COMENIUS UNIVERSITY, BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS



OPTIMAL PORTFOLIO RULES WITH INFORMATION FROM ASSET NETWORK

MASTER THESIS

2019 Bc. Anna KOLEGOVÁ

COMENIUS UNIVERSITY, BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

OPTIMAL PORTFOLIO RULES WITH INFORMATION FROM ASSET NETWORK

Využitie sietí pri hľadaní optimálneho portfólia

MASTER THESIS

Diplomová práca

Study programme:	Mathematical Economics, Finance and Modelling
Field of study:	1114 Applied Mathematics
Department:	Department of Applied Mathematics and Statistics
Supervisor:	Mgr. Katarína Lučivjanská, PhD.

Bratislava 2019 **Bc. Anna KOLEGOVÁ**





Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Bc. Anna Kolegová
Study programme:	Mathematical Economics, Finance and Modelling (Single degree study, master II. deg., full time form)
Field of Study:	Applied Mathematics
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Optimal portfolio rules with information from asset network

Annotation: Networks mapping relationship among assets in which we want to invest (e.g. shares) can provide useful information about their mutual dependence. We can model the strength of this dependence between two assets by correlation (a weight of the edge in network). We know several types of centrality in the network theory that help us understand which shares have the biggest effect on the portfolio performance. The aim of the thesis is to study how we can use this information to find the optimal portfolio. We will mainly focus on optimization of portfolios with bigger number of assets where we might face high parameter uncertainty. Therefore, it is suitable to look just on certain subset of assets. Network modeling will help us to identify this subset and corresponding weights.

Supervisor:	Mgr. Katarína Lučivjanská, PhD.
Department:	FMFI.KAMŠ - Department of Applied Mathematics and Statistics
Head of department:	prof. RNDr. Daniel Ševčovič, DrSc.

Assigned: 31.01.2018

Approved: 01.02.2018

prof. RNDr. Daniel Ševčovič, DrSc.
Guarantor of Study Programme

.....
Student

Lučivjanská
.....
Supervisor

Poďakovanie By these few words, I would like to thank my supervisor Katka Lučivjanská for the patience and all advices she gave me during creation of this thesis. For answers on mine (sometimes silly) questions and for not being mad at me when I rescheduled a call. I would also like to thank my friends for pushing me to finish this and enjoy the summer.

Abstract

KOLEGOVÁ, Anna: Optimal portfolio rules with information from asset network [Master Thesis], Comenius University, Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; supervisor: Mgr. Katarína Lučivjanská, PhD., Bratislava, 2019, 55 p.

Aim of this thesis is to clarify connection between network theory and portfolio strategies. At the beginning we introduce what the network is, it's basic features and centrality measures. We examine various ways to reduce number of edges in the adjacency matrix to eliminate the noise. Next step is to find out whether the network has informative value even in world of assets and portfolios. We do that based on the comparison of centrality measures and with construction of linear regression model. Here we try to proove that value of eigenvector centrality depends on market events and nevertheless from financial parameters obtained for each asset. The last part of our work is an estimation of optimal weights devided from adjacency matri which is base of the network.

Abstrakt

KOLEGOVÁ, Anna: Využitie sietí pri hľadaní optimálneho portfólia [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: Mgr. Katarína Lučivjanská, PhD., Bratislava, 2019, 55 s.

Cieľom tejto práce je objasniť teóriu sietí a jej prepojenie na portfólio stratégie. Najskôr si predstavíme čo to sieť je, aké sú jej základné vlastnosti a metriky. Preskúmame akými spôsobmi sa dajú redukovať matice susedností, aby sme odstránili neželaný šum. Ďalším krokom bude zistiť, či má sieť výpovednú hodnotu aj vo svete aktív a portfólií. To spravíme na základe porovnania centralít a zostavením lineárnej regresie, kde sa budeme snažiť dokázať, že výška hodnoty centrality vlastných vektorov sa odvíja od diania na burze a teda od finančných faktorov prislúchajúcich daným aktívam. Poslednou časťou práce bude odhad optimálnych váh odvodených na základe matice susedností, ktorá je stavebným pilierom siete.

Kľúčové slová: sieť, portfólio, centrality sietí, váhy, optimalizácia

Contents

List of Figures **9**

List of Tables **10**

Introduction **11**

1 Networks **13**

1.1 What they are and why we are interested in them 13

1.2 Basic networks attributes and types 14

1.3 Network centralities 17

2 Basic Portfolio Strategies **22**

2.1 Markowitz 22

2.2 Portfolio strategies 23

2.3 Relationship between centralities and optimal portfolio weights 24

2.4 Eliminate number of edges in network 26

2.5 Portfolio characteristics 27

3 Centralities and Networks More Practical **28**

3.1 Industry portfolios 28

3.1.1 Dataset 28

3.1.2 Network and Centrality Measures 29

3.2 Country indices 33

3.2.1 Dataset 33

3.2.2 Network and Centrality Measures 33

3.3 S & P 500 34

3.3.1 Dataset 34

3.3.2 Dataset variables 35

3.3.3 Network and Centrality Measures 36

4 Linnear regression **38**

4.1 Industry Portfolios 38

4.1.1 Cross-sectional analysis 39

4.1.2 Dimensional time series analysis 41
4.2 S&P 500 43
4.2.1 Cross-sectional analysis 44
4.2.2 Dimensional time series analysis 47

5 Optimal weights in portfolio **49**
5.1 Industry Portfolio and Country Indices 49
5.2 S&P 500 50

Conclusion **52**

Bibliography **54**

Appendix A **55**

List of Figures

1 Undirected and unweighted network 16
2 Directed network 16
3 Weighted network 16
4 Network of industry portfolios 29
5 Network of industry portfolios with reduced edges 30
6 Network of industry portfolios with reduced edges 31
7 Network of industry portfolios with reduced edges 32
8 Network of country indices with reduced edges 33
9 Network visualization for 100 assets 36
10 Coefficient β_0 from linear regression (26) 39
11 Coefficient β_1 from linear regression (26) 40
12 Coefficient β_2 from linear regression (26) 40
13 Coefficient β_0 from linear regression (27) 41
14 Coefficient β_1 from linear regression (27) 42
15 Coefficient β_2 from linear regression (27) 42
16 Coefficient β_3 from linear regression (27) 42
17 Coefficient β_4 from linear regression (27) 43
18 Coefficient β_0 from linear regression (28) 45
19 Coefficient β_1 from linear regression (28) 45
20 Coefficient β_2 from linear regression (28) 45
21 Coefficient β_3 from linear regression (28) 46
22 Coefficient β_4 from linear regression (28) 46
23 Portfolio Measures 51

List of Tables

1 Table of node centralities 21
2 Data Bases used in the chapters 28
3 Table of (not)important industries centralities in the industry portfolios . . 32
4 Table of centrality measures in country indices 34
5 Results of cross-sectional linear regression 40
6 Results of time series dimension linear regression 43
7 Results of cross-sectional linear regression 46
8 Results of time series dimension linear regression 48
9 Optimal weights in Country Indices dataset 49
10 Optimal weights in Industry Portfolio dataset 49
11 Optimal weight in S&P 500 dataset 50
1 Explanation of industry portfolio abbreviations 55
2 Explanation of MSCI Country Indices abbreviations 55
3 Explanation and abbreviation of variables in S&P 500 Dataset 55

Introduction

In this thesis we cover basic network theory and try to apply it in the finance sphere. We explain how to construct such a network, its' basic attributes and also the information hidden in the system. The key subjects of our analysis are network centrality measures and their relationship with the portfolio weights.

As we are talking about the investing and construction of ideal portfolios which is constantly changing environment, people are trying to find out the best solution for them. Because optimal weights and optimal model means that we can have, depending on our criteria, bigger profit or we can minimize our risk.

In the last few years, as the network theory was developing, group of mathematicians and economist tried to used this systems in portfolio weights. First there was a article [2] which came out back in 2007. DeMiguel compared naive portfolio with other different approaches to the portfolio strategy. Later, in 2016, Peralta [1] continued in the research of relationship between network and portfolio assets. In the article he examined turnovers and Sharpe Ratio, came with the optimal weights divided from minimum- and mean-variance strategies and the most interesting part for us was an attempt to explain company characteristics via its eigenvector centrality in the system. The last article [3] was published just last year, in 2018 here in Slovakia. They tried to examine reduction of network edges in the adjacency matrix and use the network systems in the evaluation of different portfolio strategies.

In this thesis we want to continue in the previous research and try to find optimal weights for the portfolio construction connected with finding a best way how to reduce the correlation matrix. Apart from that, we want to check whether we will find any more additional information in the system like using centralities of surrounding nodes in the network to explain asset position.

In the first chapter we briefly introduce networks and their centralities. In the second chapter we will continue with the finance world, where we discuss basic portfolio strategies and how the network is related to this topic. Third chapter is devoted to the illustration and visualization of networks for three different datasets. Next, fourth chapter, studies how the network change in the time and on the other hand, what remains stable and what effect it has on the cebtrality measures. The last chapter tells about optimal weights and

the resulting portfolio performance.

1 Networks

1.1 What they are and why we are interested in them

When you hear word 'network' in these modern times first thing that comes to your mind is internet. Computers connected with lines usually layout in specific shape. If you transform this very concrete image to its simpler version you can talk about dots and lines. But it is not just some piece of modern art - position and size of this objects have a meaning. In network theory we call these "dots" **vertices** or **nodes** and lines joining them **edges**.

Every system with objects and existing relationships could be transformed into network - objects will become nodes and relationships will be labeled by simple line pairing of two objects. Visualization of the network is very helpful in most of the cases and we can have better idea about the system then from looking at the numbers. Package `igraph` in R provide very useful tools for network graphs. You are able to make more important object larger and also strong relationships can be marked with wider lines.

Imagine that you are starting a new job after school. Whole work environment can be transformed to a network which you can see on the table in the building lobby. Before you will enter office premises there is an opportunity to see connections between employees and who is a strong member of this society. Such a simple thing can provide you a lot of useful information. You can find out members of specific teams; friends within the coworkers; how much will consequences of a problem hit specific chain of people; hierarchy of the firm; who is important in decision making and who are big players in the game.

In previous paragraph we have described example of social network. These type of networks are best for truly understanding how they are created and how the relationships work in this kind of environment. When you get the basic idea of networks its easy to apply them also in other fields once you have data suitable for this type of analysis. Many times its helpful to describe relationships between objects and take a look at them from other perspective.

This thesis are going to mainly focus on economical and financial, not social networks. In the world of finance, optimal distribution of weights in portfolio investments is very crucial. Exactly this is the place where we could use advantages of network. Nodes in

this system are assets included in portfolio and weights of edges are calculated from their yields. When the companies are from the same, subsequent or competitive sectors we expect stronger relationships, on the other hand when companies have absolutely nothing in common connection should be weak. Subject matter is whether we can ignore this weak connections when constructing a network. Another thing we can observe in the network is the importance of nodes. This gives us information about size of the damage in case of one of the companies bankrupt. On the one hand, we can say that investing huge part of our capital in the very central company can lead us to money loss. On the other hand, bankrupt of this company will also affect companies which are strongly connected to this strong individual.

To summarize previous paragraphs our portfolio assets can be transformed into the network. Information included in the nwly created system can be used to calculate weights for out investments and set the strategy. When it comes to aesthetic we need to keep in mind that big amount of data usually means a messy picture with no information value.

1.2 Basic networks attributes and types

We need to specify several concepts when we want to properly apply network theory to our problem. We have defined nodes and edges, but we have not defined the way how to mark links between the nodes. For this we use **adjacency matrix** which structure differ based on network type we are using. However, one thing they have in common is that all elements of the matrix are from the interval $[-1, 1]$. Notation of adjacency matrix is Ω and elements of the matrix are Ω_{ij} (We will mark elements of adjacency matrix with big Ω to avoid confusion with the correlation matrix). Size of the matrix is $n \times n$, where n is number of all nodes in the network.

