FINANCIAL CORRELATION NETWORKS

by

Umut Kirmit

Submitted to the Graduate School of Sabancı University
in partial fulfillment of the requirements for the degree of
Master of Science

Sabanci University

August, 2008

FINANCIAL CORRELATION NETWORKS

Umut Kirmit

IE, Master's Thesis, 2008

Thesis Supervisor: Ali Rana Atılgan

Keywords: Correlation network, financial correlation, spectral properties, rewired model

Abstract

We construct a financial network based on the correlations of the assets. Referred to as a correlation network, its nodes are the assets and its edges are the pair-wise correlations. The network is inspected by both spectral and statistical analyses. We find that these analyses provide complementary information regarding the interactions of securities in the market and their clustering as well as a hierarchy in the market structure. Market portfolio dominating behavior indicates scale free type of interactions where a small number of assets linked to many others accounts for most of the activities. We further introduce a pricing model that uses the interrelations of emerging dominant correlational motifs. Those that are in accord with piecewise stationary behavior are found to successfully define the future price bounds.

FINANCIAL CORRELATION NETWORKS

Umut Kirmit

IE Master Tezi, 2008

Thesis Supervisor: Ali Rana Atılgan

Keywords: Korelasyon ağları, finansal korelasyon

Özet

Tez çalışmasında finansal varlıkların birbirleri ile oluşturdukları karşılıklı ilişkileri gösteren bir ağırlıklandırılmış ağ yapısı oluşturulmaktadır. Bu ağ yapısında varlıklar düğüm olarak ve bu düğümler arasındaki bağ da karşılıklı ilişkileri, istatistiki terim ile korelasyonu göstermektedir. Ağ yapısı istatistiksel ve fiziki tayf olarak incelenmiştir. Yaptığımız çalışmalar istatistiksel ve fiziki tayf analizlerinin herhangi bir finans marketinin hiyerarşik ve grup yapısını ortaya çıkarmakta ve birbirini tamamlayan ve doğrulayan analizler olduğunu göstermektedir. Market porföyünün baskın hareketi az sayıda varlığın diğer çok sayıda varlık ile ölçümden bağımsız bir etkileşimi işaret etmektedir. Aynı zamanda çalışmada karşılıklı olarak gelişen baskın korelasyon ağ yapılarının kullanılması ile oluşturulmuş yeni bir getiri üretme ve fiyatlama modeli tanıtılmaktadır. Bu modelin parçalı durgun hareket eden ağ yapılarının gelecek fiyat aralıklarını daha iyi belirlediği gösterilmektedir.

Acknowledgements

First of all, I wish to express my gratitude to my thesis adviser Ali Rana Atılgan for his guidance, motivation and inspiration throughout my research. His sincerity, valuable advice and unlimited support motivated me in every study that I have performed during my master of science. I really consider myself fortunate and privileged as his student and assistant.

I am grateful to Koray D. Şimşek who is also in my thesis committee for helping and guiding me in computational finance stages of the work. I would like to acknowledge the advice and guidance of Vedat Akgiray, Reha Yolalan and Nilay Noyan, graduate committee. Their valuable review and comments on the dissertation is a great motivation and support for me to complete this thesis. I would like to thank to my roommate Isa Kemal Pakatçı is the one who have important discussions for my thesis and also for sharing his experience and practical knowledge on programming. I would also like to thank to Cem Kesici for his special and important support in standing surety for my scholarship. I would also like to express my great gratitude to TUBİTAK for spectacular scholarship. I am greatly grateful to my family for their concern, infinite support and trust.

Finally, I want to express my strong gratitude to my love Sevilay Gökduman for her friendship and endless love. I would like to thank her valuable support for grammar check in completion stage of my thesis.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Objectives

Recent years have witnessed the growing interest in the motions of financial markets as a research area, the interest arising from the fact that they involve many sub-systems. We can define a financial market as a union or intersection of sub financial markets. Stock exchange markets, currencies, bonds, derivatives can be listed as examples of sub-financial markets. We can classify the motions of a sub-financial market into two main groups. The first one is affected by the motion of the whole financial market and the interactions between the other sub-financial markets. The second one is the clustering motion. There are some clusters within each sub-market which move together and are affected by the same kind of information. This phenomenon constitutes a hierarchical structure in the market.

Many researchers have tried to figure out this hierarchical structure of the financial market by approaching the problem from different point of views. When Mantegna put forward his studies by working on a network, the concepts started to be constructed on a more tangible basis[16]. First, he transformed the financial market into a network concept by representing the assets by nodes and the correlation coefficients by edge weights. Correlation networks give useful structural properties to analyze the market. Bouchaud et al.and Stanley et al. and Boccaletti et al. who have worked on correlation matrix found that the correlation matrix is differentiated from random matrix by having hierarchical structures.[1, 3, 19]

These two studies motivated us to direct our studies to discovering the hierarchical structures of the financial market. We therefore aim to find out the special properties and enriched parameters as alternatives to the simple mean-variance parameters. In our analysis we utilized several tools to join the different perspectives. For instance, we used physical statistics and network statistics. During our research we recognized that some properties of networks change by the frequency of data. We also realized that using different time periods cause significant changes in the correlation networks, and the resulting properties In the light of the foregoing, we developed a scenario generation model.

1.2 Data Set

During our research, we combine our theoretical studies with real life data. We utilized from two data samples. The first is obtained from Istanbul Stock Exchange Market (ISE) and is used for constructing a matrix which consists of all the correlation coefficients. The set of data is formed by 211 assets (stocks) which includes daily closing prices from 18.10.1999 to 31.05.2007. We use ISE data in different intervals. We use 1,2,5,10 and 20 day price intervals to calculate the rate of returns which are used to construct the correlation networks.

The second data set is used in the last part of thesis these are we also use data of some commodities, a currency and the index of New York stock market. We will define the second data set in more detail in the corresponding chapter.

1.3 Organization of the Thesis

The thesis is divided into three parts. In the first part, we first construct the correlation matrix which will be used in both the spectral and statistical analyses and followed by the study of the spectral properties of the correlation network. In the second part, we form the minimum spinning tree and study its statistical behavior in addition to that of correlation network. To discover the underlying properties that may have been masked by either the market portfolio or the noise, we reconstruct the correlation matrix and correlation network. We combine the results of spectral and statistical analyses to obtain more accurate results. The final part, we develop a pricing scenario model based on generating networks of cross-correlations between assets. The novelty is in using equally divided, stationary pieces of the time series so as to allow for the rewiring of the networks, instead of concentrating on the data available as a whole. This approach enables us to observe the emergence of certain motifs that dominate the cross-correlations, leading to a more reliable portfolio generation. The implications of our findings as well as an outline of future work are presented in the final chapter.

2 SPECTRAL ANALYSES

In this chapter, spectral properties of the correlations which are obtained by data with different frequencies are analyzed. First of all, probability distribution function of the eigenvalues for different intervals is determined. Secondly, the eigenvector – sector analysis is conducted.

2.1 Basic Terminology and Literature Review

The focus of this chapter is the spectral properties of the data correlations. First of all, the correlation matrix and its properties are examined. Then the probability distribution function of the eigenvalues is determined for the data with different frequencies. Finally, the eigenvector – sector analysis is conducted. Some of the key concepts from statistical physics have been used in the explanation of financial systems. In order to analyze the financial markets, we started our studies by using the spectrum analysis concept of statistical physics. Recently, spectrum analysis is used for complex financial systems especially spectra of correlation matrix has been studied in detail with motivating results.[3, 10, 17, 19]

The correlation coefficient ρ_{ij} of two random variables such as i and j, with expected values of rate of return and standard deviations σ_i and σ_i is defined as follows:

$$C_{ij} = \frac{\langle r_i^t r_j^t \rangle - \langle r_i^t \rangle \times \langle r_j^t \rangle}{\sqrt{\left[\langle (r_i^t)^2 \rangle - \langle r_i^t \rangle^2 \right] \times \left[\langle \left(r_j^t \right)^2 \rangle - \langle r_j^t \rangle^2 \right]}}$$

In this study our aim is to focus on the global behavior of the market rather than focusing on the local interactions of assets. To do that, we figure out the interactions of all assets which affect the motion of the system. Since the spectrum of correlation matrix gives constructive information about the financial assets, we directed our analysis to gaining the spectra of the correlation matrix. To obtain the motion of the market and clusters, the spectral analysis with eigenvalue decomposition method is appreciably new and useful in finance domain. The eigenvalue problem firstly came across during the rope motion studies of Johann, Bernoulli and d'Alembert in the 18th. Then Euler focused on the importance of the principal axes while his study on rotational motion and Cauchy used principal axes in explaining the quadratic surfaces and generalized it to arbitrary dimensions. Although the eigenvalue concept had been studied by several researchers since 18th century, the word "eigen" was introduced by Hilbert in 1904 to cover the meanings "own", "peculiar to", "characteristic", or "individual".[11]

Eigenvalue decomposition is a factorization method for complex matrix, used for transformation of the eigenvalues and associated eigenvectors in several applications in statistics.

The eigenvalue equation is given below:

$C \times U = U \times D$

D is a matrix which includes eigenvalues in its diagonals and the U is the eigenvectors' matrix corresponding to the eigenvalues. Eigenvalue decomposition is used by many authors to measure the correlations of stock price change and motion of the assets in these correlation concepts.[3, 19] Stanley and Rosenow applied the eigenvalue decomposition method in random matrix theory (RMT) to demonstrate the validity of the universal predictions of RMT for eigenvalue statistics. (Figure 23 in Appendix.) In all studies, firstly they gained the eigenvalues and figured out the distribution of the eigenvalues.[19] Bouchaud plotted the distribution of the eigenvalues and suggested that the largest eigenvalue corresponds to the motion of the whole market. [3] In an ineffective market the largest eigenvalue gets larger values in comparison to the more effective markets. We tried to observe this relationship between ineffective markets and the biggest eigenvalue by changing the frequency of the data. (figure 2). The researchers have focused on the distribution of eigenvalues to explain this spectrum. [14, 3, 10] They have worked to analyze the main motion of dynamic systems by considering the noises. Bouchaud stated that the eigenvector corresponding to the biggest eigenvalue shows the main dimension of the market and Pafka et al put up the argument of small eigenvalues symbolize the noises. Whereon Kertesz et al [10] discussed the eigenvalue spectrum by review of Bouchaud and Pafka's studies and developed a model for the explanation of the market behavior by spectral properties of eigenvalues. The model proposes that the largest eigenvalue defines the motion of market; the small eigenvalues are associated with noisy of the interactions of nodes, and intermediate eigenvalues carry important information about assets moving together, such as a cluster which is corresponding to a sector. Saramaki et al [17], integrated spectral properties with approach of clique of percolation which is used to interpret set of communities (clusters) by Palla et al.[7] As it can be seen form the wealth of literature review, spectral analysis is a promising topic for new researches.

2.2 Spectral Analyses

First of all we obtained the correlation matrix and decomposed it into eigenvalues and eigenvectors by eigenvalue decomposition technique. Eigenvalues are the attributes that characterizes the spectra of the financial data. Then we exposed the probability density function of these eigenvalues which is shown in Figure 1.

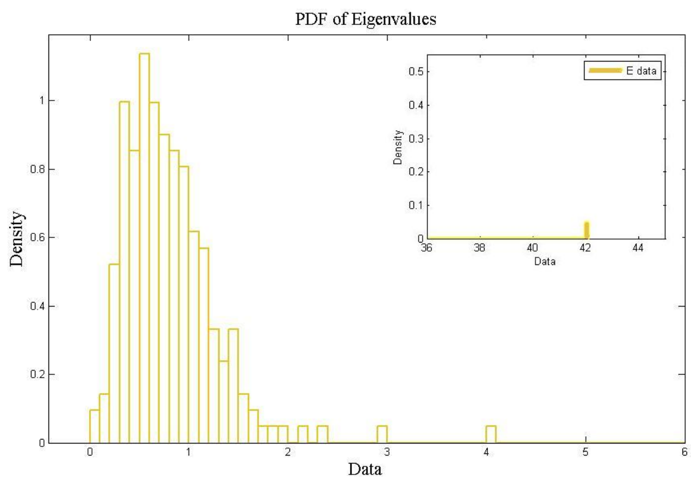


Figure 1: The probability density function of the eigenvalues.
These eigenvalues obtained by decomposing cross-correlation of matrix of 211 companies which are trading in ISE during the period 11.1999-05.2007. The inset shows the largest value among all eigenvalues.

Below, the classification of the eigenvalues according to the probability density function analysis is presented[10]:

1. There is a quasi continuum of small eigenvalues which can be described by random matrix theory corresponding to noise and majority of them fall into this category

2. The largest eigenvalue is far from rest and it corresponds to the global

behavior of the market.

3. The discrete spectrum of intermediate eigenvalues carries important information about the correlation that can be related to market taxonomy.

As mentioned above, we will investigate the areas that will show us the clustering information on the motion of the market. It is obvious that the largest eigenvalue is distinguished and this picture will lead us to obtain the entire market's correlation. [3] Market's behavior is examined by the plot of probability distribution functions of different frequencies of data which are shown in Figure 2.

