence (usually fewer than 3 loops). Moreover, filling empty values by the linear interpolation is in line with situations in the real stock market. We perform the algorithm to all three equity stock indexes and the statistics is presented in Table 3.1.

Note that the first column(risky asset) in each of the three equity stock indexes represents the corresponding market index. Although parts of the stocks are eliminated by data cleaning, based on Table 3.1 we still obtain a well-organized data which consists of a relatively large number of stocks trading among almost 2000 days.

¹“A stock's closing price on any given day of trading that has been amended to include any distributions and corporate actions that occurred at any time prior to the next day's open. The adjusted closing price is often used when examining historical returns or performing a detailed analysis on historical returns.”(see [INVESTOPEDIA](#))

²Data Source: Bloomberg.

Algorithm 1: Algorithm of Data Cleaning

Input: Raw Data

Output: Cleaned Data

- 1 Count the number of empty values in all columns;
 - 2 Set a threshold level at η ;
 - 3 Delete the columns whose nulls exceed η ;
 - 4 Replace abnormal values with nulls for each remaining column;
 - 5 Use interpolation to find substitutes for all nulls in each column;
 - 6 Check,
 - (I) If abnormal values still exist, then repeat *Steps* 4 to 5;
 - (II) If no abnormal values exists any more, then move on;
 - 7 Summarize statistics;
 - 8 Terminate algorithm.
-

Table 3.1: Data statistics of all three indexes after cleaning. “Untouched Data” refers to number of originally existing values which are not influenced by interpolation and “Similarity Percentage” indicates the portion of those values.

Market	SHASHR	SZASHR	SHSZ300
Kept Rows (Trading Days)	1917	1917	1917
Kept Columns (Stocks)	593	340	153
Empty Values	0	0	0
Abnormal Values	0	0	0
Total Data Points	1136781	651780	293301
Untouched Data	1079495	614759	278031
Similarity Percentage	94.9607%	94.3200%	94.7937%

Table 3.2: Statistics of Estimated Linear Correlations and Rank Correlations.

	SHASHR	Cols:	593	Rows:	593	Pairs:	175528
Coefficient	Min	Max	Mean	Median	Std	Kurtosis	Skewness
Linear	0.1198	0.8458	0.4157	0.4203	0.0745	2.9322	-0.1475
Kendall's	0.0862	0.6769	0.2869	0.2887	0.0521	3.1927	-0.0014
Spearman's	0.1270	0.8345	0.4065	0.4101	0.0688	3.0677	-0.1185
	SZASHR	Cols:	340	Rows:	340	Pairs:	57630
Coefficient	Min	Max	Mean	Median	Std	Kurtosis	Skewness
Linear	0.1337	0.8311	0.4248	0.4269	0.0736	3.2799	0.0135
Kendall's	0.0942	0.6496	0.2902	0.2905	0.0516	3.6772	0.1917
Spearman's	0.1403	0.8172	0.4114	0.4129	0.0676	3.4676	0.0690
	SHSZ300	Cols:	153	Rows:	153	Pairs:	11628
Coefficient	Min	Max	Mean	Median	Std	Kurtosis	Skewness
Linear	0.1318	0.8561	0.3988	0.3985	0.0869	4.2821	0.4678
Kendall's	0.0857	0.6903	0.2751	0.2730	0.0639	5.5494	0.7881
Spearman's	0.1267	0.8540	0.3912	0.3900	0.0836	4.6012	0.5388

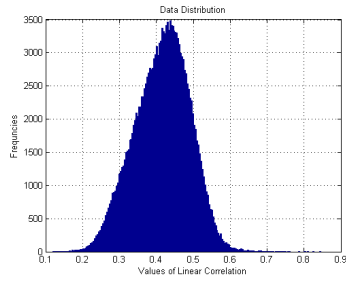
3.2 Modeling Dependence Structures via Copulas

In this section, we present the details of numerical experiment in modeling various dependence structures among assets' returns. The theoretical framework is built upon Chapter 2.

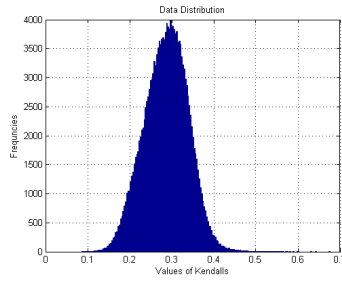
For convenience purposes, we denote the time series' features of the cleaned data using the same notations from Section 2.3. We utilize Equation 2.60 for all of the adjusted closing prices in order to acquire the corresponding continuously compounded (or log) returns.

One of the simplest dependence measures is the linear correlation given by Equation 2.29. Meanwhile, though closely related to copulas, the rank correlation Spearman's $\rho_{X,Y}^S$ can be directly estimated from the data by Equation 2.34. With Spearman's $\rho_{X,Y}^S$ given, we calculate another rank correlation Kendall's $\tau_{X,Y}$ by combing Equations 2.31 and 2.33. That is to say, we are able to empirically describe linear relationships and concordances between all pairs of time series from the data, bypassing the procedures of actually fitting copulas to the data. Hence, we first look into these three dependence measures.

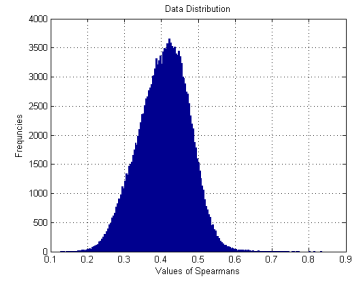
The coefficient estimates are presented by Figure 3.1 and Table 3.2. The minimum values of the linear correlation are positive in all three markets, indicating that the linear relationships exist among all pairs of returns. Although some display weak linear ties with



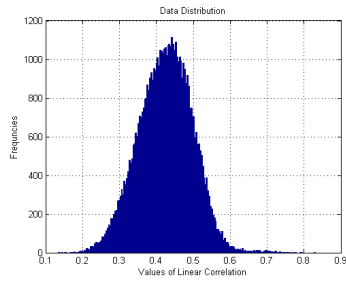
(a) Linear Correlations



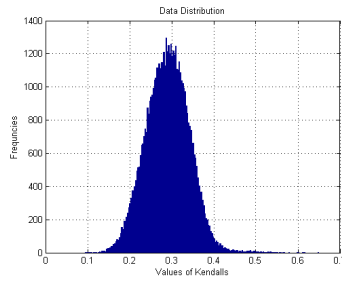
(b) Kendall's $\tau_{X,Y}$



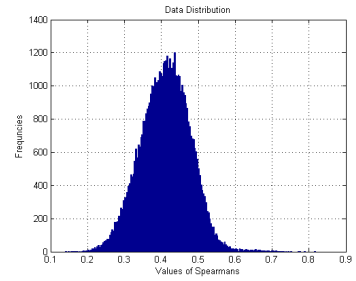
(c) Spearman's $\rho_{X,Y}^S$



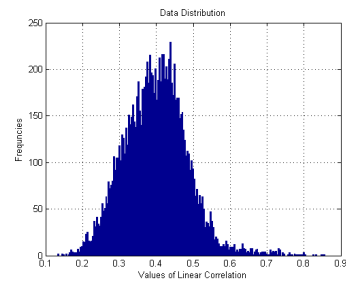
(d) Linear Correlations



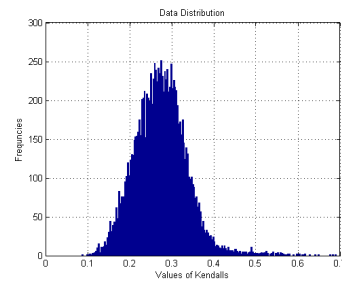
(e) Kendall's $\tau_{X,Y}$



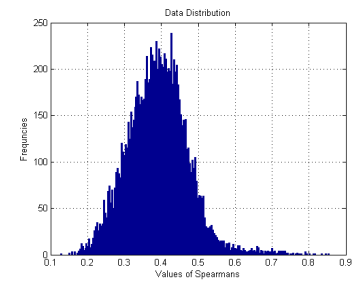
(f) Spearman's $\rho_{X,Y}^S$



(g) Linear Correlations



(h) Kendall's $\tau_{X,Y}$



(i) Spearman's $\rho_{X,Y}^S$

Figure 3.1: Distributions of Estimated Linear Correlations and Rank Correlations. (a), (b) and (c) refer to the Market Data of SHASHR. (d), (f) and (g) refer to the Market Data of SZASHR. (g), (h) and (i) refer to the Market Data of SHSZ300.

each other, the majority of securities are moderately positively correlated. We conjecture that the unanimity of positive values comes from the length of investment horizon which lasts almost 8 years (1917 trading days in total). In such a long duration it is not entirely unexpected that we do not observe negative correlations. Under a much shorter period which could be 1 year, we observe negatively correlated returns. On the other hand, even with a long investment horizon, some securities are still extremely correlated with the maximum values of correlation being around 0.85. As we see from the figure, all distributions of correlations have fat tails on the right-hand side, especially SHSZ300. This asymmetric phenomenon is consistent with the results reported in [Ang and Chen \(2002\)](#) and [Hong et al. \(2007\)](#). The securities abstracted from such tail distributions demonstrate high consistency in price movement, which is considered as a risky signal in the portfolio construction due to the lack of diversification. The reasons for the asymmetrical and extreme correlations could be trading activities as explained by [Chordia et al. \(2011\)](#).

Next, we consider Kendall's τ and Spearman's ρ^S . The shapes of their distributions are similar to those of the linear correlations while their meanings are completely different. The rank correlations aims at capturing concordance which is more concerned about non-linear relationships. As we mentioned earlier, the two dependence measures differ in their expressions but both of them are constructed independently of the marginal distributions and reflect the ties of the extreme values in pairs of returns to some degree. From [Table 3.2](#) we see that concordance suggested by Kendall's τ and Spearman's ρ^S are widely observed. In addition, Spearman's ρ^S always expresses such ties stronger.

The above three measures imply basic dependence structures among market data, but we are more concerned with the joint structure of the tail distributions of equities' returns. It leads to one of the major work in this thesis: modeling tail dependence measures. We mentioned previously that by [McNeil et al. \(2005\)](#) the coefficient of tail dependence is a concept of "limiting conditional probabilities". Yet empirically, such coefficients must be estimated through fitting copulas to data.

Following the steps described in [Section 2.1.4](#), we will need to acquire smooth and accurate density estimation of marginal distributions first. Among various techniques, *Kernel Density Estimation* is probably the best choice as its estimators are smoother and converge to the true density faster. [Wasserman \(2004\)](#) and [Wasserman \(2007\)](#) explain the Kernel methods in details while we present a basic review of them.

Kernel (Wasserman, 2007) refers to any smooth function K such that

$$\left\{ \begin{array}{l} K(x) \geq 0, \\ \int K(x)dx = 1, \\ \int xK(x)dx = 0, \\ \sigma_K^2 \equiv \int x^2K(x)dx > 0. \end{array} \right. \quad (3.1)$$

Some of the frequently used Kernels are:

the boxcar kernel: $K(x) = \frac{1}{2}I(x)$,

the Gaussian kernel: $K(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$,

the Epanechnikov kernel: $K(x) = \frac{3}{4}(1 - x^2)I(x)$,

the tricube kernel: $K(x) = \frac{70}{81}(1 - |x|^3)^3I(x)$.

where

$$I(x) = \begin{cases} 1, & |x| \leq 1, \\ 0, & |x| > 1. \end{cases}$$

Normally we can obtain a general description of empirical data by *histograms*, nonetheless such nonparametric estimation is always discontinuous or not smooth. Kernels solve the smoothing problem properly by taking local averages. A formal definition of kernel density estimation is given below:

Definition 3.2.1. (Wasserman, 2004) Given a kernel K and a positive number h , called the *bandwidth*, the *kernel density estimator* is defined as:

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - X_i}{h}\right). \quad (3.2)$$

Hence, a kernel density estimator $\hat{f}_n(x)$ is the average value of the kernels spread around observation X_i with respect to each x .

Although both the choice of kernel function K and bandwidth h could affect the smoothness of density estimation, the impact of h is more important. As h becomes smaller, the smoothness of estimation decreases and \hat{f}_n displays a lot of spikes around each observation X_i . While increasing h makes estimation smoother with the extreme case of uniform density when h goes to infinity. Roughly speaking, we may obtain an accurate estimation close to real density distribution when $h \rightarrow 0$ and $n \rightarrow \infty$. However, this is neither empirically nor computationally tractable and thus of a little practical use. In fact, we make a trade off between the smoothness and the accuracy of kernel density estimation.

In practice, one possible tool to select the optimal bandwidth h is the so-called *Normal Reference Rule* (Wasserman, 2007). It states that if we assume the real density function f to be very smooth, then the bandwidth of a normal kernel is given by:

$$h_n = \frac{1.06\hat{\sigma}}{n^{1/5}}, \quad (3.3)$$

where $\hat{\sigma}$ is the sample standard deviation of the data.

When the reference rule is not applicable due to the non-smooth assumption, h is selected to minimize a *Cross-Validation Score* (Wasserman, 2007):

$$\hat{J}(h) = \int \hat{f}^2(x)dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{-i}(X_i), \quad (3.4)$$

where f_{-i} is the kernel estimator with X_i omitted.

In our applications, the histograms of most equities' returns are non-smooth so that the reference rule is abandoned. On the other hand, we choose not to use cross-validation either because we empirically find no material impact on the kernel density estimates of our data when the bandwidth h approaches 0. More specifically, the experiments with h taking value of 0.001 or 0.002 may differ a little bit in their shapes of density curve but they are of little impact to the step of fitting copulas. To computationally fit copulas to the data, we need the marginal density estimates from a large quantity of data points. Considering the global size of market environment, the influence caused by some variations in very small h is numerically immaterial in the estimation of copula parameters and thus can be omitted.

We illustrate the kernel density estimation on empirical distribution in Figure 3.2. The blue bars in the plot represent the histogram of empirical returns of SHASHR Index. Note that we choose the index of SHASHR because it is capable of reflecting the whole market movement. As the index embraces a tremendous stock market, it exhibits relative robustness comparing to those of the individual stocks. Indeed, the kurtosis and skewness

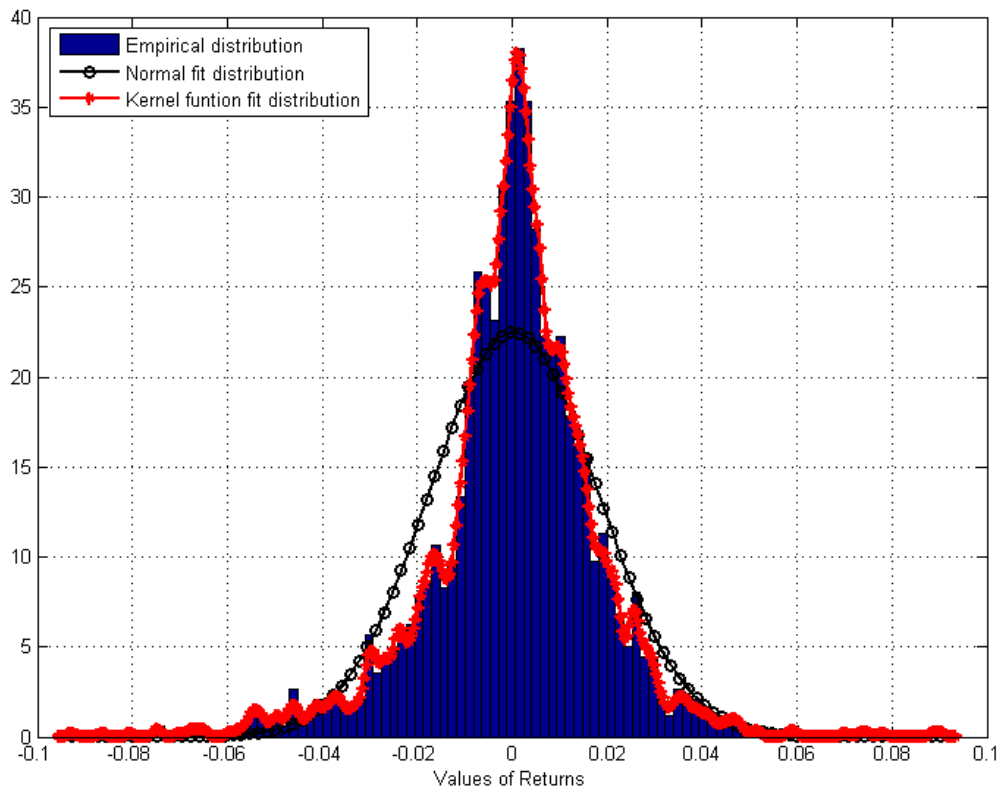
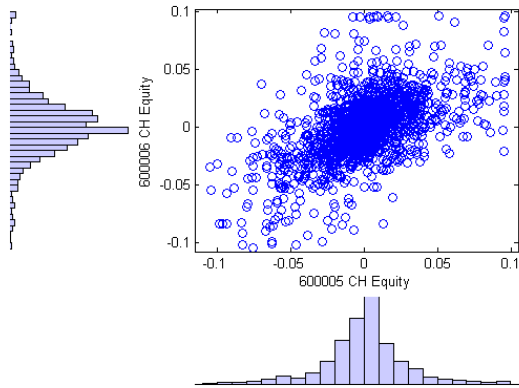


Figure 3.2: Kernel Estimation on Empirical Distribution of SHASHR Index.

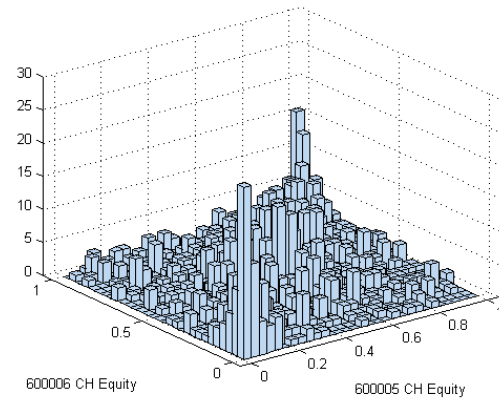
of the empirical distribution are 6.195 and -0.392 respectively. We observe a number of extreme values on both ends of the distribution, indicating “bull days” or “bear days” when most stocks experience great price movements. However, those movements can not go over 10% threshold as it is a protection mechanism in Chinese stock market. Based on the statistics of empirical data we plot the Gaussian fitted distribution in black curve. Not surprisingly, the curve could not well capture the histogram in its peak due to unreconciled kurtosis. Likewise, the histogram displays steeper and rougher slopes on both sides. In contrast, the kernel density estimation curve in red exhibits a similar structure as the histogram. In this experiment, we choose the Gaussian kernel function with a bandwidth of 0.001 for the estimation.

After performing the density estimation of the marginal distributions, we can estimate the parameters of copulas via MLE. For instance, the bivariate Gumbel copula is defined in Equation 2.22. It is quite obvious that there only exists one parameter θ which dominates the Gumbel copula. We set up the log-likelihood function following Equation 2.45 with the marginal density estimation. Then the MLE $\hat{\theta}$ of Gumbel copula is obtained by maximizing the log-likelihood function as in Equation 2.46. In some simplified cases, we can also use the CML method (2.47) for estimation of the copula parameters. Hence, we obtain all parameters of the fitted copulas.

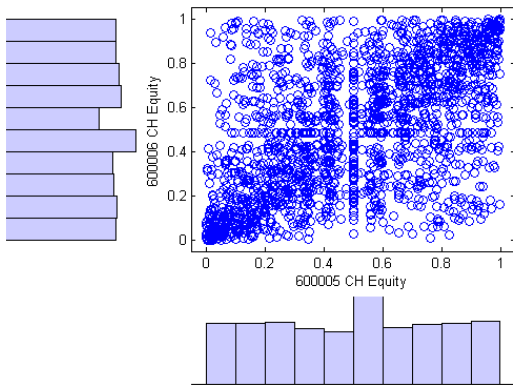
We randomly pick a pair of securities’ returns and illustrate the result of the experiment relative to fitting copulas in Figure 3.3. In this experiment, the two stocks are “600005 CH” and “600006 CH” from SHASHR. We present their marginal distributions in 3.3a and a histogram plot of their cumulative distributions in 3.3b. In the “3D” version 3.3b, we observe an approximate probability scatter. Next we perform the kernel density estimation and transform the distributions into cumulative scale as displayed in 3.3c. At this stage, we use MLE to fit all copulas to our data. In order to visually interpret results, we present results from the simulations with samples generated from five commonly used copulas we have estimated. Among these simulations, the t-copula in 3.3e perhaps best captures the original data structure. It shows a symmetric pattern in both tails of the empirical distribution. Likewise, the Clayton copula in 3.3g and the Gumbel copula in 3.3h capture particular dependence at one tail of the empirical distribution respectively.