Another attribute of the network is the weight of the nodes. We can differentiate two basic types - **un-weighted** and **weighted** network. Difference between these two types is that in the first case all connections inside the network are equally important whereas in the second scenario all edges of specific node have various strength of relationship remaining nodes. Adjacency matrix will defer in the elements from which it consists. When we talk about un-weighted network, matrix is build from set of 0 and 1, while in weighted networks strengths are noted by numbers from interval $[0, 1]$. For both cases we

can observe (Fig. 1 and Fig. 3) that $\Omega_{ij} = \Omega_{ji}$ so the matrix Ω will be symmetric $\Omega = \Omega^T$.

For better imagination what an (un)weighted network displays we explain it on social enviroment. Imagine previously mentioned situation when you are starting a new job. If you just wonder who knows who and you do not care about how well they know each other in the firm then we are talking about unweighted network. Once you start to be interested in the strength of the relationships between employees this can be portrayed in the weighted network. In the business language we either map cognate companies or we also map how much they affect each other. More described exmaples can be found in [7].

Last attribute of the network we discuss in this thesis is whether the system is **directed** or **undirected**. If a network is directed it means that nodes are behaving as magnets and edges are oriented, ending with an arrow on one of the sides. We can say that the relationship is the type superior-subordinate while in the undirected network it is the equal type. Parallel in the company hierarchy are boss and employee versus co-workers and in business it could be producer and supplier versus two cooperating companies.

Information about direction of connection could be found in the adjacency matrix. Edge goes from one node i to another one j which implies that Ω need to include valuation of relationship twice - on position Ω_{ij} and also Ωji. To indicate who is superior and who is subordinate we use positive values for superior and negative for subordinate. When we include also information about existence of weighted network, we can claim that mathematically $\Omega_{ij} \in [-1, 1] \forall i, j = 1, \ldots, n$. One of the triangles (upper or lower) of adjacency matrix contains negative value ($\Omega_{ij} < 0$) and the other one positive ($\Omega_{ji} > 0$) but both will have the same absolute value ($|\Omega_{ij}| = |\Omega_{ji}|$) so we can keep also information about strength of the edge.

In graph theory we can then define network as pair of sets $G = \{N, \Omega\}$, where $N = 1, 2, ..., n$ is a set of nodes and Ω is adjacency matrix. For further facts reader can see [9].

In the following paragraph we can see various examples of network together with the adjacency matrices:

$$G_1 = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

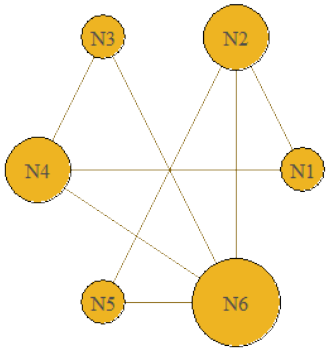


Figure 1: Undirected and unweighted network

Example of adjacency matrix and graph of unweighted and undirected network with 6 nodes. Size of the nodes is proportional to their degree centralities.

$$G_2 = \begin{pmatrix} 0 & 1 & 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & -1 & -1 & -1 & -1 & 0 \end{pmatrix}$$

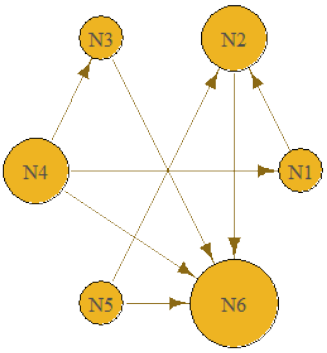


Figure 2: Directed network

Example of adjacency matrix and graph of directed network with 6 nodes where direction between nodes is labeled with arrow.

$$G_3 = \begin{pmatrix} 0 & 0.2 & 0 & 0.4 & 0 & 0 \\ 0.2 & 0 & 0 & 0 & 0.3 & 0.8 \\ 0 & 0 & 0 & 0.1 & 0 & 0.4 \\ 0.4 & 0 & 0.1 & 0 & 0 & 0.2 \\ 0 & 0.3 & 0 & 0 & 0 & 0.9 \\ 0 & 0.8 & 0.4 & 0.2 & 0.9 & 0 \end{pmatrix}$$

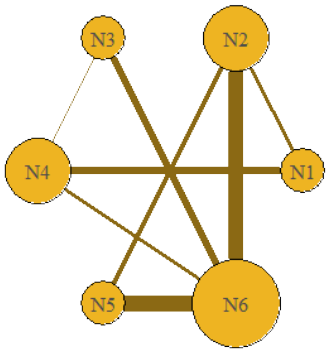


Figure 3: Weighted network

Example of adjacency matrix and graph of weighted network with 6 nodes. Width of edges is proportional to strength of connections.

1.3 Network centralities

In section 1.1 we have described what the network is, how and where we can use it. We talked about the specific nodes and their emplacement and importance inside the network. But we have not specified how to measure this.

Centrality measures are very advantageous tools when we study networks. Importance of the node is not easy to define. That is why we have measures which help us to gauge it and take a look at the importance of the node or company from different angles. The four most relevant aspects according to [4] in network theory we need to focus on are:

- **degree** - number of direct connections which the specific node have,
- **closeness** - position of node in the whole network,
- **betweenness** - importance of node as an intermediary point,
- **influence** - how many good connections a node have.

Degree centrality

This is the one most natural and simple type of centrality. In society general unspoken rule applies - the more people you know, the more possibilites appear. Whether it is a job opportunity or introduction of a new interesting hobby. In network theory these connections could be measured with degree centrality. It is the sum of all edges with other nodes of the specific node. Formula for standardized degree centrality is

$$C_{degree}(i) = \frac{\sum_{j \neq i}^{n} p_{ij}.e_{ij}}{n-1}, \tag{1}$$

where p_{ij} is edge between two nodes i and j and e_{ij} is existence of this edge. More specifically $e_{ij} \in \{0, 1\}$, where 0 means that edge between i and j does not exist and opposite to this is noted by the number 1. When we want to have $C_{degree}(i) \in (0, 1)$ we need to standardize the sum of links by the number of all existing nodes in the network. This way the degree centrality will express the ratio between the actual existing edges with nodes and the maximum it could possibly have.

Degree centrality is not the best way of finding out the importance of the node inside the system. In this type we do not know where exactly in the network the node stands and

we can assume that vertex which is more in the center of the system is more important than the one which stands at the surrounding. In conclusion this type of centrality will tell us just something about close surrounding of the node but nothing about the fact how much can node influence others.

Closeness centrality

As the name of the centrality says we want to know what is the position od the node in the system - how close it is to others. When we talk about parallel in social life we can explain this centrality as some human mediator which can introduce us to each orher. A person is more "central" when he can introduce friend to another friend straightly, but he is less "central" when he want to introduce friend to friends' sister which he does not know. In the language of networks centrality is higher when the node has a lot of straight connections so if we want to connect vertex j and vertex k all we need to do is join them via node i.

Standardized closeness centrality is proportion of sum of lengths of all shortest paths to other nodes it can achieve (which is obviously $n-1$) to the sum of actual lengths of this paths. Formula looks ak follows

$$C_{closeness}(i) = \frac{n-1}{\sum_{j \neq i}^{n} l_{ij}}, \tag{2}$$

where l_{ij} states for the longest path between nodes i and j. This centrality scales together with the distance which means that once I will move from the node twice as before, centrality will be twice lower. This type will tell us a little bit more than the degree centrality but it's not exactly what we want.

Decay centrality

This type of centrality is a combination of two previously mentioned. We introduce a new parameter δ which reflects how important is the node position comparing to others and also how much it can reach from indirect connections. It tells us which centrality - degree or closeness - is more important for specific node. When the δ is small, close to

zero, we consider node as marginal and the degree centrality (number of connections) is more important for it. Opposite, the big δ means that we should focus on the closeness centrality.

Formula for the decay centrality is

$$C_{decay}(i) = \frac{\sum_{j \neq i}^{n} \delta^{l_{ij}}}{(n-1)\delta}. \tag{3}$$

Betweenness centrality

In a group of friends we can usually indicate person who is mostly "in between". In many cases he is the one who put the crew together and who knows others the best. He is solving problems and starting a conversation between members of group. Applied to networks, node with high betweenness centrality is the one which connects most of the nodes in surrounding. So if we want to join nodes j and k we need to go through node i and it is our most efficient way. According to this importance of the node is then measured by all shortest paths which are crossing this node.

Formula for this type of centrality is

$$C_{betweenness}(i) = \frac{\sum_{i \neq j \neq k}^{n} \frac{P_i(j,k)}{P(j,k)}}{\frac{(n-1)(n-2)}{2}}, \tag{4}$$

where $P(j,k)$ denotes number of shortest paths between nodes j and k. When we add subscript i, $P_i(j,k)$ will pick just those paths of this set which cross the node i. If the node i appears in most of the routes the probability that we will have to pass it will be higher and the node becomes more central in the system.

To summarize previous statements we have looked at the three aspects of network. We know how to count quality of our relationships inside the community. We can compare us to others in three points - how many people do we directly know, how much we are in a centre of attention and how important we are in connecting people.

Eigenvector centrality

One can wonder if there is a way how to connect all these areas and measure some kind of general centrality. What should we do if we have two vertices with the same degree

centrality? First thought would be that centrality of node which has neighbours with degree centrality 6 and 7 could not be equally important as the node with centralities of neighbours equal to 2 and 3.

As we can predict we are not the first ones who thought about this problem (see [7]). To answer our question we will use eigenvector centrality. From previous observations we can try to count centrality of node as a linear combination of centralities of surrounding vertices. We will get equation

$$C_i = \sum_{j \neq i}^{n} \Omega_{ij} C_j, \tag{5}$$

where C_i is the centrality of our node of interest, C_j are centralities of other vertices, Ω_{ij} is an element of adjacency matrix of an unweighted and undirected network. It is equal to 1 when there is existing edge betweens nodes i and j and 0 if it does not. When we rewrite the equation (5) in the matrix notation we get

$$C' = \Omega C, \tag{6}$$

which lead us to the theory of eigenvalues and eigenvectors. We can rewrite C as a linear combination of eigenvectors $v_i i = 1, \ldots, n$ of adjacency matrix Ω

$$C = \sum_i a_i v_i, \tag{7}$$

where a_i are constants. If we put it back into (6) we get

$$C' = \Omega \sum_i a_i v_i = \sum_i a_i \lambda_i v_i = \lambda_1 \sum_i a_i \frac{\lambda_i}{\lambda_1} v_i, \tag{8}$$

where $\lambda_i i = 1, \ldots, n$ are eigenvalues of Ω while λ_1 is the biggest one of them. Therefore $\frac{\lambda_i}{\lambda_1} < 1 \ \forall i \neq 1$ which implies that most of the addends of sum limitly goes to 0. Considering this fact vector of centralities is proportional to the leading eigenvector λ_1. Thus

$$C' \to \lambda_1 a_1 v_1 \tag{9}$$

and we can rewrite (6) as

$$\lambda_1 C = \Omega C \tag{10}$$

The final formula for eigenvector centrality evaluation is

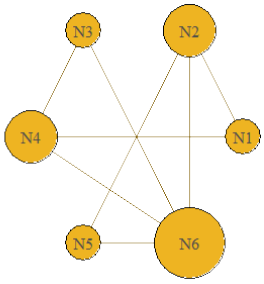
$$C_{eigenvector}(i) = \frac{1}{\lambda_1} \sum_{i \neq j}^{N} \Omega_{ij} C(j). \tag{11}$$

We can demonstrate centrality measures on the following table:

Table 1: Table of node centralities

	Centrality measures			
Node	Degree	Closeness	Betweenness	Eigenvector
N1	0.4	0.625	0.05	0.551
N2	0.6	0.714	0.15	0.776
N3	0.4	0.625	0.00	0.631
N4	0.6	0.714	0.15	0.776
N5	0.4	0.625	0.00	0.631
N6	0.8	0.833	0.35	1.000

Note: Centrality measure for undirected and unweighted network shown in Fig. 1.