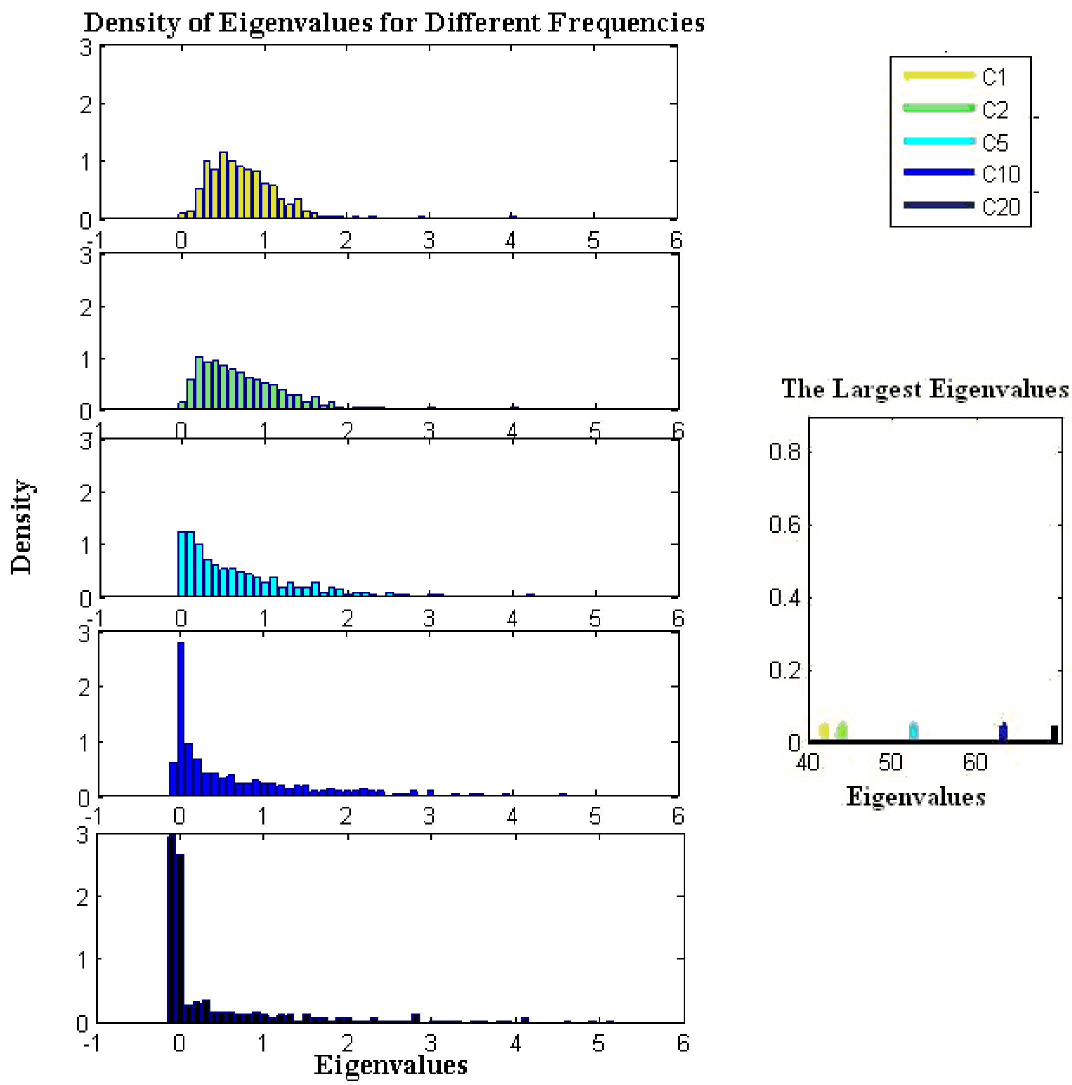


Figure 2: The probability density function of the eigenvalues
We construct our cross-correlation of matrices by using different intervals of price time series. For instance, while we form the 5-day correlations we use changes on price in 5-day interval.

The reason that we observe the behavior of different frequencies is to understand whether our data is informative or not. We conclude that as we increase the frequency, the largest eigenvalue gets smaller and reaches to equilibrium and at the same time the ratio of smaller eigenvalues gets larger. For example for our data the largest eigenvalue was 69 and after we increased the frequency it converged to 42. This results shows that the dynamics of the system behaves as expected.

2.3 Clustering/Sector Analysis with Eigenvectors

Financial markets show different types of motions. In order to observe the source of these different motions, it is important to understand the dynamics of the market. One of those main motions is the dynamic of co-moving and clusters. Spectral analysis suggests a method to analyze the clusters of the market. In eigenvalue decomposition, it is stated that the largest eigenvalues except for the largest one, can be correlated to the clusters of the market. It is also stated that eigenvectors of those eigenvalues can be informative about the clusters. Onnela et al. who studied on eigenvalue and clustering relation. They proposed a methodology to relate the clusters to eigenvalues and eigenvectors.

The methodology is as follows:

1. Correlation matrix is decomposed into eigenvalues matrix and eigenvectors matrix with eigenvalue decomposition method.(n*n)

2. The sector vector is created by sector numbers which companies belong to. (1*n)

3. Normalize the sector vector to which is composed by 13 rows refers to sectors and 211 columns refers to companies and normalize variables refer to the relation between each company and each sector.

4. Obtain the inner products of normalized sector matrix and eigenvectors matrix. (k*n)

Normalized sector matrix is a k*n matrix, corresponds to k different sectors in rows and n companies in columns. The matrix includes normalize variables such that, there is $\sqrt{1/m^2}$ in each column, m identifies number of nodes in each sector.

To sum up the methodology; on eigenvectors matrix, components of eigenvectors are associated with each asset and then we take the summation of sector-specific assets' components times normalization term. Finally we reach the weight of the sectors on the eigenvectors. This analyze is only applicable for the eigenvectors corresponding to the largest eigenvalues. Sector-specification is one of the main parts of our analysis. Since the data we used starts in November 1999 and lasts in May 2007, we found assets which were trading continuously. To cluster the assets, we applied ISE sector-specific indexes. ISE has 20 sector-specific indexes, which are shows in Figure 25 in appendix. Since there are very few assets in some clusters we merged them and decreased to 13 sectors.

EigenValues	42.07	4.08	2.91	2.31	2.13	1.96	1.86	1.73	1.64	1.63	1.60	1.56
EigenVectors	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Machine Industry	**0.39**	**-0.19**	0.02	0.14	-0.16	-0.16	0.01	0.01	**0.22**	0.06	0.04	-0.05
Building and Construction	**0.38**	0.11	-0.15	0.06	**-0.25**	0.05	0.21	**-0.28**	-0.15	0.09	-0.01	0.09
Finance	**0.43**	-0.14	-0.13	**-0.26**	0.11	**0.31**	**-0.38**	0.07	-0.04	0.01	0.02	0.16
Textile	**0.27**	**-0.20**	0.05	-0.14	-0.12	-0.14	0.05	0.17	-0.06	-0.03	-0.09	0.00
Travelling and Tourism	**0.24**	-0.02	0.08	-0.02	0.11	0.02	0.10	-0.03	0.14	-0.06	0.04	-0.07
Food and Agriculture	**0.22**	0.03	-0.01	0.02	**0.26**	0.02	0.02	-0.02	0.06	0.01	0.00	-0.07
Consumer durables	**0.23**	0.07	0.11	0.07	0.02	-0.01	0.03	-0.02	-0.01	0.12	0.01	0.00
Groups	**0.17**	-0.16	0.02	0.02	0.04	-0.09	0.09	0.02	-0.24	0.11	**0.21**	-0.08
Otomotive	**0.15**	0.08	**-0.20**	0.03	0.20	-0.03	-0.02	-0.03	0.02	0.04	0.03	-0.07
Energy	**0.31**	-0.05	0.12	-0.11	-0.07	-0.08	0.18	-0.09	0.14	**-0.20**	0.08	**0.19**
Medicine	**0.02**	0.01	0.01	0.01	0.00	-0.01	0.00	-0.01	0.00	0.00	-0.01	0.00
Communication and Media	**0.21**	0.08	0.05	0.08	0.19	0.10	0.01	0.13	0.02	-0.03	-0.08	0.05
Retailing	**0.09**	0.07	0.02	0.08	0.00	-0.05	0.02	0.18	-0.09	0.03	-0.12	0.08

Table 1: The inner product of normalized sector matrix and eigenvectors matrix.
Normalized sector matrix is a 13*211 matrix, corresponds to 13 different sectors in rows and 211 companies in columns. Normalized sector matrix is a k*n matrix, corresponds to k different sectors in rows and n companies in columns. The matrix includes normalize variables such that, there is $\sqrt{1/m^2}$ in each column, m identifies number of nodes in each sector. The last operation is to take inner products of these two matrix.

We determine the matching between the eigenvectors and sectors. The eigenvector which belongs to the biggest eigenvalue is affected by large sectors in a uniform way. It is an expected result. On the other hand, the eigenvectors which belong to other discrete big eigenvalues are expected to be affected by only a unique sector. The idea says that, the largest eigenvalue refers to the variance of its own eigenvector. The largest eigenvalue corresponds to the global behavior of the market, and then it must be affected by sectors of the market by the proportion of their size. Additionally, other large eigenvalues correspond to eigenvectors for each sector and it is expected to be affected only one sector.

Large Eigenvalues for Correlation Matrix of 1-Day Price Change

EigenValues	42.07	4.08	2.91	2.31	2.13	1.96
EigenVectors	X1	X2	X3	X4	X5	X6
Machine Industry	**0.39**	**-0.19**	0.02	0.14	-0.16	-0.16
Building and Construction	**0.38**	0.11	**-0.15**	0.06	**-0.25**	0.05
Finance	**0.43**	-0.14	-0.13	**-0.26**	0.11	**0.31**
Textile	**0.27**	**-0.20**	0.05	-0.14	-0.12	-0.14

Large Eigenvalues for Correlation Matrix of 5-Day Price Change

EigenValues	52.42	4.28	3.17	3.04	2.78	2.69
EigenVectors	X1	X2	X3	X4	X5	X6
Machine Industry	**0.40**	-0.18	0.00	-0.03	0.07	-0.19
Building and Construction	**0.37**	**0.25**	-0.04	**-0.14**	**0.11**	-0.22
Finance	**0.45**	-0.10	**-0.23**	0.07	0.05	**0.31**
Textile	**0.29**	-0.19	-0.03	-0.09	-0.01	0.08

Large Eigenvalues for Correlation Matrix of 20-Day Price Change

EigenValues	69.13	6.40	5.13	4.94	4.70	4.20
EigenVectors	X1	X2	X3	X4	X5	X6
Machine Industry	**0.39**	**-0.18**	-0.01	0.14	**-0.08**	-0.12
Building and Construction	**0.38**	0.13	0.04	0.05	-0.04	**-0.18**
Finance	**0.46**	-0.10	**-0.09**	**-0.28**	-0.06	0.05
Textile	**0.31**	-0.14	0.01	0.06	0.04	0.03

Table 2: Eigenvector analysis results for different frequencies of data

When we analyze the data with different frequencies, we aimed to find more informative intervals. Therefore we applied the clustering analysis and achieved important results. The bold values show the biggest values which are obtained by the summation of sector specific components of each eigenvector. It shows the effect of each eigenvector to the corresponding sector. As we decrease the frequency of the data the bold values starts to have more distinct values from remaining values and we see that each sector starts to match an eigenvector. Getting more distinct values shows that our data gets rid of noisy which is a desired result for our analysis. The details of this finding will be studied in statistical properties chapter. In order to identify visually apparent structure, we have utilized the minimal spanning tree method introduced by Mantegna by defining a new parameter which will be mentioned in the next chapter.

3 STATISTICAL ANALYSIS OF THE MARKET

3.1 Basic Definitions and Literature Review

Network concept is an effective representation technique to visualize the set of connections within complex systems. The Financial market is one of these complex and dynamic systems since it has so many difficulties in modeling of financial systems. To be able to model financial systems, correlation is a fundamental variable to define the system and interactions in multivariate model. Covariance is a multivariable which depends on the correlation. It can be used to model the whole system whereas some forecasting methods have local variables such as moving averages models and autoregressive conditional heteroscedasticity process. A correlation network can also be defined as a weighted network. In network terminology, the assets are the nodes and the correlations between the assets are the weights of the arcs (links). The weighted network is constructed by studying different methods. The purpose of using different methods during the formation of the network is to observe the relationship of clusters and find the most informative illustration of the network.

Correlation network was defined by Mantegna in 1999 for the first time. He wanted to demonstrate the whole market and the interactions of the assets. The Correlation network was constructed with using cross – correlations of the changing stock prices on the same interval. There are several financial networks which are constructed by the help of correlation. Some of the recent methods which are used to illustrate the financial networks are a minimal spanning tree [16], an asset graph with a threshold [10, 9], a maximal spanning tree [17] and a planar maximally filtered graph [21]. According to correlation network models, a minimal spanning tree differentiates by a new parameter which is used while constructing the network. Mantegna has introduced a new parameter “distance” which is in inverse proportion to the correlation coefficient. The distance parameter represents the high correlation between the assets with a short distance in the network.[16]

A financial network branches into two main parts called tree network and graph. The difference between them is that the graph is allowed to include cycles while the tree is not allowed. To visualize the clustering, tree network concept is more efficient and apparent. We have used minimal spanning tree (MST) to visualize the correlation network of Istanbul Stock Exchange market.