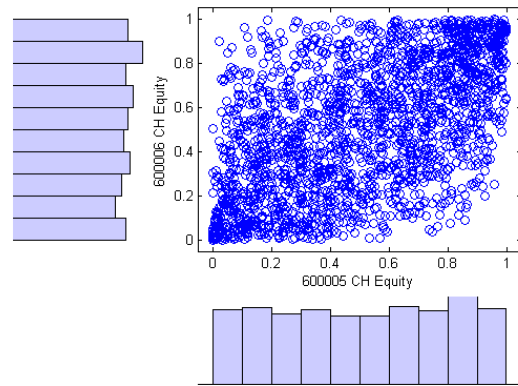
(a) Empirical Scatter Plot with Marginal Distributions



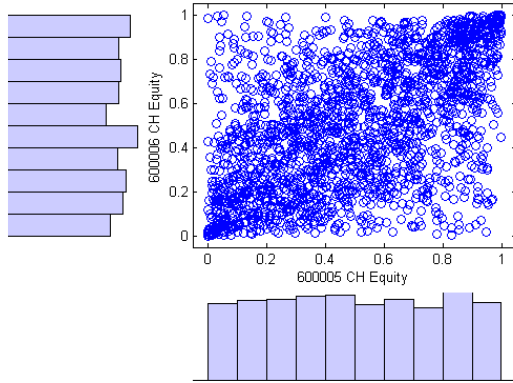
(b) Bivariate Histogram



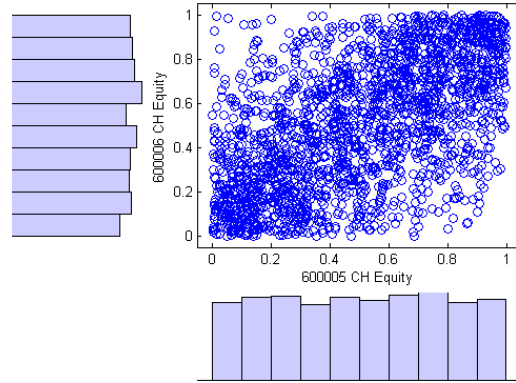
(c) Transformation into Cumulative Probability Scatter Plot



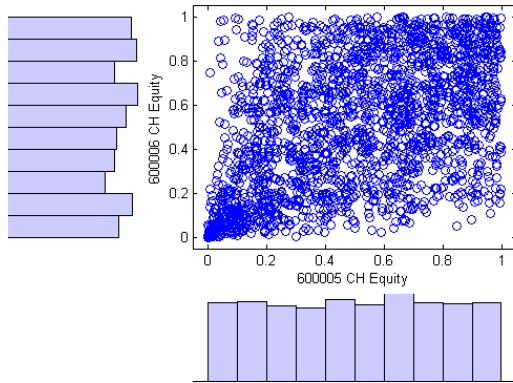
(d) Gaussian Copula Fitted Probability Scatter Plot, $\rho^{Ga} = 0.5636$



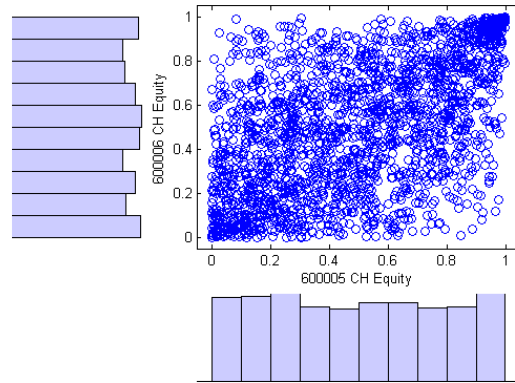
(e) t Copula Fitted Probability Scatter Plot,
 $\rho^t = 0.5831, \nu^t = 3.5981$



(f) Frank Copula Fitted Probability Scatter Plot,
 $\theta^{Fr} = 4.2926$



(g) Clayton Copula Fitted Probability Scatter Plot,
 $\theta^{Cl} = 1.0452$



(h) Gumbel Copula Fitted Probability Scatter Plot,
 $\theta^{Gu} = 1.5864$

Figure 3.3: Fitting Different Copulas to the Returns of 600005 CH Equity and 600006 CH Equity.

It would be questionable if we just use these copulas to describe the entire data structure. As shown in the plots, none of those could fully restore the scatter of original data. In fact, even sophisticated models could not fit the data sufficiently well. The deviation always exists due to certain restrictions in the model. However, in this thesis we focus on dependence structures of the empirical distribution, and not on the entire structure. We only use these copulas to capture the data structure in its tails. In other words, the degrees

for measuring such dependence are reflected by the scale of coefficients λ_U (2.35) and λ_L (2.36). In this case, the estimates of the copula parameters will be utilized to compute tail dependence measures.

The tail dependence measures written in terms of a copula are defined in Equation 2.37 and 2.39. More specifically, we can measure the extent of the tail dependence with Clayton, Gumbel and t copulas as defined in 2.41, 2.42 and 2.43 respectively. The extent could be regarded as the similarity of association of extreme gains or losses in the clustering context.

Note that when we push the quantile value t of conditional probability defined in 2.1.28 to the limit (0^+ or 1^-), Gaussian and Frank copulas admit no tail dependence. However, we still observe some degree of points clustering near both ends in 3.3d and 3.3f. It implies that these two copulas also capture some associations in tail distributions¹. We will not discard such features as they can be used to describe the dependence structure in a new way called *Mutual Information* (MI).

The concept of mutual information is closely linked to entropy. In information theory, mutual information represents the communication rate in the presence of noise while entropy indicates the complexity of a random variable (Cover and Thomas, 2006). While in Equation 2.85 we deem entropy as a measure of uncertainty of a single random variable. For simplicity, we denote $H(X)$ as the entropy of random variable X . Then we denote $H(X|Y)$ as the *conditional entropy* of X given the knowledge of random variable Y .

Definition 3.2.2. (Cover and Thomas, 2006) For random variables X, Y , the reduction in uncertainty of X due to the knowledge of Y is called the *mutual information*, given by:

$$I(X, Y) = H(X) - H(X|Y). \tag{3.5}$$

The mutual information $I(X, Y)$ is a measure of the dependence between two random variables. Moreover, it is always symmetric and nonnegative.

As Cover and Thomas suggest, mutual information speaks of the amount of information one random variable contains about another and in this sense entropy is seen as the self-information of a random variable. Both of these two terms can be interpreted as functions of the probability distributions and for continuous random variables X, Y we have:

$$H(X, Y) = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \ln f(x, y) dx dy, \tag{3.6}$$

¹In such a case, the conditional probabilities defined in 2.1.28 exist when their quantile values t are not pushed to the limit.

$$I(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \ln \frac{f(x, y)}{f(x)f(y)} dx dy. \quad (3.7)$$

In addition,

$$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y). \quad (3.8)$$

This is the connection between mutual information and entropy. Furthermore, mutual information measures dependence among random variables. Such a structure based on entropy is nonlinear. It also differs from the measures we have previously presented. However, computing mutual information in Equation 3.7 is intractable without knowing the joint density function which is usually difficult to describe. [Zhao and Lin \(2011\)](#) and [Ma and Sun \(2011\)](#) tackle this problem from another route. Both of them verify two important facts:

Theorem 3.2.3. *The Mutual information of some random variables is equivalent to the negative copula entropy of these random variables.*

Theorem 3.2.4. *The joint entropy of some random variables can be decomposed as two parts: the sum of entropy of individual random variables and the corresponding copula entropy.*

Proof. The mutual information of a general d -dimensional random vector $\mathbf{X} = (X_1, \dots, X_d)'$ is given by:

$$I(\mathbf{X}) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_d) \ln \frac{f(x_1, x_2, \dots, x_d)}{\prod_{i=1}^d f_i(x_i)} dx_1 \dots dx_d. \quad (3.9)$$

Using Equation 2.44,

$$\begin{aligned} I(\mathbf{X}) &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} c(F_1(x_1), F_2(x_2), \dots, F_d(x_d)) \cdot \prod_{i=1}^d f_i(x_i) \\ &\quad \cdot \ln c(F_1(x_1), F_2(x_2), \dots, F_d(x_d)) dx_1 \dots dx_d \\ &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} c(u_1, u_2, \dots, u_d) \cdot \prod_{i=1}^d f_i(x_i) \\ &\quad \cdot \ln c(u_1, u_2, \dots, u_d) dx_1 \dots dx_d \\ &= \int_0^1 \dots \int_0^1 c(u_1, u_2, \dots, u_d) \cdot \ln c(u_1, u_2, \dots, u_d) du_1 \dots du_d \\ &= -H_C(u_1, u_2, \dots, u_d) \\ &= -H_C(\mathbf{U}). \end{aligned} \quad (3.10)$$

Hence, mutual information $I(\mathbf{X})$ is equivalent to a negative copula entropy $-H_C(\mathbf{U})$.

Then, by Equation 3.8,

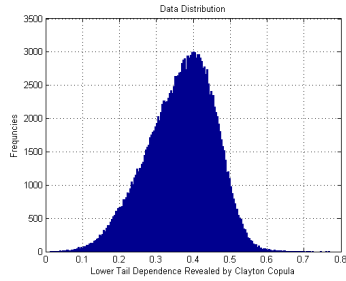
$$\begin{aligned} H(\mathbf{X}) &= \sum_{i=1}^d H(X_i) - I(\mathbf{X}) \\ &= \sum_{i=1}^d H(X_i) + H_C(\mathbf{U}). \end{aligned} \tag{3.11}$$

Thus the joint entropy $H(\mathbf{X})$ is composed of the sum $\sum_{i=1}^d H(X_i)$ and the copula entropy $H_C(\mathbf{U})$. □

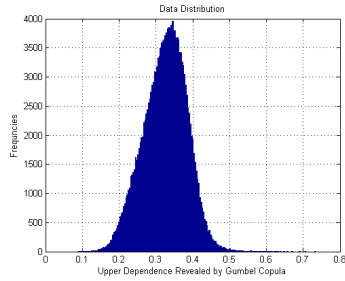
Since we are able to empirically fit copula functions to data via maximum likelihood, the corresponding copula entropy on pairs of returns can be calculated. In this sense, we reduce the computation complexity in the estimation of mutual information. Moreover, the equivalence revealed by Equation 3.10 provides an alternative way to capture mutual information as it takes advantages of both the copula and the entropy. For the entropy we can quantify the amount of uncertainty in the data while for the copula we can depict the associations among different variables in the data. When the negative copula entropy achieve its maximum, the uncertainty of one random variable decreases to its minimum due to the knowledge of another. In fact, if the random variables are mutually independent, then there exists maximum uncertainty and mutual information displays 0. On the contrary, if we fully understand the behavior of one random variable given the knowledge of another one, then there exists no uncertainty among them with mutual information holding value 1. As a result, mutual information can be applied to measure the dependence among returns in all orders and such ties will grow stronger with more mutual information being revealed.

Introducing mutual information helps us complete the steps in modeling a dependence structure. Following the previous estimation results of the linear and the rank correlations (which are presented in Figure 3.1 and Table 3.2), we illustrate the estimation of tail dependence and mutual information in Figure 3.4 and Table 3.3. We relegate additional results to Appendix A.1, where the estimation from Market SZASHR and SHSZ300 are displayed in Figure A.1 and A.2 respectively.

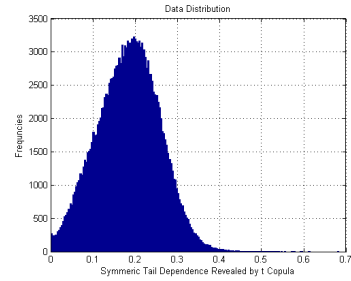
The results reveal some features of price co-movement. We start presenting our findings from the distributions of the lower tail dependence and the upper tail dependence. In all three markets, the distributions of the upper tail dependence show higher kurtosis than



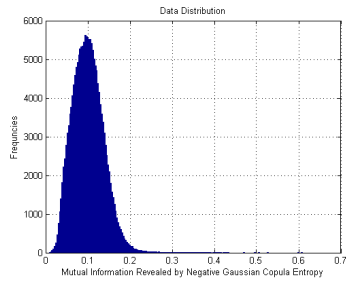
(a) Lower Tail Dependence Revealed by Clayton Copula



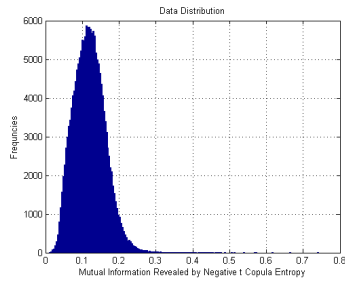
(b) Upper Tail Dependence Revealed by Gumbel Copula



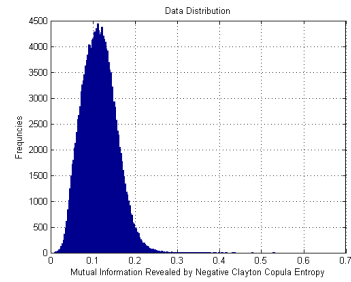
(c) Tail Dependence Revealed by t Copula



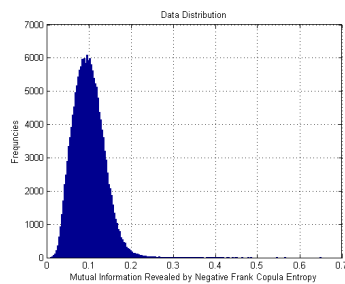
(d) Mutual Information Revealed by Negative Gaussian Copula Entropy



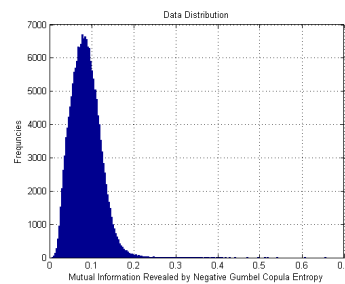
(e) Mutual Information Revealed by Negative t Copula Entropy



(f) Mutual Information Revealed by Negative Clayton Copula Entropy



(g) Mutual Information Revealed by Negative Frank Copula Entropy



(h) Mutual Information Revealed by Negative Gumbel Copula Entropy

Figure 3.4: Distributions of Estimated Tail Dependence and Mutual Information of Market Data of SHASHR.

Table 3.3: Statistics of Estimated Tail Dependence and Mutual Information.

	SHASHR	Cols:	593	Rows:	593	Pairs:	175528
Coefficient	Min	Max	Mean	Median	Std	Kurtosis	Skewness
Lower Tail (Clayton)	0.0116	0.7702	0.3660	0.3746	0.0908	2.9078	-0.3551
Upper Tail (Gumbel)	0.0889	0.7335	0.3252	0.3288	0.0590	3.1432	-0.1181
Tail (t Copula)	0.0010	0.6828	0.1846	0.1866	0.0734	2.8624	0.0276
MI (Gaussian)	0.0084	0.6092	0.1006	0.0986	0.0371	5.5248	0.7250
MI (t Copula)	0.0103	0.7425	0.1198	0.1177	0.0437	5.2237	0.6820
MI (Clayton)	0.0081	0.5318	0.1153	0.1134	0.0406	3.7804	0.4718
MI (Frank)	0.0084	0.6508	0.0999	0.0975	0.0368	5.6946	0.7747
MI (Gumbel)	0.0054	0.6574	0.0862	0.0839	0.0348	7.0852	0.9292
	SZASHR	Cols:	340	Rows:	340	Pairs:	57630
Coefficient	Min	Max	Mean	Median	Std	Kurtosis	Skewness
Lower Tail (Clayton)	0.0340	0.7515	0.3697	0.3763	0.0909	2.9525	-0.2475
Upper Tail (Gumbel)	0.0828	0.7112	0.3292	0.3313	0.0581	3.4504	0.0248
Tail (t Copula)	0.0010	0.6140	0.1797	0.1817	0.0769	2.6574	0.0462
MI (Gaussian)	0.0108	0.5838	0.1040	0.1007	0.0384	7.9744	1.1544
MI (t Copula)	0.0126	0.6670	0.1222	0.1190	0.0450	6.4506	0.9407
MI (Clayton)	0.0147	0.4990	0.1175	0.1145	0.0422	4.9161	0.7441
MI (Frank)	0.0100	0.5885	0.1021	0.0987	0.0378	7.8450	1.1445
MI (Gumbel)	0.0046	0.6093	0.0891	0.0859	0.0357	9.3246	1.2934
	SHSZ300	Cols:	153	Rows:	153	Pairs:	11628
Coefficient	Min	Max	Mean	Median	Std	Kurtosis	Skewness
Lower Tail (Clayton)	0.0186	0.7876	0.3235	0.3271	0.1065	3.2948	0.0974
Upper Tail (Gumbel)	0.0941	0.7454	0.3155	0.3141	0.0715	4.8958	0.6055
Tail (t Copula)	0.0011	0.6841	0.1540	0.1472	0.0823	4.6104	0.7723
MI (Gaussian)	0.0090	0.6711	0.0961	0.0890	0.0497	18.6588	2.6898
MI (t Copula)	0.0124	0.7751	0.1114	0.1035	0.0565	18.0800	2.5790
MI (Clayton)	0.0111	0.5864	0.0990	0.0933	0.0473	12.6130	1.9826
MI (Frank)	0.0083	0.6864	0.0938	0.0867	0.0495	19.6374	2.7691
MI (Gumbel)	0.0069	0.7063	0.0857	0.0781	0.0485	23.3240	3.1011

those shown by the lower tail dependence, indicating a concentration of ties of extreme gains among pairs of securities. However, although more concentrated, the degree of the upper dependence are generally lower than those of the lower dependence. Hence, we conjecture that the associations of extreme loss among pairs of securities are stronger than the ones of extreme gains. In addition, the ties in the lower tails could vary a lot depending on the selection of securities. Secondly, we look at the tail dependence revealed by t Copula. Due to the endogenous property of capturing symmetric tail structures, t Copulas have to take both ends into account simultaneously. As a result, the strength of ties shown by them is weaker than the previous two copulas. We interpret this phenomenon as a compromise between describing structures on both tails at the same time and utilizing only one value for explaining the degree of association. These distributions assume that the probabilities of getting pairwise extreme loss and pairwise extreme gains are the same. Still, such chances vary a lot depending on the chosen pair of securities. Thirdly, we look at the distributions of mutual information. As we discussed earlier, mutual information is also a measure of dependence and we use the negative copula entropy value to quantify it. We observe similar shapes in distributions of all five types MI estimation in Figure 3.4. Theoretically, there should only exist one unique mutual information to finish the job. However, from the computational point of view, none of the copulas could perfectly capture the entire dependence structure of data, thus resulting in different negative copula entropy estimations.

So far, we have presented results based on linear correlation, rank correlations, tail dependence and mutual information. Also, so far we have discussed the dependence structure of data. However, the measures from different categories have not been introduced for the purpose of a comparative study. We intent to analyze the stock networks constructed by a data clustering technique where the dependence measures play key roles. Ultimately, we make comparisons among the diversification of the stock networks. In this sense, we can make conjecture as to which of the dependence measures has more impact on the portfolios.

Before we can proceed to the clustering section of social network analysis, there is one more issue which needs to be covered: market phases. The market data we study in the thesis holds a duration of almost 8 years (which contains 1917 observations for each asset). Empirically, the market would undergo multiple states¹ in long-terms². The dynamics of general market movements are importantly distinguished between periods of “bulls” and “bears”. As a result, the estimation of dependence measures also varies in different market

¹For instance, the states can represent upside and downside movements of the market.

²The Shanghai Shenzhen Csi 300 Index (SHSZ300) has been calculated since April 8, 2005. We extract the adjusted daily closed price of SHSZ300 Index ranging from 1/4/2006 to 11/28/2013. In this sense, we call it “long-term” data.

phases which will impart further influence on the clustering results or in other words the stock networks. Hence, it is necessary to identify market phases to us. We will perform a *Markov Regime Switching* estimation in the next section.

3.3 Markov Regime Switching Analysis

The idea of considering time series with changes in regime is that “many variables undergo episodes in which the behavior of the states seems to change quite dramatically.”(Hamilton, 1994). In the past decade, the implications of regime shifts are studied in a variety of financial literatures. Hardy (2001) defines a regime-switching log-normal model to study long-term stock returns and compares it with other switching models. Ang and Bekaert (2002) characterize the regime switching process of correlations and volatilities in international equity market, pointing out both variables grow stronger in bear states. The same results are also demonstrated for individual stock returns (Ang and Chen, 2002). Kritzman et al. (2012) specially discuss the impact of regimes on investment strategies.

In our thesis, we aim to identify regime shifts in entire markets rather than individual equities so that we can globally observe its impact on dependence structure. Considering our markets’ sizes, we speculate that regime changes in single equities are unlikely to be significant. To accomplish the task, we implement the methods designed by Perlin (2014)¹. Perlin’s Markov regime switching models have great flexibility and allow for handling processes with a variety of statistical specifications.

In next section, we briefly introduce Perlin (2014)’s framework of Markov regime switching models.

3.3.1 The Switching Model

The theoretical framework of Perlin (2014)’s model is built upon Hamilton (1994) and Hamilton (2005). Hamilton (1989), Kim and Nelson (1999) and Tsay (2010) also provide useful details. For instructional purposes, we first introduce *Markov Chains* and the *Transition Matrix*.

Definition 3.3.1. (Hamilton, 1994) Let s_t be a random variable that takes an integer value from set $\{1, 2, \dots, N\}$. If $s_t = j$, then the process $\{s_T, T = 1, 2, \dots, \}$ is said to be in

¹Available at <http://ssrn.com/abstract=1714016>. The package of source code is offered by the author on his website: <https://sites.google.com/site/marceloperlin/>.

state (or regime) j at time t . Suppose that the probability that s_t equals some particular value j depends on the past only through the most recent value s_{t-1} :

$$\mathbb{P}[s_t = j | s_{t-1} = i, s_{t-2} = k, \dots] = \mathbb{P}[s_t = j | s_{t-1} = i] = p_{ij}. \quad (3.12)$$

Such a process is described as an N -state Markov chain with transition probabilities $\{p_{ij}\}_{i,j=1,2,\dots,N}$. The transition probability p_{ij} gives the probability that state i will be followed by state j . Since the probabilities are nonnegative and since the process must make a transition into some states, we have

$$\sum_{j=1}^N p_{ij} = 1. \quad (3.13)$$

Let \mathbf{P} denote the matrix of one-step transition probabilities p_{ij} . Such matrix \mathbf{P} is known as the *transition matrix*:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix}. \quad (3.14)$$

For simplicity, the transition matrix of a two-state Markov chain is:

$$\mathbf{P} = \begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix}, \quad (3.15)$$

where for 2 states, a numerical transformation ([Hamilton, 1994](#)) is applied for p_{11}, p_{22} , so that $p_{12} = 1 - p_{11}$ and $p_{21} = 1 - p_{22}$.