In the Table 1 we see all four centrality measures of 6 nodes from undirected and unweighted network with adjacency matrix from Fig. 1. All centrality measures point out that node N6 is the most central one. Degree centrality is 0.8 so the node is connected to 80% of remaining vertices, more specifically 4 out of possible 5 in our network. Such a high number of connections implies also node's central position in the system. Vertices with lower degree centrality have also lower closenness centrality, because they are more far away from the system. This direct connections also affects shortest paths going through this node which has an impact on value of betweenness centrality. For example we can compare node N6 and N4 where N6 is a part of shortest paths between nodes N2-N3,N2-N4,N3-N5,N4-N5 while N4 is included only in paths N1-N3 and N1-N6. As we can see N6 has higher betweenness centrality than the node N4. Last type of measure, eigenvector centrality, supports the fact that node N6 is the best in the system. Interesting observation we want to point out is that the node N1 has lower eigenvector centrality than nodes N3 and N5 even though they have the same degree and closeness centrality. Reason for this that N1 is connected to two nodes with degree centrality 0.6 while N3 and N5 are connected to vertices with degree centrality 0.6 and 0.8.

2 Basic Portfolio Strategies

In this section we will talk about basic principles and theorem used in portfolio choice and investments. Every portfolio consists from group of assets. Such a set can be transformed into the network which can lead us to useful findings. In paragraphs below we deduce formulas for weights in the desired portfolio with selected strategy using centrality measures.

2.1 Markowitz

Probably everyone who had at some point of time interest in portfolio strategies has heard about the Markowitz model. We have various types of investors and every one of them has different goals when it comes to finding the most suitable portfolio. On the one hand, we want to maximize our expected return, but on the other one we want to minimize risk. When we mark expected return of whole portfolio as $\bar{r}_p$ and the risk of the portfolio as σ_p we get the following problems

$$\begin{aligned} \bar{r}_p &\longrightarrow max, \\ \sigma_p &\longrightarrow min. \end{aligned} \tag{12}$$

All information we have at the beginning returns of all n assets $r_i \ \forall i = 1, \ldots, n$ from which we estimate an expected return of an asset $\bar{r}_i$. But with estimation comes also probability of mistake that is expressed by standard deviation of return σ_i.

Two set goals from (12) could not be met at once and we get multiple solutions. To eliminate this, Markowitz focused on just one of thess problems which is minimization of portfolio risk. We can express the expected return of a portfolio as

$$\bar{r}_p = \sum_{i=1}^{n} w_i \bar{r}_i \tag{13}$$

and the standard deviation which represents risk of the portfolio as

$$\sigma_p^2 = E\big[\big(r_p - \bar{r}_p\big)^2\big] = \sum_{i,j} w_i w_j \sigma_{ij} \tag{14}$$

where parameter σ_{ij} is covariance between returns of assets i and j and w_i, w_j are weights of assets in portfolio.

If we fix the return $\bar{r}_p$ we get the information about investor's view of risk. If he is not afraid of loss and is more tolerant to risk for price of higher portfolio's returns we get bigger $\bar{r}_p$. In opposite case when investor want to feel safe and covered, his expected return is lower. After fixing $\bar{r}_p$ and using (13) and (14) we get an optimalization problem [8]:

$$\begin{aligned} \min \frac{1}{2} \sum_{i,j \neq 1}^{n} w_i w_j \sigma ij \\ \sum_{i=1}^{n} w_i \bar{r}_i = \bar{r}_p \\ \sum_{i=1}^{n} w_i = 1, \end{aligned} \tag{15}$$

Formula (15) is known as **Markowitz problem**. To find one solution you have two fulfill few more conditions:

1. Returns of assets $r_i \ \forall i = 1, ..., n$ are linearly independent - if one of the returns is created by linear combination of other returns it can be removed from the portfolio and replaced by others;
2. Covariance matrix $\Sigma = (\sigma_{ij})$ is regular and positive semi-definite - it is implied by the first condition;
3. $\exists i, j : 1 \leq i, j \leq n : \bar{r}_i \neq \bar{r}_j$ - if actives would have same expected return $\bar{r}$ then also the portfolio would have $\bar{r}_p = \bar{r}$ and (15) would not have solution for $\bar{r}_p \neq \bar{r}$.

2.2 Portfolio strategies

In the previous subsection 2.1 we have introduced Markowitz model and it's basic principles. Further we introduce two possible scenarios.

First one of them is a **minimum-variance strategy**. Problem is similar to the basic Markowitz (15) but we add one extra condition which is $w^T \mathbf{1} = 1$. In this model we completely ignore predictions of expected returns and work just with the covariance matrix. Our original model changes to

$$\begin{aligned} \min \sigma_p^2 = \min_w w^T \Sigma w \\ w^T \mathbf{1} = 1, \end{aligned} \tag{16}$$

where Σ stands for covariance matrix of returns. By solving problem (16) we get optimal weights for assets:

$$w^*_{minv} = \frac{1}{\mathbf{1^T}\Sigma^{-1}\mathbf{1}}\Sigma^{-1}\mathbf{1}. \tag{17}$$

Second scenario is **mean-variance strategy**. In this strategy we need to include also risk-free asset with returns r_f resulting in increasing number of assets from n to $n+1$. Excess return of asset i is $r^e_i = (r_i - r_f)$ and we denote vector of expected excess returns as $\bar{r}^e$. New model is

$$\begin{aligned} \min \sigma_p^2 &= \min_w w^T \Sigma w \\ w^T \bar{r}^e &= R^e, \end{aligned} \tag{18}$$

where R^e is a given level of excess return. Optimal weights for this strategy are

$$w^*_{meanv} = \frac{R^e}{\bar{r}^{e^T}\Sigma{-}1\bar{r}^e}\Sigma^{-1}\bar{r}^e \tag{19}$$

2.3 Relationship between centralities and optimal portfolio weights

In the chapter 1 we have introduced an adjacency matrix and centrality measures of network. In this chapter we want to connect network information to portfolio theory by using eigenvector centrality for estimating optimal weights.

If we define matrix Ω as correlation matrix of asset returns in our portfolio, standard deviation of returns as $\sigma_i \ \forall i = 1, 2, ..., n$ and $\Delta = diag(\sigma_i)$ then we can express connection between correlation and covariance matrix as

$$\Sigma = \Delta\Omega\Delta. \tag{20}$$

Using the equation (20) we can change formula for optimal portfolio weights of individual strategies (17) and (19) to

$$\begin{aligned} \hat{w}^*_{minv} &= \varphi_{minv}\Omega^{-1}\epsilon \\ \hat{w}^*_{i,minv} &= w^*_{i,minv}\sigma_i \\ \varphi_{minv} &= \frac{1}{\mathbf{1^T}\Sigma^{-1}\mathbf{1}} \\ \epsilon_i &= \frac{1}{\sigma_i} \end{aligned} \tag{21}$$

for minimum variance strategy and

$$\begin{aligned} \hat{w}^*_{meanv} &= \varphi_{meanv}\Omega^{-1}\hat{r}^e \\ \hat{w}^*_{i,meanv} &= w^*_{i,meanv}\sigma_i \\ \varphi_{meannv} &= \frac{R^e}{\bar{r}^{e^T}\Sigma^{-1}\bar{r}^e} \\ \hat{r}^e_i &= \frac{\bar{r}^e_i}{\sigma_i} \end{aligned} \tag{22}$$

for mean variance strategy.

Because of the fact that the correlation matrix express dependency of variables we can replace adjacency matrix in weighted portfolio network by it. We modify it a little and set the diagonal to zero to avoid self-loops. In the article [1] in the appendix we can find a proof that setting diagonal of adjacency matrix would not change the structure of eigenvectors neither the order of eigenvalues of this matrix. This proposition is important for evaluation of eigenvector centralities of the vertices.

Proposition 2.1. *[1] Consider a capital market network* $CM = \{N, \Omega\}$ *and* $\{v_1, v_2, ..., v_n\}$ *and* $\{\lambda_1, \lambda_2, ..., \lambda_n\}$ *the sets of eigenvectors and eigenvalues (in descendent order) of* Ω*, respectively. Then, the optimal portfolios weights (21) and (22) could be written as:*

$$\begin{aligned} \bar{w}^*_{minv} &= \varphi_{minv}\epsilon + \varphi_{minv}\left(\frac{1}{\lambda_1} - 1\right)\epsilon_M v_1 + \Gamma_{minv} \\ \bar{w}^*_{meanv} &= \varphi_{meanv}\hat{r}^e + \varphi_{meanv}\left(\frac{1}{\lambda_1} - 1\right)\hat{r}^e_M v_1 + \Gamma_{meanv} \end{aligned} \tag{23}$$

where

$$\begin{aligned} \Gamma_{minv} &= \varphi_{minv}\left[\sum_{k=2}^{n}\left(\frac{1}{\lambda_k} - 1\right)\right]\epsilon \\ \Gamma_{meanv} &= \varphi_{meanv}\left[\sum_{k=2}^{n}\left(\frac{1}{\lambda_k} - 1\right)\right]\hat{r}^e \ , \\ \epsilon_M &= v_1^T\epsilon \text{ and } \hat{r}^e_M = v_1^T\hat{r}^e \end{aligned}$$

Corollary 2.2. *[1] Assume that* $\lambda_1 > 1$ *and that* ϵ_M *and* $\hat{\mu}^e_M$ *are positive number. Then, a large eigenvector centrality of stock* i *is consistent with a low optimal portfolio's weight in both, minimum variance and mean variance strategy.*

Proofs for the proposition and corollary mentioned above could be find in the article

[1]. Using proposition 2.1 we can find a clear connection between network and portfolio theory, which is the main aim of our thesis.

Last thing we want to mention here is difference between individual and systematic performance of an asset. As a name suggests, individual performance provide an information of asset behaviour in the certain period of time independent on the other assets in the constructed portfolio. Systematic performance is looking at the whole portfolio when also relationships with other assets are included. That means that if we take a look at the portfolio as a whole we can identify which assets would be reach some in case of sudden decrease on the market. If we study this problem from network perspective we can predict that investor puts his capital into marginal assets that are not strongly connected to some other ones. This way the risk is diversified even though in optimal scenario we can invest into assets that has no correlation which is very unlikely to happen.

2.4 Eliminate number of edges in network

Considering the fact that we are replacing adjacency matrix with the correlation matrix, we need to think about the way to reduce number of connection between nodes. We have multiple options how to do this. In the following list is a brief description of the methods which we compare in the Chapter 3.

- **Threshold -** Based on the statistics of the correlation matrix we can set a value called threshold to eliminate weak connections.

- **Correlation significance test -** Is a statistic tool that help us to detect whether correlation between two assets is significant and has some informative value. We test the hypothesis on set seignifance level noted as α:

$$H_0\colon\ \rho_{ij} = 0 \text{ vs. } H_1\colon\ \rho_{ij} \neq 0$$

  where ρ_{ij} stands for correlation between assets i and j. If the correlation is too small we do not reject H_0 and the edge can be removed from the matrix.

- **Minimum spanning tree -** This method has origin in the network graph theory. It creates acyclical, connected subgraph of all assets where we measure distances

between nodes with $d_{ij} = \sqrt{1 - \rho_{ij}}$. The bigger the correlation is the smaller the distance is. In the end we solve an optimilazition problem which tries to minimalize sum of these distances.