Although network visualization is an informative method for searching statistical properties, we have also analyzed the network in order to reach meaningful statistical metrics or enriched properties of assets-nodes on the network. We applied some network metrics used on concrete network such as air-transportation network and scientific collaboration network which were studied by Vespignani and Barrat.[22] Strength, neighbors degree, next neighbors degree and clustering coefficients are some of the statistical parameters which have been used for analyzing complex networks. [9] In addition to these studies, we conducted enriched statistical parameters. We have also studied network properties, such as topologically different growth types, number and size of clusters and also growth types of clusters cited by Onnela and Kertesz.[9]

3.2 Minimal Spanning Tree and Application for Istanbul Stock Exchange Data

The minimum spanning tree is an effective tool for discovering the hierarchical structure of the network. In a fully network, some information cannot be seen easily. Many links between nodes makes it harder to obtain the required results. Using the minimum spanning tree algorithms, we can figure out the basic structure of the data. It can be considered as a filtering tool.

The study started by constituting a matrix consisting of all the correlation coefficients for a set of the Istanbul Stock Exchange Market (ISE). The set of data is formed by 211 assets (stocks) which include daily closing prices between (October 18, 1999) and (May 31, 2007). The assets are separated into 13 main sectors. The sectors, number of companies corresponding to each sector and color of the nodes referring to the companies is given in Figure 3:

Sectors	Number of Company	Color of nodes
Machine Industry	40	Red
Building and Construction	29	Green
Finance	38	Yellow
Textile	29	Dark Blue
Consumer durables	16	Turquoise
Groups	8	Lila
Automotives	10	White
Traveling and Tourism	9	Purple
Energy	4	Mustard Yellow
Food and Agriculture	21	Pink
Medicine	11	Orange
Media and Communication	4	Grey
Retailing	2	Dark Green

Figure 3: The sectors, number corresponding to each sector and color of the nodes which referring to the sectors in the minimal spanning tree.

A minimal spanning tree was used for financial implementation for the first time by Mantegna in 1999. MST has been applied to Istanbul Stock Exchange Market sample data.

After constructing the correlation matrix, we used the distance metric of Mantegna to construct the distance matrix, d which is defined as the distance between node i and node j.

$$d_{ij} = \sqrt{2 \times (1 - \rho_{ij})}$$

This metric is expected to fulfill the metric axioms which are explained below:

1. $d(i,j) = 0$ if and only if $i = j$; The diagonal of distance matrix is zero because $\rho_{ii} = 0$.

2. $d(i,j) = d(j,i)$; Distance matrix is a symmetric matrix .

3. $d(i,j) \leq d(i,k) + d(j,k)$; The third axiom is proved by [13].

3.3 Minimum Spanning Tree Calculations and Results

A minimum spanning tree is a sub-graph which ensures that all nodes connect to each other. Among many algorithms to generate sub-graphs from network, Kruskal algorithm is the most general and simple algorithm to construct a MST from a fully connected graph.[4, 15] The principle of the Kruskal algorithm is to add an edge with minimum weight as long as it does not create a cycle. We use Kruskal algorithm to convert a financial network which is a connected and undirected graph into a MST and find out the statistical properties our network.

Algorithm 1 Pseudo code of Kruskal algorithm which is used in financial correlation network

```
function Kruskal(G)
  for each  node n in G do
     Define an elementary cluster C(n) ? {n}.
   Initialize a priority queue Q to contain all edges in G, using
the weights as keys.
  Define a tree T ? Ø       //T will ultimately contain the edges
of the MST
   // N is total number of nodes
  while T has fewer than N-1 edges do
    // edge u,n is the minimum weighted route from/to n
    (u,n) ? Q.removeMin()
    // prevent cycles in T. add u,n only if u and n are not on the
same cluster
    if C(n) ? C(u) then
      Add edge (n,u) to T.
      Merge C(n) and C(u) into one cluster, that is, union C(n) and
C(u).
  return tree T
```

The minimum spanning tree is obtained by implementing the Kruskal algorithm in Matlab software and displayed by YED graph visualization program. In addition to the Kruskal MST algorithm, we also differentiate the nodes with respect to their sectors and sizes. The color of a node shows the sector that node belongs to and the size of a node shows the strength of that node. The first input of the algorithm is the distance matrix which is obtained by correlation matrix according to Mantegna distance metric. Second input is the vector of assets. It takes numbers between 1 and 13 which represents the sector that a node belongs to.

By applying the Kruskal algorithm in Matlab and YED visualization program to our data we have obtained the minimum spanning tree given in Figure 4. The contact number of each node is given in the inset of the graph. We will use this contact number for comparison of minimum spanning trees which are gained by data with different frequencies.

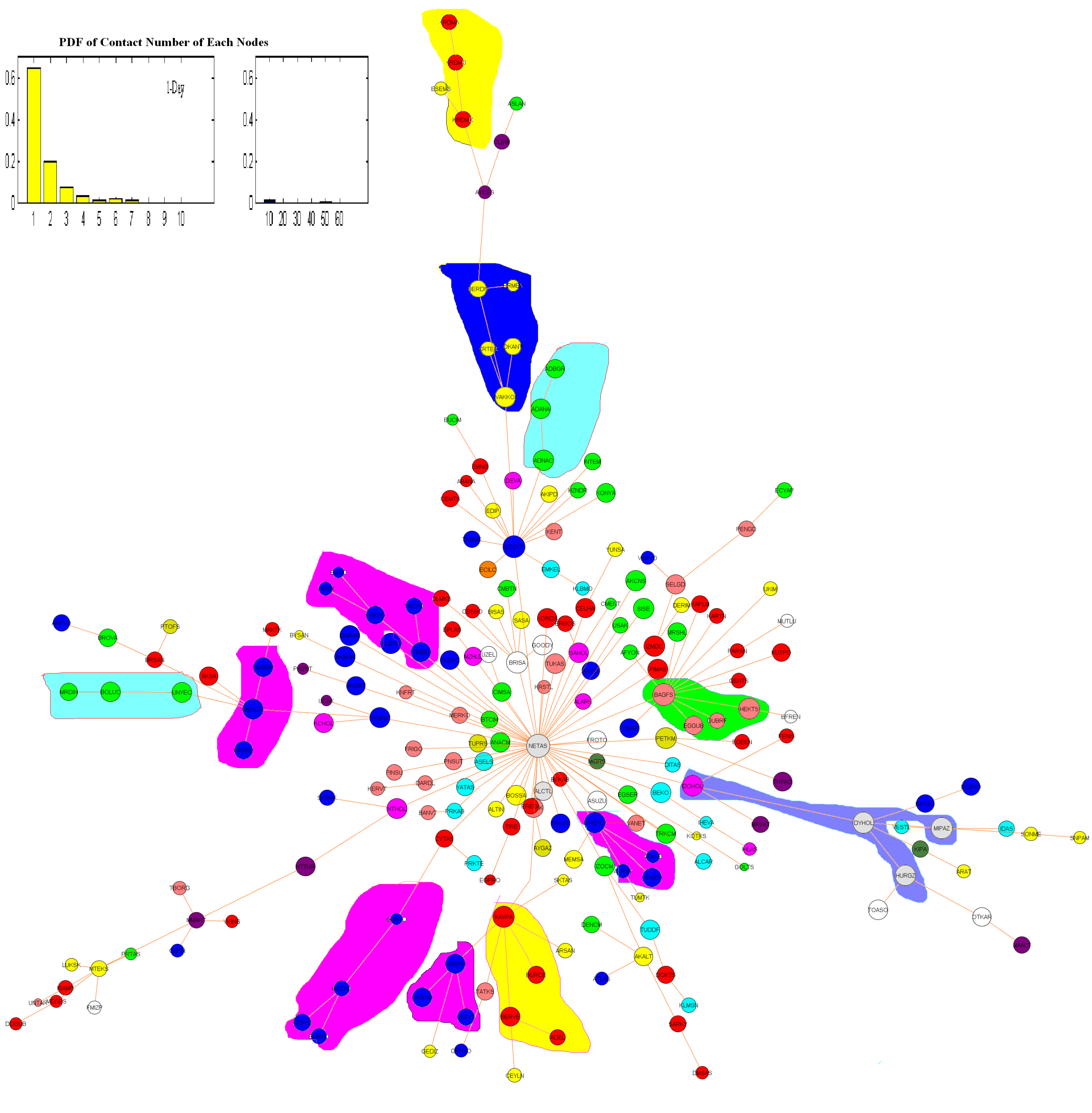


Figure 4: Minimum spanning tree of Istanbul Stock Exchange Market
Minimum spanning tree of Istanbul Stock Exchange Market's correlation network by daily price change. There are some sector-specific groups colored in a pool which are indicated strong relation with their position in the tree. (Inset: Probability distribution function of contact number of each node in minimum spanning tree)

A minimal spanning tree figures out the dynamic system of the market. The first result is the sector-specific clusters are not shown clearly. The tree has a centralization motion and very few clusters, the clusters are also not sector-specific. In the previous chapter it was stated that the spectral analysis of 1-Day data did not give us clear clustering information. This result is proved by the tree, too. We can also see the subsystems which can be called clusters are moving in same the dynamics and are affected by each other.

Another important result is about central nodes and strength of the nodes. The tree provides us to see the big system, clusters in the system, central of the clusters and motion from the central to the branch. It also gives us motivation to research if there is a lag or effect from central to the branches.

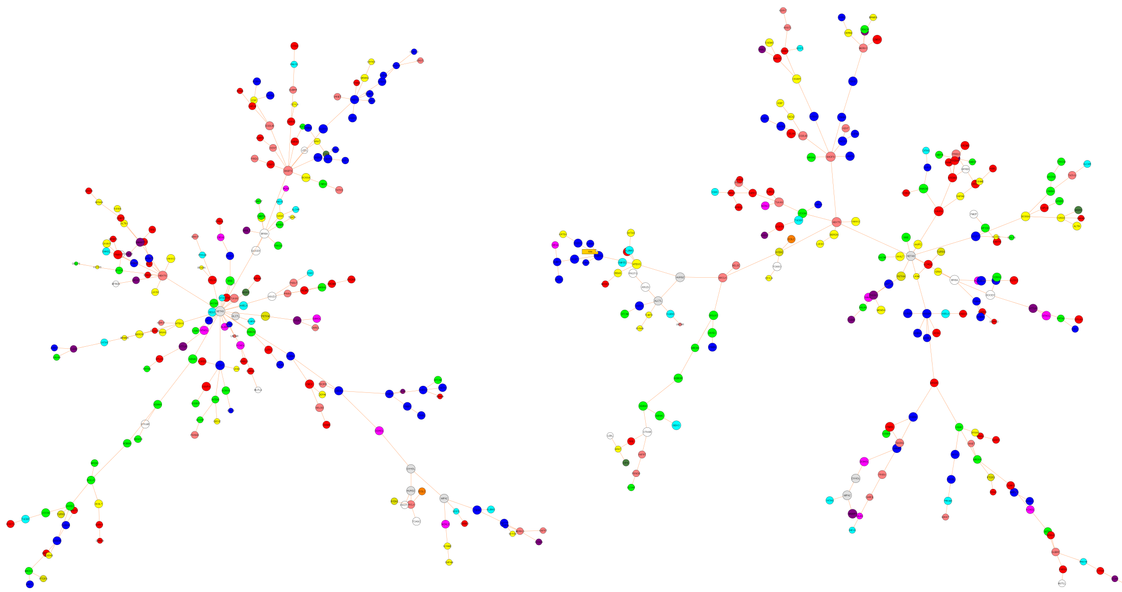


Figure 5: Comparison of MST's of different frequencies of same data
Left:Istanbul Stock Exchange Market Correlation Network by 1-Day Price Change, Right:By 20-Day Price Change

The Figure 5shows the minimum spanning trees, which are obtained by using different time intervals. In our analysis, we used 1-Day, 2-Day, 5-Day, 10-Day, 20-Day price change data but only the results of 1-Day and 20-Day analysis are displayed. We can see that the tree reduces to decentralized system and includes more hubs when we decrease the frequency of data. This result shows in right and middle parts of the Figure 7 by the probability distribution of contact numbers. The contact number of the node had the largest contact number is decreased by decreasing frequency of data. We obtained a similar result with the eigenvector analysis. The common result of those two studies is that the larger intervals give us more information about clusters and interaction in these clusters or blocks. For instance, when we increase the interval of data, the probability density function of contact numbers can be changed dramatically. The number of nodes which have only one neighbor and the maximum contact number reduce, while the number of nodes with neighbors different than 1 gets larger. (Figure 7). This situation explains the increasing hub structure of the trees which is shown in the visualization of MST.

When we observe the minimum spanning trees, we can see the discrete hubs while decreasing frequency of data. We will analyze the statistical properties of these hubs such as the clustering coefficient and strengths in the next sub-chapters.

3.4 Re-Construction of the Correlation Matrix

In recent research the market behavior has been symbolized by the largest eigenvalues. We want to filter this main motion to look into the inside of the hierarchical structure in detail. We started by decomposing the C correlation Matrix of the market by eigenvalue decomposition and then we equaled to zero the largest eigenvalues and discrete eigenvalues. We constructed the new C correlation matrix by only small eigenvalues. After we obtained a new C matrix which was filtered the market behavior, we apply to minimum spanning tree algorithm and figure out the new hierarchical structure and graphs for different frequencies of data. We see in Figure 6 that, the tree is decentralized and there are much more than a few hubs and small groups. We can see the same result in the left panel of Figure 7. When we filter the dominant motion the market we can see the blocking and clustering motion as well as background noises.