[Perlin \(2014\)](#) assumes that a time series y_t follows a generalized Markov switching model of the form:

$$y_t = \sum_{l=1}^{N_{nS}} \beta_l x_{l,t}^{nS} + \sum_{m=1}^{N_S} \phi_{m,s_t} x_{m,t}^S + \epsilon_t, \quad (3.16)$$

$$\epsilon_t \sim \mathbb{P}(\Phi_{s_t}). \quad (3.17)$$

This generalized model could cover a number of Markov switching specifications. In Equation 3.16, the notations S and nS globally represent whether or not a parameter contains a switching effect. Then, the term N_S (or N_{nS}) is the total number of coefficients which have (or do not have) switching effects. $x_{m,t}$ (or $x_{l,t}$) are some explanatory variables. The terms ϵ_t are the innovations (or white noises) with probability density function $\mathbb{P}(\Phi_{s_t})$.

In practical applications, we usually consider the time series of continuously compounded returns of assets as *autoregressive processes* (Hamilton, 1989) or simply as *mixture distributions* (Hamilton, 1994).

Here, we briefly express a model considering 2-states mixture distributions as an illustration of how the Markov switching model¹ works. In this case, the model can be written as:

$$y_t = \begin{cases} \mu_1 + \epsilon_t, & \text{if } s_t = 1, \\ \mu_2 + \epsilon_t, & \text{if } s_t = 2. \end{cases} \quad (3.18)$$

$$\begin{cases} \epsilon_t \sim N(0, \sigma_1^2), & \text{if } s_t = 1, \\ \epsilon_t \sim N(0, \sigma_2^2), & \text{if } s_t = 2. \end{cases} \quad (3.19)$$

Hence, the density of y_t conditional on the random variable s_t taking on regime j is:

$$f(y_t | s_t = j; \Theta) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(y_t - \mu_j)^2}{2\sigma_j^2}\right\}, \quad (3.20)$$

where $j = 1, 2$. Note here Θ is a vector of population parameters that includes $\mu_1, \mu_2, \sigma_1, \sigma_2$.

Additionally, assume that the regime variable s_t is generated by some probability distribution function. The unconditional probability that $s_t = j$ is denoted by:

$$\mathbb{P}[s_t = j; \Theta] = \pi_j \quad \text{for } j = 1, 2. \quad (3.21)$$

Hence, the population parameters Θ can be rewritten as:

$$\Theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \pi_1, \pi_2)'. \quad (3.22)$$

Then, by the rules of conditional probability, we have:

$$\mathbb{P}[y_t, s_t = j; \Theta] = f(y_t | s_t = j; \Theta) \cdot \mathbb{P}[s_t = j; \Theta] \quad (3.23)$$

$$= \frac{\pi_j}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(y_t - \mu_j)^2}{2\sigma_j^2}\right\}. \quad (3.24)$$

The unconditional density of y_t is thus given by:

$$f(y_t; \Theta) = \sum_{j=1}^2 \mathbb{P}[y_t, s_t = j; \Theta]. \quad (3.25)$$

¹Further details on deductions of the model are presented in (Chapter 22, Hamilton (1994)).

In this case, we estimate the model parameters by **MLE**. The log likelihood function of the model is given by:

$$\ln L(\Theta) = \sum_{t=1}^T \ln f(y_t; \Theta) \quad (3.26)$$

$$= \sum_{t=1}^T \ln \sum_{j=1}^2 \mathbb{P}[y_t, s_t = j; \Theta]. \quad (3.27)$$

The full likelihood function can be perceived as a weighted average of the likelihood function for each s_t . The estimates of Θ are obtained by maximizing Equation 3.26. [Hamilton \(1994\)](#) provides an analytical solution of the **MLE** based on *EM Algorithm*¹:

$$\hat{\mu}_j = \frac{\sum_{t=1}^T y_t \cdot \mathbb{P}[s_t = j|y_t; \hat{\Theta}]}{\sum_{t=1}^T \mathbb{P}[s_t = j|y_t; \hat{\Theta}]}, \quad \text{for } j = 1, 2, \quad (3.28)$$

$$\hat{\sigma}_j^2 = \frac{\sum_{t=1}^T (y_t - \hat{\mu}_j)^2 \cdot \mathbb{P}[s_t = j|y_t; \hat{\Theta}]}{\sum_{t=1}^T \mathbb{P}[s_t = j|y_t; \hat{\Theta}]}, \quad \text{for } j = 1, 2, \quad (3.29)$$

$$\hat{\pi}_j = T^{-1} \sum_{t=1}^T \mathbb{P}[s_t = j|y_t; \hat{\Theta}], \quad \text{for } j = 1, 2. \quad (3.30)$$

The above is the illustration of Markov switching framework in the case of mixture distributions. In general, a Markov switching model assumes that the time series y_t follows a process such as an *auto-regression*.² A 2-states **MSA** model is given as follows:

$$y_t = c_{s_t} + \sum_{i=1}^p \phi_{s_t, p} y_{t-i} + \epsilon_{s_t, t}, \quad (3.31)$$

$$\epsilon_{s_t, t} \sim N(0, \Sigma_{s_t}), \quad (3.32)$$

¹Alternatively, [Perlin \(2014\)](#) provides an iterative algorithm for updating the estimates of the model.

²The Markov switching autoregressive (**MSA**) model ([Hamilton, 1989](#)). Further details can also be found at [Hamilton \(1994\)](#), [Hamilton \(2005\)](#), [Tsay \(2010\)](#). We only present a general representation of the model here.

$$\Sigma_{s_t} = \begin{pmatrix} \sigma_{11,s_t}^2 & \sigma_{21,s_t}^2 \\ \sigma_{12,s_t}^2 & \sigma_{22,s_t}^2 \end{pmatrix}, \quad (3.33)$$

$$s_t = 1, 2. \quad (3.34)$$

In the next section, we present our regime switching results based on [Perlin \(2014\)](#)'s method¹.

3.3.2 Regime Switching Results

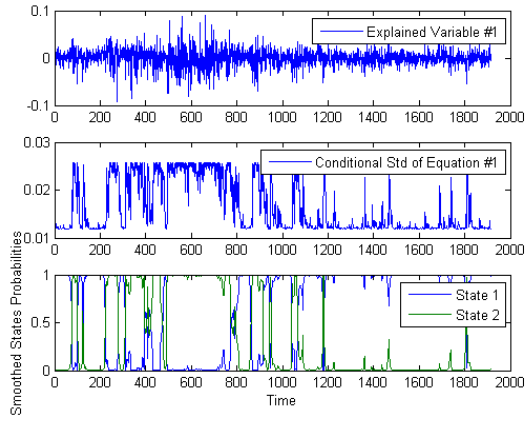
We perform Markov regime switching experiment on all three market indexes (which are SHASHR, SHZSHR and SHSZ300) to identify phases. Initially, we choose the simplest model, i.e. 2-states mixture distributions. The details of the switching model is expressed in Equation 3.18,3.19. Then, the switching model is enriched with 3-states mixture distributions. Moreover, for comparative purposes, we also implement a Markov switching autoregressive model with specifications introduced by Equation 3.31.

We put the graphical results of all experiment on SHASHR Index in Figure 3.5 whereas some comparative analysis between markets are displayed in 3.6. Additional graphics are illustrated in A.3, A.4 and A.5 . Lastly, full details of the model parameters are stated in Table 3.4.

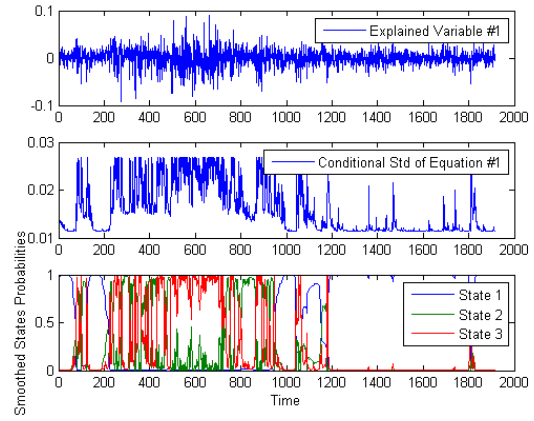
The regime switching experiment reveals some important properties of market phases. Taking Figure 3.5a as an example, the top row represents the continuously compounded returns of market index during the entire period. The middle row displays the conditional standard deviation of Equation 3.18 with respect to different market states. Afterwards, the bottom row which is the most important one, indicates the regime changes with respect to time.

We discuss the detail of each experiment. For 2-states mixture distributions, we observe that the probability of state 2 holds a firm value of 1 lasting from time 500 to 700, indicating state 2 can represent the market phase during this period. From Table 3.4, we see the estimates of switching parameter in state 2 is -0.0013 with innovations $\epsilon_t \sim N(0, 6.79 \times 10^{-4})$. As showed in Figure 3.5d, it captures the sharp decreasing phase of SHASHR Index. In other words, the returns during a rapid market contraction period can approximately be characterized by a normally distributed random variable $\hat{y}_t \sim N(-1.3 \times 10^{-3}, 6.79 \times 10^{-4})$ while this contraction period can be observed by states probabilities. The experiment using

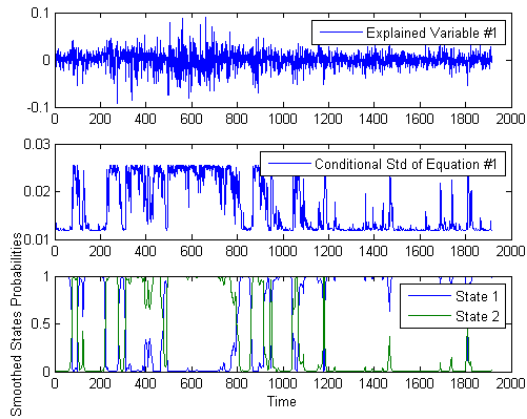
¹Note that Perlin's method does not allow for time varying transition probabilities or state space models with Markov switching effects. In addition, the estimation of model parameters are achieved by directly maximizing the log likelihood function rather than implementing the EM algorithm.



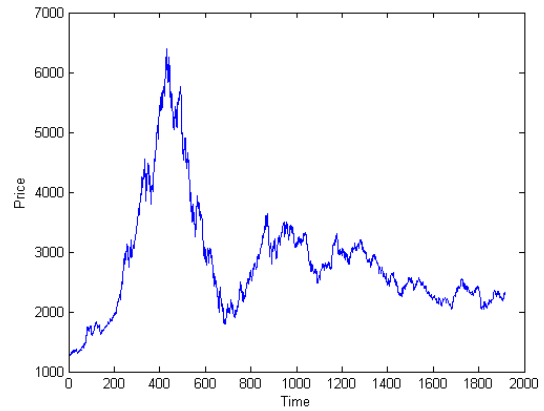
(a) 2-states Mixture Distributions Regime Switching Result



(b) 3-states Mixture Distributions Regime Switching Result



(c) 2-states Autoregressive Regime Switching Result



(d) Historical Price Dynamics of SHASHR Index

Figure 3.5: Markov Regime Switching Results on SHASHR Index. We present the price dynamics for comparison.

an autoregressive model in 3.5c suggests that the time period [500, 700] can be identified as a special market phase.

All three switching experiment identify the time period [1200, 1800] as another stable market phase, especially the 3-states mixture distributions. In Figure 3.5b, this period attributes to state 1 which is characterized by $\hat{y}_t \sim N(2 \times 10^{-4}, 1.29 \times 10^{-4})$. Indeed, from Figure 3.5d we see that after undergoing “bulls” and “bears” periods, the market remains

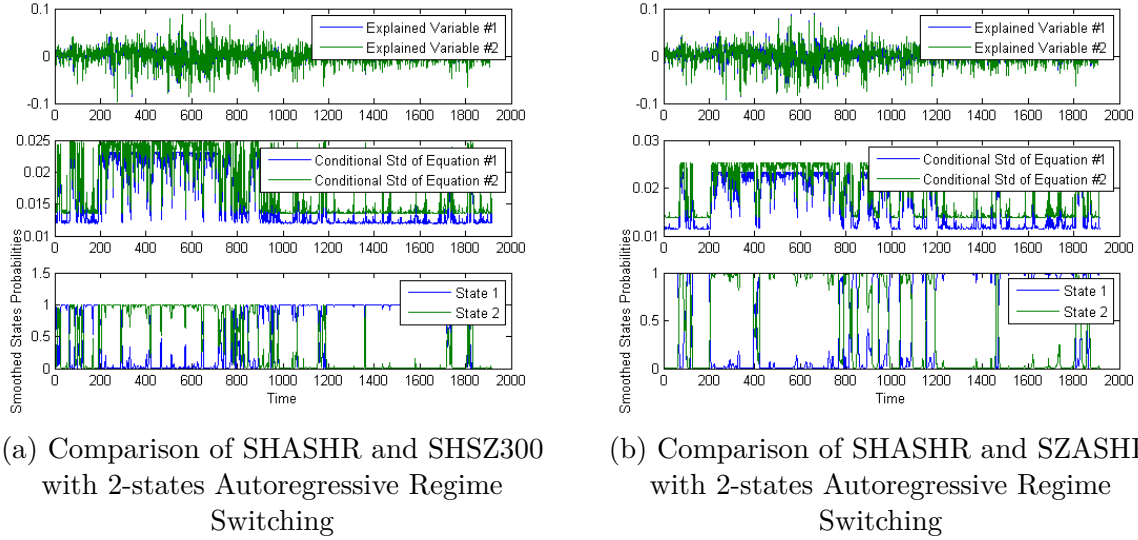


Figure 3.6: Markov Switching Autoregressive Model with Two Market Indexes. The MSA model can handle two time series together.

low volatility during this period.

Here comes another question: “Can the fast increasing market phase be identified by a Markov regime switching estimation?”. The answer is “Yes, but with some conditions.” In all three experiments, we do not observe that some states entirely dominate the growing period roughly consisting of $[0, 500]$. We conjecture that it is because this rising phase internally consists of multiple regimes each of which can be expressed by a special state of the model. Based on the model specifications adopted in this thesis, the states transform to each other in order to capture some very tiny changes of returns. As a result, although we notice a generally upward trend of the market, it essentially experiences some regime transitions during time $[0, 500]$ based on our models. A more sophisticated switching model perhaps could better explain what happened at that time but it is beyond the purpose of this thesis. After all, the regime of downside markets imposes more challenges on portfolio construction and its diversification.

In addition to the analysis based on a single market, Figure 3.6 exhibits two comparative switching results between markets. The dynamics of conditional standard deviations and states probabilities imply great similarity of the regime changes in all three markets. Though some fluctuations are observed, the transitions of states in the other two markets (SZASHR and SHSZ300) are consistent with those in SHASHR. Such inference is also sup-

Table 3.4: Statistics of Markov Switching Results.

		2-states Mixture Distributions					
Market Index	Switching Parameters				Transition Matrix		
SHASHR	State 1	$\hat{\mu}_1$:	0.0011	$\hat{\sigma}_1^2(\epsilon_t)$:	0.000135	0.99 0.02	
	State 2	$\hat{\mu}_2$:	-0.0013	$\hat{\sigma}_2^2(\epsilon_t)$:	0.000679	0.01 0.98	
SZASHR	State 1	$\hat{\mu}_1$:	0.0023	$\hat{\sigma}_1^2(\epsilon_t)$:	0.000177	0.98 0.04	
	State 2	$\hat{\mu}_2$:	-0.0023	$\hat{\sigma}_2^2(\epsilon_t)$:	0.000789	0.02 0.96	
SHSZ300	State 1	$\hat{\mu}_1$:	0.0013	$\hat{\sigma}_1^2(\epsilon_t)$:	0.000166	0.98 0.03	
	State 2	$\hat{\mu}_2$:	-0.0012	$\hat{\sigma}_2^2(\epsilon_t)$:	0.000767	0.02 0.97	
		3-states Mixture Distributions					
Market Index	Switching Parameters				Transition Matrix		
SHASHR	State 1	$\hat{\mu}_1$:	0.0002	$\hat{\sigma}_1^2(\epsilon_t)$:	0.000129	1.00 0.00 0.01	
	State 2	$\hat{\mu}_2$:	0.0059	$\hat{\sigma}_2^2(\epsilon_t)$:	0.000166	0.00 0.91 0.08	
	State 3	$\hat{\mu}_3$:	-0.0068	$\hat{\sigma}_3^2(\epsilon_t)$:	0.000793	0.00 0.09 0.91	
SZASHR	State 1	$\hat{\mu}_1$:	0.0017	$\hat{\sigma}_1^2(\epsilon_t)$:	0.000176	0.98 0.10 0.00	
	State 2	$\hat{\mu}_2$:	0.0093	$\hat{\sigma}_2^2(\epsilon_t)$:	0.000167	0.02 0.00 0.39	
	State 3	$\hat{\mu}_3$:	-0.0051	$\hat{\sigma}_3^2(\epsilon_t)$:	0.000884	0.00 0.90 0.61	
SHSZ300	State 1	$\hat{\mu}_1$:	0.0010	$\hat{\sigma}_1^2(\epsilon_t)$:	0.000019	0.02 0.30 0.00	
	State 2	$\hat{\mu}_2$:	0.0014	$\hat{\sigma}_2^2(\epsilon_t)$:	0.000242	0.95 0.70 0.02	
	State 3	$\hat{\mu}_3$:	-0.0017	$\hat{\sigma}_3^2(\epsilon_t)$:	0.000804	0.03 0.00 0.98	
		2-states Autoregressive Model (Lag = 1)					
Market Index	Switching Parameters				Transition Matrix		
SHASHR	State 1	$\hat{\phi}_1$:	0.03	$\hat{\sigma}_1^2(\epsilon_t)$:	0.00014	0.99 0.02	
	State 2	$\hat{\phi}_2$:	-0.01	$\hat{\sigma}_2^2(\epsilon_t)$:	0.00066	0.01 0.98	
SZASHR	State 1	$\hat{\phi}_1$:	0.09	$\hat{\sigma}_1^2(\epsilon_t)$:	0.00019	0.99 0.03	
	State 2	$\hat{\phi}_2$:	0.06	$\hat{\sigma}_2^2(\epsilon_t)$:	0.00079	0.01 0.97	
SHSZ300	State 1	$\hat{\phi}_1$:	0.04	$\hat{\sigma}_1^2(\epsilon_t)$:	0.00017	0.99 0.03	
	State 2	$\hat{\phi}_2$:	0.01	$\hat{\sigma}_2^2(\epsilon_t)$:	0.00075	0.01 0.97	

ported by the illustration of Figure A.4 and A.5. Moreover, from a statistical point of view, the variances of innovation terms ϵ_t take much greater values when the switching parameters display negative than positive. Smaller positive values of $\hat{\mu}_j$ always associate smaller

variances. It is also consistent with the empirical fact in the market that the volatility soars as market index falls rapidly and remain mild in a relatively stable market period.

Hence, the Markov regime switching results lead to a further partition of the entire investment period. The analysis indicates that we can identify market phases representing an increasing regime, a decreasing regime and a stable regime. As a result, we identify three corresponding time intervals:

1. Upside Market-movement Phase: $t \in (0, 500]$ (approximately from 1/4/2006 to 1/4/2008);
2. Downside Market-movement Phase: $t \in (500, 700]$ (approximately from 1/4/2008 to 10/26/2008);
3. Mild Market-movement Phase: $t \in (1200, 1800]$ (approximately from 1/11/2011 to 6/7/2013).

3.4 Social Network Clustering Experiment

We have modeled the dependence structure among equities' returns using a set of alternative measures and identified different market phases by using the regime switching technique. We are now in the position to present and discuss a social network analysis (**SNA**). This section plays a key role in this thesis as it provides a bridge connecting dependence structures and diversification.

Recall from Section 2.2.1, we interpret **SNA** as the mapping and measuring of relationships and flows between entities. The equities in our data are treated as those entities and the relationships among them are revealed by dependence measures. Worthy of further explanation, all equities are seen as nodes or vertexes on a network with a lot of edges as the linkage. Such a linkage represents a certain pattern of the dependence structure among equities and can be quantified by a measure of distance. In our thesis, the distance $s(i, k)$ between two nodes (equities) i, k is computed as:

$$s(i, k) = \sqrt{2(1 - a_{i,k})}, \quad (3.35)$$

where $a_{i,k}$ denotes the coefficient of a certain dependence measure between i and k . In Section 3.2, we compute such $a_{i,k}$ among all equities with various dependence measures so that we can quantitatively depict the equity networks with some similarity matrices. A network also turns into a valued graph in this sense.

Then, we locate all “central nodes” on an equity network as these nodes build the foundation of its structure. The network will be viewed as a collection of clusters each of which contains only one “central node”. Naturally, the set of the “central nodes” becomes our selection of equities for a portfolio. This goal is achieved by *affinity propagation clustering* (**AP**) (Frey and Dueck, 2007) (see Section 2.2.2). In our study, we also slightly improve the **AP** algorithm to reduce the computational intensity.