2.5 Portfolio characteristics

To compute whether our constructed portfolio have good performance we can calculate following measures

- **Sharpe ratio -** Is a ratio between return of the portfolio compared to it's risk. We can tell whether the risk in the portfolio is properly distributed. Formula for calculation is as follows

$$SR_i = \frac{r_p^e}{\sigma_p},$$

where r_p^e is an excess return of the portfolio and σ_p is the calculated risk.

- **Turnover -** Is a measure which captures amount or size of the operations needed for rebalancing of the strategy. Formula is as follows

$$Turnover_i = \frac{1}{T - M - 1} \sum_{t=M}^{T-1} \sum_{j=1}^{N} \left| w_{j,t+1} - w_{j,t+} \right|,$$

where T and M are measures of the period windows, $w_{j,t+1}$ is weight asset j in the beginning of the period after recalibration of weigths and $w_{j,t}$ is weights of asset j before calibration.

3 Centralities and Networks More Practical

In following three chapters we work with three data sets:

Table 2: Data Bases used in the chapters

Data Bases						
			Periods			
Name	Description	Number of objects	From	To	Number of periods	Frequency
Country indices	Monthly returns of country indices	19	30/01/1970	30/12/2016	564	Monthly
Industry portfolios	Monthly average value weighted returns and ratio od Book and Market Equities (BE/ME) of industry portofolios	30	07/1926	11/2018	1109	Monthly
S&P 500	Daily returns, ask, bid, number of trades, volume, price of assets which has been part of the index calculation at certain period of time	2374	02/01/2004	30/10/2015	3021	Daily

As we can see in Table 2 above first dataset consists just from matrix of returns whereas the remaining two include also data from market. This data are used as parameters of linear regression in Chapter 4 where we try to explain value of centrality measures of nodes in networks created from datasets.

In this chapter we just introduce datasets and networks created from them and describe their centrality measures. As mentioned in Chapter 4 we focus on the evaluation of market parameters to study relationship with eigenvector centrality and in Chapter 5 we focus on demonstration of weights optimalization problem.

3.1 Industry portfolios

3.1.1 Dataset

This dataset is downloaded from well-known Fama&French webpage:
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
Data set gathers multiple tables of observations on 30 industry portfolios. In our empirical part we use two of them - 'Average Value Weighted Returns – Monthly' and 'Value-Weighted Average of BE/ME'. BE stands for Book Equity and ME for Market equity. When we look at the ratio between these two parameters we obtain a value of Book-

to-Market of an industry. This provides us a comparison between performance of the industry in accounting and financial sphere.

Dataset was last updated in November 2018 and oldest data are dated in 1926. Description of the abbreviations used in the data could be found in Appendix A in Table 1.

3.1.2 Network and Centrality Measures

Starting point in transformation of data to a network is setting an adjacency matrix. Returns are not only used to calculate weights and build a portfolio model, but we can also use them to create a adjacency matrix. Correlation of returns provide us an insight into the strength of relationships between assets - the bigger the correlation is the more powerful the connection is. Precisely that is why we use a correlation matrix of returns as an initial input.

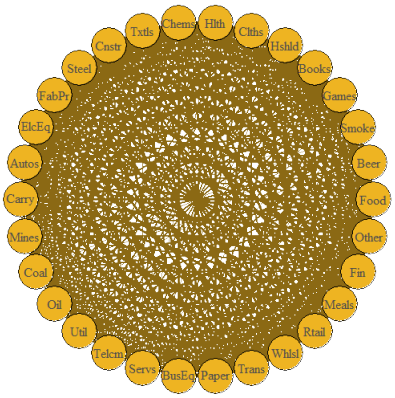


Figure 4: Network of industry portfolios
As an adjacency matrix we consider full correlation matrix of 30 industry portfolio's returns.

In Fig. 4 we can not see almost anything. The network is folded from all existing edges between nodes which means that nodes are connected between each other, only relationship that is missing is a selfloop. This results in the same degree centrality for each vertex. Figure is messy and we can not see widths of edges that can provide us information about the strength of the relationship.

Therefore, we rather prefer another approach to modelling the network of assets to reduce number of edges between vertices as now the network is not very informative. First,

we have tried to reduce edges based on basic correlation test, where we tested hypothesis whether correlation between nodes i and j is significant

$$H_0\colon\ \rho_{ij} = 0 \text{ vs. } H_1\colon\ \rho_{ij} \neq 0 \tag{24}$$

with simple Pearson's Method. We have removed those edges where we could not declined Hypothesis 0.

As we can see on Fig. 5 this step did not help us a lot and the network is similar to one showed in Fig. 4. Number of edges is still very high and the centrality measures obtained from such a network would barely give us any additional information to the prior analysis of returns. When we compared adjacency matrices we saw that they are equal and no edge has been removed.

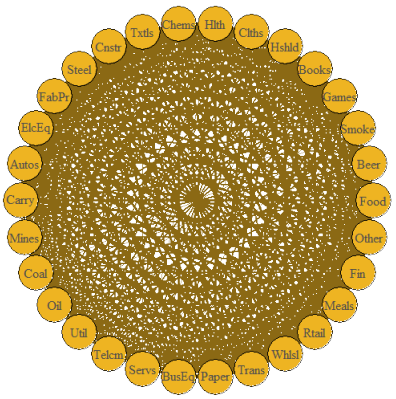


Figure 5: Network of industry portfolios with reduced edges
Edges in adjacency matrix are reduced with correlation test.

What is interesting to mention is that correlation matrix of returns was positive definite. All correlations between industries were positive where minimum was 0.29 and maximum was 0.88. When we want to diversify the risk ideal option is to invest to industries with correlation approximately equal to 0. Once there is some value greater than 0 we expect that if price of one industry portfolio is collapsing also the correlated one would follow it. As we have all industries correlated positively diversification of risk is really hard choice for the investor. But in general basic idea is to choose fields with the lowest correlation to protect the portfolio even a little so crash of one industry would not

cause massive degrade of whole portfolio value.

Our another approach based on the Chapter 2.4 was to set a threshold for the correlation value. Mean of all correlations was 0.65 so we guessed that critical value of 0.75 could remove uninformative edges that produce unwanted noise and keep just considerable connections. This step has brought some changes and the figure looked different.

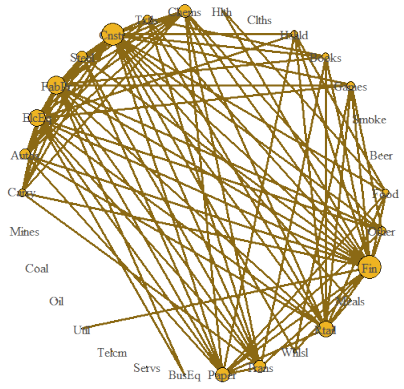


Figure 6: Network of industry portfolios with reduced edges

Edges in adjacency matrix are reduced with threshold of 0.75. Nodes size is proportional to their degreee centrality and the width of edge is multiple of edge weight.

The visualization of network in Fig. 6 supports our decision to use threshold as the 'edge-cuter'. We can identify industries with no strong relationships and on the other hand also those that are strongly connected to many others.

If we want to analyze position of individual nodes just by looking at the networsk visualisation we see that the most significant are industries of Construction and Finance, then the Fabricated Products, Electrical Equipment, Paper and Retail. Some of the industries did not have even one important connection with other industry as Mines, Coal, Oil, Telecommunication, Services, Beer, Tobacco Products and Clothes. This industry fields are the most independent according to network. All of the edges of the remaining vertices seems almost equally wide which means that connection with other industries are approximately the same.

Table 3: Table of (not)important industries centralities in the industry portfolios

Industry	Degree Cent	Betweenness Cent	Closeness Cent	Eigenvector Cent
Finance	0.655	0.133	0.1123	0.982
Construction	0.620	0.022	0.1118	1.000
Fabricated Products	0.517	0.012	0.1109	0.922
Electrical Equipment	0.483	0.044	0.1107	0.860
Retail	0.448	0.079	0.1105	0.737
Paper	0.414	0.020	0.1101	0.782
Chemicals	0.379	0.010	0.1098	0.770
Mines	0.000	0.000	0.033	0.033
Coal	0.000	0.000	0.033	0.033
Oil	0.000	0.000	0.033	0.033
Telecommunication	0.000	0.000	0.033	0.033
Services	0.000	0.000	0.033	0.033
Beer	0.000	0.000	0.033	0.033
Tobacco Products	0.000	0.000	0.033	0.033
Clothes	0.000	0.000	0.033	0.033

In the Table 3 centralities of the important and not important nodes in the system are reported. We can see that Finance industry has the biggest values except the eigenvector centrality which is a measure used to calculate optimal weights. This is caused by the fact that Construction industry has more convenient neigbors than the Finance insustry. In the second half of the table are nodes that has no edges. All of them have same centrality measures which are pretty small compared to the first half of the table. All of the industries with low centrality measures look like self-sustaining as they produce and sell their own products. The most central industries are either suppliers or they need a supplier to operate. Finance as the most central node depends on economical situation in the country which caused it's big depenedency on economy of remaining industries.

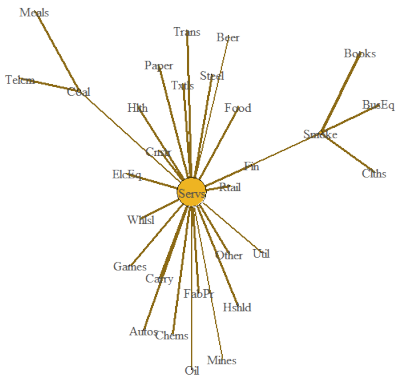


Figure 7: Network of industry portfolios with reduced edges

Edges in adjacency matrix are reduced Minimmum spanning tree method.

The last method we consider for edge-cutting is Minimum spanning tree. We plot the resulting network for this method in Fig. 7. The method seems to be quite restructive, as too many relationships have been ignored. Coal, Tobacco Products and Services are the only nodes with more than one edge. For future case studies we have chosen to use just threshold method.

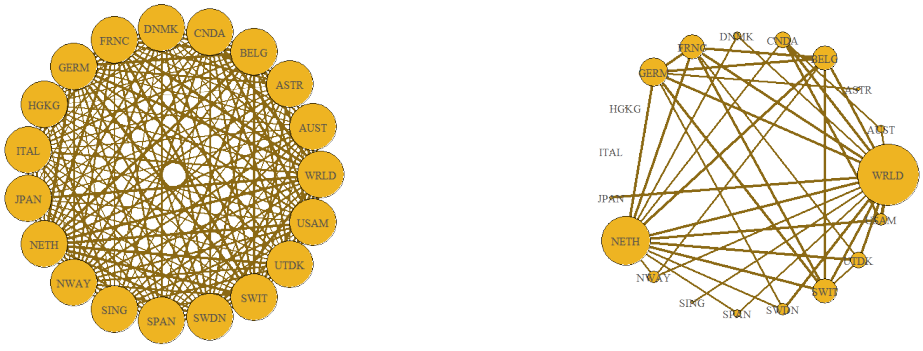
3.2 Country indices

3.2.1 Dataset

Dataset in this section consist from end-of-month returns of 18 countries + extra column for the world index. First date mentioned in the set is January 1970 and the last one is from December 2016. Together we have 564 observations which we can test. This dataset does not contains any information related to financial markets, only matrix of returns. Purpose of inclusion of this dataset is to confirm our decision to use threshold as an edge-cuter compared to simple correlation test.

Mean value of correlations was 0.55, minimum value was 0.29 and maximum value was 0.89. We have chosen threshold equal to 0.6 which is less strict that in the network of industry portfolios.