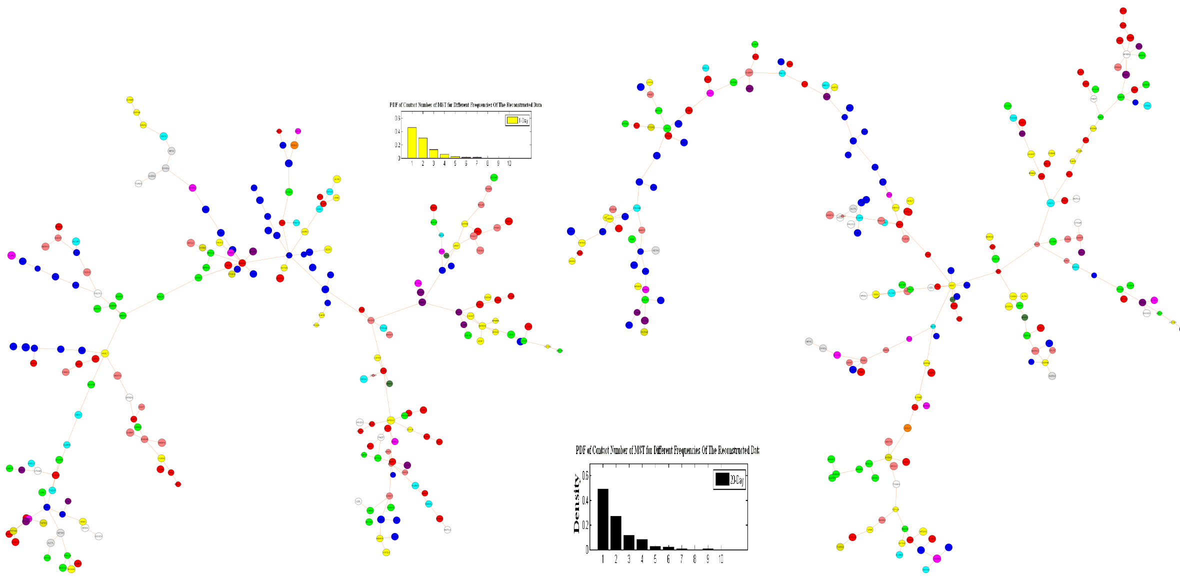


Figure 6: Comparison of MST's of different frequencies of the reconstructed data.
Left:Istanbul Stock Exchange Market Correlation Network by 1-Day Price Change. Right: By 20-Day Price Change

The decentralization and the secession attribution are repeated in reconstructed correlation of the minimum spanning tree of different intervals of data.

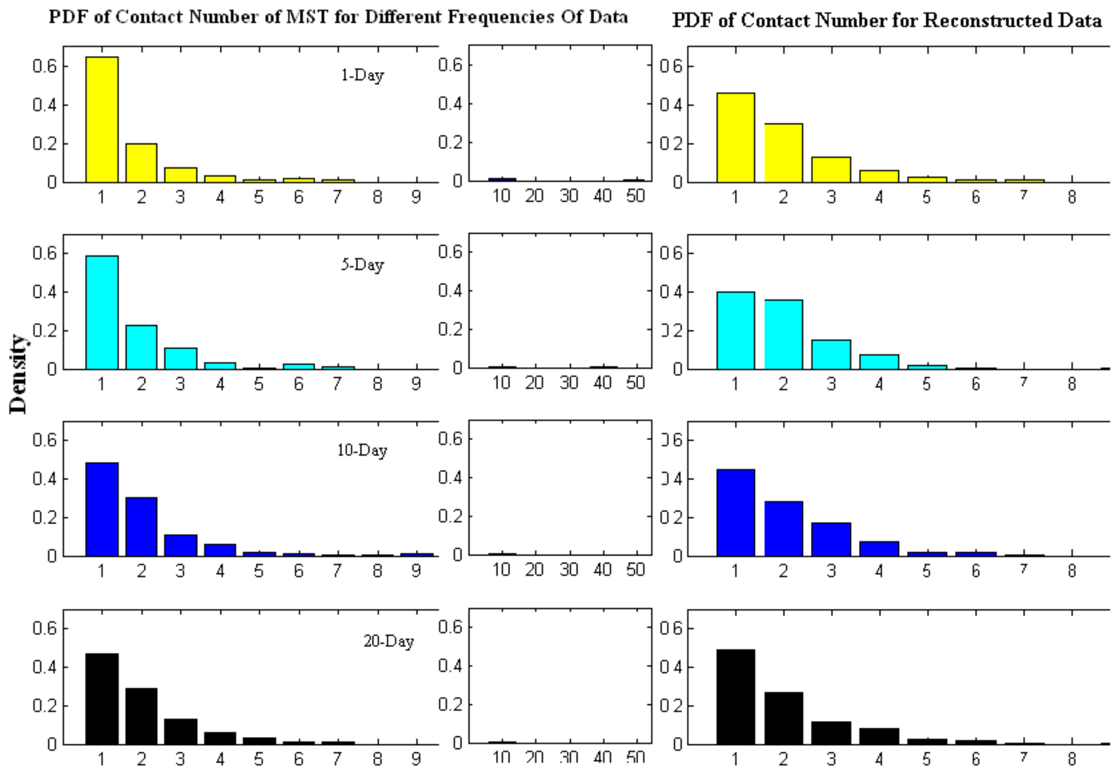


Figure 7: Comparison of the pdf of contact degrees of nodes in minimum spanning tree. (Left: Original data, right: Reconstructed data).

In addition we can reconstruct the C matrix by eigenvalue decomposition and then equal to 0 all eigenvalues except the largest eigenvalues and construct C matrix again. When we draw the minimum spanning tree of this new correlation network we get the centralized minimum spanning tree which is shown in Figure 8. By this method we have filtered all the details of the hierarchical structure. Figure 8 shows that a node symbolizes the capital asset market portfolio and this node acts like the first neighbor of all other nodes.

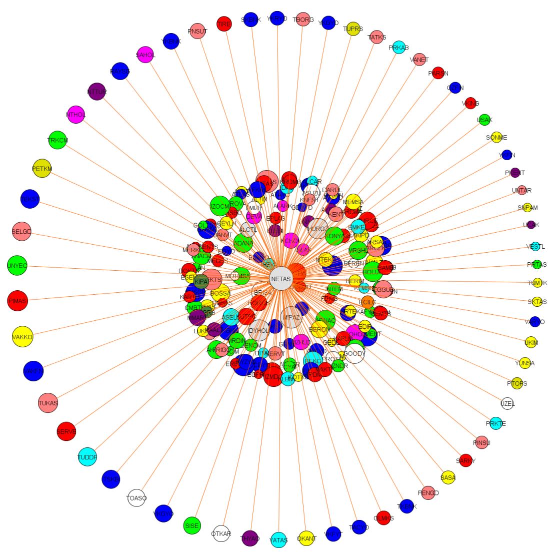


Figure 8: Minimum spanning tree of Istanbul Stock Exchange Market's correlation network reconstructed by only the largest eigenvalue.

We will study to discover the statistical properties of all correlation networks and the minimum spanning trees. We will combine our results of statistical and spectral properties at the end of the chapter.

3.5 Enriched Parameters of the Statistical Properties

On the networks, we need some parameters to assign the structural properties. In some physical real networks such as supply chain network or science collaboration network, some appropriate metrics have been defined to merge weighted and topological evidence that enable us to characterize the complex statistical properties of links and nodes. Aim of the study on statistical properties to obtain the information about groups and clusters and also define topological structure of the clusters. We expect to find out enriched properties of clusters and compare with minimum spanning tree. Strength, neighbor degree, next neighbor's degree and clustering coefficients are some of the statistical parameters which have been used for analyzing the complex financial correlation networks. [9, 8, 18]

In addition, to filter information from noise in correlations, a threshold value, p is used in our analysis. The p value is a ratio of the number of edges which are greater than the threshold correlation value to the number of all possible edges. We use this threshold obtained by spanned graph order analyses.

In our study we also analyze statistical properties by changing the frequency of the data. We use 5 different frequencies of data. The first one is 1 day interval and the others stand as 2, 5, 10 and 20 days interval.

3.5.1 Strength of Nodes

Strength distribution varies according to the network's nature. For instance, on a transportation network, some nodes would be covered more than others or some settling areas would be visited more than others. Another example can be a science collaboration network. The number of references given from the papers would not be the same for all. The nature of the network would be structured by the characteristics of the case. In the financial network case, strengths increase with high correlations and also in central nodes (assets). The strength of the node is defined as;

$$S_i = \sum_j^N a_{ij}\rho_{ij}$$

This quantity measures the strength of nodes in terms of the total correlations between their connections.[22, 9]

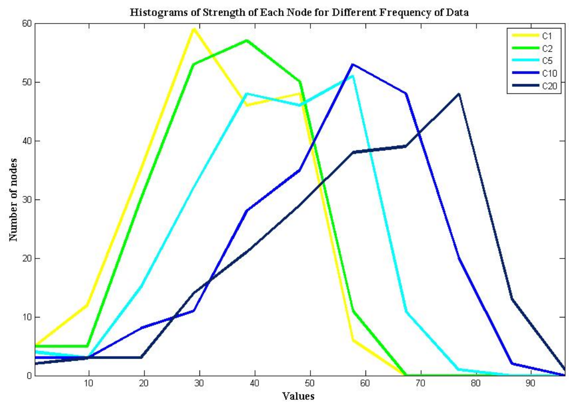


Figure 9: Probability disribution functions of strengths in different data

The ρ_{ij} is the correlation between node i and node j. The a_{ij} matrix is the binary matrix for a special p value as a threshold, it has specified the existing links after threshold. Strength is an effective measure and can be used without any threshold. Strength generally used for providing the standardization with remaining statistical properties and to eliminate the nodes has low strengths and they can be accepted as noisy. In our study we figure out probability distribution function of strength with threshold and without threshold. Figure 9shows the probabilty distribution of the strength of the nodes. This figure obtained by without threshold. This measure shows the common increase on the whole correlations when we decrease the frequency of data. The distribution becomes low and spreading case in 20-Day interval data.

3.5.2 Neighbors and Weighted Average Nearest Neighbors Degree Of Nodes

A node degree or defined as neighbors degree of node is

$$k_i = \sum_{j}^{N} a_{ij}$$

which is the number of the nodes connected to i in specific p value. In fully network which is without a threshold the k is the total number of nodes minus one for all nodes. To reach informative results, threshold has to be used. This threshold gives us the information about central nodes. We can also compare the largest strength of the node and the largest degree of the node.

To perform a local weighted average of the nearest-neighbor degree according to the normalized weight of the connected edges, the weighted average nearest the neighbor degree is introduced as; [22]

$$k_{nn,i}^{w} = \frac{1}{S_i} \sum_{j}^{N} a_{ij} \rho_{ij} k_j$$

In Figure 10 we can obtained the result that, while nearest the neighbors degree are increasing the strength of nodes are decreasing. Their neighbors have less connection and at the same time large strengths. It is similar for all intervals of data but this reverse pattern gets lost in big intervals such as 20-day data.

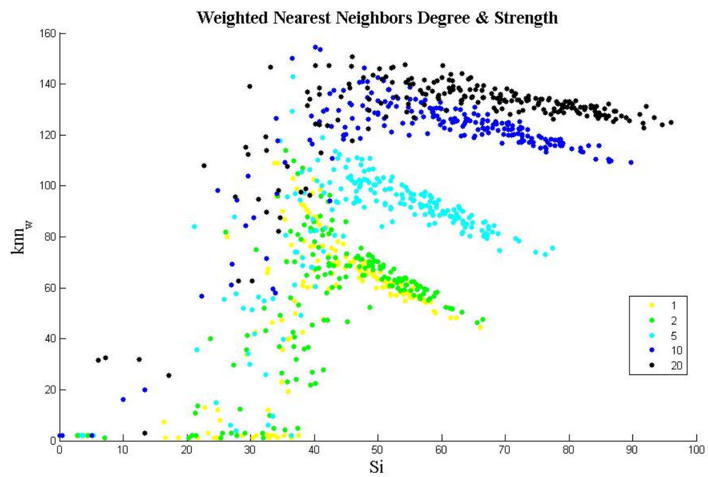


Figure 10: Scatter plot of weighted average nearest neighbors degree of nodes to strengths.

3.5.3 Clustering Coefficient and Weighted Clustering Coefficient

The clustering coefficient is a metric of strength and local cohesiveness. The cohesiveness around node i can be observed by the clustering coefficient C_i, defined as the ratio between the number of triangles of node i and the maximum possible number of such triangles:[1]

$$C_i = \frac{\triangle_i}{k_i(k_i-1)/2} = \frac{\sum_{j,h} a_{ij}a_{jh}a_{hi}}{k_i(k_i-1)}$$

a is binary adjacency matrix. Hence $C_i = 0$ if none of the neighbors of a node are connected, $C_i = 1$, if all of the neighbors are connected.