Suppose that we have N equities in total including the market index. Then the similarity matrix which consists of entries $s(i, k)$ is of size $N \times N$ ¹. In Frey and Dueck (2007), the “responsibility” $r(i, k)$ and “availability” $a(i, k)$ are characterized by Equation 2.49:

$$\begin{cases} r(i, k) \leftarrow s(i, k) - \max \{a(i, k') + s(i, k')\}, & k' \text{ s.t. } k' \neq k, \\ a(i, k) \leftarrow \min \{0, r(k, k) + \sum \max \{0, r(i', k)\}\}, & i' \text{ s.t. } i' \neq i, k. \end{cases}$$

In order to update a certain $r(i, k)$ (where the message is sent from node i to a potential exemplar k), the running time is of order $O(N)$ (in search for $\max\{a(i, k') + s(i, k')\}$). Considering the complexity of order $O(N^2)$ for simply going over all $r(i, k)$, the full running time of updating “responsibility” is of order $O(N^3)$ currently. Now, we tackle the algorithm from another view. For a fixed node i , the choice of $\max\{a(i, k') + s(i, k')\}$ remains the same with respect to all potential exemplars k (still, $k' \text{ s.t. } k' \neq k$). We then define $w(i)$ as:

$$w(i) = \max \{a(i, k') + s(i, k')\}, \quad k' \text{ s.t. } k' \neq k, \quad (3.36)$$

and assume that the maximum for node i corresponds to a certain exemplar $k'_{max}(i)$, so that

$$w(i) = a(i, k'_{max}(i)) + s(i, k'_{max}(i)). \quad (3.37)$$

Now for fixed node i and exemplar k , it still takes a running time of order $O(N)$ to find its maximum, i.e. the value of $w(i)$, or more explicitly the exemplar $k'_{max}(i)$ ($k'_{max}(i) \neq k$). But we can skip such repetitions for different exemplars k since the maximum remains unchanged for them. In this way, the substitute term $w(i)$ can be inserted directly into the following equation:

$$r(i, k) \leftarrow s(i, k) - \max \{a(i, k') + s(i, k')\}, \quad k' \text{ s.t. } k' \neq k.$$

Hence, the new running time of updating “responsibilities” for all nodes i with their exemplars k is of order $O(N^2)$.

¹In this section, $i = 1, 2, \dots, N$ and $k = 1, 2, \dots, N$.

Accordingly, such an improvement can be applied to the “availability” $a(i, k)$ (where the message is sent from a potential exemplar k to a node i) with a substituting term:

$$q(k) = \sum \max \{0, r(i', k)\}, \quad i' \text{ s.t. } i' \neq i, k. \quad (3.38)$$

As a result, the total computational complexity of updating the messages of “responsibility” and “availability” reduces from $O(N^3)$ to $O(N^2)$.

In practical applications, we should not only consider the algorithm’s computational efficiency but also the flexibility of the clustering scheme in meeting the investors’ needs. A typical need of the investors is to control the size of portfolio. For instance, individual investors may prefer holding a small number of equities in a time with more fundamental analysis whereas professional fund managers could manage some portfolios which consist of a large number of equities due to their access to additional capacity of money and information.

Therefore, we expect to obtain some clustering solutions with their number of clusters (or cluster centers) located in some desired intervals. In the original **AP** method, the number of clusters is influenced by the values of the input “preferences” $p(i)$ which indicate preference that node i is chosen as a cluster center (Frey and Dueck, 2007). However, such values of $p(i)$ are always difficult to prescribe and do not necessarily lead to desired numbers of clusters (Wang et al., 2008). To solve this problem, we plan to integrate *Between-Within Proportion* (**BWP**) (Zhou et al., 2011) into the **AP** method. As we introduced in Section 2.2.3, **BWP** works as a criterion to evaluate the clustering result. The solutions with high **BWP** values will be considered as good ones. In this integrated clustering framework, our full scheme is described in Algorithm 2.

With the improved clustering technique, the construction of the social network structure based on various dependence measures is accomplished. We choose one item from each of the following categories to implement Algorithm 2 and generate stock networks,

1. **Market Source:** SHASHR, SZASHR and SHSZ300;
2. **Market Phase:** Full Length (from 1/4/2006 to 11/28/2013), Bull Phase (from 1/4/2006 to 1/4/2008), Bear Phase (from 1/4/2008 to 10/26/2008) and Mild Phase (from 1/11/2011 to 6/7/2013);
3. **Dependence Measures:** Linear Correlations, Kendall’s, Spearman’s, Lower Tail Dependence indicated by Clayton Copula, Upper Tail Dependence indicated by Gumbel Copula, Tail Dependence indicated by t Copula, Mutual Information (or Negative Copula Entropy) indicated by Gaussian, t, Clayton, Frank and Gumbel Copula.

The entire experiment is conducted through all combinations of the above items and leads to a large amount of results, so we choose some examples to illustrate our analysis. Figure 3.7 illustrates the equity networks constructed by the social network clustering. It contains all equities from SHASHR and its dependence structure is revealed by the linear correlation. On the plot, each node represents an unique equity. Every node belongs to a certain cluster and such a relationship is described by an edge linking the node and its cluster center. The centers indicate the target equities to be inserted into a portfolio. We display some locally optimal solutions to control the size of a portfolio based on different needs.

In Figure 3.8, we fix the portfolio size and observe the equity networks constructed by different dependence measures. Note the true meaning of cluster centers behind these dependence measures. When we obtain a network based on a certain structure, e.g. a linear correlation, it is not that the cluster centers are more linearly correlated to each other in terms of their returns. On the contrary, they are chosen with a prediction of weak correlation among them. This is due to the transform of all nodes' similarity (which are denoted by the coefficients of dependence measures) so that the nodes located in different clusters are "distantly correlated" to each other and the cluster centers indeed act as the "leading roles". Accordingly, in terms of lower tail dependence, we expect the equities to be less likely to suffer a pairwise significant loss. In this sense, a cluster center will hold weak associations with all other cluster centers and we expect a better diversification from these equities, especially in the case when the market undergoes an extreme downside-movement (e.g. SHASHR from 1/4/2008 to 10/26/2008).

The statistics of some social networks are presented in Table 3.5 and Table 3.6, with the former displaying long-term results and the latter summarizing the results associated with the bear phase, respectively. Based on various dependence measures, the statistics reveals the optimal solutions located in continuous intervals with corresponding **BWP** values¹. In particular, Figure 3.7 corresponds to the 1st row in Table 3.5 whereas Figure 3.8 corresponds to the 4th column of Table 3.6. Furthermore, Table 3.7 presents a full description of the chosen equities as an expansion of Table 3.5's 1st row.

Therefore, with the modeling of dependence structure and the construction of social network clustering framework, we are able to achieve the first main goal of this thesis: to find a strategy of selecting equities for portfolios. The cluster centers in the final solutions represent those target equities to be inserted into a portfolio. In the next two chapters we discuss our portfolios under a mean-variance framework and measure their diversification.

¹Note that in brackets we time all **BWP** values by 100 for better views.

Algorithm 2: Algorithm of Social Network Clustering

Input: Cleaned Data (from Algorithm 1)

Output: Cluster Centers and Values of **BWP**

- 1 Model the dependence structure of market data,
 - (I) Estimate the coefficients $a_{i,k}$ for all dependence measures on full time scale,
 - (II) Estimate $a_{i,k}$ for all dependence measures on different market phases;
- 2 Compute the distances $s(i, k)$ among all pairs of data points,

$$s(i, k) = \sqrt{2(1 - a_{i,k})} ;$$

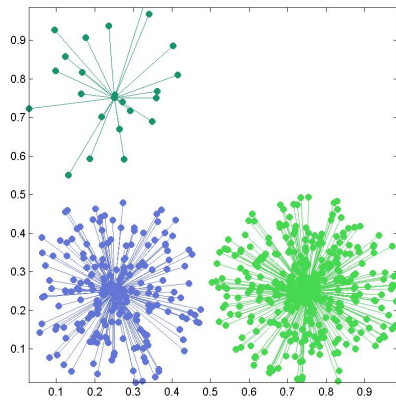
- 3 Prescribe number of clusters k and execute *Steps* 4 to 8 for all $k \in \{2, 3, \dots, 30\}$;
- 4 Set the range of “preferences” p ,

$$p \in \left[\frac{1}{8} \min\{s(i, k)\}, \max\{s(i, k)\} \right] ;$$

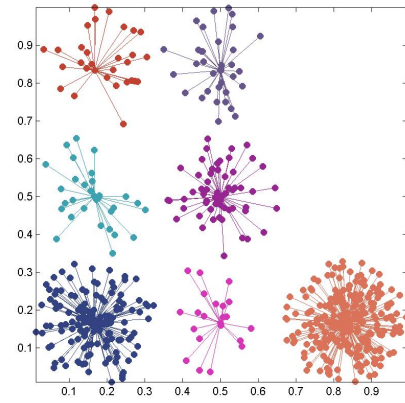
- 5 Pick a value for p from its range;
- 6 Apply affinity propagation clustering algorithm with the given value of p , obtaining a solution with n clusters;
- 7 Check,
 - (I) if n equals k , then retain the solution,
 - (II) if n does not equal k , then apply bisection method to the range of p in *Step* 4 and redo *Steps* 5 to 7;
- 8 Compute the solution’s **BWP** value: $\text{BWP}(k)$;
- 9 Set 6 intervals,

$$[2, 5], [6, 10], [11, 15], [16, 20], [21, 25], [26, 30];$$

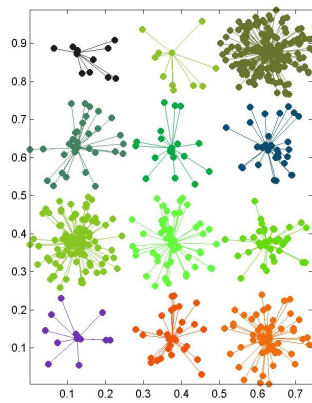
- 10 Find a k in each of the intervals (in *Step* 9) with the maximum $\text{BWP}(k)$ and such corresponding solutions will be chosen at last.
-



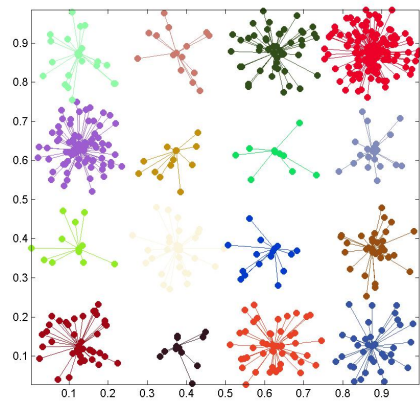
(a) 3 Clusters



(b) 7 Clusters

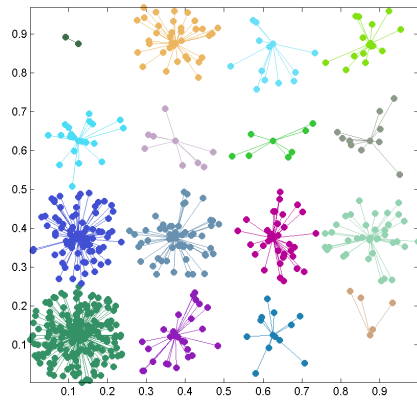


(c) 12 Clusters

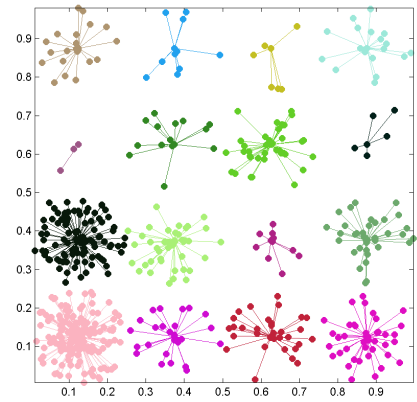


(d) 16 Clusters

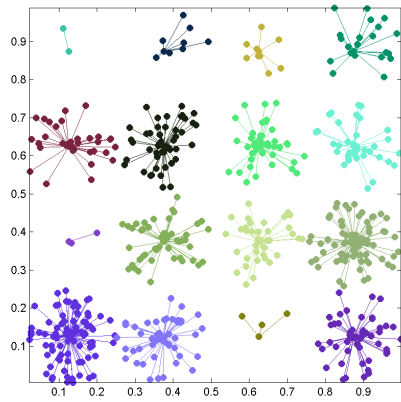
Figure 3.7: Social Network Clustering on Linear Correlations. Data: Market SHASHR from 1/4/2006 to 11/28/2013 (Entire Length).



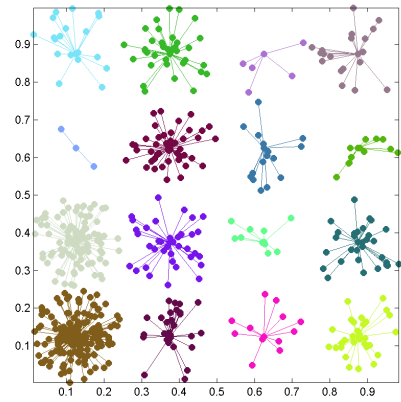
(a) Pearson's Linear Correlation



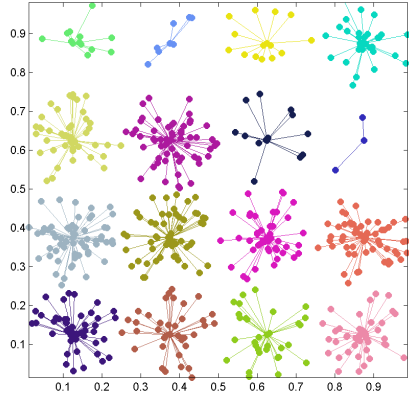
(b) Kendall's



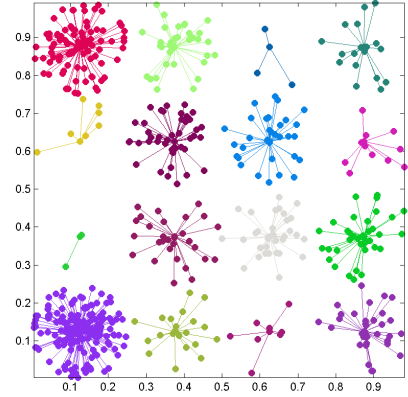
(c) Lower Tail (Clayton Copula)



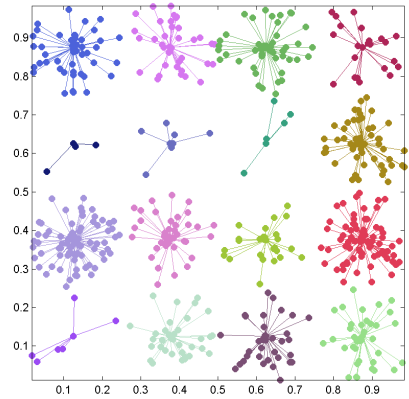
(d) Upper Tail (Gumbel Copula)



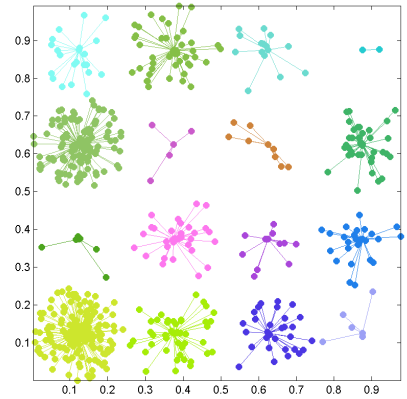
(e) Tail Dependence (t Copula)



(f) MI (Gaussian)



(g) MI (Clayton)



(h) MI (Gumbel)

Figure 3.8: Social Network Clustering on Multiple Dependence Measures with Illustration of 16 Clusters. Data: Market SHASHR from 1/4/2008 to 10/26/2008 (Bear Phase).

Table 3.5: Statistics of Social Network Clustering Results. Data: Market SHSZ300 from 1/4/2006 to 11/28/2013 (Entire Length).

Dependence Measures	Optimal Number of Clusters (with $BWP \times 10^2$) in Different Ranges					
	2 – 5	6 – 10	11 – 15	16 – 20	21 – 25	26 – 30
Linear	3(0.63)	6(-0.96)	11(-3.83)	20(-6.83)	24(-6.03)	30(-5.35)
Kendall's	2(0.38)	6(-1.01)	11(-3.69)	16(-6.08)	24(-6.04)	30(-5.28)
Spearman's	3(0.80)	6(-0.61)	11(-4.01)	16(-7.23)	25(-6.75)	30(-4.90)
Lower Tail (Clayton)	3(0.96)	6(-1.15)	11(-3.37)	17(-5.73)	25(-3.89)	28(-2.82)
Upper Tail (Gumbel)	3(0.38)	6(-0.85)	11(-3.36)	16(-6.11)	24(-5.81)	30(-5.20)
Tail (t Copula)	2(0.50)	6(-1.41)	11(-3.37)	16(-4.62)	25(-3.53)	30(-1.38)
MI (Gaussian)	2(-0.07)	6(-1.26)	11(-2.60)	16(-4.48)	25(-5.13)	30(-3.87)
MI (t Copula)	2(0.17)	6(-0.74)	11(-2.44)	18(-4.20)	25(-2.94)	30(-1.74)
MI (Clayton)	2(-0.02)	6(-1.32)	11(-2.93)	20(-4.18)	25(-3.06)	28(-2.36)
MI (Frank)	2(0.08)	6(-1.09)	11(-2.44)	16(-4.32)	25(-4.06)	30(-3.43)
MI (Gumbel)	2(-0.01)	6(-1.00)	11(-2.38)	16(-4.06)	25(-4.63)	28(-3.69)

Table 3.6: Statistics of Social Network Clustering Results. Data: Market SHASHR from 1/4/2008 to 10/26/2008 (Bear Phase).

Dependence Measures	Optimal Number of Clusters (with $BWP \times 10^2$) in Different Ranges					
	2 – 5	6 – 10	11 – 15	16 – 20	21 – 25	26 – 30
Linear	2(1.14)	6(-0.93)	12(-1.62)	16(-2.33)	22(-4.38)	26(-4.39)
Kendall's	2(0.30)	6(-0.50)	11(-0.83)	16(-1.51)	21(-2.30)	26(-3.07)
Spearman's	2(1.05)	7(-1.12)	11(-1.15)	17(-2.29)	21(-3.91)	26(-4.00)
Lower Tail (Clayton)	2(0.59)	6(-1.03)	12(-1.99)	16(-2.84)	22(-3.33)	26(-3.51)
Upper Tail (Gumbel)	2(0.30)	6(-0.85)	11(-3.36)	16(-6.11)	21(-5.81)	26(-5.20)
Tail (t Copula)	5(-0.13)	6(-0.13)	11(-0.92)	16(-1.68)	22(-2.66)	26(-3.23)
MI (Gaussian)	2(0.30)	6(-0.34)	11(-0.90)	16(-1.25)	21(-2.26)	26(-3.16)
MI (t Copula)	2(0.25)	7(-0.32)	11(-0.61)	16(-1.30)	21(-1.63)	27(-2.56)
MI (Clayton)	2(0.22)	6(-0.53)	11(-1.10)	16(-1.56)	21(-2.21)	26(-2.73)
MI (Frank)	2(0.20)	6(-0.30)	11(-1.10)	16(-1.36)	21(-1.66)	26(-2.50)
MI (Gumbel)	2(0.22)	6(-0.28)	12(-1.43)	16(-1.80)	21(-2.17)	26(-2.47)

Table 3.7: Selection of Stocks by Social Network Clustering on Linear Correlations. Data: Market SHSZ300 from 1/4/2006 to 11/28/2013 (Entire Length).

Number	Equity Tickers Corresponding to Cluster Centers					
	2 – 5(3)	6 – 10(6)	11 – 15(11)	16 – 20(20)	21 – 25(24)	26 – 30(30)
1	000630 CH	000002 CH	000002 CH	000002 CH	000002 CH	000002 CH
2	600085 CH	000528 CH	000528 CH	000422 CH	000422 CH	000063 CH
3	600642 CH	000630 CH	000630 CH	000528 CH	000528 CH	000528 CH
4		600030 CH	600000 CH	000630 CH	000630 CH	000629 CH
5		600085 CH	600019 CH	000656 CH	000656 CH	000630 CH
6		600642 CH	600029 CH	000729 CH	000729 CH	000656 CH
7			600030 CH	000983 CH	000983 CH	000729 CH
8			600085 CH	002001 CH	002001 CH	000792 CH
9			600188 CH	600000 CH	600000 CH	000983 CH
10			600642 CH	600019 CH	600019 CH	002001 CH
11			600809 CH	600029 CH	600029 CH	600000 CH
12				600030 CH	600030 CH	600019 CH
13				600085 CH	600085 CH	600029 CH
14				600104 CH	600104 CH	600030 CH
15				600118 CH	600118 CH	600085 CH
16				600489 CH	600315 CH	600104 CH
17				600642 CH	600340 CH	600118 CH
18				600694 CH	600415 CH	600132 CH
19				600703 CH	600489 CH	600315 CH
20				600809 CH	600598 CH	600340 CH
21					600642 CH	600415 CH
22					600694 CH	600489 CH
23					600703 CH	600550 CH
24					600809 CH	600598 CH
25						600642 CH
26						600694 CH
27						600703 CH
28						600783 CH
29						600809 CH
30						600832 CH

Chapter 4

Portfolio Selection Evaluation

This chapter is dedicated to the performance evaluation of the portfolios consisting of the selected equities based on the social network clustering. The numerical analysis is performed through both the mean-variance framework (which is introduced in Section 2.3) and the mean-diversification framework (introduced in Section 2.4). With such an analysis, we are looking to acquire a good understanding of the portfolio selection strategies in terms of the dependence measures and the market phases.