3.2.2 Network and Centrality Measures



(a) Edges reduced by correlation test **(b)** Edges reduced via threshold 0.6

Figure 8: Network of country indices with reduced edges

As we can see on the comparison showed in Fig. 8 also in this dataset cutting edges with threshold was more efficient than cutting it using the correlation test even though we set the threshold less strict. Most central node is a MSCI World Price Index which is pretty predictable as it is calculated from countries. Other significant country was Netherland which is also connected with most of the other countries.

Table 4: Table of centrality measures in country indices

Country	Degree Cent	Betweenness Cent	Closeness Cent	Eigenvector Cent
World	0.833	0.314	0.361	1.000
Netherlands	0.667	0.209	0.347	0.931
Germany	0.389	0.111	0.324	0.670
Switzerland	0.389	0.020	0.321	0.657
France	0.333	0.000	0.322	0.662
Belgium	0.333	0.026	0.322	0.640
United Kingdom	0.222	0.000	0.314	0.471
Canada	0.222	0.013	0.314	0.372
United States	0.222	0.000	0.305	0.381
Sweden	0.167	0.000	0.315	0.367
Norway	0.167	0.000	0.312	0.344
Spain	0.111	0.000	0.310	0.254
Denmark	0.111	0.000	0.308	0.256
Australia	0.111	0.000	0.307	0.191
Singapore	0.056	0.000	0.305	0.128
Japan	0.056	0.000	0.300	0.144
Austria	0.056	0.000	0.276	0.093
Italy	0.000	0.000	0.053	0.139
Hong Kon	0.000	0.000	0.053	0.139

3.3 S & P 500

3.3.1 Dataset

S&P 500 dataset consist of returns and market parameters od 2374 assets which were part of S&P 500 index in any time period between January 2004 and October 2015. For each asset we have 3021 daily observations which we proceed. Data contains from ID of the firm, date to which variables are bound, bid and ask price of the share, returns of the asset, number of trades which have been made during the day together with their price and volume of trades. In further analysis we construct also ratios based on them that might have predictive power with respect to our research question. Information about variables and abbreviations we used can be found in Appendix A in Table 3.

Before we start with analysis we have to modify raw data matrix and put it in the

form that suits us. Column `permno` stands for the ID of an asset which implies that matrix obtain many rows as we are watching such a long period of time (139 months) for so many assets (2374). Raw matrix was built from 5 617 623 rows but if we split it to sub-tables based on the asset ID this number decrease and matrices are more suitable for manipulation. To reach the desired working enviroment we have rather modified initial matrix as follows:

1. Split data into multiple matrices based on the asset ID so each matrix have information about all previously mentioned parameters just for one ID.
2. Create separate matrices for parameters where rows will represent dates and columns assets.
3. Exclude rows with missing (NA) values.

3.3.2 Dataset variables

Aim of this sub-chapter is to take a closer look on the variables we need for our empirical analysis - both initial and calculated. This can help us understand how we can further work with data and interpret the results.

List below consists of the brief characteristics of variables and their impact on our research question:

- **Asset ID** (`permno`) - ID used for identification of an asset;
- **Bid-Ask Spread** (`spread`) - calculated each day using formula

$$spread_{daily} = \frac{ask - bid}{\frac{ask+bid}{2}}$$

  is the ratio between what we are willing to pay and what we are willing to accept which could serve as a scale of transaction costs;
- **Number of trades** (`numtrd`) - daily number of trades which tells us how much was the asset attractive in that day.
- **Turnover** (`volpri`) - is calculated as

$$turnover_{daily} = volume * price$$

  on the daily basis so we have a better idea about volume of trades. Turnover provides

us more efficient information as simple volume of trades that can be the same for several days but price of a share can change and so does the turnover.

3.3.3 Network and Centrality Measures

Similar as in the previous two cases we constructs a network based on the correlation matrix of returns. In this particular scenario in the beginning we had 2374 assets which was hard to put in the Figure and see something there. That is why we demonstrate our theory on a small sample of these assets. We have randomly chosen 100 IDs which were in the dataset for the whole period of time so we can reach maximum information value. Reason for a random selection was to have various assets in the network so we can examine their centrality measures and their position in the network.

Minimum value of correlation in our extracted matrix was 0.04, maximum was 0.76 and mean value was 0.76. As we have 100 assets in the network we set a threshold pretty strict which of value 0.5 to get rid of the noise.

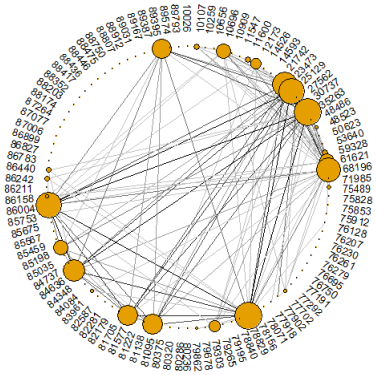


Figure 9: Network visualization for 100 assets

The resulting network with edges and centralities is visualized on the Fig. 9 above. We have modified size of the nodes so they are proportional to their eigenvector centrality and also color of the edge so it reflects weight. Previously we have used the width of the line as an identificator of strength of the relationship, nevertheless in this network we used colors instead so the results are more visible to the reader. Weak connections are in the gray color and the stronger they are the darker they become.

Constructed network seems to consist from few nodes which are very central among others and nodes which have almost zero centrality. We have identified that circa 10 nodes had big centrality measures. As we are working with data of assets these nodes represent companies which are strongly dependent on others. Unfortunately, we can not identify the industry or anything specific about them while we have in hand only anonymous assets' IDs. More darker edges signalized that more central companies have important connections.

When we look at the figure from investors point of view a lot of connections of the company is not good for us. Sudden downflow on a market of one company would affect also many others. This is what we call a systematic risk. To eliminate it investor can consider to distribute his capital into multiple industries. As we can not tell which assets belong to which industry we can only analyse information involved in the network. Being supposed that connected firms creates one industry we may invest in several so created groups. For example if we invest in asset 25129 than our next choice is asset 89395 as it they are not connected. Other approach to diversify the risk is investing in companies with low centrality as they have the lowest impact on the surrounfing notes. In the Chapter 5 we examine these theories with calculation of the optimal weights with help of eigenvector centrality obtained from the network. This centrality measure give us one of the best information about company and their connection within the portfolio.

4 Linnear regression

Our further aim in this chapter is to study determinants of the eigenvector centrality. As we mentioned earlier this measure is the most suitable to examine strong connections of the companies. From Table 2 in Chapter 3 we see that only two datasets contained also financial variables - 'Industry Portfolios' and 'S&P 500'. We construct linear regression between them and the value of the centrality. Based on the β-coefficients we are able to say which of them explain the value of measure and which do not.

Each analysis consists of two parts - **cross-sectional analysis** and **dimensional time series analysis**. Main purpose of cross-sectional analysis is to compare behaviours of the industry/asset to its pears. We set a time window based on the periodicity of the data and move it across whole time period. For wach window we construct a model and evaluate coefficients together with their confidence intervals, standard deviation and p-value. This way we obtain a matrix with behaviours of each industry/asset in a given period and we can study how the affect of individual parameters changed during the time.

Dimensional time series analysis is focusing on one asset/industry and its individual behaviour. We construct linear model based on sequence of numerical data points for each asset/industry and we can compare their perfomances. We track changes of the financial identificators and their affect on eigenvector centrality.

If we want to identify significance of β-coefficient we need to take a look on the p-value appertaining to the coefficient. We test the hypothesis

$$H_0\colon\ \beta_i = 0 \text{ vs. } H_1\colon\ \beta_i \neq 0 \tag{25}$$

on significance level $\alpha = 5\%$. Big p-value means that H_0 can not be rejected. Consequently, β_i is important in the regression and ommittance of the variable is not necessary. Conversely, if p-value is too small we can ommit this parameter from the regression as it is not important for explanation of the centrality value.

4.1 Industry Portfolios

In this dataset we work with matrices of returns and ratio between Book and Market Equity. Returns have monthly frequency whereas ratio BE/ME has annual frequency.

For each year we construct small submatrix of returns we calculate values of eigenvector centralities of each industry included in the network together with the standard deviation of returns. In the end we obtain data about performance of all 30 industries in all 93 periods.

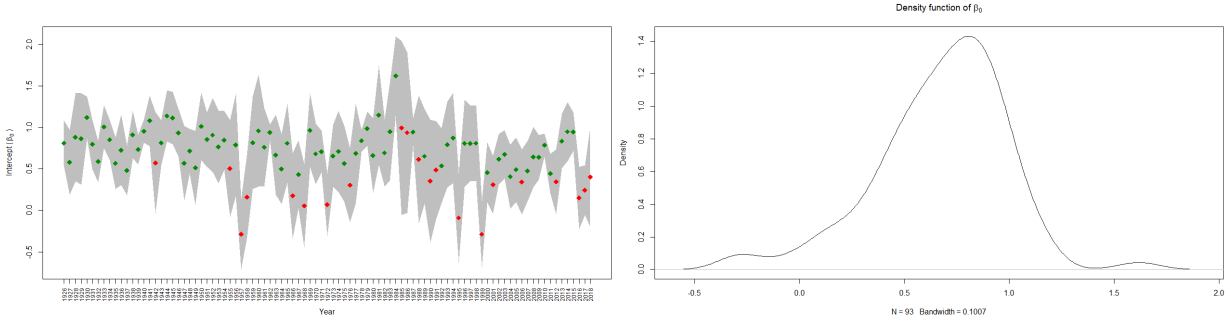
4.1.1 Cross-sectional analysis

To study individual perfomance of industries, for each year we contruct following linear model:

$$\begin{aligned} Eigen_centrality_i = &\beta_0 + \beta_1 Standard_deviation_of_returns + \\ &+ \beta_2 Average_value_weighted_BE/ME + \epsilon \\ &\forall i = 1, \ldots, 93. \end{aligned} \tag{26}$$

Coefficient β_0 is known as an intercept of the model. It tells about significance of the centrality - how much of it is self-explained without usage of other parameters entering the regreesion. Coefficients β_1 and β_2 explains whether value of centrality would be changed in case that either standard deviations of returns or ratio of BE/ME will increase or decrease.

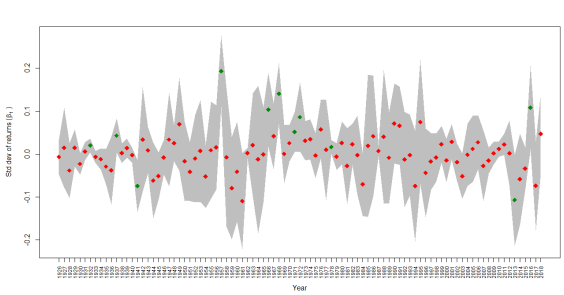
On the figures below we can see how individual coefficients changed during the time. Left side of the figures illustrate time structure of coefficient values together with their confidence intervals. On the right side we see density function of the regression coefficients. Green dots are identificators of signifficant betas while the red dots designate those who have not passed the test (25). Instead of table with individual results we provide Table 5 where is a summarization of regression and their parameters.