In some cases all nodes or links do not have the same features. Some nodes have more strength and can be located in central of the network. So, the parameter of Vespignani et al. provides that some of nodes are more important.[22]

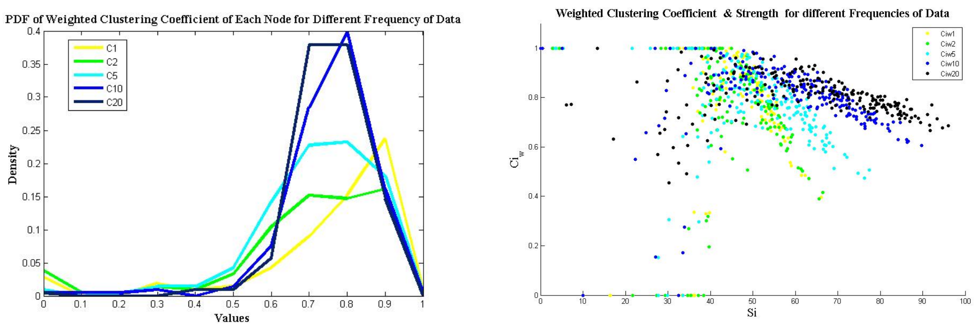


Figure 11: PDF of weighted clustering coefficients in different frequencies and Scatter plot of weighted clustering coefficient and strength

The weighted clustering coefficient C_i^w measures the local group cohesiveness and is defined for any node i as the fraction of connected neighbours of i.[18, 9, 22]

$$C_i^w = \frac{1}{S_i(k_i - 1)} \sum_{j,h}^{N} \frac{(w_{ij} + w_{ih})}{2} a_{ij} a_{ih} a_{jh}$$

Clustering and also weighted clustering coefficients are normalized on the interval[0,1]. If $C_i > C_i^w$ then, the topological clustering of the network is generated by links with low weight. On the contrary, $C_i < C_i^w$ the interconnected triples, are more likely to be formed by the links with larger weights.

In Figure 11 the value of weighted clustering coefficients increases by decreasing the frequency. The other important result is the p ratio value which is connectivity, is increased. The most important result is visible in $C_i^w \& S_i$ figure. While Clustering Coefficients are getting bigger the strength of nodes are decreasing.

It has shown that, clustering coefficient and weighted average neighbors' degree increase at the same proportion, while the clustering coefficient increases the strength decreases but nearest the neighbors degree increases.

3.6 Synthesis of Spectral and Statistical Analyses

The correlation matrix is differentiated from the random matrix by having hierarchical structures and special properties[3]. Minimum spanning tree and statistical properties of a correlation network are used to figure out these special properties. Meanwhile, eigenvalue analysis enabled us to expose the hierarchical structures of correlation network. As expected, we observed that the correlation matrix is differentiated from a random matrix by eigenvalues distribution. By considering the results of the analysis, we reconstructed our correlation network by using two different methods, and tried to find out the special properties. In the previous subsection we introduced the statistical properties of original network which have spectacular patterns. Figure 10 Now, we will demonstrate the statistical properties of the reconstructed correlation networks.

First of all to compare the results easily knn_i^w versus S_i figured in this sub-section.

The correlation matrix firstly decomposed and then composed with only largest eigenvalue and statistical properties and minimum spanning tree are achieved, at the same time.

We see patterns in Figure 12clearly. Statistical properties of reconstructed network are separated from original network by definite patterns.

We applied second reconstruction method which is composed with only small eigenvalues. We have almost different properties figured out in right parts of the Figure 12.

The most important result is that the largest eigenvalue carries all information about the patterns and properties. Minimum spanning tree instances

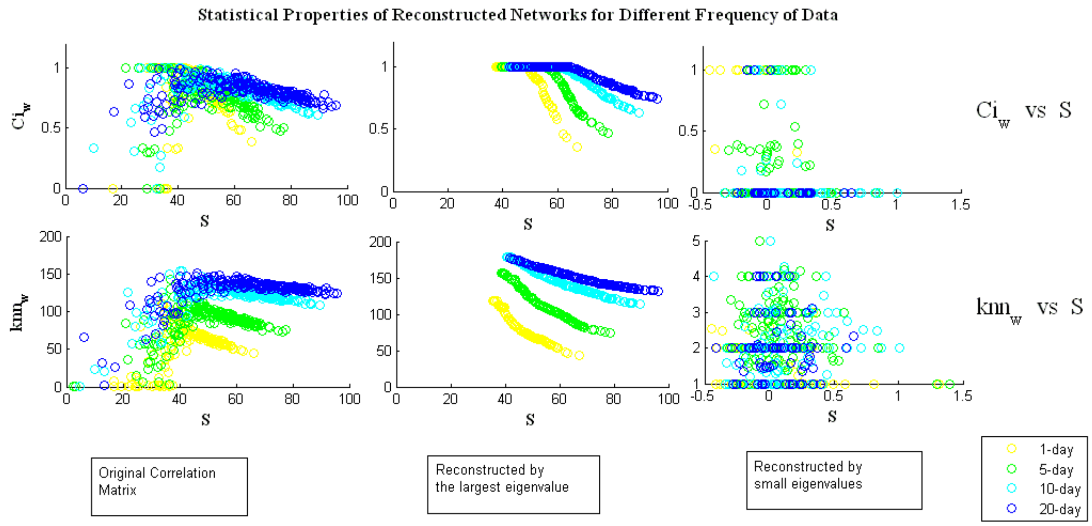


Figure 12: Top figures: Weighted average nearest neighbors degree versus Strength for original network and reconstructed networks. Bottom figures: Weighted clustering coefficient versus Strength for original network and reconstructed networks.

support our findings about the largest eigenvalue.(Figures 8, 6). We reconstruct the correlation network to purify the information about clustering and local interactions from general behavior of the market. Our last figure says that without market behavior there is not any pattern in statistical properties. This result can be helpful to mention about the market efficiency. (Figure 12)

4 REWIRED PRICING MODEL

4.1 Basic Terminology and Literature Review about Estimation Methods and Rewired Networks

Estimation of the future has always been an enigmatic topic. Uncertainty of future is intrigued by scientist from different disciplinary such as economics and finance. Prediction of uncertainty in dynamic systems is a more difficult topic compared to the static systems. Finance is an example of dynamic and multi-disciplinary systems that can be affected from different type of information. Financial market is a huge and union system which can be defined by the union of sub financial markets and intersection of them. Stock markets, derivative markets, commodities, bond and futures are main and well-known markets which form the financial market. Since each market is adequately huge, analysis usually conducted on the clusters or groups of the sub-market which is an open system and can be affected by other sub systems. When we started to research on stock market we easily realized that the sub market assets or nodes behave as an integrated or non-stationary time series. This result shows us that to the sub-systems' behavior differs from Gaussian motion which is a desideratum condition by finance analysts, economists and all forecasters. A time series is stationary if it is probability distribution function invariant under time shift [13] and data fluctuates around a constant value. However time series of nodes are integrated or as we called in our study they are piecewise-stationary. When we move the asset from the cluster or sub-financial market to the union market, the system becomes a closed system and more stationary. So the studies which are about time series shows that the predictability of time series become easier when the system is stationary. Nobel laureate scientist Grange emphasizes that many pairs of macroeconomic series seem to be stationary on linear combination of them, as it is suggested by economic theory.[1] This finding certifies our conjecture which states that union financial market called closed system behaves more stationary according to sub financial markets which are open systems. Analysts usually work on time series which are only a node or clusters or a stock market which we have studied in previous chapters. Estimation of time series of assets which are accepted non-stationary is relatively more difficult to predict. Although estimation of the future is not an easy topic in finance, we can use some basic forecast methods such as auto-regressive moving average models, generalized auto-regressive conditional heteroskedasticity, decision trees, artificial neural networks and other several methods which are recently have been researched. Time series forecasting is a harder problem because of the fact that it is nonlinear and integrated. Since integrated time series are predicted relatively in a more difficult way, performances of forecasting methods are affected detrimentally. On the contrary the stationary series can be estimated easier by the help of the trend or seasonality. Unfortunately, different to product demand series or detective product series, financial time series which is formed by daily closing price of assets have integrated motions such as random shocks, short trends and lots of noises. Auto-regressive moving average model (ARMA) sometimes called Box-Jenkins is a time series prediction model that was pioneered by George Box and Gwilym Jenkins. It includes two parts. First one is AR part which principal is p and the second part is MA part which is represented by q. When p=1, weight is in part p equal 1 and all q part is equal to 0 then the last state become random walk formulation. Other univariate forecasting model which considers the volatility of returns was first introduced by Engle who is the Nobel laureate in 2003[5, 6]. Autoregressive conditional heteroskedasticity (ARCH) model considers the variance of the current error term to be a function of the variances of the previous time period's error terms. ARCH relates the error variance to the square of a previous period's error. It is employed commonly in modeling financial time series that exhibit time-varying volatility clustering, i.e. periods of swings followed by periods of relative calm. Generalized form of ARCH, GARCH, considers the previous time period's variances at the same function.

Financial time series must be considered as multivariate models, since corelation, co-integration and co-movement terms are the central part of the time series in financial markets. Researchers have taken into consideration the structure of market while modeling a forecasting method to increase the accuracy. Therefore it is important to extend the considerations to multivariate GARCH (MGARCH) models. Multivariate GARCH models are mostly used by Bollerslev et al[2],for asset pricing based on the covariance of the assets in a portfolio, and risk management.

4.2 Principle of Rewiring

For the studies which are explained in detail in the previous chapters, we always used the correlation coefficient and statistical properties of the correlations. However the correlation coefficient is only a value and it changes by the scale of data range and also by frequency. For instance, when we take two time series and find the linear relation of these time series such as correlation between them and if we divide each time series into two equal parts and find the linear relation of each part of time series, we can find two different correlation coefficients. The coefficients can be different and may be reverse of each other. It is our point of origin. We will explain this point by principal of rewired network or Small World which was defined by Strogatz S.H. and Watts D.J.[20]. They states that in a network if a link between node i and node j had rewired and resulted in a new link between node i and node k, hierarchical properties of network will changed dramatically. We can compare this principal to our correlation coefficients. By using the information above, we can say that if there is a correlation coefficient in set of earlier time series it can be changed in the later part of time series or the relation may disappear. Correlation coefficient is a matter which can be utilized for a lot of statistical properties. But taking only a value which is changing in sub intervals of time series can be decreased forecasting accuracy. We can consider it like principal of small world theory. Since, change on the structure of network triggers modification on enriched properties of network.

The above example is used to explain the small world theory on a ring. In the first network, edges are in a regular sequence and the weight is assumed to be 1. In the second network, the edge between node 1 and node 4 disappears and an edge between node 1 and node 3 arises. We can expound this situation as there is a relation between two time series. The co-relation between two nodes can be perceived as the relationship between two time series. If the co-relation between two nodes disappears then we expect that there is a new co-relation between two other nodes or the co-relation is disappeared

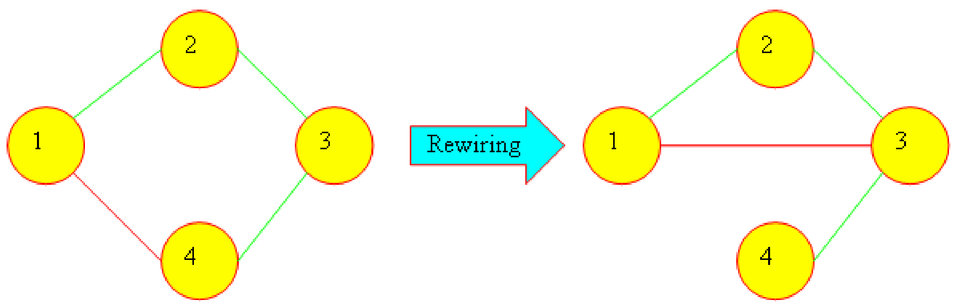


Figure 13: Sample rewired network. The 1-4 link is flip ped to 1-3.

forever. If the co-relation between two nodes does never disappear then we can say that correlation between two nodes is same and high all the time. This condition is accepted as stationary behavior or co-integration between two nodes. Constant high relation means that two time series are moving together and linear combinations of time series must be closer to the stationary condition. The assumption of stationary behavior makes the prediction of time series easier. Unfortunately, in many associated time series the correlation changes in different part of time series. We will call to this condition as 'rewiring' in the rest of the study. In this chapter we will define some of networks which are rewired in length of time. Then we search for the windows length to define rewiring clearly, such as to define state more efficient. We will use the states to introduce the length of interval in which correlation coefficient is constant or in low volatility. Setting the states is main part of our Rewired Pricing Model. We must take into consideration smooth correlation coefficients in this interval and also find this condition in all correlations which take place on networks. Then we introduce our pricing model which works with multivariate forecasting parameters and rewiring principle. We will apply our model to three networks from different financial markets or clusters. We compare results and performance of model. At the end of chapter we will discuss our lacking parts and also why the model behaves dissimilar in various networks.

4.3 Rewired Pricing Model

Time series prediction is a hard and interesting topic for research. The literature includes many studies which predict the future values by using only one time series. Especially financial analysts use different methods such as technical analysis, univariate models, historical Monte Carlo and multivariate models. We will try to develop a historical Monte Carlo method to generate future time series by using rewiring principals. As we mentioned in the previous topic, we want to identify the relationship of several time series and develop a model which reflects the relations between those time series. The model is based on the fact that the correlation network changes in different time intervals as figure out in time series which are below.

For instance, we have 4 price time series. When we divide into 4 equal intervals these time series and construct the correlation network, we gained different correlation values between the assets in Figure 14.