4.1 Mean-Variance Analysis

In the classical mean-variance framework, the optimization goal is to maximize a portfolio's return given a acceptable level of risk (denoted by variance or standard deviation of returns), whose optimization structure is presented in 2.67. Alternatively, we can choose to minimize a portfolio's risk with a targeted level of return as presented in 2.66. In both applications we aim at achieving the optimal portfolio allocations.

In ideal situations where we do not consider the portfolio constraints but only a general budget constraint¹, the efficient frontier could reach the optimal level of returns or risk. However, such a situation is not applicable to most real investors as they are usually restricted by further constraints. Here, we describe the full portfolio constraints for the MV analysis:

¹The portfolio weights ω_i follows $\sum_{i=1}^N \omega_i = 1$, $i = 1, 2, \dots, N$.

- Long-Short constraint: $-0.1 \leq \omega_i \leq 1$, $i = 1, 2, \dots, N$;
 1. We allow for short sell on a certain risky asset up to 10% of total wealth;
 2. The money allocated to a certain risky asset should not exceed total wealth.
- General budget constraint: $\sum_{i=1}^N \omega_i = 1$.

Next, we describe the historical return of a certain risky asset by Equation 2.60. The risky assets in one portfolio selection are the cluster centers coming from the social network clustering solution based on some specific settings. For instance, the pool of equities will be the 5th column of Table 3.7 with the following settings:

Market Source for Clustering: SHSZ300;
 Time Series Length for Clustering: 1/4/2006 to 11/28/2013 (Entire Length);
 Dependence Measure for Clustering: Linear Correlation;
 Portfolio Size Preference: 21 – 25 (Picking the one with largest BWP).

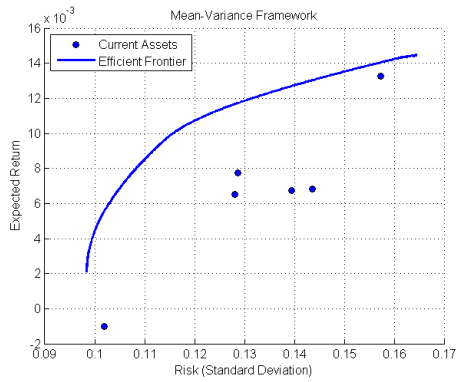
In the following context, all of the portfolio selections will follow a similar pattern of market settings as the above example. The mean-variance efficient allocation strategies are then performed on our selections based on a variety of specific settings, e.g. choices of market source, time series length, dependence measures and so on.

Figure 4.1 illustrates the performance of the MV allocations¹ on some selections from Market SHASHR². In each of the sub-figures, the blue curve represents the efficient frontier based on the current stock selection given the portfolio constraints whereas the blue points scattered under it denote the individual stocks from the selection with the corresponding mean returns and the corresponding volatilities³. In each row of the sub-figures, we compare the allocation performance with the same dependence measure but a distinct number of stocks. The result indicates that with a given level of risk, the expected returns

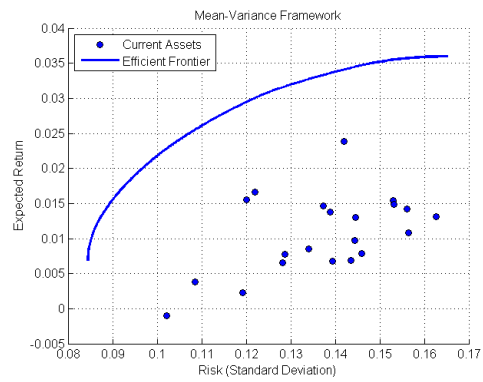
¹To obtain MV efficient allocations, we solve 2.66 or 2.67 for optimal portfolio weights. Equation 2.68 describes the efficient frontier in a σ (standard deviation), μ (expected return) space.

²*Shanghai Stock Exchange* is one of the major stock exchanges operating in China. *A shares* are issued for the domestic currency trading.

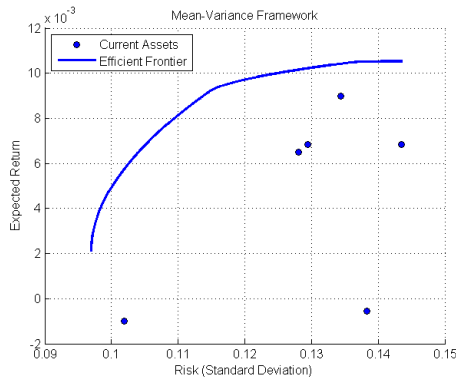
³Note that these points are not portfolios. Any two distinct portfolios on the frontier can generate the entire frontier. However, the individual stocks can not replace the efficient portfolios. As a result, given a pool of some risky assets and some allocation constraints, there will exist a unique mean-variance efficient frontier.



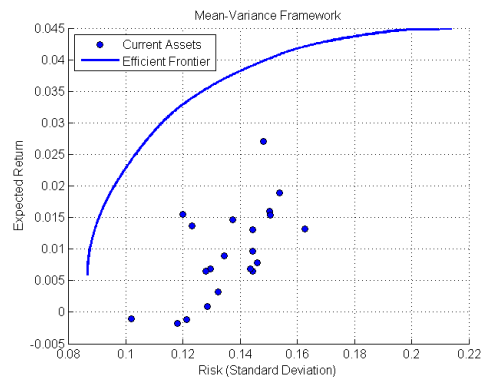
(a) Spearman 6 Equities



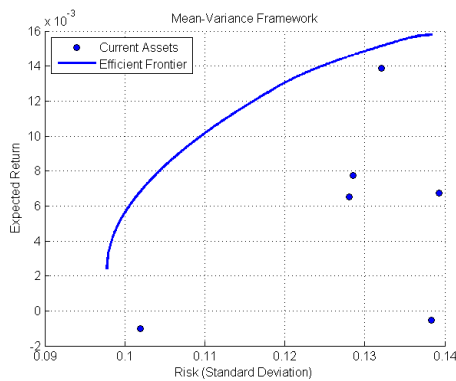
(b) Spearman 21 Equities



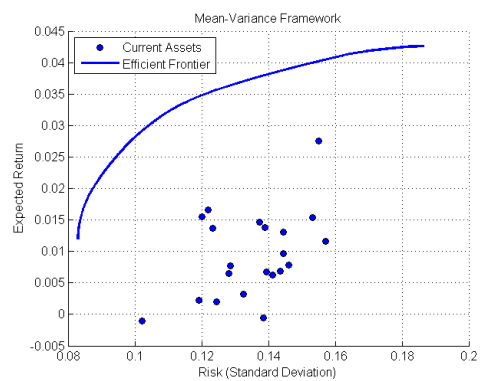
(c) Tail Dependence (t Copula) 6 Equities



(d) Tail Dependence (t Copula) 21 Equities

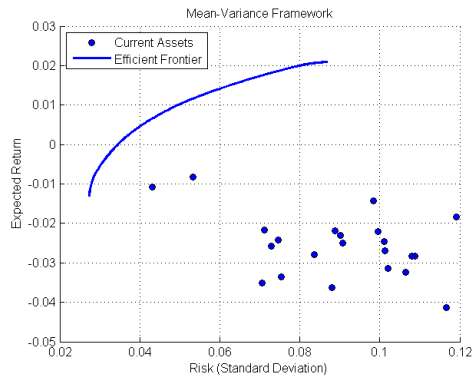


(e) MI (Clayton) 6 Equities

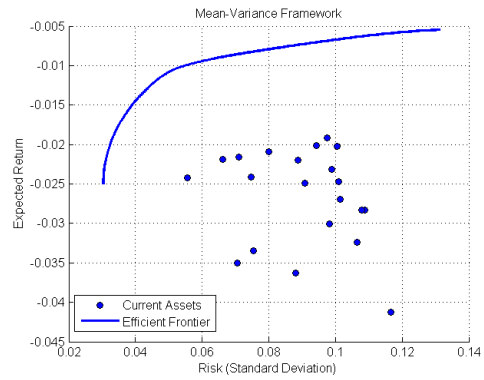


(f) MI (Clayton) 21 Equities

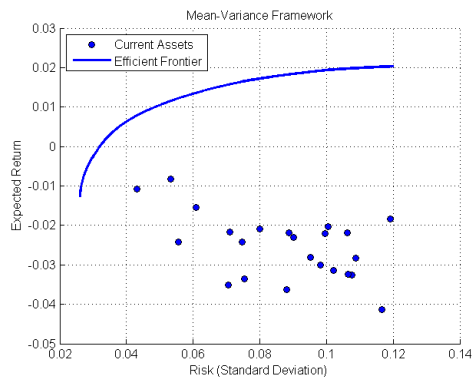
Figure 4.1: MV Efficient Frontiers Consist of Small or Large Number of Equities based on Various Dependence Measures. Data: Market SHASHR from 1/4/2006 to 11/28/2013 (Monthly Scale, Entire Length).



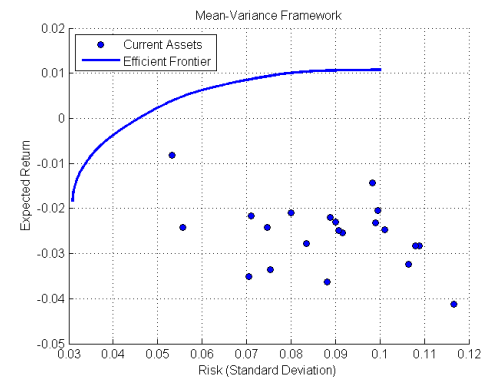
(a) Linear Correlation 22 Equities



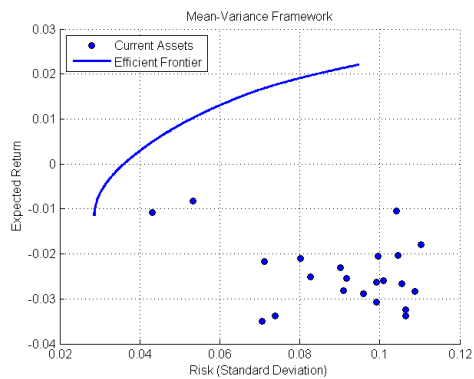
(b) Kendall's 21 Equities



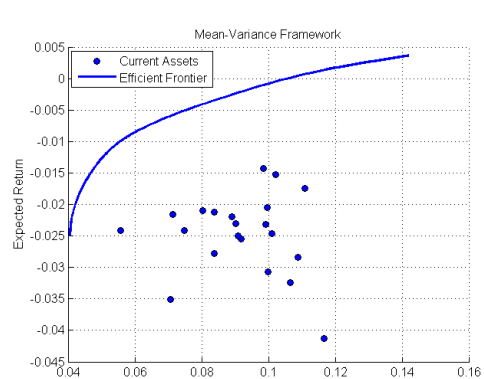
(c) Lower Tail Dependence 23 Equities



(d) Upper Tail Dependence 21 Equities



(e) Tail Dependence (t Copula) 22 Equities



(f) MI (Gumbel) 21 Equities

Figure 4.2: MV Efficient Frontiers Consist of Large Number of Equities based on Various Dependence Measures. Data: Market SHASHR from 1/4/2008 to 10/26/2008 (Weekly Scale, Extreme Bear Phase).

significantly improve as we insert more stocks into the allocations. Vertically viewing, the stock selections based on the tail dependence (the stocks are distantly scattered on the social network in terms of the tail dependence structure revealed by the t copula) have better performance in both the small and the large portfolios comparing to the remaining two portfolios. Considering the length of historical data, we conjecture that the tail dependence indicated by the t copula will better capture the association structure among assets than the Spearman’s rank correlation. Thus, the corresponding stocks chosen by the data clustering technique are less associated in the movement of their returns and can achieve a higher level of expected return given the same level of risk. In contrast, we observe slight differences in the performance of the equity selections between the tail dependence and the mutual information.

To gain another perspective of the equity selections based on various dependence measures, we apply the MV approach to the large portfolios on an extreme downside market phase. The corresponding efficient frontiers are illustrated in Figure 4.2. The Markov regime switching model identifies 1/4/2008 to 10/26/2008 as a “bear” phase where the SHASHR index sharply decreases from 4951.35 to 2083.77. The market is highly volatile with a depressing weekly expected return of -0.0217 . Such a severe depression is not only caused by the global financial crisis but also due to the violent bursts of the deceptive bubbles inflated by the Chinese real estate industry¹. During the depression, most equities admit negative expected returns and even the MV efficient allocations fail to produce profits. As we see in Figure 4.2b, all the efficient points (with the current portfolio constraints) of the equity selection by the Kendall’s τ imply a loss and those of the selection by the mutual information (Gumbel copula) (i.e., the negative Gumbel copula entropy) can hardly make through the break-even level. This suggests a recommendation to withdraw wealth from the equity market during this period as there do not exist a fully feasible short selling option². Even in such bad times, we still observe some favorable portfolio selections. The equity selection plotted in Figure 4.2c are the cluster centers based on the linear correlation which pursues the lowest degree of such dependence among the selected equities. With a proper allocation, these less correlated equities are able to provide a good premium in the MV approach. Accordingly, the selection by the lower tail dependence which focuses on weakening the association of the extreme loss also proves to be impressive in the bear phase. Such equity selections provide choices for the investors who can not fully withdraw

¹The SHASHR Index has been climbing up before the bursts, breaking through 6000 in 2007. With the confidence in the local economic blooming and the faith in the Government policy control for the 2008 Olympic Games hosted in the city of Beijing, some researchers “predict” SHASHR Index would hit 10000.

²The Chinese futures market was not reopened at that point. Not to mention that any hedging strategies by longing the put options can not work as they are not available for trading at all.

themselves from the market.

We further present a cross-sectional performance of the MV efficient allocations with respect to all the dependence measures. In Table 4.1, the first 3 columns on the left represent the monthly volatilities corresponding to the expected returns in the small equity selections. The 3 columns in the middle display the scenarios of the large equity selections. We use the long-term dependence data for the social network clustering in order to obtain the above selections. As we can observe, with the current portfolio constraints, some levels of the expected returns are not attainable by the MV efficient allocations. Removing the constraints might bring a chance for reaching a target of returns, but it is at the cost of greatly increasing the portfolio's volatility. In the case of the small portfolios, the selection with the linear correlation turns out to be the best. The 2nd place is taken by the selection with the lower tail dependence and the third by the MI (Mutual Information, Clayton Copula). However, it is rather obscure to see if these small portfolio selections could beat the market in long terms. We raise the target levels of expected returns for the large portfolios. In this case, the selection with the linear correlation seems rather unpredictable when we require high premiums. We conjecture that the investment length should partially have influence on it. As shown in Figure 3.1, all equities are positively correlated to each other in longer investment periods. The corresponding correlation matrix might not necessarily lead to a clustering solution in which the cluster centers are all distantly located. In the market view, it is possible to find some stocks which are less mutually correlated in long terms but difficult to do so with a large number of equities. In contrast, the selections with the measures representing the non-linear dependence structure give an impressive performance. Among them the selections by the lower tail dependence and the MI (Clayton Copula) outrun the rest. The tail structure and the mutual information indicated by t copulas also lead to good selections. Their dependence data always emphasize the ties lurked in the tail distributions, leading to excellent social networks and clear clustering solutions even in the long run. With less uncertainty, the selections by the tail dependence and the mutual information admit good risk premiums.

The previous selections are chosen based on the data over a long time. In the right 3 columns of Table 4.1, we examine the performance of the equity selections from the bear market phase. Note that we refer the "bear" phase (from 1/4/2008 to 10/26/2008) to the market data for studying the dependence structures and the clustering results. Then we apply the MV approach to the equity selections with the same data (from 1/4/2006 to 11/28/2013) as in the previous studies. Hence, these selections hold particularly weak ties in terms of their dependence during the extreme downside market phase. To our surprise, with respect to all the dependence measures, the equity selections from the bear phase outrun those from the longer term. We observe the lower level of volatilities given

the same level of expected returns, especially in the linear correlation case. Indeed, during this extremely depressing and relatively short period, the discrepancies among the pairwise correlations (or other dependence measures) can be better observed and appreciated. Thus, the corresponding social network clustering solutions contain more meaningful information. On the other hand, the equity selections with the lower tail dependence and the MI (Clayton Copula) still admit more robust performance. The MI (Frank Copula)'s selection slightly outrun the others.

With more interest in studying the “bear” phase, we present the equity selections with the MV approach in Table 4.2. In this table, both the equity selections and the backtests are built upon the bear market data. Due to the sharp decrease, the returns and the corresponding volatilities are weekly scaled. In this case, the small selections are basically not profitable, though the ones with the tail dependence (indicated by Clayton copula and t copula) and the MI (t Copula) are less risky than the market portfolio. They will be slightly profitable when switching to a large size, however the corresponding risk is not substantially reduced.

As a result, the equity selections by the social network clustering based on the lower tail dependence prove to be more robust than the equity selections by other dependence measures within a mean-variance framework. The large equity selections with the tail dependence indicated by the t copula also show good performance. The large equity selections with the linear correlations are outstanding if the data is abstracted from the bear market phase.

In the next section, we study the equity selections from the perspective of diversification.

Table 4.1: The Mean-Variance Analysis of the Equity Selection from Different Market Settings. The equities (cluster centers) are selected from the market data of the full investment period and the bear phase respectively. All MV analysis is performed based on the data consisting of returns from the full investment period (“>” and “<” mean the MV allocation on the frontier can not exactly achieve the target return due to the portfolio constraints, but it is very close. “UA” means unattainable, indicating that under the current portfolio constraints, the target return is unattainable by the MV allocation on the frontier).

	Data Source: Market SHASHR (1/4/2006 to 11/28/2013, Entire Length)					
	Volatilities σ Corresponding to Returns μ					
	Equities from Full Period		Equities from Full Period		Equities from Bear Phase	
Dependence Measures	Small Portfolio (6 – 10) with μ (monthly) equals	Large Portfolio (21 – 25) with μ (monthly) equals	Small Portfolio (6 – 10) with μ (monthly) equals	Large Portfolio (21 – 25) with μ (monthly) equals	Small Portfolio (6 – 10) with μ (monthly) equals	Large Portfolio (21 – 25) with μ (monthly) equals
Linear	0.005	0.010	0.015	0.010	0.020	0.030
Kendall's	< 0.0903	0.0910	0.0969	0.0828	0.0963	> 0.1811
Spearman's	0.1009	0.1156	> 0.1643	0.0854	0.0966	UA
Lower Tail (Clayton)	0.1009	0.1158	> 0.1650	0.0851	0.0965	0.1218
Upper Tail (Gumbel)	< 0.0930	0.0965	0.1074	< 0.0838	0.0895	0.1106
Tail (t Copula)	0.1011	0.1167	> 0.1683	0.0852	0.0965	UA
MI (Gaussian)	0.1002	0.1249	UA	0.0874	0.0959	0.1125
MI (t Copula)	0.0954	0.1145	0.1674	0.0852	0.0950	0.1504
MI (Clayton)	0.1011	0.1161	> 0.1640	0.0837	0.0916	0.1142
MI (Frank)	0.0993	0.1095	0.1310	< 0.0828	0.0872	0.1039
MI (Gumbel)	0.1009	0.1156	0.1643	0.0838	0.0916	0.1147
	0.1009	0.1156	0.1643	0.0852	0.0947	0.1383

Table 4.2: The Mean-Variance Analysis of the Equity Selection from Different Market Settings. The equities (the cluster centers) are selected from the market data of the bear phase. All MV analysis is performed based on the data consisting of returns from the bear phase (“>” and “<” mean the MV allocation on the frontier can not exactly achieve the target return due to the portfolio constraints, but it is very close. “UA” means unattainable, indicating that under the current portfolio constraints, the target return is unattainable by the MV allocation on the frontier).

	Market SHASHR (1/4/2008 to 10/26/2008, Bear Phase)	
	Volatilities σ Corresponding to Returns μ	
	SHASHR Index (Weekly) — $\mu = -0.0217, \sigma = 0.0665$	
Dependence Measures	Small Size (6 — 10 Equities)	Large Size (21 — 25 Equities)
	with μ (weekly) equals	with μ (weekly) equals
	—0.025 —0.0215 —0.018	—0.01 0.000 0.01
Linear	0.0630 0.0665 UA	0.0279 0.0345 0.0496
Kendall's	0.0512 0.0527 0.0594	0.0550 UA UA
Spearman's	< 0.0508 0.0519 0.0561	0.0515 > 0.1501 UA
Upper Tail (Gumbel)	0.0632 0.0662 0.0772	0.0338 0.0458 0.0792
MI (Gaussian)	0.0634 0.0672 0.0768	0.0761 UA UA
MI (Clayton)	0.0661 0.0708 0.0789	0.0323 0.0429 0.1203
MI (Frank)	0.0634 0.0672 0.0768	0.0338 0.0481 UA
MI (Gumbel)	0.0633 0.0662 0.0770	0.0556 0.1052 UA
	with μ (weekly) equals	
	—0.015 —0.0113 —0.0075	
Lower Tail (Clayton)	< 0.0470 0.0475 0.0496	0.0265 0.0320 0.0487
Tail (t Copula)	0.0499 0.0509 0.0537	0.0288 0.0360 0.0528
MI (t Copula)	0.0377 0.0403 0.0469	0.0294 0.0353 0.0520

4.2 Portfolio Diversification Analysis

As we documented in Section 2.4, the diversification is another important key to achieve a robust portfolio selection. Meucci (2009) proposes to decompose market risk sources into uncorrelated ones using the principal component analysis (**PCA**). Then, the portfolio risk can be further characterized as some additive risk contributions which are denoted by a diversification distribution (see Equation 2.82). Through an entropy risk measure, we are able to quantify the degree of the total diversification in some portfolios. Such a methodology also provides a framework for proper reallocations in order to obtain the efficient portfolios considering the maximum diversification or the maximum expected returns (see Equation 2.87). Hence, analogous to the mean-variance analysis, a mean-diversification efficient frontier comes to life.