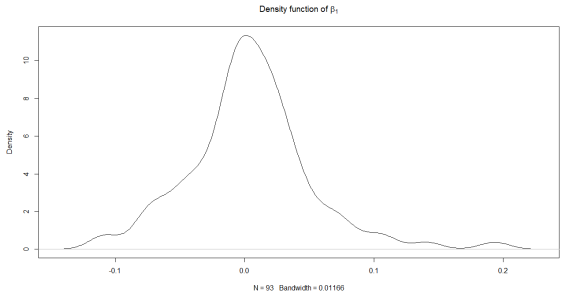


(a) Difference of coefficient values with confidence intervals

(b) Density function

Figure 10: Coefficient β_0 from linear regression (26)

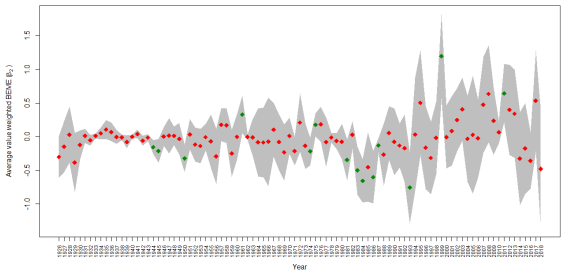


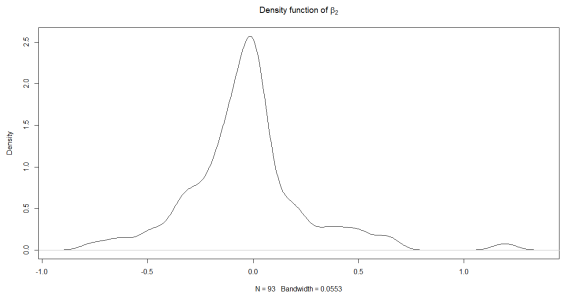


(a) Difference of coefficient values with confidence intervals

(b) Density function

Figure 11: Coefficient β_1 from linear regression (26)





(a) Difference of coefficient values with confidence intervals

(b) Density function

Figure 12: Coefficient β_2 from linear regression (26)

Table 5: Results of cross-sectional linear regression

		Coefficient			Standard Deviation			P-value			Significant (in %)		
Coeff	Parameter	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Yes	No	NA
						Independent variables							
β_0	Intercept	-0.288	1.619	0.673	0.047	0.512	0.202	0.000	0.825	0.064	77.42	22.58	0.00
β_1	Standard deviation of returns	-0.109	0.193	0.004	0.006	0.084	0.036	0.000	0.991	0.431	11.83	88.17	0.00
β_2	Average value weighted BE/ME	-0.760	1.195	-0.028	0.012	0.429	0.175	0.000	0.990	0.430	15.05	84.95	0.00

Summarization of linear regression results of cross-secitonal analysis for 'Industry Prtfolios' dataset.

Interesting fact we can notice in Table 5 is that even intercept coefficients of the regression were not significant in 100% cases. This means that if we remove the coefficient from regression and both standard deviation and portio BE/ME are zeros, eigenvector centrality have to be zero too. Possible reason for this is that returns in the observed period did not have strong connections to each other.

In most of the cases even financial variables were not significant and they had zero

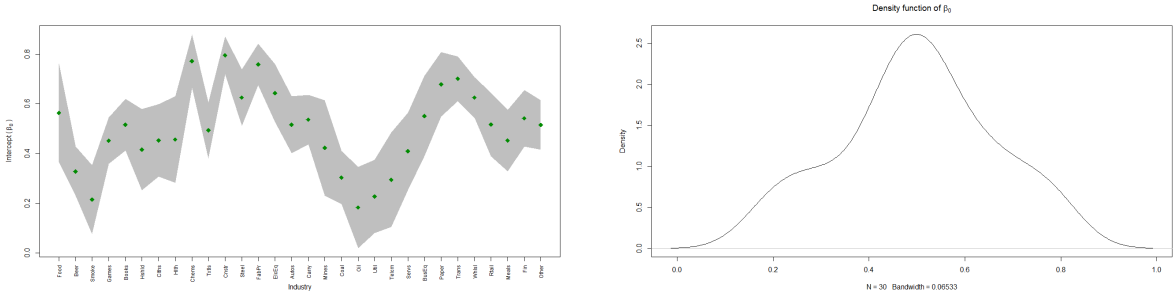
effect at the value of the measure. This can be predicted from the mean value of p-value which was approximately 0.43 in both cases.

4.1.2 Dimensional time series analysis

This time we took a look at the performance of industries separately. Our model has changed to

$$\begin{aligned} Eigen_centrality_i = &\beta_0 + \beta_1 Standard_deviation_of_returns + \\ &+ \beta_2 Average_value_weighted_BE/ME + \\ &+ \beta_3 Standard_deviation_of_returns_{centr} + \\ &+ \beta_4 Average_value_weighted_BE/ME_{centr} + \epsilon \\ &\forall i = 1, \ldots, 30. \end{aligned} \tag{27}$$

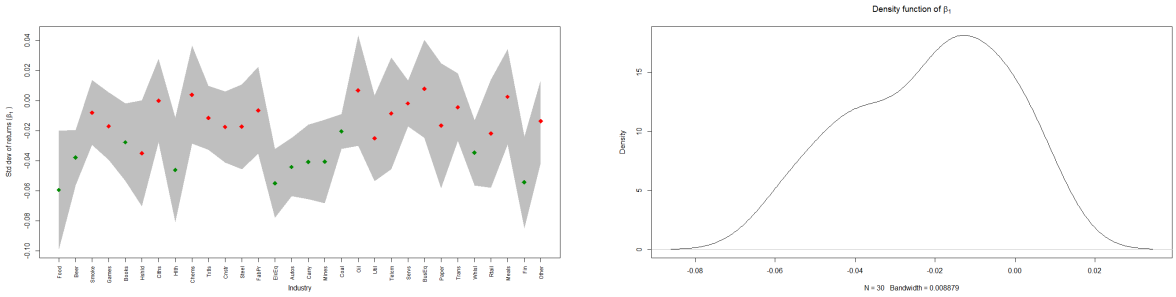
In this model we observed also variables with subscript 'central'. Reason for this is that we want to examine whether industries with strong connection can affect also value of eigenvector centrality. In each time window we then calculate eigenvector centrality of each industry, it's standard deviation of the returns and portio of BE/ME, plus we calculate also mean of standard deviations of returns and BE/ME of industries which have been strongly connected in that time period. We set value 0.75 as an identificator of strong connection.



(a) Difference of coefficient values with confidence intervals

(b) Density function

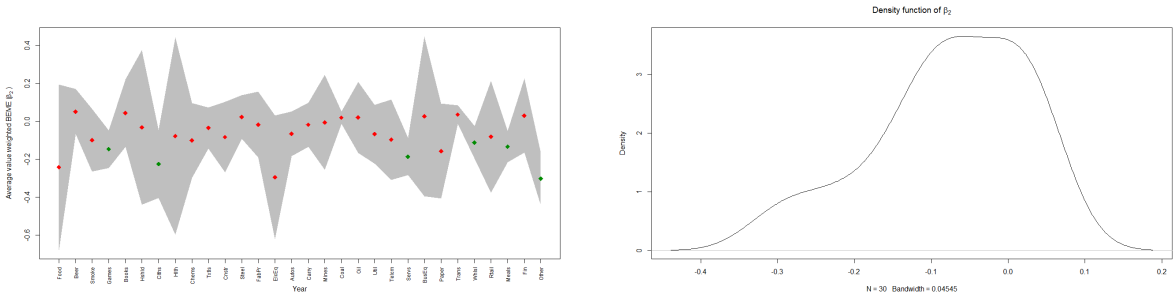
Figure 13: Coefficient β_0 from linear regression (27)



(a) Difference of coefficient values with confidence intervals

(b) Density function

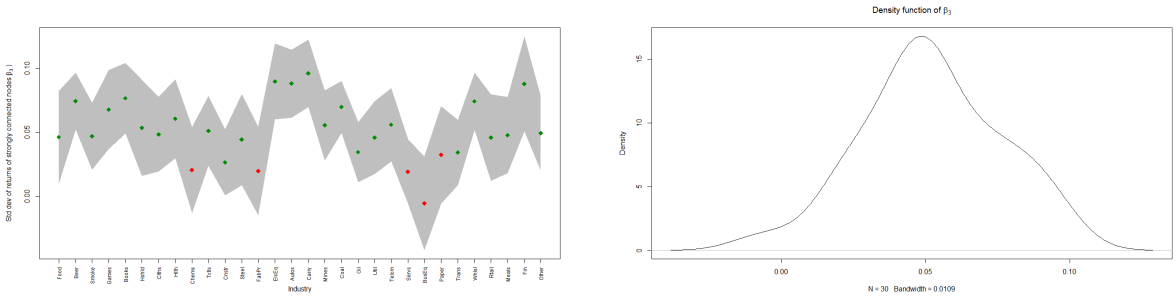
Figure 14: Coefficient β_1 from linear regression (27)



(a) Difference of coefficient values with confidence intervals

(b) Density function

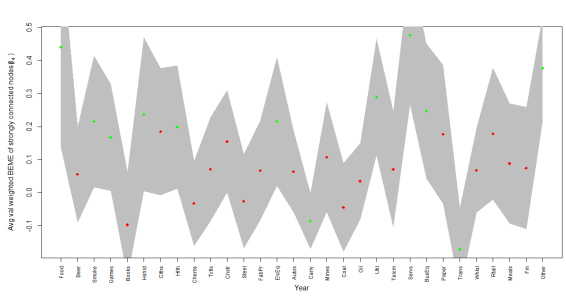
Figure 15: Coefficient β_2 from linear regression (27)



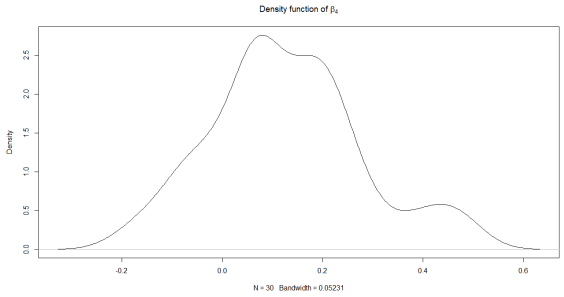
(a) Difference of coefficient values with confidence intervals

(b) Density function

Figure 16: Coefficient β_3 from linear regression (27)



(a) Difference of coefficient values with confidence intervals



(b) Density function

Figure 17: Coefficient β_4 from linear regression (27)

Table 6: Results of time series dimension linear regression

		Coefficient			Standard Deviation			P-value			Significant (in %)		
Coeff	Parameter	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Yes	No	NA
						Independent variables							
β_0	Intercept	0.183	0.796	0.499	0.038	0.099	0.064	0.000	0.029	0.001	100.00	0.00	0.00
β_1	Standard deviation of returns	-0.059	0.008	-0.022	0.006	0.021	0.014	0.000	1.000	0.309	36.67	63.33	0.00
β_2	Average value weighted BE/ME	-0.302	0.052	-0.077	0.017	0.262	0.099	0.000	0.969	0.415	20.00	80.00	0.00
						Centrality dependent variables							
β_3	Standard deviation of returns	-0.006	0.096	0.052	0.010	0.019	0.015	0.000	0.757	0.053	83.33	16.67	0.00
β_4	Average value weighted BE/ME	-0.171	0.476	0.127	0.043	0.154	0.085	0.000	0.716	0.213	40.00	60.00	0.00

Summarization of linear regression results of dimensional time series analysis for 'Industry Prtfolios' dataset.

In this type of analysis we did not observe any unsignificant intercept. Also ratio between unsignificant and significant coefficient β_1 and β_2 slightly improved. But what is the most important thing to notice is that parameters calculated from strongly connected industries had quite big impact of the value of centrality. In case of standard returns it was 83% of signficant coefficients and in case of BE/ME it was 40%. This example nicely show that created network was informatively relevant.

4.2 S&P 500

Dataset 'S&P 500' is quite larger compared to previous one. Not only we have dataset with higher frequency but we also have more market data that can be included in the regression. We split the data on month periods for which we calculated following values:

- **Eigenvector centrality** of asset in the network in the current month;

- **Standard deviation (volatility) of returns** is calculated as mean of squared returns;
- **Monthly number of trades** is represented by sum of daily trades;
- **Monthly spread** is mean of daily spreads;
- **Month turnover** is represented by sum of products of daily turnovers.