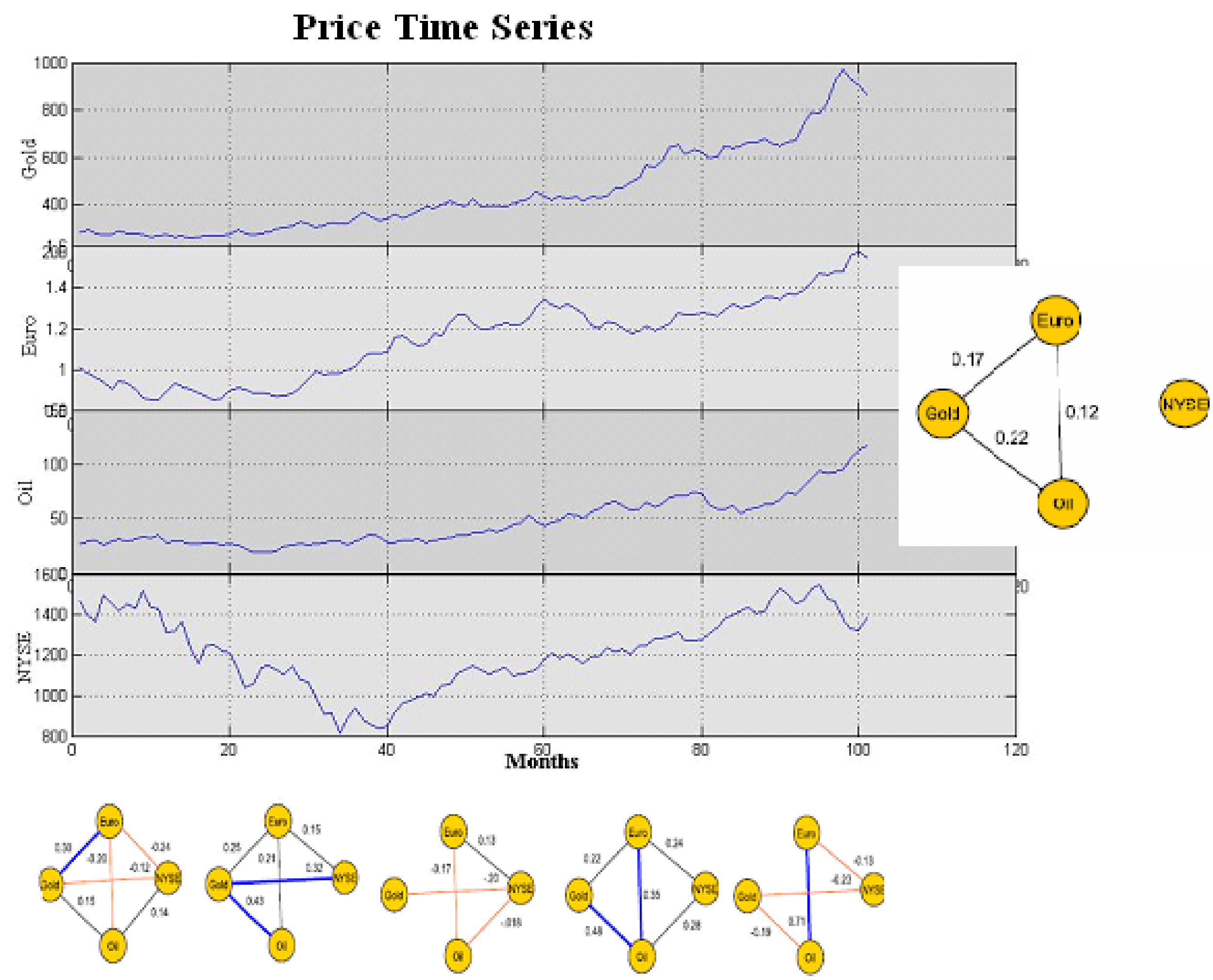


Figure 14: Asset price time series and corresponding correlation networks in different time intervals. Red edge denotes the negative correlation, bold blue is strong positive correlation (C>30) and black one refers to correlations between 0.29 and 0.10

In the figure the blue lines represents the strong correlation, black lines represents the weak correlation and red lines represents the negative correlation between the assets. In the second network we see that there is a strong correlation between gold and oil but in the third network this relation between those assets disappears.

We raised some interesting questions; by answering those questions we try to develop our model. The questions can be summarized as follows:

1. How can we identify the time intervals or as we will call windows length hereafter?
2. How can we define periods which we call states hereafter?
3. Which factor of states can affect the model of accuracy?
4. Which multivariate data generation model can we use?

4.4 Factors Affecting Information Content of the Analysis

4.4.1 Stationarity in Correlation Time Series

Multivariate data generation methods use covariance matrix and the correlations between variables. For gaining sufficient results from simple multivariate data generation, first we have to be sure that the correlation stays constant during the time. Therefore we can estimate the adequacy of the forecasting by the stationary of the correlation time series. Stationary in a correlation time series can be identified in two ways. We can either find the parameters, mean and variance, by looking at the data partially or by moving on the data. We will use moving correlation of time series which has been a critique issue since the late 1920s. [12]

4.4.2 Window Length and State Length in Optimal Piecewise Stationarity

Volatility is estimated by sample standard deviation of returns over a short period. However determining this short period requires a lot of work. Engle said that: 'But, what is the right period to use? If it is too long, then it will not be so relevant for today and if it is too short, it will be very noisy. Furthermore, it is really the volatility over a future period that should be considered the risk, hence a forecast of volatility is needed as well as a measure for today".[6] This problem is very similar to the problem that we face for the measurement of linear relation of two time series which is the correlation coefficient. We will try to find an adequate window length by analyzing the relation of two time series. The window length should be as small as possible in order to ignore the convergence of the correlation to a mean variable and it should be as large as possible in order to prohibit the noisy.

We started our analysis by calculating the correlation of first ten data points of each time series. Then we added the next data point and found the corresponding correlation. This procedure is applied until we found the correlation of 100 data points. In other words, we first found the correlation of the variables 1 to 10. Then we found the correlation of the variable 1 to 11 and it continued until we found the correlation between the variables 1 to 100. The above figure shows the correlations which were obtained by this technique. This figure shows that the statement of Engle is viable for correlations, too. When we look at the earlier part of each curve, the part that we use small window length, we observed a lot of noisy. We see that when we

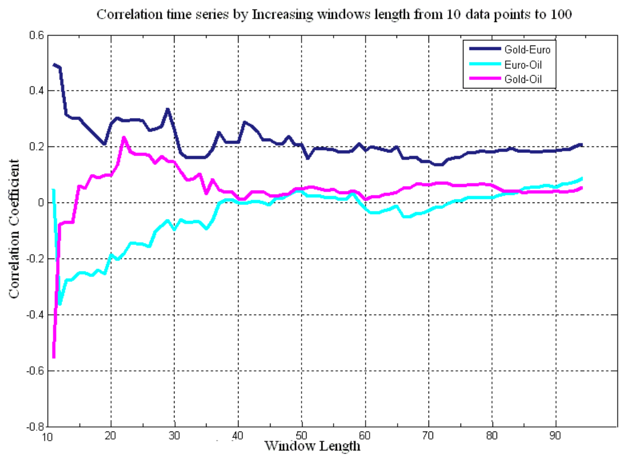


Figure 15: Correlation time series by increasing the window length of time series which is considered.

add a data point to the window length the correlation changes dramatically. In an efficient time interval, the effect of adding one data point should not be as large as we observe in the earlier part. For example the correlation of gold-oil was calculated as nearly -0.5 for the first window length but it becomes almost zero for the second window length. We search for a window length which provides us the change in confidence bounds. Since, our model will consider the changes on correlations, we do not want to a steady attitude, too. For example if the window length is so large then a correlation can lose sensitivity of adding the data points. It may not reflect the current relation of time series.

As an alternative to increasing window length, we also conducted our analysis with constant window length. We will utilize the moving correlation concept which is an old topic studied earlier than 1930(Simon Kuznets). We move on time series by our window length and get the moving correlation

Figure 16: Moving correlation time series for different window length

time series. For example, for window length 20, we calculated the correlation of the data 1 to 20, then 2 to 21 and continued until we found the correlation of all data. The above figure shows the results of the analysis that we performed for different window lengths: 10, 20, 30 and 40. When we look at the curves of moving correlation time series, we observe the same behavior with the previous analysis. Although the earlier part of the curves is very sensitive to addition of a new point, the curves are getting more stable as we move through the data. We call the behavior of the curves as unrewire after position 30. In topmost figure, the volatility of correlation time series is huge, but the volatility is reducing as we get down of the figure. In most bottom figure which is 40-window-length, curves reached the stable form. By combining the results of these two analyses, we decided to use 20 data point as the window length of our studies. After we define the window length we directed our studies to the definition of the state and state length which gives us the rewiring period clearly. We search for some piecewise stationary in the correlation time series. Our aim is to recognize the rewiring between these stationary pieces and pick the corresponding time interval which has the lowest volatility and more stable in the correlation network. For defining an interval as stable, we have to make sure that all the correlations defined in that interval should show similar behavior to stable correlation series. For our example, gold-euro, euro-oil, gold-oil correlations should be all stable in a time interval. As long as the correlations between the assets stay constant we keep in the same time interval but when we observe a change in the relations of the assets or as we called when we observe rewiring we move to a different time interval. To obtain the technical information we also apply autocorrelation of moving correlation time series in Figure 17. For instance we took moving correlation time series of 5, 10, 20, 30, and 40 window length and find out the autocorrelations of the each series from lag 1 to 20. We try to figure out auto memory of each moving correlation time series. We want to pick a window length to have a memory in lenght of window which it corresponds.

As the first step of our analysis, we performed the moving correlation time series with 20-Month-Window-Length. We analyzed the correlation of all time series in the network. In the below figure, the blue points reflect the correlation values for 80 month data. When we analyze the graph, we figured out that the correlation coefficients take similar values during 5 months period. For the second step, we calculated the average correlation coefficients

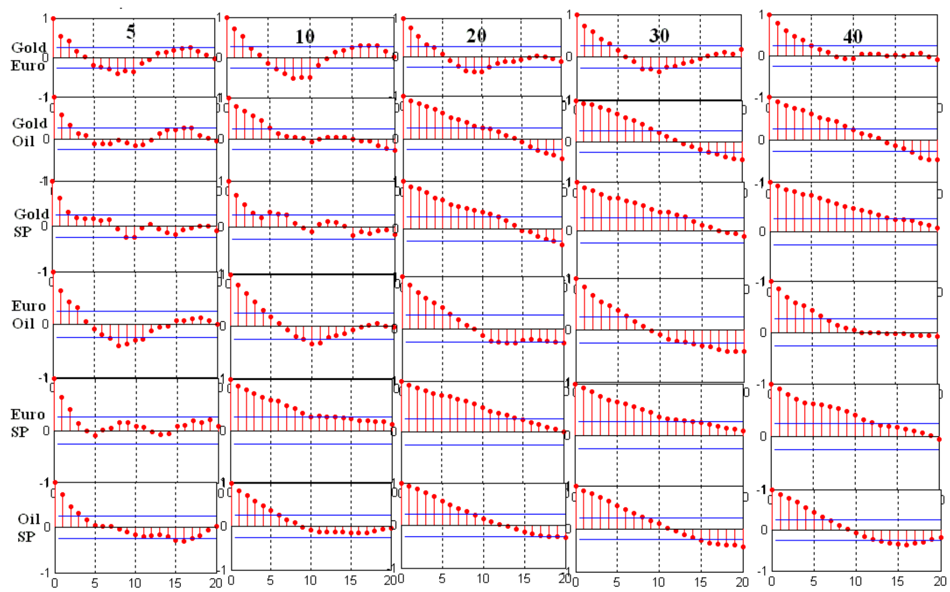


Figure 17: Autocorrelations of moving correlation time series for assets in different window length.

of the groups that include 5-months data and highlighted them with pink. Finally we pooled 4 of those data groups under one unit and defined those units as state. So we came up with 4 states which include 4 data groups. We characterized the states with a correlation value that cover the average correlation coefficients of all four groups within the state. In the figure, those values are shown with yellow lines.

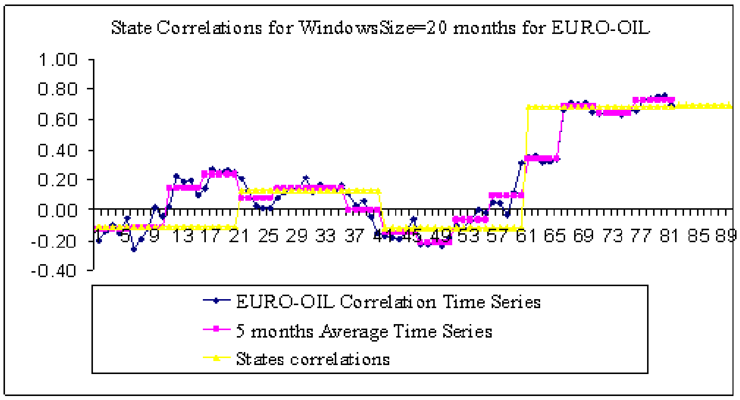


Figure 18: Euro-Oil Correlation time series and defining the states

To improve the aim we will constitute the moving correlation time series by consider the 20-Month-Window. We analysis for all correlation time series in network, it is 6 for our simple network, and pick the stationary parts as states. To define states we develop a metric, it is 5-month average time series. It considers the average of 5 data point for helping in defining of states. Then to inspect an average correlation value which cover the 5-month time series. In our study we have choose the state length which has 20 correlation data point. We set the state correlation network by considering the visual average (not numerical) of 5-month average time series.