The diversification analysis is initialized with our MV results. The portfolio constraints stay the same as in the previous section. Here, we add a specified reallocation constraint:

- reallocation constraint: $\sum_{i=1}^N \Delta\omega_i = 0$.

1. We allow for re-allocation on all equities none of which is suspended;
2. We do not allow for refinancing.

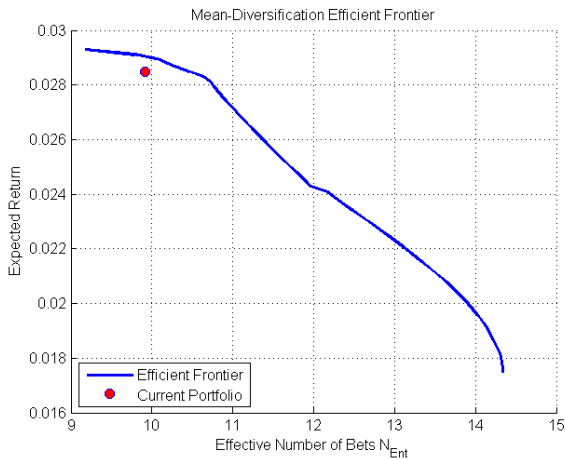
We illustrate the MD analysis of an equity selection in Figure 4.3. The equities in the selection (the portfolio is denoted by the red dot in Figure 4.3a) come from a clustering solution built upon 8 years' linear correlations. The portfolio consists of 21 equities from the Market SHASHR with their covariance matrix denoted by Σ (which is 21×21). Its current allocation is given by the MV model with a monthly expected return of 2.85%. Figure 4.3b describes this portfolio in details with its top row denoting the MV allocation weights ω on all equities. The second row in 4.3b represents the “substituting weights” $\tilde{\omega} = E^{-1}\omega$ on the principle portfolios where E is obtained by applying the **PCA** to the original portfolio (see Equation 2.70, 2.78). E holds the same dimensions as Σ , so that we have exactly 21 principal portfolios. The third row in 4.3b displays the principal portfolios' volatilities which are implied by E 's eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_{21}$. With the properties of the **PCA**, the leading eigenvector always holds the largest eigenvalue so that the first principal portfolio accounts for the major volatility in total (shown as the first bar in 4.3b). In this sense, the uncorrelated principal portfolios can be treated as the additive risk sources with the first one denoting the dominating risk factor in the market. It also builds up a connection to a diversification measure. Due to $\tilde{\omega} = E^{-1}\omega$, we can denote the weight-adjusted uncorrelated

risks by $v_i \equiv \tilde{\omega}_i^2 \lambda_i$ (see Equation 2.79). Then the volatility concentration curve and the diversification distribution (see Equation 2.81, 2.82) are presented in 4.3b by its 4th and the last row, respectively. These two bar plots imply how well the uncorrelated risk source are diversified by the current portfolio allocation. In a perfect diversification situation, all bars shall be of the same height. Apparently, the current MV efficient allocation does not produce the portfolio at that level.

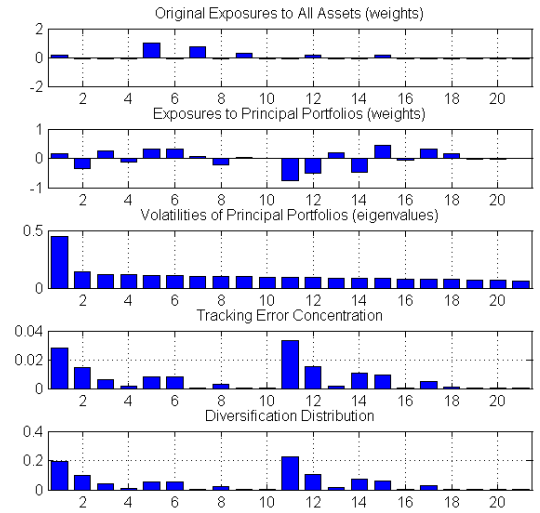
A more straightforward view of the diversification is given by the “number of effective uncorrelated bets” \mathcal{N}_{Ent} (see Equation 2.86). Meucci (2009) quantifies the level of the portfolio diversification by taking the exponential of its diversification distribution’s entropy. In Figure 4.3a, the current MV portfolio allocation displays $\mathcal{N}_{Ent} \approx 10$ which implies that it is well diversified on 10 risk source. Considering that the market is characterized by 21 uncorrelated risk sources, the diversification of the current portfolio is not optimal but still meaningful.

Following the principal of the maximum entropy (MEP), a further portfolio optimization framework is built upon the entropy diversification measure \mathcal{N}_{Ent} . Meucci (2009) names the optimization framework as “mean-diversification”(MD). It generates an MD efficient frontier by solving the optimizations proposed in 2.87 and 2.88. The ideas of the MD and the MV efficient frontiers share some similarities. In both frameworks the optimal solutions are obtained by adjusting the allocations subject to some constraints, e.g. the portfolio constraints and a target level of returns. On the other hand, the differences in both frameworks are obvious as well. Since \mathcal{N}_{Ent} represents the level of diversification, it is aimed to be as large as possible (which is exactly opposite to the volatility). Unlike in the MV, the preference between the expected return and the diversification in the MD is controlled by a parameter φ in 2.88. Xiong (2009) gives a more intuitive expression of the parameter in 2.89.

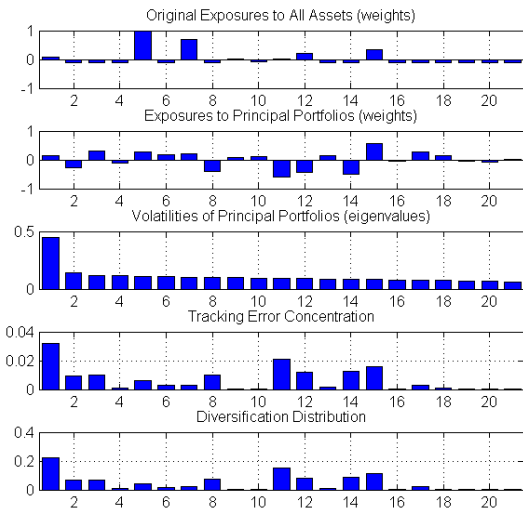
The MD efficient frontier for the equity selection is illustrated in Figure 4.3a. The frontier consists of all the efficient points each of which maximizes both expected returns and \mathcal{N}_{Ent} with respect to φ . The efficient point on the top left end represents the allocation considering the maximum expected returns while the one at the bottom right end denotes the allocation considering the maximum diversification. The red dot represents our current MV allocation on the equity selection with the corresponding expected return and the level of the diversification.



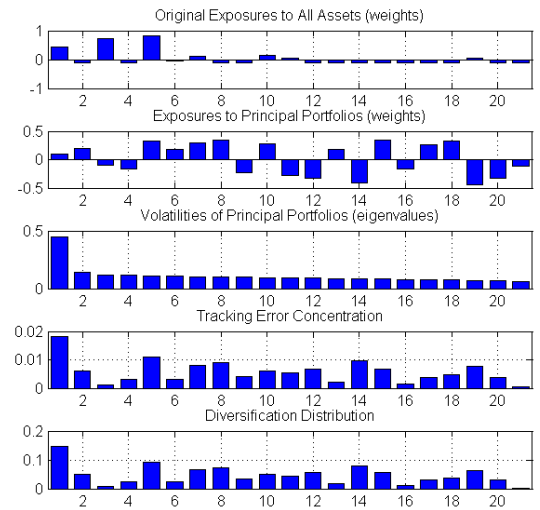
(a) MD Efficient Frontier



(b) Diversification of Current Portfolio



(c) Portfolio with Maximized Expected Return



(d) Portfolio with Maximized Diversification

Figure 4.3: The Mean-Diversification Analysis of the Equity Selection (21 Equities) based on the Linear Correlation. The original portfolio weights are given by the MV efficient allocation with a monthly expected return of 0.0285. Data: Market SHASHR from 1/4/2006 to 11/28/2013 (Monthly Scale, Entire Length).

In an ideal situation, the MD efficient frontier in Figure 4.3a should be able to smoothly stretch across \mathcal{N}_{Ent} approximately from 1 to 21. “1” means that the allocation is fully concentrated on a risk source and “21” means that the allocation is fully diversified on all risk sources. However, such a frontier can only be achieved without any constraints, so that maximizing the entropy will lead to a uniform diversification distribution with $p_i = 1/21$. In our case, the existence of portfolio constraints restricts the investors from buying or short selling assets as much as they want. Hence, a perfectly diversified allocation that all principal portfolios are exposed with the same volatility adjusted proportion on the total volatility (d-fine GmbH, 2011) might not be attainable. In our previous MV analysis, the portfolio constraints also restrict the MV efficient frontier to be part of the entire parabola. Additionally, Meucci (2009)’s experiment comes with an assumption that $\boldsymbol{\mu} \equiv 0.5\boldsymbol{\sigma}$ so that the values of expected returns are forced to be half of the corresponding volatilities. It potentially leads to a smooth and convex frontier. Xiong (2009) replicates Meucci (2009)’s work with the market data of real returns and the corresponding covariance matrix. He finds discontinuities on the efficient frontiers. He conjectures that the discontinuity is mainly due to the similarity in some largest volatilities of the principal portfolios. In our experiment, the real historical data does not suggest $\boldsymbol{\mu} \equiv 0.5\boldsymbol{\sigma}$ either, which leads to some bumps on the frontier curve. We conjecture that another possible reason which contributes to the non-smoothness could be the choice of the initial allocation¹ (in our case it is the MV efficient allocation). In real applications, the MD optimization initializes with an existing portfolio selection and adjusts its allocation in searching for the efficient points. Different choices of the initial allocations could generate a lot of ambiguities in the process of finding the optimal solutions. In our experiment, random MV allocations mostly lead to identical frontiers. Sometimes, slight differences at both ends are observed with similar overall trends. Occasionally, a few allocations lead to strange looking curves.

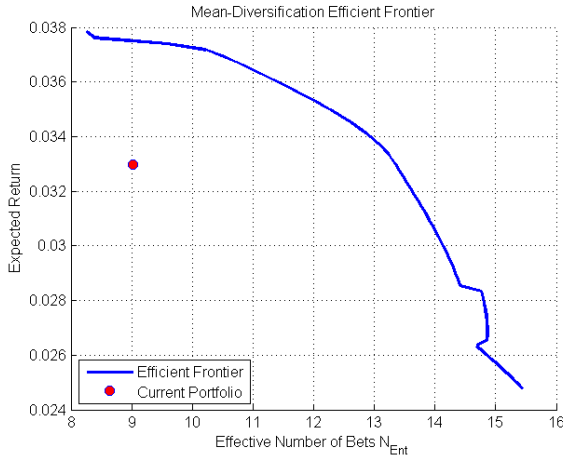
The current MV allocation implies $\mathcal{N}_{Ent} \approx 10$ and $\mu = 0.0285$. In view of diversification, it is pretty close to the MD efficient frontier. To guarantee this level of expected return given the portfolio constraints, the MV allocation almost achieves the optimal diversification (in which $\mathcal{N}_{Ent} \approx 10.5$).

Analogous to Figure 4.3b, Figure 4.3c and Figure 4.3d represent the maximum expected return and maximum diversification on the MD frontier, respectively. We can observe the difference between them in their diversification distributions. The distribution with the maximum diversification reduces the large concentrations of the risky principal portfolios and increases the allocations on the less risky principal portfolios, thus making itself more like a “uniform distribution”. Once again, due to the portfolio and the reallocation con-

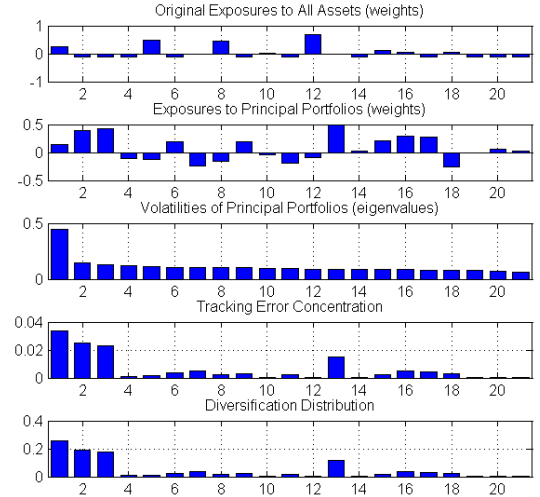
¹The initial allocation in Meucci (2009) is a naive strategy ($\boldsymbol{\omega} \equiv \mathbf{1}/N$) relative to a benchmark with its weights proportional to all equities’ market capitalization.

straints, an ideal uniform with $\mathcal{N}_{Ent} \approx 21$ is not attainable. On the other hand, we observe how the allocation changes from 4.3b to 4.3c. To achieve the maximum expected return, the reallocation increases weights on some stock holdings and short more on the remaining ones. The diversification distribution is more concentrated on some principal portfolios in 4.3c than in 4.3d.

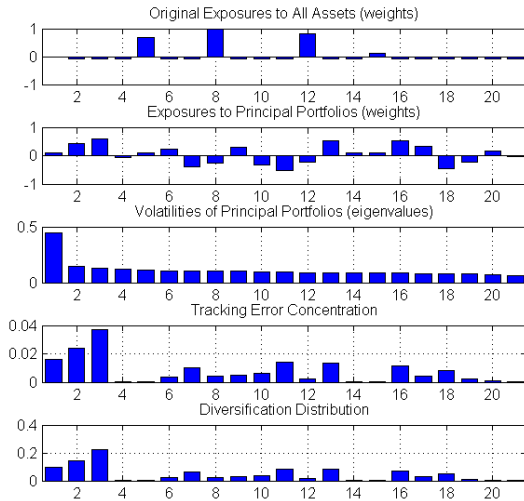
Figures 4.4, 4.5 and 4.6 illustrate more MD analysis. These cases vary in the dependence measures for the equity selections, the initial MV allocation and the market phase. Due to the reasons presented earlier, we observe non-smoothnesses in their MD efficient frontiers. In these cases, the initial MV allocations are way off the frontier curves and thus heavily suboptimal. As shown in 4.4b, 4.5b and 4.6b, the diversification distributions of the MV allocations are strongly concentrated on a few principal portfolios. In contrast, the maximum diversification distributions effectively mitigate such concentrations by spreading them among the other principal portfolios. From the initial MV allocation to the maximum diversification reallocation, the weights on both the selected equities and the transitioned principal portfolios vary significantly.



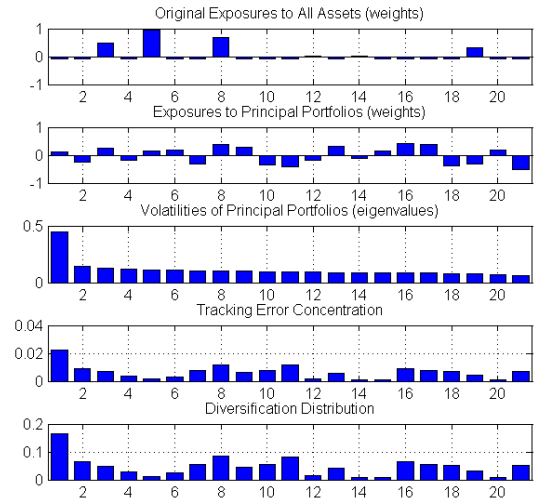
(a) MD Efficient Frontier



(b) Diversification of Current Portfolio

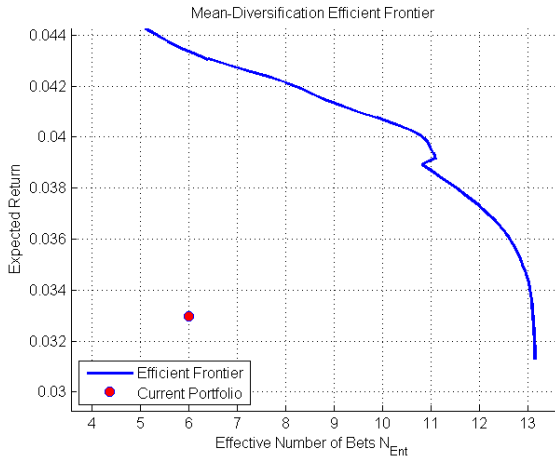


(c) Portfolio with Maximized Expected Return

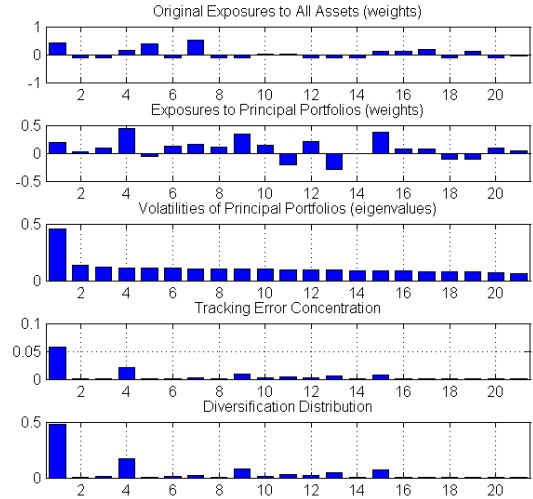


(d) Portfolio with Maximized Diversification

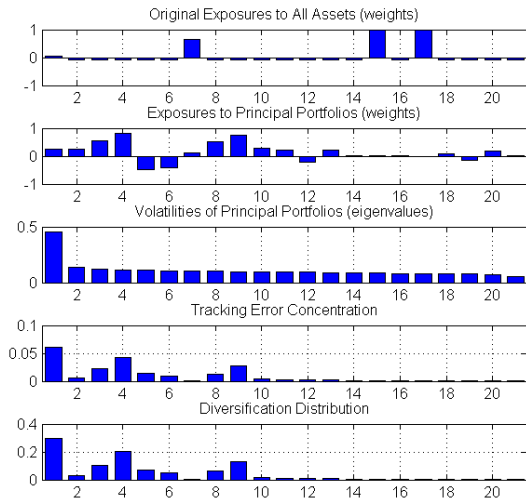
Figure 4.4: The Mean-Diversification Analysis of the Equity Selection (21 Equities) based on the Lower Tail Dependence. The original portfolio weights are given by the MV efficient allocation with a monthly expected return of 0.033. Data: Market SHASHR from 1/4/2006 to 11/28/2013 (Monthly Scale, Entire Length).



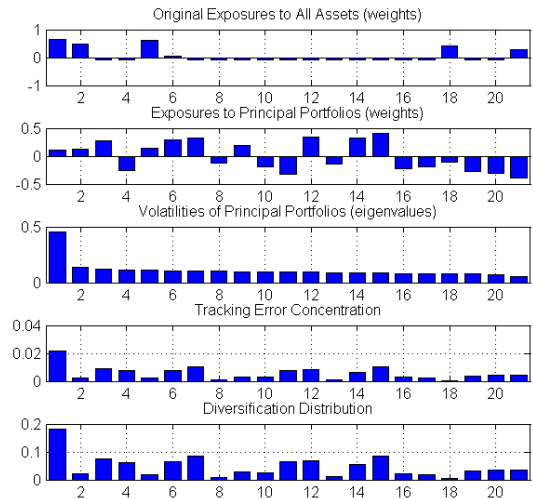
(a) MD Efficient Frontier



(b) Diversification of Current Portfolio

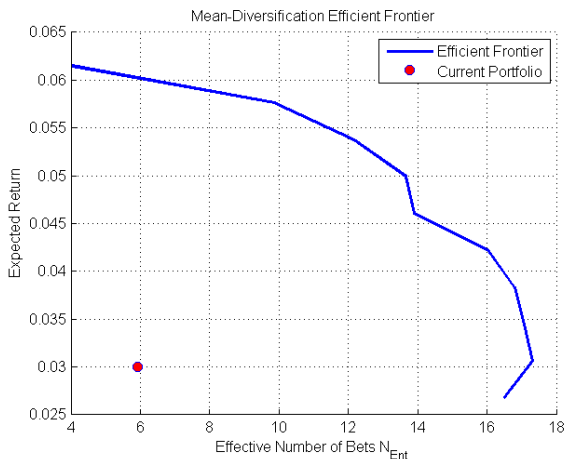


(c) Portfolio with Maximized Expected Return

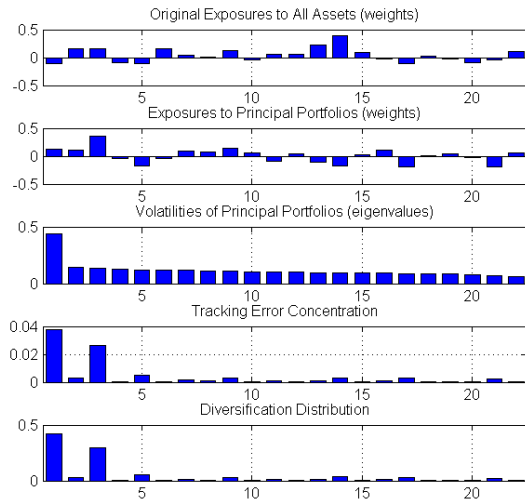


(d) Portfolio with Maximized Diversification

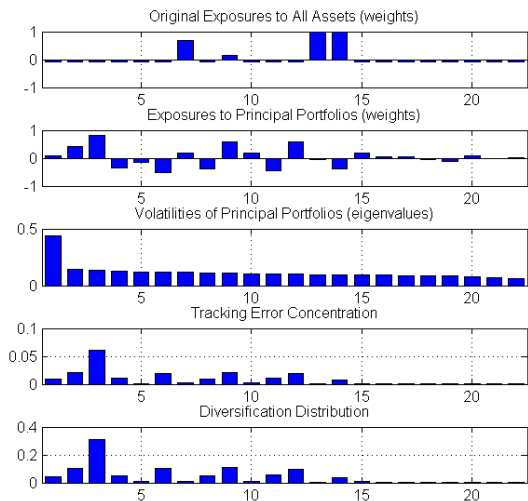
Figure 4.5: The Mean-Diversification Analysis of the Equity Selection (21 Equities) based on the Tail Dependence Indicated by t Copula. The original portfolio weights are given by the MV efficient allocation with a monthly expected return of 0.033. Data: Market SHASHR from 1/4/2006 to 11/28/2013 (Monthly Scale, Entire Length).