We need to note that unlike in the previous case we did not have the same number of objects entering the regression each period. We eliminated assets that started to be included in the index in the middle of the month period as we are not able to calculate correlation between vectors which contain 'NA' values.

4.2.1 Cross-sectional analysis

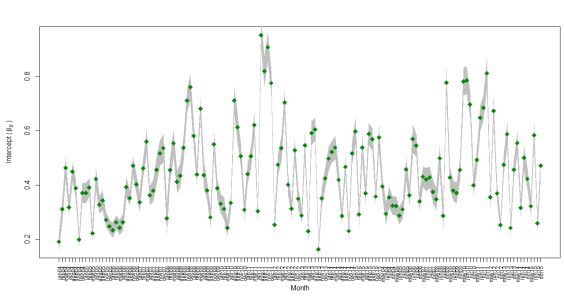
Similar as with the 'Industry Portfolios' dataset we wanted to take a look at the systematic performance of the assets. For each month we calculated performance of each asset that has been part of the index in that time and did the linear regression to look at the performance of assets in this period.

Eigenvector centrality is the y in the regression and rest of the financial variables mentioned above would be our x's. Main motivation of this investigation is to find out whether centrality of the asset in the constructed network somehow depends on the behaviour of the asset at the market. In other words we will try to explain small or high centrality of the asset by these factors - how many trades has been made during the month (number of trades), how much did the returns changed in the period (volatility), what are cost of the trades (spread) and what is the volume of sold shares (turnover). A founded connection between this observations would introduce us new look on the construction of optimal portfolios.

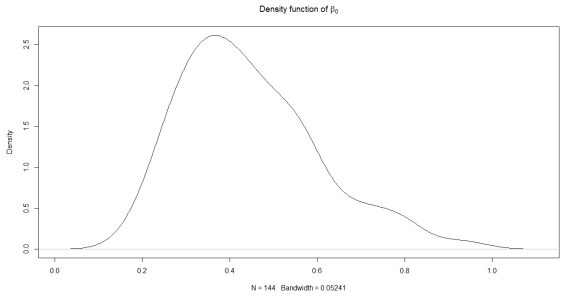
Model formula is

$$\begin{aligned} Eigen_centrality_i = &\beta_0 + \beta_1 Standard_deviation_of_returns + \\ &+ \beta_2 Number_of_trades + \beta_3 Spread + \beta_4 Turnover + \epsilon \\ &\forall i = 1, \ldots, 144. \end{aligned} \tag{28}$$

Together we had 144 months in the dataset. Values of β-coefficients and their confident intervals are pictured in the figures below on the left side and on the right side is again density function of the coefficient.

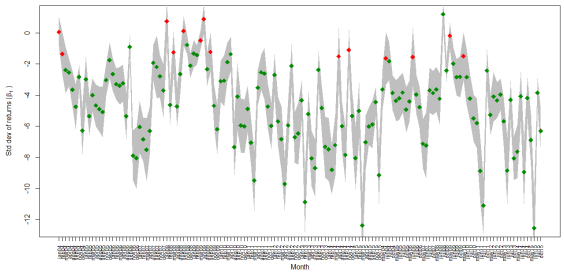


(a) Difference of coefficient values with confidence intervals

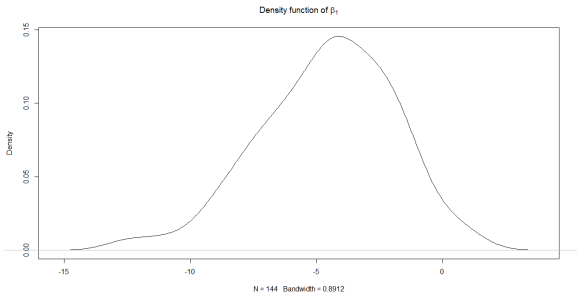


(b) Density function

Figure 18: Coefficient β_0 from linear regression (28)

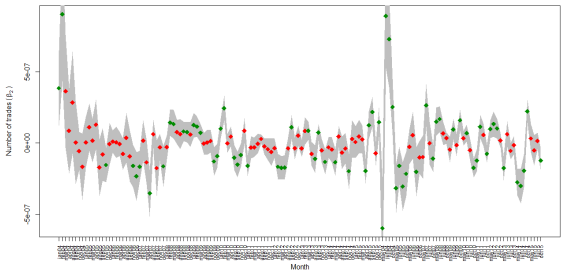


(a) Difference of coefficient values with confidence intervals

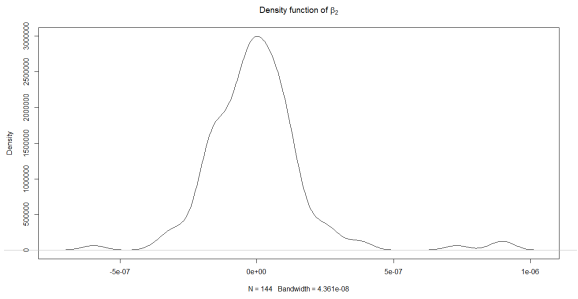


(b) Density function

Figure 19: Coefficient β_1 from linear regression (28)

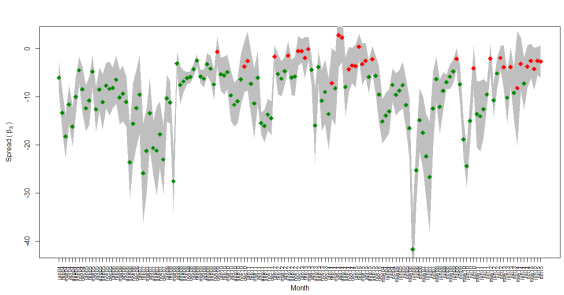


(a) Difference of coefficient values with confidence intervals

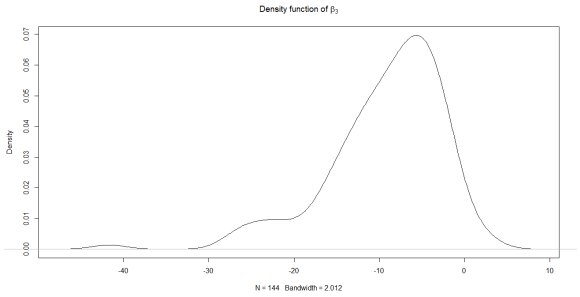


(b) Density function

Figure 20: Coefficient β_2 from linear regression (28)

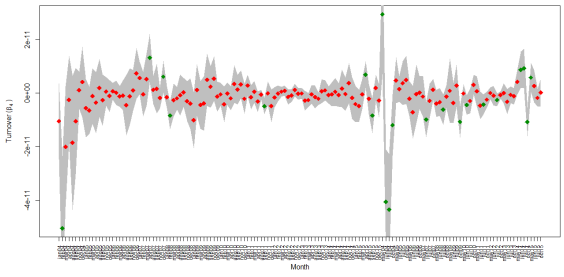


(a) Difference of coefficient values with confidence intervals

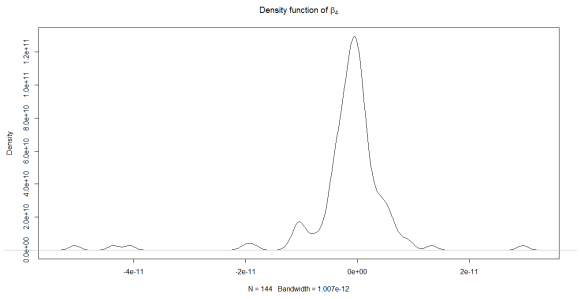


(b) Density function

Figure 21: Coefficient β_3 from linear regression (28)



(a) Difference of coefficient values with confidence intervals



(b) Density function

Figure 22: Coefficient β_4 from linear regression (28)

Table 7: Results of cross-sectional linear regression

		Coefficient			Standard Deviation			P-value			Significant (in %)		
Coeff	Parameter	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Yes	No	NA
						Independent variables							
β_0	Intercept	0.164	0.952	0.445	0.010	0.030	0.020	0.000	0.000	0.000	100.00	0.00	0.00
β_1	Standard deviation of returns	-12.565	1.203	-4.568	0.302	1.161	0.737	0.000	0.887	0.025	90.28	9.72	0.00
β_2	Number of trades	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.996	0.227	45.83	54.17	0.00
β_3	Spread	-41.667	2.808	-9.240	0.636	6.034	2.419	0.000	0.956	0.062	77.78	22.22	0.00
β_4	Turnover	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.993	0.439	14.58	85.42	0.00

Summarization of linear regression results of cross-sectional analysis for 'S & P 500' dataset.

Table 8 are summarized results of out cross-sectional linear regression for 'S&P 500' dataset. Intercept is significant in all time periods we compared. If we look at the parameters from market standard deviation of returns and spread are important in majority of cases, even number of trades is significant in almost 50% of periods. Only turnover seems to be a factor that does not have almost any impact on the eigenvector centrality value as

we can see from the percentange of signicance values and also from the mean od p-value. If we look at the figures above confidence intervals for this variable are also pretty wide.

Intervals of the coefficients for number of trades and turnover are very small and to see them in the table we would have to multiply them at least with 10^9. This is caused by the fact that centrality is from the interval $\left[0, 1\right]$ and spread from $\left[-0.014, 0.036\right]$. Coefficient interval of β_3 is then also really wide.

In the end we can say that centrality of the asset is explained by most of the market parameters in pretty noticable way. That means that position of the asset in the network is strongly connected to events on the market which we can include in our weights analysis in further approach.

4.2.2 Dimensional time series analysis

This analysis shares the basic principles with the analysis did on the previous dataset. The only difference is that we moved from industries to assets and we have more market data points. This model also includes mean values of market variables of strongly connected assets that are marked in the formula by subscript 'centr'. As a breaking point of strong relationship we chosen value 0.85.

$$\begin{aligned} Eigen_centrality = &\beta_0 + \beta_1 Standard_deviation_of_returns + \\ &+ \beta_2 Number_of_trades + \beta_3 Spread + \beta_4 Turnover + \\ &+ \beta_5 Standard_deviation_of_returns_{centr} + \beta_6 Number_of_trades_{centr} + \\ &+ \beta_7 Spread_{centr} + \beta_8 Turnover_{centr} + \epsilon \\ &\forall i = 1, \ldots, 2374. \end{aligned} \tag{29}$$

In this scenario we do not provide any figures as we have 2374 assets which is hard to visualize on the graph. This would not bring any informative value. Instead of that all results are summarized in Table 8. Table also includes subtable named 'Number of observations'. Here we can find minimum, maximum and average value of months when assets have been active and also number of assets that had under or above 50 observations. Assets with low number of observations had tendency to achieve bad model coefficients nad in most of the cases they were NAs.

Table 8: Results of time series dimension linear regression

		Coefficient			Standard Deviation			P-value			Significant (in %)		
Coeff	Parameter	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Yes	No	NA
					Independent variables								
β_0	Intercept	-0.594	1.153	0.308	0.005	1.034	0.081	0.000	0.969	0.051	84.84	15.08	0.08
β_1	Standard deviation of returns	-28.180	48.944	-1.862	0.110	50.477	3.399	0.000	1.000	0.353	22.37	77.55	0.08
β_2	Number of trades	0.000	0.000	0.000	0.000	0.001	0.000	0.000	1.000	0.369	8.55	33.70	57.75
β_3	Spread	-2602.8	2448.2	-42.84	0.156	1844.6	56.76	0.000	1.000	0.334	25.91	74.01	0.08
β_4	Turnover	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.282	32.94	66.98	0.08	
					Centrality dependent variables								
β_5	Standard deviation of returns	-100.76	49.96	6.928	0.753	66.48	3.538	0.000	0.996	0.156	61.33	38.58	0.08
β_6	Number of trades	-0.008	0.000	0.000	0.000	0.001	0.000	0.000	1.000	0.224	39.34	55.73	4.93
β_7	Spread	-867.71	3807.30	47.63	2.027	2264.5	55.55	0.000	0.998	0.331	25.91	73.93	0.17
β_8	Turnover	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.198	50.42	49.16	0.42

Number of observations				
Min	Max	Mean	≤ 50	>50
5	144	112	376	1998

Summarization of linear regression results of dimensional time series analysis for 'S& P 500' dataset.