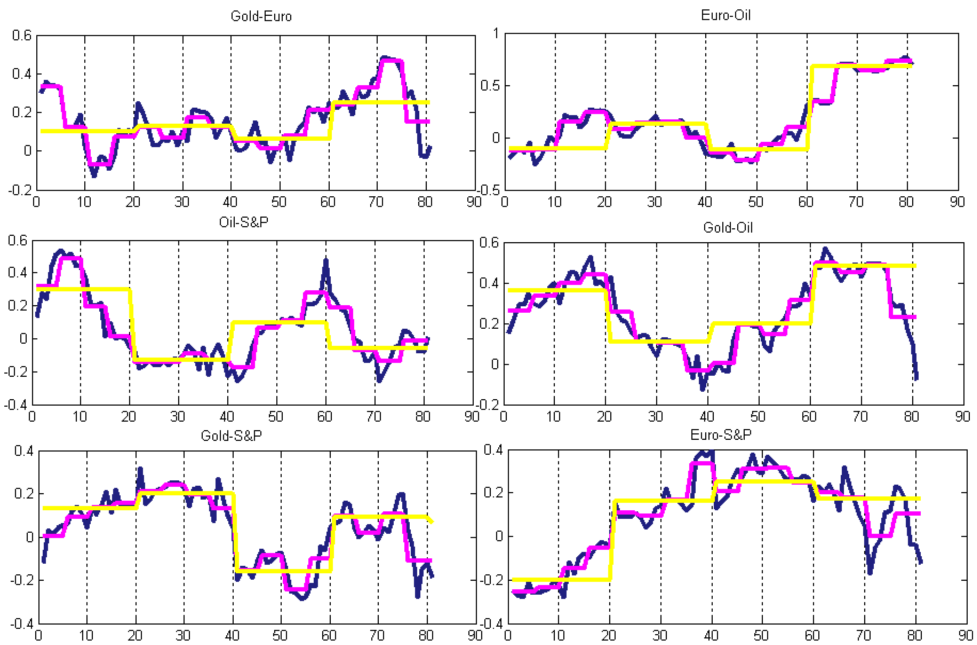


Figure 19: Moving corelation time series of assets in 20-month-window-length To define states we use 5-month average metric. It considers the average of 5 data point for helping in defining of states. Then we try to inspect an average correlation value which cover the 5-month time series. In our study we have choose the state length which has 20 correlation data point.

4.4.3 Multivariate Data Generation Method and Random Walk

The states which are defined by above analyses can be used for forecasting. Our idea says that correlation networks can rewire at any moment, but we have only historical states. Another important point is that we want figure out the piecewise stationary and construct a historical simulation monte carlo model by using predictability of stationary. Our model is a risk management measure it is not a forecasting model yet. However we try to improve we can not develop an algorithm to predict the future state. We only say that these states have realized in historical data. If these states expand to the future, the scenarios which are generated will be implemented.

To generate scenarios we will use this pricing model.

P(t)=P(t-1) + P(t-1) * R(t)) and

R(t) = $\sigma_{ij} \times E(t)\ E(\mu, 1)$

To get this form we will use Cholesky Decomposition method.

Data generation steps can be numbered such as below;

1. Decompose Covariance Matrix by Cholesky Factorization for each state.

$\sigma_{ij} = L \times L^T (bycholesky)$

$\tilde{\sigma}_{ij} = Covariance(RateReturnTimeSeries) = Cov(R)$

$\tilde{\sigma}_{ij} = Cov(GeneratedRateReturns) = Cov(L \times E)\ E(\mu, 1)$

$\tilde{\sigma}_{ij} = Cov(L \times E) = L \times Cov(E) \times L^T$

$\tilde{\sigma}_{ij} = Cov(L \times E) = L \times I \times L^T = L \times L^T$

$\tilde{\sigma}_{ij} = \sigma_{ij}$

then, the future rate of returns can be generated by $L \times E$

2. Generate random rate of returns for each states which covariance matrix behaves as original rate of returns.

3. Pricing according to generated rate of returns

P(t)=P(t-1) + P(t-1) * R(t))

and repeat this algorithm in 500 runs. We will have 500 different forecasts by each state. We apply to our model two different sample data set.

4.5 Rewired Model Applications

4.5.1 Assets from Different Sub Financial Market

The first data has 4 assets from different sub financial markets:.

Data set;

- Prices of assets from January 2000 to may 2008
- Gold, Euro, Oil, NYSE (S&P 500)
- Total 100 months
- First 80 months have been used to predict the months of 81-90

Initialization

1. Obtain the monthly rate of returns.
2. Construct Correlation Network
3. Analyze the rewiring

When we divide the data into 4 equal intervals and construct the correlation network, we gained different correlation values between the assets in Figure 14. In the figure the blue lines represents the strong correlation, black lines represents the weak correlation and red lines represents the negative correlation between the assets. In the second network we see that there is a strong correlation between gold and oil but in the third network this relation between those assets disappears. After we observed such rewiring on the correlations of assets, we continued with the application of our model. Our aim is to predict the correlation coefficients of the time interval which includes the months between 80th month and 90th month with 4 historical states. The below figure shows the monthly forecast distribution of the future oil price time series generated by state 4. These distributions are obtained by repeating the forecasting algorithm for 500 times. We claim that the pick values of those distributions can be regarded as the future prices of oil. The red line shows the exact prices of oil which eventuated in 80-90 month-intervals. As we see from the figure, our model gives close forecasted values to the realized prices.

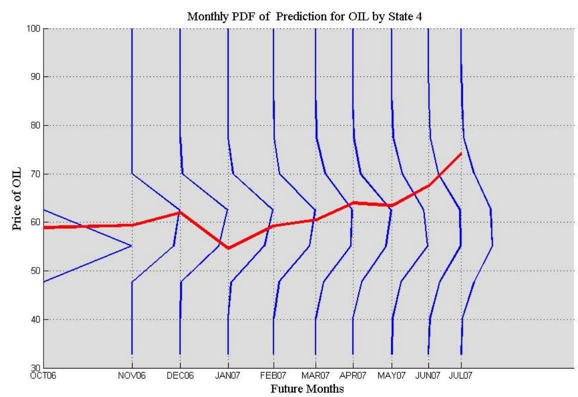


Figure 20: Monthly predictions of Oil by State 4. The blue lines show the vertical probability distributions obtained by 500 runs for each month.

In Figure 20 the red line shows the exact time series which have eventuated in 80-100 month-interval. In Figure 21 the probability distribution function of runs is illustrated.

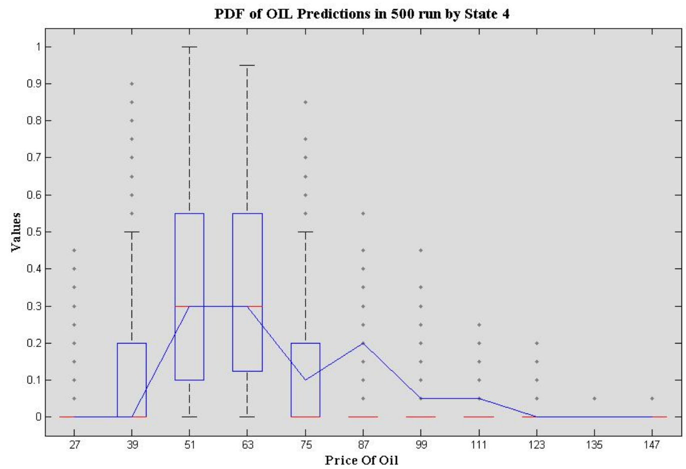


Figure 21: Probability distribution function of the predictions of OIL by 500 runs of state 4

In addition to the above analysis, we also compared the probability distributions functions of the forecasted values and real prices. The red line in the middle of boxes shows the median of the probability distribution function for the value which forms in horizontal axis, and the boxes cover the 25% of the data which correspond to value in horizontal axis. The black dash boundaries have the 75 % data. The black bold points correspond to the outliers.

4.5.2 Assets from the Same Cluster of the ISE

The second data set is selected by using the minimum spanning tree of correlation network, developed in section three. Our assets are from same cluster in the second dataset. We use the assets from the finance cluster and they are also national banks.

Data set;

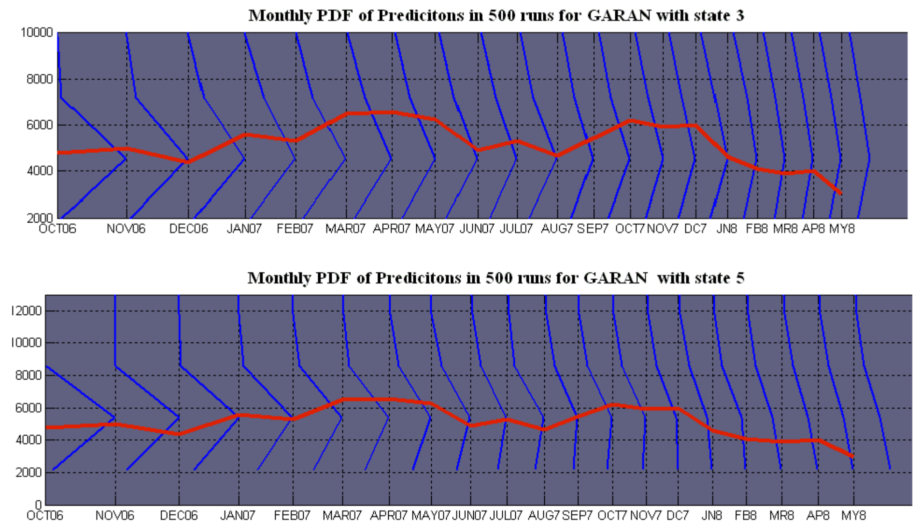


Figure 22: Monthly predictions of GARAN by State 3 and future state which is state 5. The blue lines show the vertical probability distributions obtained by 500 runs for each month.

- Prices of assets from January 2000 to may 2008
- AKBNK, FINBN, GARAN, YKBNK
- Total 100 months
- First 80 months have been used to predict the months of 81-100

The Figure 22 shows the results of the analysis that we made for GARAN by state 3 and state 5. We see that our model give good near future predictions when we conduct it with state 3 which is a historical state. However the results become insufficient as time passes. On the other hand, our predictions behave similar to the real prices when we use state 5 which is a future state. In contrast to state 3, the predictions of state 5 get better as time passes. It is a result of stationary behavior of the correlation time series.

5 SUMMARY AND FUTURE WORK

This study aims to understand the hierarchical structure of the financial markets and develop a model that would be used for forecasting of future relations of assets. During our analysis, we attempted to integrate the spectral and statistical analyses, which were shown to be complementary tools. In addition we use the minimum spanning tree as a conjugate of these analyses.

We have shown that a hierarchical structure exists in the financial market, by using both the analysis tools and visualization on an MST. Our first important result is represented in figure 10. There are two kinds of patterns between weighted average nearest neighbors' degree and strength of nodes. We name the first as the dominant reverse pattern, which implies that as the affinity of clustering increases the strength of nodes decreases. It shows that there is a blocking structure in the network and supports our claim that the networks involve clustering. The second pattern shows that the weighted average nearest neighbors' degree and strength shifts are directly proportional. Although the latter is a general property of all networks, the former is a special property that causes the differentiation between the correlation matrix and a random matrix.

Another important finding of the current study is about the effects of frequency of data used in the analysis on the structure of the market. Statistical properties and the behavior of MST change when we use different frequencies. As we increase the length of the interval, the effect of the pattern that we observe in the statistical properties decrease, since the correlation coefficients of all assets increase significantly. In MST, we observe decentralization as a result of long intervals.

In the light of the foregoing, we have next made an attempt to investigate the dynamics of the network by searching for the underlying reasons of the changes in the statistical properties. Hence we reconstructed the correlation network by using the particular components of their eigenvalues distributions. We find that there is a single motion in the market that dominates all the other motions within that market. This information about the structure is carried in the largest eigenvalue. This analysis can be used to determine the efficiency of the market by comparing with the structures of the reconstructed matrices. Furthermore, when the correlation matrix is reconstructed with the largest eigenvalue to obtain the probability distribution of strengths, the same results are attained in all reconstructions, independent of the time interval. The upper and lower bounds as well as peak values of the distributions do not change. Thus, the reconstruction inhibits the information loss due to changes in sampling intervals of returns.

Another contribution of this thesis is the development of a rewiring principle for the study of the correlations between assets. In addition to providing an insight into the structure of the systems, this information is further used to generate forecasts. We benefited from the rewiring principle concept for scenario generation in a historical Monte Carlo model. We applied our scenario generation model with two different portfolios and obtained different results. While the first portfolio includes different types of assets whose correlations rewire many times, the second portfolio involves assets from the same cluster. We find our model to be more effective for the latter type of portfolio selection. A factor that increases the efficiency of our model is defining the state length with piecewise stationarity, because it is easier to predict the future of a stationary time series.

In this thesis we have observed the structure of the financial markets and developed a scenario generation model. In the future, all the calculations carried out for ISE will be applied to other stock exchange markets, both from developed and emerging markets. The aim is to evaluate the applicability of the current approaches by selecting the parameters that best differentiate the characteristics of these markets. Another extension of the current study will be based on the rewired pricing model. We have identified certain motifs dominating the cross-correlations. However, this development lacks an understanding of the order and the combinations of the motifs that emerge. Based on these, our ultimate aim is to develop a model that best describes the possible correlational patterns of the future.

References

[1] Moreno Y. Chavez M. Hwang D.U Boccaletti S, Latora V. Compleks networks: Structure and dynamis. *Physics Reports,Elsevier*, 424,p. 191, 2006.