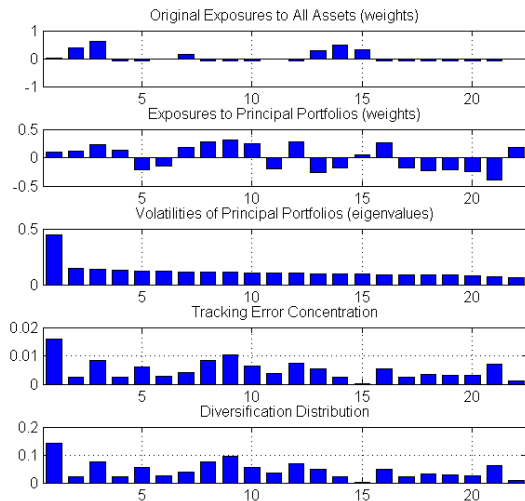
(a) MD Efficient Frontier



(b) Diversification of Current Portfolio



(c) Portfolio with Maximized Expected Return



(d) Portfolio with Maximized Diversification

Figure 4.6: The Mean-Diversification Analysis of the Equity Selection (22 Equities) based on the Linear Correlation in the Bear Market Phase. The original portfolio weights are given by the MV efficient allocation with a monthly expected return of 0.030. Data: Market SHASHR from 1/4/2006 to 11/28/2013 (Monthly Scale, Entire Length).

The previous MD analysis focuses on studying a certain equity selection. Below we present a comprehensive analysis covering a variety of market settings, e.g. different dependence measures, time series length and so on. By applying Meucci (2009)’s methodology, we verify the diversification of the equity selection with all the dependence measures implemented in our study. Each of the equity selections is given an initial MV efficient allocation with the same identical level of expected returns. Such portfolio selections are compared in terms of \mathcal{N}_{Ent} and the volatility σ . Then, we reallocate the wealth on each of the equity selections to achieve its maximum expected return and the maximum diversification on the MD efficient frontier. Such portfolio selections take \mathcal{N}_{Ent} , the expected return μ and the volatility σ into account. In practice, the new allocations are given by the MD optimal solutions, e.g. the first row illustrated in Figure 4.3c or 4.3d.

Table 4.3 presents the result of the MD analysis for the long-term market (SHASHR). The dependence measures for the social network clustering are built upon the data lasting 8 years with daily scaled returns. The equities are selected based on the long-term market performance. The size of the equity selections is chosen as “large” (21 – 25 equities). The MV allocation and the MD optimization are operated on the same data but with monthly scaled returns. The initial MV efficient allocations sets a target at a monthly expected return of 2.75%.

All equity selections result in 21 equities. In terms of the initial MV allocations, the equity selections by the Kendall’s rank correlation and the upper tail dependence (Gumbel copula) show greater \mathcal{N}_{Ent} . The portfolios have already effectively diversified the risk concentrations on 11 principal portfolios out of 21. However, their total volatilities are also much greater than the other portfolios. Taking both the diversification and the total volatility into consideration, the equity selections by the linear correlation and the MI (negative Gaussian copula entropy) represent the compromised solutions.

We next move to the MD efficient reallocation targeting at the maximum expected returns. The equity selection by the tail dependence (t copula) reaches the highest monthly return of 4.13% at the cost of bearing the largest total volatility 0.2076 and the lowest level of diversification 7.7916. Comparing with the other equity selections, this one seems too risky, thus it is not recommended. The proper dependence choices for the equity selection that pursues high risk premiums while suitably suppresses the total risk with impressive diversification are the Spearman’s rank correlation and the MI (negative Gaussian copula entropy). Through the MD efficient allocation, these two portfolios preserve the higher end of the expected return (above 3.1%) and the relatively moderate level of the total volatility (about 0.17). Moreover, they indicate outstanding diversification in which $\mathcal{N}_{Ent} \approx 14$. In this sense, they effectively diversified the volatility concentrations among 14 principal

portfolios which account for the majority of the total risk. If the investor still prefers lower total risk, the equity selections by the Kendall’s rank correlation and the upper tail dependence (Gumbel copula) will be the preferred choices.

The third session focuses on the MD efficient reallocation targeting at the maximum diversification. In this case the priority switches from producing high profits to mitigating the portfolio’s risk. If the investor merely pursues diversifying the portfolio’s risk, the equity selections by the linear correlation and the Kendall’s rank correlation are the best. The values of \mathcal{N}_{Ent} come up to almost 17. With some slight sacrifice of the diversification and the total volatility, the performance of the equity selections by the lower tail (Clayton copula) and the upper tail dependence (Gumbel copula) are also impressive. If the investor has more bias on the expected return in the maximum diversification (e.g. 2% monthly), the equity selections by the Spearman’s rank correlation, the MI(t copula) and the MI(Clayton) are recommended. The MI(Gaussian) is not preferred in this case as it goes against the priority of mitigating the total risk.

Table 4.4 also presents the result of the MD analysis for the long-term market (SHASHR). The only difference is that the equities are selected from the extreme bear market phase (1/4/2008 to 10/26/2008). The size of the equity selections is chosen as “large” (21 – 25 equities). The MV allocation and the MD optimization are operated on the same long-term data as Table 4.3.

Analogously, we evaluate the initial MV allocations first. In this case, none of the equity selections indicates a good diversification. The “best” diversification belongs to the equity selection by the MI(Gaussian copula) with \mathcal{N}_{Ent} of only 6.8 out of 21. Taking both the total risk and the diversification into considerations, the equity selections by the linear correlation and the MI(Frank copula) slightly outperform the others. Without the MD efficient reallocation, the equity selections from the bear phase generally show a poor diversification.

Then, we observe obvious differences between Table 4.4 and Table 4.3 in their maximum expected returns of the MD reallocations. The equity selections from the extreme bear phase generate much higher profits than those from the longer term. This finding is consistent with our MV analysis presented in Table 4.1. In exchange for higher premiums, lots of equity selections turn out to be riskier with a lower degree of diversification than in Table 4.3. The equity selection by the tail dependence (t copula) indicates an exceptional monthly expected return of 7% with the middle-class diversification and the volatility overall. The selection by the MI (Gumbel copula) is very well diversified with lowest total volatility. Its neighbor by the MI (Frank copula) seems slightly more volatile, but the improvement in the expected return is more substantial. These three equity selections

with the MD reallocations are good choices for the investors. They beat those selections by the same dependence measures presented in Table 4.3 in terms of all three indicators, i.e. \mathcal{N}_{Ent} , μ and σ .

Lastly, we observe the MD reallocation for the maximum diversification in Table 4.4. In this case, most selections hold higher \mathcal{N}_{Ent} than those reported in Table 4.3, indicating that the equity selections from the bear phase are better diversified by the MD efficient allocations. In 4.4, the selection by the linear correlation is overall a great choice with a top diversification, a bottom volatility and a good expected return. The selection by the MI (Frank copula) preserves the robust performance, showing a top diversification and a leading monthly expected return of 3.58%.

In conclusion, we examine the performance of the equity selections under a mean-diversification framework. The equity selections are built on the social network clustering with various dependence measures. The focus of this section is to verify the relationship between the dependence measures and the portfolio selections under different market settings. In the case when the dependence structure¹ is described by the long-term data, the Spearman's rank correlation and the Kendall's rank correlation prove to be preferred choices for the equity selections with the general market settings. The selection by the linear correlation can be well diversified. On the other hand, when the dependence structure is constructed by the extreme downside market data, the corresponding equity selections lead to better diversification. In this case, the linear correlation and the MI (Frank copula) give rise to good results. Overall, we recommend the selections extracted from extreme downside market data as they have the ability to be well diversified in "bear" market and give robust performance in long terms.

¹In other words, it refers to the estimation of dependence measures.

Table 4.3: The Mean-Diversification Analysis of the Equity Selection from Different Market Settings. The equities (cluster centers) are selected from the market data of the full investment period. The portfolio sizes are chosen as “large” (21-25 equities in each). All MD analysis is also performed based on the market data of the full investment period.

	SHASHR from 1/4/2006 to 11/28/2013 (Monthly Scale, Entire Length) Uncorrelated Bets \mathcal{N}_{Ent} Corresponding to Returns μ and Volatilities σ SHASHR Index (Monthly) — $\mu = 0.0069, \sigma = 0.0814$					
Dependence Measures	MV Efficient Portfolio High Premium		MD Efficient Portfolio Maximum Return		MD Efficient Portfolio Maximum Diversification	
	\mathcal{N}_{Ent}	μ σ	\mathcal{N}_{Ent}	μ σ	\mathcal{N}_{Ent}	μ σ
Linear	8.9692	0.0275 0.1284	10.7671	0.0280 0.1425	16.8439	0.0157 0.1232
Kendall's	10.9252	0.0275 0.1407	12.8204	0.0271 0.1457	16.8308	0.0142 0.1226
Spearman's	5.2766	0.0275 0.1139	13.4494	0.0344 0.1688	16.4932	0.0202 0.1388
Lower Tail (Clayton)	5.3135	0.0275 0.1024	10.7918	0.0358 0.1760	16.0771	0.0177 0.1245
Upper Tail (Gumbel)	10.7789	0.0275 0.1558	12.2594	0.0278 0.1574	15.7245	0.0167 0.1255
Tail (t Copula)	4.5690	0.0275 0.1075	7.7916	0.0413 0.2076	15.4163	0.0083 0.1248
MI (Gaussian)	7.5567	0.0275 0.1157	14.5796	0.0308 0.1706	15.1022	0.0299 0.1586
MI (t Copula)	6.4849	0.0275 0.1058	9.7754	0.0381 0.1975	14.1406	0.0247 0.1232
MI (Clayton)	5.8513	0.0275 0.0986	10.8726	0.0407 0.1738	15.8187	0.0236 0.1282
MI (Frank)	6.3867	0.0275 0.1060	9.8862	0.0376 0.1793	14.8371	0.0147 0.1319
MI (Gumbel)	7.1734	0.0275 0.1142	10.2988	0.0325 0.1840	14.7959	0.0261 0.1485

Table 4.4: The Mean-Diversification Analysis of the Equity Selection from Different Market Settings. The equities (cluster centers) are selected from the market data of bear phase. The portfolio sizes are chosen as “large” (21-25 equities in each). All MD analysis is performed based on the market data of the full investment period.

		SHASHR from 1/4/2006 to 11/28/2013 (Monthly Scale, Entire Length) Uncorrelated Bets \mathcal{N}_{Ent} Corresponding to Returns μ and Volatilities σ SHASHR Index (Monthly) — $\mu = 0.0069, \sigma = 0.0814$					
Dependence Measures	MV Efficient Portfolio High Premium		MD Efficient Portfolio Maximum Return		MD Efficient Portfolio Maximum Diversification		
	\mathcal{N}_{Ent}	μ σ	\mathcal{N}_{Ent}	μ σ	\mathcal{N}_{Ent}	μ σ	
Linear	5.9151	0.0300 0.0899	9.2985	0.0580 0.1961	17.6273	0.0268 0.1122	
Kendall's	5.2815	0.0300 0.1164	9.5817	0.0423 0.2043	15.1913	0.0182 0.1274	
Spearman's	6.1159	0.0300 0.1199	9.0431	0.0409 0.2071	16.7570	0.0113 0.1206	
Lower Tail (Clayton)	4.7866	0.0300 0.0942	7.7704	0.0569 0.1858	17.1752	0.0225 0.1389	
Upper Tail (Gumbel)	4.3109	0.0300 0.0944	9.1806	0.0545 0.2032	17.7509	0.0110 0.1195	
Tail (t Copula)	4.4046	0.0300 0.0945	10.0749	0.0694 0.1898	15.4801	0.0268 0.1228	
MI (Gaussian)	6.8021	0.0300 0.1183	9.6652	0.0373 0.1862	14.6849	0.0113 0.1269	
MI (t Copula)	4.6521	0.0300 0.0939	9.9179	0.0564 0.2038	14.6855	0.0263 0.1372	
MI (Clayton)	4.4844	0.0300 0.0945	10.6045	0.0556 0.1878	16.0878	0.0238 0.1339	
MI (Frank)	6.1395	0.0299 0.0894	13.8961	0.0577 0.1624	17.0448	0.0358 0.1290	
MI (Gumbel)	5.8782	0.0300 0.1119	14.5113	0.0340 0.1577	15.5899	0.0242 0.1270	

Chapter 5

Conclusion

In this thesis, we have studied a portfolio construction technique and proposed a new methodology to select the equities from the large trading markets. The methodology considers various dependence structures in the market. The equity selections are realized by the social networks and a data clustering technique. We have also considered the influences of market regime changes on the equity selections. Then we have evaluated the selection strategy to complete the portfolio construction. The evaluation analysis was performed within the mean-variance optimization framework. At last, we have examined the diversification of the relevant portfolio selections within the mean-diversification framework. The evaluation analysis has suggested the proper equity selections in different market settings. All numerical results and relevant analysis were obtained based on three major Chinese equity markets.

The motivation of the equity selection methodology came from the concept of the dimension reduction technique. A large equity market can be treated as high dimensional data with each of its dimensions representing an asset. Intuitively, we used the dependence to properly describe the organization of the market data and the associations among its components. Lots of literatures indicated asymmetric and extreme correlations between financial assets. In our work, the dependence structure was modeled through the linear correlation and the copula related dependence measures so that it accounts for both the linear and the non-linear associations among the assets. More explicitly, the estimation of the relevant dependence measures included the linear correlation, the rank correlations, the tail dependence (revealed by copulas) and the mutual information (revealed by negative copula entropies). Each of the measures represented a unique pattern of the dependence structure. In our applications, they were interpreted as the numerical estimation of relevant

coefficients.

The dependence structure among the data was further described by the social networks. In the social networks, the individual assets were denoted by the nodes with their scatterings which explain the associations. More explicitly, we used the distances between pairs of the nodes to measure the strength of the relevant dependence structure. The construction of the social networks based on the dependence laid the foundation of the equity selections as it provided us with a platform to apply the specifically designed dimension reduction techniques. In our work, we chose to use the data clustering technique for dimension reduction in order to find all cluster centers as the selected equities. We studied and implemented the popular *Affinity Propagation Clustering* technique (**AP**) proposed by [Frey and Dueck \(2007\)](#), due to its various advantages in handling large data. We slightly improved the **AP** algorithm to reduce the computational intensity and integrated the **BWP** into it for a better control of the clustering solutions. As a result, we built up a full social network clustering framework to perform the equity selection based on various patterns of the dependence structures. We also studied and applied the Markov Regime Switching Model ([Perlin, 2014](#)) to identify different market phases in which the social network clustering for equity selections were used.

We examined the equity selection strategies under the mean-variance framework ([Markowitz, 1952](#)) as well as the mean-diversification framework ([Meucci, 2009](#)). With such an analysis, we acquired a good understanding of a reasonable selection strategy in terms of the dependence measures and the other factors.

In the classical mean-variance framework, we observed the influence of the dependence measures, the equity selection sizes and the market phases on the portfolio returns and volatilities. We obtained some empirical findings based on the mean-variance analysis. In the social network clustering framework we built up, the equity selections by the lower tail dependence prove to be more robust than the equity selections by other dependence measures. In addition, the large equity selections by the tail dependence indicated by the t copula have outstanding performance. In the case when the data is abstracted from the bear market phase, the large equity selections by the linear correlations also show good performance.

Diversification also plays a key role in the robust portfolio construction. In the mean-diversification framework, [Meucci \(2009\)](#) unitizes the **PCA** to decompose the correlated components of the market into the uncorrelated principal portfolios and expresses the diversification with an entropy related measure. We studied this methodology and applied the mean-diversification optimization for the efficient asset allocations. In the mean-diversification framework, the equity selections are built on the social network clustering

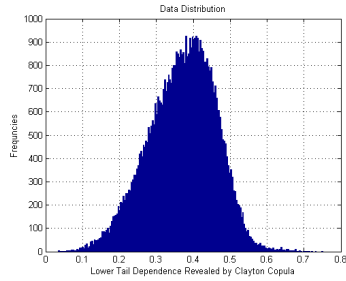
with various dependence measures. We analyzed the salient features of the equity selections based on the consideration of the mean-diversification performance, the dependence measures and the market phases. We reached some conclusions based on our mean-diversification analysis. Long-term data indicate that the selections by the Spearman's rank correlation and the Kendall's rank correlation are more preferred. The selection by the linear correlation can also be well diversified. From another point of view, when the dependence measures are estimated by the extreme downside market data, the obtained equity selections lead to better diversification. In this case, the equity selections by the linear correlation and the MI (Frank copula) show robust performance.

APPENDICES

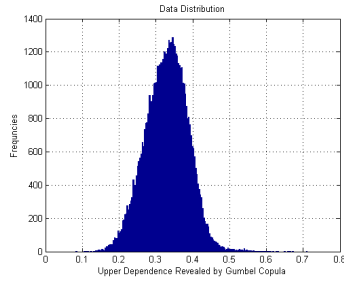
Appendix A

Additional Figures and Tables

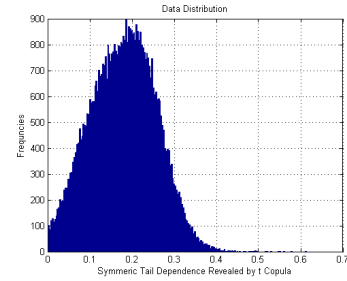
A.1 Estimation of Tail Dependence and Mutual Information on Other Markets



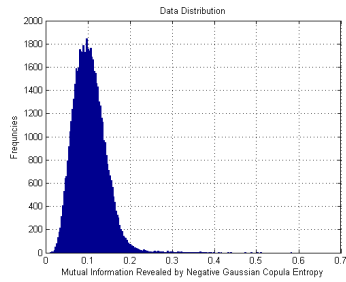
(a) Lower Tail Dependence Revealed by Clayton Copula



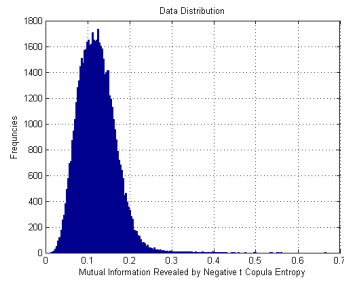
(b) Upper Tail Dependence Revealed by Gumbel Copula



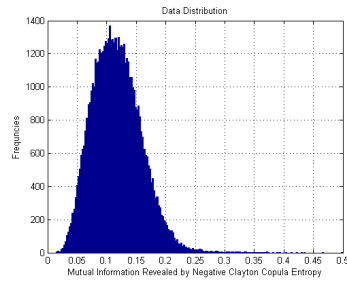
(c) Tail Dependence Revealed by t Copula



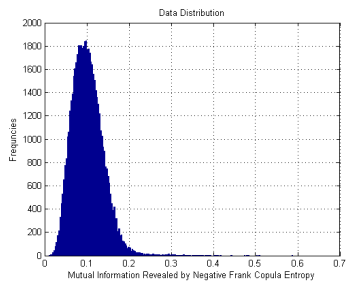
(d) Mutual Information Revealed by Negative Gaussian Copula Entropy



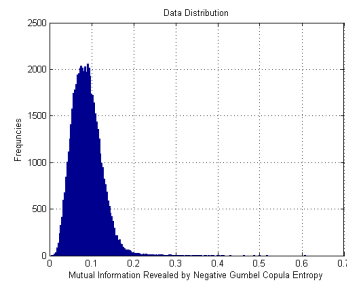
(e) Mutual Information Revealed by Negative t Copula Entropy



(f) Mutual Information Revealed by Negative Clayton Copula Entropy

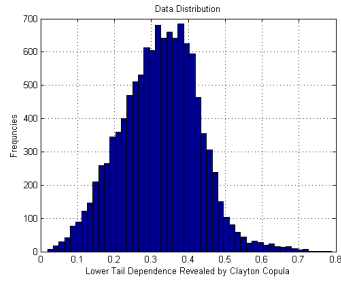


(g) Mutual Information Revealed by Negative Frank Copula Entropy

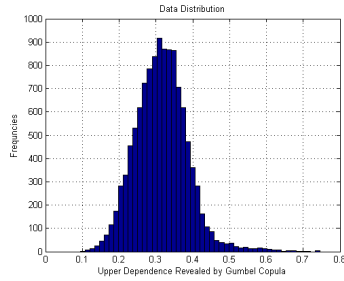


(h) Mutual Information Revealed by Negative Gumbel Copula Entropy

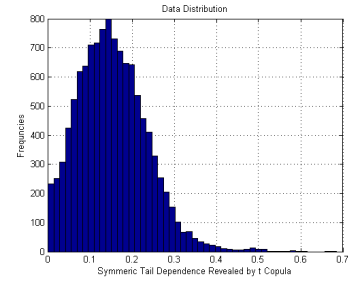
Figure A.1: Distributions of Estimated Tail Dependence and Mutual Information of Market Data of SZASHR.