[2] R.F. Engle ve J.M. Wooldridge Bollerslev, T. A capital-asset pricing model with time-varying covariances. *Journal of Political Economy*, 96, 1998.

[3] Cizeau P. Potters M. Bouchaud J-P., Laloux L. Noise dressing of financial correlation matrices. *PRL*, 83, 1469, 1999.

[4] West D.B. Introduction to graph theory. *Prentice-Hall*, 1996.

[5] Robert Engle. The use of arch/garch models in applied econometrics. *The Journal of Economic Perspectives*, 15, 2001.

[6] Robert Engle. Risk and volatility: Econometric models and financial practice. *Nobel Lecture*, 2003.

[7] Illes Farkas Tamas Vicsek Gergely Palla, Imre Derenyi. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435, 2005.

[8] Precup O. Iori C. Weighted network analysis of high-frequency cross-correlation measures. *Physical Review E*, 2007.

[9] Kaski K. Kertesz J., Onnela J. P. Clustering and information in correlation based financial networks. *Eur. Phys. J*, B 38, 353, 2004.

[10] Onnela J.P. Saramaki J. Kaski K. Kertezs J., Tibely G. Spectrum, intensity and coherence in weighted networks of a financial market. *Eur. Phys. J*, B 38, 353:178, 2004.

[11] Morris Kline. Mathematical thought from ancient to modern times. *Oxford University Press*, p.706-707, 1972.

[12] Simon Kuznets. On moving correlation of time sequences. *Journal of the American Statistical Association*, 23, 1928.

[13] Stanley H. E. Mantegna R. N. Introduction to econophysics 'correlations and complexityin finance'. *Cambridge University Press*, 2000.

[14] Nowak M. A. Kondor I. Papp G., Pafka S. Random matrix filtering in portfolio optimization. *Acta Physica Polonica*, 36, 2005.

[15] James B. Orlin Ravindra K. Ahuja, Thomas L. Magnanti. Network flows: theory, algorithms, and applications. *Prentice Hall*, 1993.

[16] Mantegna R.N. Hierarchical structure in

nancial markets. *THE EUROPEAN PHYSICAL JOURNAL B*, 11,193-197, 1999.

[17] Kaski K. Saramaki J, Onnela J.P. Spectral and network methods in the analysis of correlation matrices of stock returns. 2007.

[18] Onnela J.P. Kaski K. Kertezs J. Saramaki J, Kivela M. Generalizations of the clustering coefficient to weighted complex networks. 2006.

[19] V. Gopikrishnan P. Rosenow B. Amaral L. Stanley H.E., Plerou. Universal and nonuniversal properties of cross correlations in financial time series. *PRL*, 83,no.7 1471-1474, 1999.

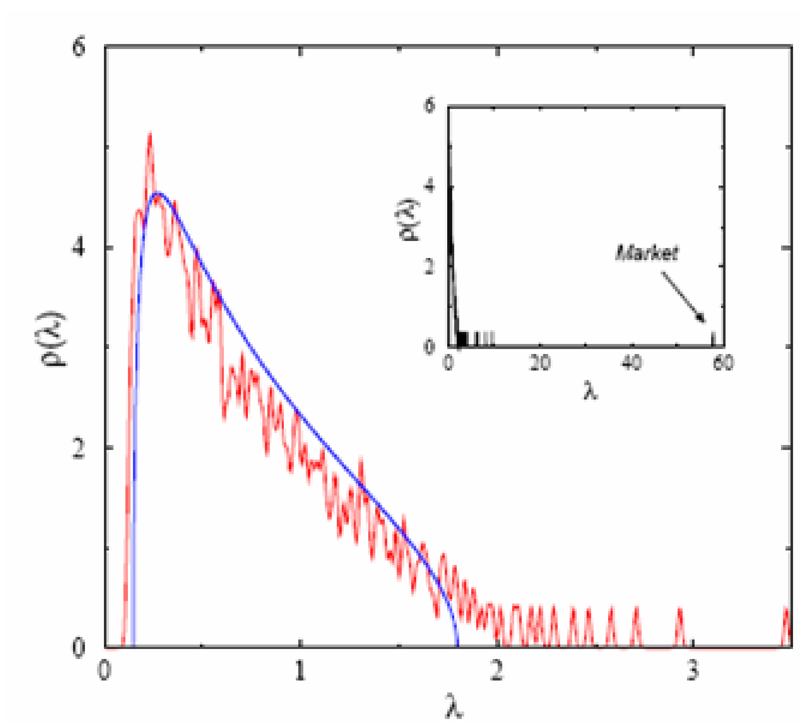
[20] Watts D.J. Strogatz S.H. Collective dynamics of small-world networks. *Nature,Macmillan Publishers*, 393, 1998.

[21] Di Matteo Mantegna R.N Tumminello, Aste T. A new tool for filtering information in complex systems. 2005.

[22] Barthelemy M Pastor-Satorras R Vespignani A., Barrat A. The architecture of complex weighted network. *PNAS*, 101, 2003.

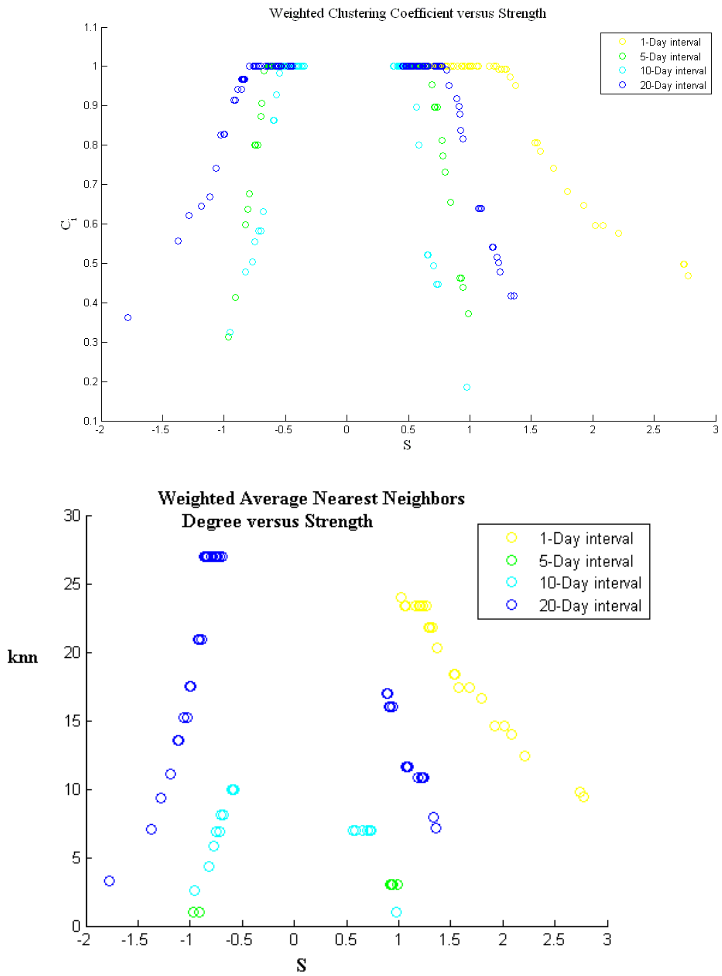
6 APPENDIX

Figure 23: Eigenvalue distribution of New York Stock Exchange Market and fitting by Random Matrix Theory. [3] Inset: Including the highest eigenvalue corresponding to the market



Density of the eigenvalues of C, where the correlation matrix C is extracted from N 406 assets of the S&P 500 during the years 1991–1996. In the figure a better fit can be seen in the interval of (0,1.74] (solid line) This result increases the importance of the discrete eigenvalues.

Figure 24: Weighted Average Nearest Neighbors Degree versus Strength and Weighted Clustering Coefficient versus Strength for reconstruction correlation network by only second largest eigenvalue.



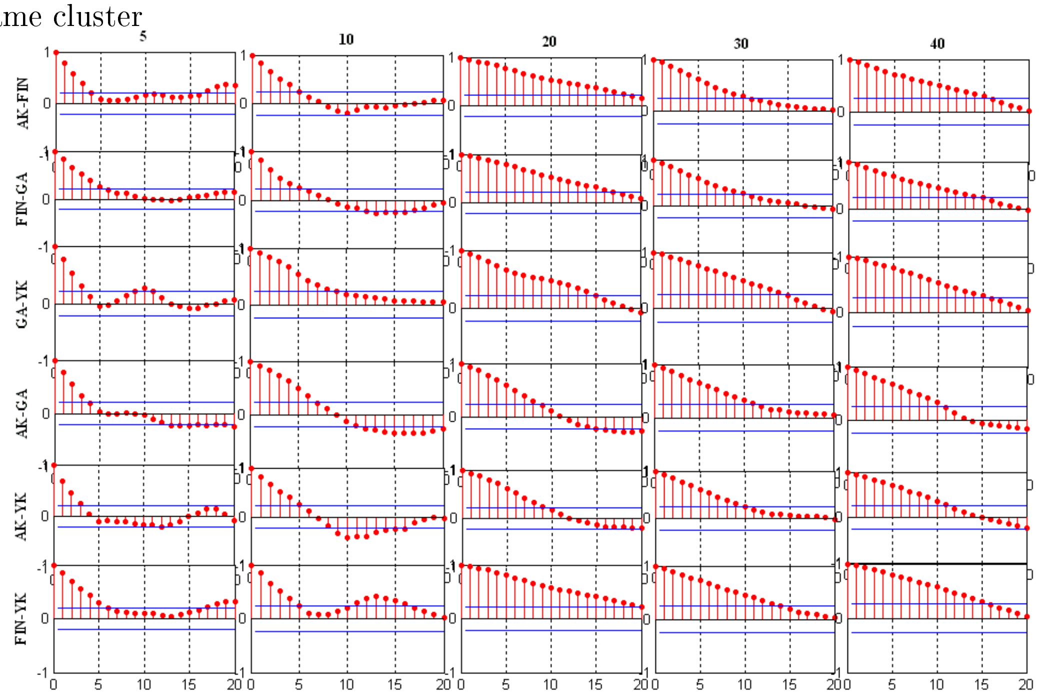
The figures were obtained by reconstruction network which constructed by only the second largest eigenvalue by the important result of Figure 23. k_{nn}&S_i and C_i&S_i plots have same patterns. We see that it has a pattern as largest eigenvalue and for negative strengths it has a symmetric pattern to the largest eigenvalue. Discrete eigenvalues are also mentioned as the future research topics identified by this study.

Figure 25: Sector Classification

Thesis Classificaiton	ISE Classification
Food and Agriculture	Food
Textile	Textile, leather
Energy	Oil, Chemistry
Retailing	Commercy
Construction and Cement	Rock, Stone, land
Consumer durables	Machine, material durables
	Raw material
Machine Industry	Electricity
Traveling and Tourism	Tourism
	Paper, Publication
Media and Communication	Communication
	Bank
	Insurance
	Leasing and Factoring
Finance	Investment
Groups	Groups and investments
Medicine	
Automotive	
	Spor
	Technology
	Information technology

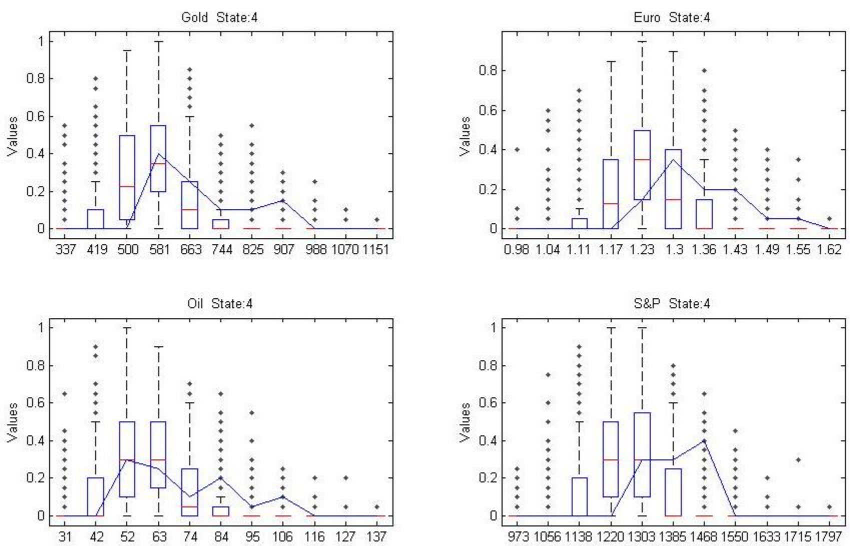
Sector classification used in the study is based on the 20 ISE sector-specific indexes. However these indexes are clustered into 13 groups due to the structure of the data used in the analysis.

Figure 26: Autocorrelations of moving correlation time series for assets from same cluster



The figure illustrates the autocorrelations of moving correlation time series for the 5, 10, 20, 30 and 40 day window-length for the assets from the same cluster. The analysis shows that the 20-window-length has the best memory is better than other window- lengths for this cluster as well as the analysis made for Gold, Euro, Oil, and S&P.

Figure 27: Boxplot of predicted prices by state 4



The bounds of the prices as illustrated in the graphs show that the predicted prices are sufficient.
The red line in the middle of boxes shows the median of the probability distribution function for the value which forms in horizontal axis, and the boxes cover the 25% of the data which correspond to value in horizontal axis. The black dash boundaries have the 75 % data. The black bold points correspond to the outliers.