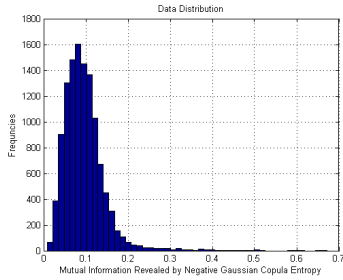
(a) Lower Tail Dependence Revealed by Clayton Copula



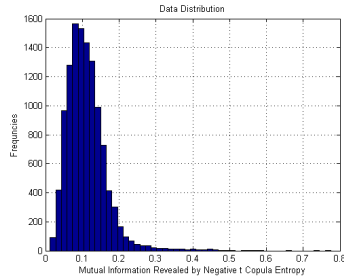
(b) Upper Tail Dependence Revealed by Gumbel Copula



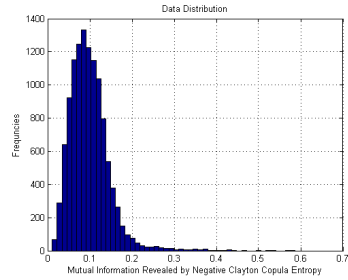
(c) Tail Dependence Revealed by t Copula



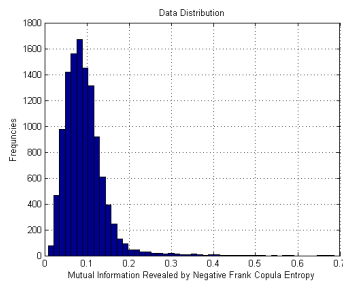
(d) Mutual Information Revealed by Negative Gaussian Copula Entropy



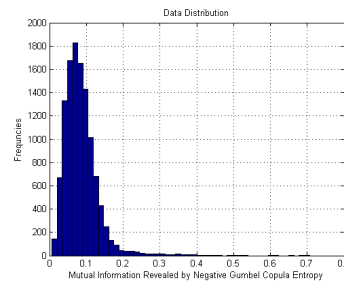
(e) Mutual Information Revealed by Negative t Copula Entropy



(f) Mutual Information Revealed by Negative Clayton Copula Entropy



(g) Mutual Information Revealed by Negative Frank Copula Entropy



(h) Mutual Information Revealed by Negative Gumbel Copula Entropy

Figure A.2: Distributions of Estimated Tail Dependence and Mutual Information of Market Data of SHSZ300.

A.2 Markov Regime Switching Results on Other Markets

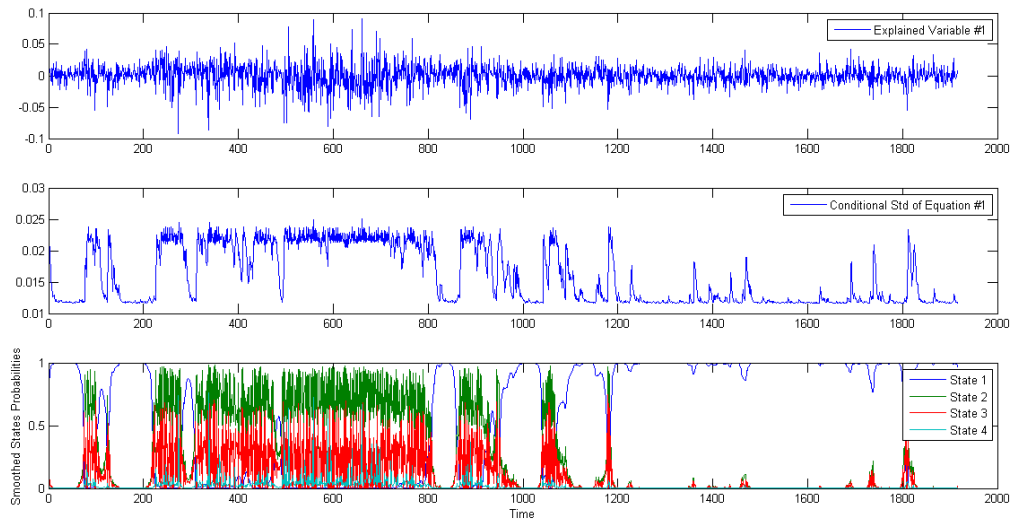
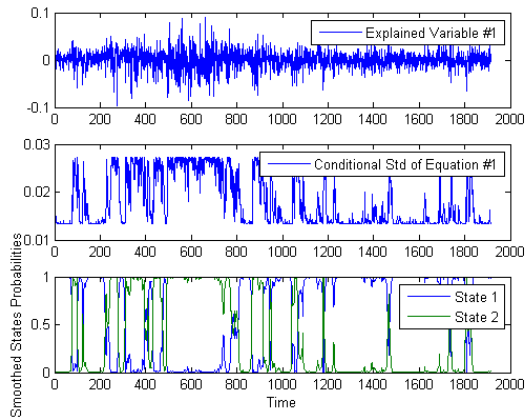
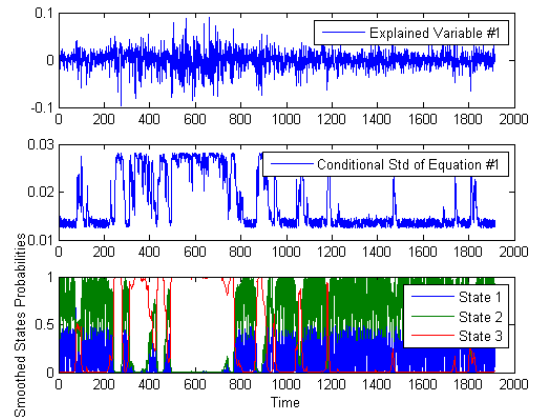


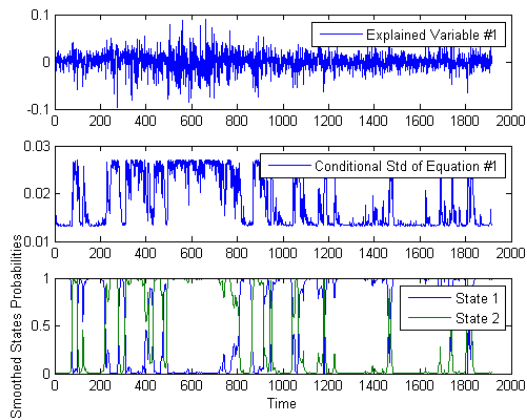
Figure A.3: Markov Regime Switching on SHASHR Index with 4-states Mixture Distributions.



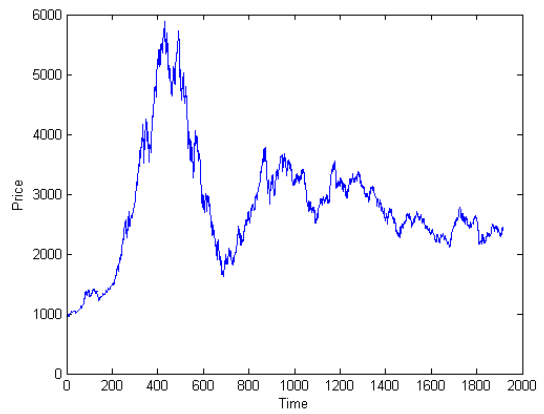
(a) 2-states Mixture Distributions Regime Switching Result



(b) 3-states Mixture Distributions Regime Switching Result

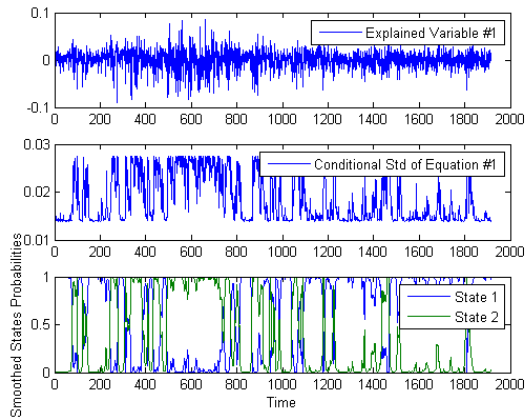


(c) 2-states Autoregressive Regime Switching Result

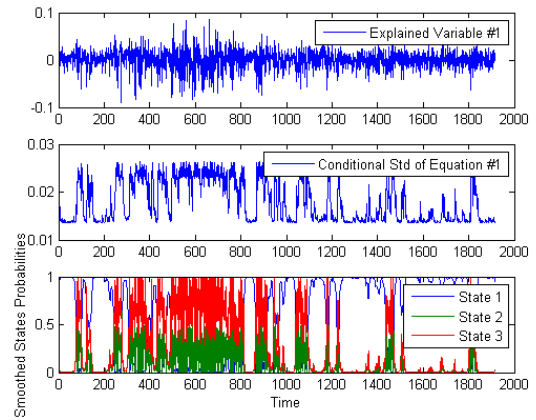


(d) Historical Price Dynamics of SHSZ300 Index

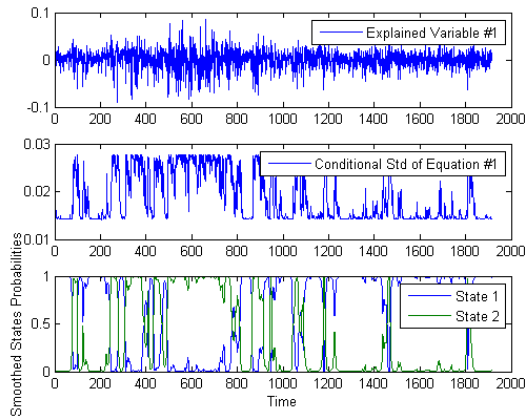
Figure A.4: Markov Regime Switching Results on SHSZ300 Index.



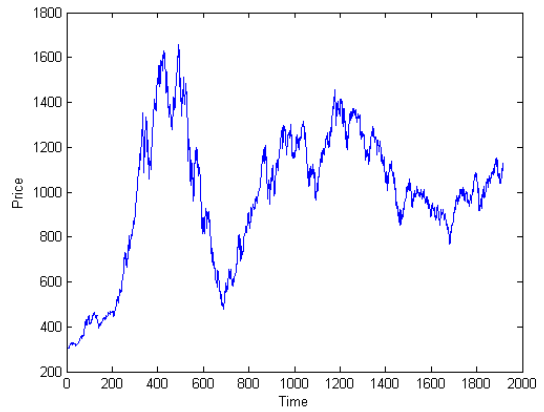
(a) 2-states Mixture Distributions Regime Switching Result



(b) 3-states Mixture Distributions Regime Switching Result



(c) 2-states Autoregressive Regime Switching Result



(d) Historical Price Dynamics of SZASHR Index

Figure A.5: Markov Regime Switching Results on SZASHR Index.

References

- Andrew Ang and Geert Bekaert. International asset allocation with regime shifts. *Review Of Financial Studies*, 15(4):1137–1187, 2002.
- Andrew Ang and Joseph Chen. Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3):443–494, 2002.
- Jushan Bai and Shuzhong Shi. Estimating high dimensional covariance matrices and its applications. 2011.
- Anil K. Bera and Sung Y. Park. Optimal portfolio diversification using the maximum entropy principle. *Econometric Reviews*, 27(4-6):484–512, 2008.
- Giovanni Bonanno, Guido Caldarelli, Fabrizio Lillo, and Rosario N. Mantegna. Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, 68(4), 2003.
- Phelim Boyle, Junichi Imai, and Ken Seng Tan. Computation of optimal portfolios using simulation-based dimension reduction. *Insurance: Mathematics and Economics*, 43(3): 327–338, 2008.
- Cristina Butucea, Jean-François Delmas, Anne Dufloy, and Richard Fischer. Maximum entropy copula with given diagonal section. *arXiv preprint arXiv*, 1312(5219), 2013.
- Umberto Cherubini, Elisa Luciano, and Walter Vecchiato. *Copula Methods in Finance*. Wiley, 1 edition, July 2004.
- Umberto Cherubini, Sabrina Mulinacci, Fabio Gobbi, and Silvia Romagnoli. *Dynamic Copula Methods in Finance*. Wiley, Nov 2011.
- Tarun Chordia, Amit Goyal, and Qing Tong. Pairwise correlations. *working papers series Available at SSRN: <http://ssrn.com/abstract=1785390>*, 4 2011.

- Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience, 2 edition edition, Jul 2006.
- d-fine GmbH. Asset allocation under a conditional diversification measure. Master’s thesis, Kellogg College University of Oxford, September 2011.
- Magnus Dahlquist, Ádám Faragó, and Roméo Tédongap. Asymmetries and portfolio choice. 2014.
- Donglei Du. Social network analysis: Introduction, 2014.
- Cheoljun Eom, Gabjin Oh, and Seunghwan Kim. Topological properties of the minimal spanning tree in korean and american stock markets. *arXiv preprint physics*, 2006. 0612068.
- Cheoljun Eom, Gabjin Ohb, Woo-Sung Jung, Hawoong Jeong, and Seunghwan Kim. Topological properties of stock networks based on minimal spanning tree and random matrix theory in financial time series. *Physica A: Statistical Mechanics and its Applications*, 388(6):900–906, 2009.
- Emilio Ferrara and Giacomo Fiumara. Topological features of online social networks. *arXiv preprint arXiv*, 1202(0331), 2012.
- Brendan J. Frey and Delbert Dueck. Clustering by passing messages between data points. *science*, 315(5814):972–976, 2007.
- James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.
- James D. Hamilton. Regime-Switching Models. *The new palgrave dictionary of economics*, 2, 2005.
- James Douglas Hamilton. *Time Series Analysis*. Princeton University Press, Jan 1994.
- Mary R. Hardy. A regime-switching model of long-term stock returns. *North American Actuarial Journal*, 5(2):41–53, 2001.
- Yongmiao Hong, Jun Tu, and Guofu Zhou. Asymmetries in stock returns: Statistical tests and economic evaluation. *Review of Financial Studies*, 20(5):1547–1581, 2007.

- INVESTOPEDIA. Definition of “adjusted closing price”. URL http://www.investopedia.com/terms/a/adjusted_closing_price.asp.
- Leonard Kaufman and Peter J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience, 1 edition edition, 2005.
- Malcolm Kemp. *Extreme Events: Robust Portfolio Construction in the Presence of Fat Tails*. Wiley, Dec 2010.
- Dror Y. Kenett, Michele Tumminello, Asaf Madi, Gitit Gur-Gershgoren, Rosario N. Mantegna, and Eshel Ben-Jacob. Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. *PloS one*, 5(12):e15032, 2010.
- Chang-Jin Kim and Charles R. Nelson. State-space models with regime switching: classical and gibbs-sampling approaches with applications. *MIT Press Books*, 1, 1999.
- Mark Kritzman, Sébastien Page, and David Turkington. Regime shifts: Implications for dynamic strategies. *Financial Analysts Journal*, 68(3), 2012.
- Tze Leung Lai, Haipeng Xing, and Zehao Chen. Mean-variance portfolio optimization when means and covariances are unknown. *The Annals of Applied Statistics*, 5(2A): 798–823, 2011.
- Robert Litterman. *Hot spots and hedges*. Risk Management Series. Goldman Sachs and Co., 1996.
- François Longin and Bruno Solnik. Extreme correlation of international equity markets. *The Journal of Finance*, 56(2):649–676, 2001.
- Jian Ma and Zengqi Sun. Mutual information is copula entropy. *Tsinghua Science & Technology*, 16(1):51–54, 2011.
- Rosario N Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11(1):193–197, 1999.
- Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.
- Alexander J. McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton Series in Finance. Princeton University Press, Oct 2005.

- Robert C. Merton. An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis*, 7(4):1851–1872, 1972.
- Attilio Meucci. *Risk and Asset Allocation*. Springer, 1 edition, 1 2005.
- Attilio Meucci. Managing diversification. *Risk. Bloomberg Education & Quantitative Research and Education Paper*, pages 74–79, 2009.
- Roger B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer, 2nd edition, 2006.
- David F. Nettleton. Data mining of social networks represented as graphs. *Computer Science Review*, 7:1–34, 2013.
- J-P Onnela, Anirban Chakraborti, Kimmo Kaski, Janos Kertesz, and Antti Kanto. Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, 68(5), 2003.
- Andrew Papanicolaou. Dimension reduction in discrete time portfolio optimization with partial information. *SIAM Journal on Financial Mathematics*, 4(1):916–960, 2013.
- M. Hossein Partovi and Michael Caputo. Principal portfolios: Recasting the efficient frontier. *Economics Bulletin*, 7(3):1–10, 2004.
- Andrew J. Patton. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics*, 2(1):130–168, 2004.
- George Pennacchi. *Theory of Asset Pricing*. Prentice Hall, 1 edition, 2007.
- Marcelo Perlin. MS_Regress - The MATLAB Package for Markov Regime Switching Models. Available at SSRN: <http://ssrn.com/abstract=1714016>, Oct 2014.
- Bernhard Pfaff. *Financial Risk Modelling and Portfolio Optimization with R*. Wiley, Jan 2013.
- George C. Philippatos and Charles J. Wilson. Entropy, market risk, and the selection of efficient portfolios. *Applied Economics*, 4(3):209–220, 1972.
- Julia Piantadosi, Phil Howlett, and Jonathan Borwein. Copulas with maximum entropy. *Optimization Letters*, 6(1):99–125, 2012.

- L. Putzig, D. Becherer, and I. Horenko. Optimal allocation of a futures portfolio utilizing numerical market phase detection. *SIAM Journal on Financial Mathematics*, 1(1):752–779, 2010.
- Marina Resta. A new framework for assets selection based on dimensions reduction techniques. *Knowledge-Based and Intelligent Information and Engineering Systems. Springer Berlin Heidelberg*, pages 372–381, 2011.
- Stephen A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13:341–360, 1976.
- Ludger Rízschorf. *Mathematical Risk Analysis: Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer Series in Operations Research and Financial Engineering. Springer, Mar 2013.
- David Ruppert. *Statistics and Data Analysis for Financial Engineering*. Springer, Nov 2011.
- Frank Schweitzer, Giorgio Fagiolo, Didier Sornette, Fernando Vega-Redondo, Alessandro Vespignani, and Douglas R. White. Economic networks: The new challenges. *science*, 325(5939):422, 2009.
- C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.
- William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.
- Abe Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris (in French)*, 8:229–231, 1959.
- Benjamin M. Tabaka, Thiago R. Serra, and Daniel O. Cajueiro. Topological properties of stock market networks: The case of brazil. *Physica A: Statistical Mechanics and its Applications*, 389(16):3240–3249, 2010.
- Yuichi Takano and Jun-ya Gotoh. Multi-period portfolio selection using kernel-based control policy with dimensionality reduction. *Expert Systems with Applications*, 41(8):3901–3914, 2014.
- Ruey S. Tsay. *Analysis of Financial Time Series*. Wiley, 3 edition edition, Aug 2010.

- Michele Tumminello, Fabrizio Lillo, and Rosario N. Mantegna. Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization*, 75(1):40–58, 2010.
- Štefan Lyócsa, Tomáš Výtrost, and Eduard Baumöhl. Stock market networks: The dynamic conditional correlation approach. *Physica A: Statistical Mechanics and its Applications*, 391(16):4147–4158, 2012.
- Kaijun Wang, Junying Zhang, Dan Li, Xinna Zhang, and Tao Guo. Adaptive affinity propagation clustering. *arXiv preprint arXiv*, 0805(1096), 2008.
- Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2nd edition, Sep 2004.
- Larry Wasserman. *All of Nonparametric Statistics*. Springer, 3rd edition, July 2007.
- Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*. Structural Analysis in the Social Sciences. Cambridge University Press, 1 edition edition, Nov 1994.
- Changwei Xiong. Risk diversification in portfolios. A final project report for the course "Portfolio Theory and Applications", December 2009.
- Ning Zhao and Winston T. Lin. A copula entropy approach to correlation measurement at the country level. *Applied Mathematics and Computation*, 218(2):628–642, 2011.
- Rongxi Zhou, Ru Cai, and Guanqun Tong. Applications of entropy in finance: A review. *Entropy*, 15(11):4909–4931, 2013.
- Shibing Zhou, Zhenyuan Xu, and Xuqing Tang. Method for determining optimal number of clusters in k-means clustering algorithm. *Journal of Computer Applications (in Chinese)*, 30(8):1995–1998, Aug 2010a.
- Shibing Zhou, Zhenyuan Xu, and Xuqing Tang. New method for determining optimal number of clusters in k-means clustering algorithm. *Computer Engineering and Applications (in Chinese)*, 46(16):27–31, 2010b.
- Shibing Zhou, Zhenyuan Xu, and Xuqing Tang. Method for determining optimal number of clusters based on affinity propagation clustering. *Control and Decision (in Chinese)*, 26(8):1147–1157, Aug 2011.