Table **??** display summarization of results of dimensional time series analysis of our largest dataset. As we can notice analysis included also asset which had only 5 observations available which resulted in NA β-coefficients in the regression. That is why we 0.08% of assets had NA value in the intercept. Overall coefficient β_0 was significant in 85% of cases. If we look at the other independent variables, spread was the parameter which influenced value of the centrality the most. Number of trades were not significant in majority of cases caused by the fact that this parameter had a lot of missing values. More interesting part are variables dependent on centrality of strongly connected assets. Number of trades in this category improved it's results and became significant in 40% of cases. Standard deviation of results has improved as well while percentage of significant cases of coefficient β_7 remained the same. The only degradation could be seen in the turnover.

In the end parameter measures of strongly connected assets seems to explain centrality pretty well. It is a good sign for our decision to use networks as a tool to optain the optimal portfolio.

5 Optimal weights in portfolio

Last chapter of the thesis is studying optimal weights of min-variance strategy derived in Proposition 2.1. As we previously discovered network seems to be a useful tool for construction of optimal weights as the eigenvector centrality is able to include information from the market.

Main problem of the weights calculation was to set a threshold for the adjacency matrix of network. If we chosen poorly we ommited a lot of useful connections in the network which resulted in negative values of weights. To use choose a correct border between edge cutting and calculating weights we always took a look at the mean correlation value and then chosen something really close to it.

All of the weights placed in this chapter are evaluated according to formula (23).

5.1 Industry Portfolio and Country Indices

Table 9: Optimal weights in Country Indices dataset

Country	Eig Cen	Weight
WRLD	1	0.329
AUST	1	0.163
ASTR	1	0.121
BELG	1	0.204
CNDA	1	0.204
DNMK	1	0.204
FRNC	1	0.204
GERM	1	0.204
HGKG	1	0.124
ITAL	1	0.121
JPAN	1	0.123
NETH	1	0.287
NWAY	1	0.162
SING	1	0.122
SPAN	1	0.163
SWDN	1	0.204
SWIT	1	0.204
UTDK	1	0.288
USAM	1	0.120

Note: Optimal weights calculated for portfolio with threshold 0.4 used as an edge-cuter for adjacency matrix.

Table 10: Optimal weights in Industry Portfolio dataset

Industry	Eig Cen	Weight
Food	1	14.641
Beer	1	9.629
Smoke	1	65.973
Games	1	14.350
Books	1	15.308
Hshld	1	16.870
Clths	1	22.662
Hlth	1	18.971
Chems	1	29.175
Txtls	1	16.498
Cnstr	1	27.058
Steel	1	16.379
FabPr	1	27.445
ElcEq	1	28.587
Autos	1	16.250
Carry	1	14.604
Mines	1	42.300
Coal	1	44.546
Oil	1	17.314
Util	1	25.940
Telcm	1	9.972
Servs	1	84.277
BusEq	1	17.804
Paper	1	12.670
Trans	1	29.771
Whlsl	1	16.325
Rtail	1	13.427
Meals	1	17.710
Fin	1	11.343
Other	1	13.661

Note: Optimal weights calculated for portfolio with threshold 0.6 used as an edge-cuter for adjacency matrix.

As we can see in Tables 9 and 10 all of the nodes had eigenvector centrality equal to 1. This was caused by setting a threshold too low. On the other hand when we put it higher all weights changed from positive to negative values. Bad manipulation with small

datasets as 'Country Indices' and 'Industry Portfolios' led us to solution of not using these two datasets further.

5.2 S&P 500

This dataset provides eligible amount of observations for our study. First approach was to evaluate weights on the whole period of time to see whether exists a relationship between eigenvector centrality and these weights. For our purposes we again used randomly generated dataset of 100 assets previously used in Chapter 3 in Fig 9. Results are diplayed in the table below.

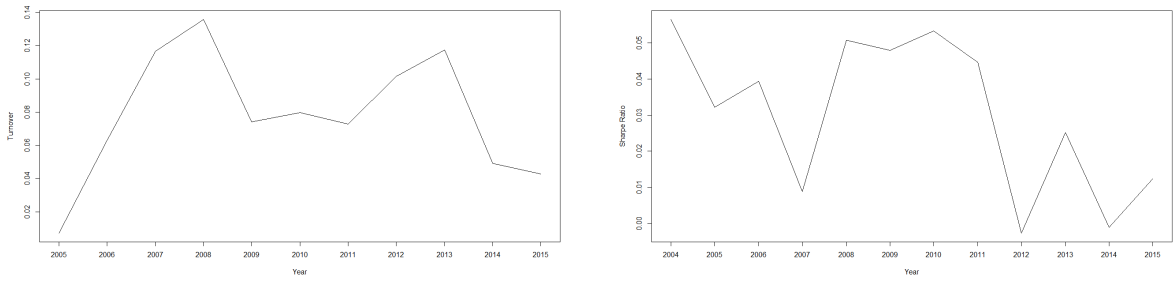
Table 11: Optimal weight in S&P 500 dataset

ID	Eig Cen	Weight	ID	Eig Cen	Weight	ID	Eig Cen	Weight	ID	Eig Cen	Weight
10026	0.777	0.340	75489	0.816	0.295	80236	0.740	0.224	86211	0.628	0.231
10107	0.801	0.379	75828	0.699	0.263	80286	0.727	0.289	86242	0.837	0.198
10259	0.522	0.173	75853	0.465	0.154	80320	0.689	0.353	86440	0.734	0.224
10656	0.686	0.218	75912	0.854	0.262	80375	0.735	0.166	86783	0.635	0.237
10696	0.915	0.404	76128	0.847	0.178	81095	0.931	0.359	86827	0.772	0.237
10909	0.678	0.321	76207	0.644	0.179	81138	0.340	0.225	86899	0.613	0.315
11547	0.760	0.314	76230	0.758	0.243	81222	0.698	0.256	87006	0.406	0.257
11600	0.858	0.424	76261	0.741	0.295	81577	0.893	0.307	87077	0.671	0.205
12173	0.854	0.217	76279	0.684	0.172	81705	0.809	0.262	87264	0.550	0.173
14526	0.843	0.217	76695	0.628	0.289	82179	0.689	0.347	88174	0.202	0.167
14593	0.683	0.289	76750	0.694	0.295	82281	0.263	0.128	88203	0.571	0.160
21742	0.721	0.308	77191	0.819	0.269	82587	0.670	0.172	88352	0.604	0.218
23473	0.974	0.359	77292	0.481	0.251	83961	0.878	0.230	88417	0.528	0.121
25129	0.973	0.378	77702	0.858	0.314	84084	0.794	0.243	88436	0.481	0.283
27562	0.814	0.288	77902	0.815	0.275	84348	0.812	0.269	88446	0.515	0.205
30737	0.747	0.250	77918	0.831	0.262	84636	0.896	0.243	88475	0.117	0.226
35263	0.998	0.268	78071	0.009	0.110	84737	0.742	0.185	88750	0.755	0.198
48486	0.821	0.249	78156	0.242	0.232	85035	0.726	0.172	88807	0.368	0.154
48523	0.815	0.424	78829	0.999	0.300	85198	0.784	0.301	88912	0.351	0.135
50623	0.805	0.282	78840	0.663	0.347	85459	0.821	0.333	89031	0.543	0.153
53640	0.822	0.243	79195	0.548	0.179	85567	0.608	0.173	89167	0.786	0.204
59328	0.871	0.346	79265	0.725	0.276	85675	0.741	0.321	89387	0.551	0.257
61621	0.904	0.443	79303	0.893	0.307	85753	0.767	0.275	89395	0.884	0.243
68196	1.000	0.300	79678	0.644	0.276	86004	0.965	0.229	89574	0.770	0.250
71985	0.630	0.166	79862	0.848	0.172	86158	0.847	0.262	89793	0.636	0.224

Note: Optimal weights calculated for portfolio of 100 randomly chosen assets with threshold 0.2 used as an edge-cuter for adjacency matrix.

We expect a negative correlation between weights and centrality measure. Reason of this idea is that we want to diversify the risk and invest to assets, which have week or none correlation with remaining assets in the portfolio. Asset with the lowest weight is 78071 which also happened to be an asset with the lowest centrality. Asset with the highest weight is 61621 that belongs to group of nodes with highest centralities. If we take a look at the correlation between the centrality and weights, it is equal to 0.53 which is absolute opposite of what we expected.

We took a next shot and try to recalibrate weights of the portfolio every year to see whether it improves our perfomance. To measure this we calculated turnover and Sharpe ratio of the portfolio.



(a) Turnover **(b)** Sharpe Ratio

Figure 23: Portfolio Measures

On the Fig.23b we see that Sharpe Ratio was pretty bad for our portfolio as it's maximum value was 5% at the beginning. Compared to this, transactions costs for calibration were too high.

To summarize we did not obtain an optimal weights for our portfolio. This could be caused by the wrong set up of threshold or wrong randomly chosen data.

Conclusion

Main object of study in our thesis were networks and their influence on the portfolio strategies. After the introduction of networks and their centrality measure we were able to apply them on set of assets to obtain optimal portfolio.

After investigation of various methods of edge cutting in the adjacency matrix we came to conclusion that threshold was the best solution for us. Correlation test did not exclude any relationships and if then really small amount. Method of Minimum spanning tree on the other hand ommited a lot connections. To avoid loosing the information or having too much information noise we dismiss both of the methods and in further analysis used just threshold.

Even though we chosen threshold as an edge-cuter it was not easy to pick the right breaking point between wanted and unwanted connections. During evaluation of optimal weights we found out that even a minor change to this value can significantly manipulate with values of weights in the portfolio.

We also compared various types of centrality measures on three datasets. Eigenvector centrality showed up to be the best choice as it is a combination of centralities of surrounding nodes. Precisely this is why we used this measure type in linear regression and optimal weights calculation.

Chapter about linear regression introduced two views on datasets: cross-sectional and dimensional time series analysis. First of them shows systematic performance of an asset - behaviours inside of the network, second an individual performance. In both we studied whether market variables have an impact on the centrality value. Analysis had an interesting twist when we found out that market variables of the individual nodes are not such an important as a mean of the variables of strongly connected assets.

Last of the chapters was less succesful. Here we tried to optimalize weights based on the information which network carry. Our expectation that assets with low centrality should have biggest weights to better diversify the risk was not really fullfiled, more to the contrary. Even if we tried to recalibrate weights every year of the time period performance of the portfolio was poor and Sharpe Ratio very low. Returns were too low compared to the risk.

To summarize - we showed that market events and centrality are connected. Un-

fortunatelly, we did not fulfill a desire to design optimal weights. For future purposes I recommend to study more choice of the edge cuting method together with the dependency of threshold to weights.

Bibliography

[1] Peralta, G., Zareei A., 2016. A Network Approach to Portfolio Selection. In *Journal of Empirical Finance*. ISSN 0927-5398, 2016, vol. 38, issue?, p. 157-180.

[2] DeMiguel, V., Gerlappi, L., Uppal, R., 2007. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?. In *The Review of Financial Studies*. ISSN ?, 2009, vol.22, i. 5, p. 1915-1953.

[3] Výrost, T., Lyócsa, Š., Baumöhl, E., 2018. Network-based asset allocation strategies. In *The North American Journal of Economics and Finance*. ISSN?, 2018, vol.?, i.?, p.?.

[4] Jackson, O. M.: *Social and Economic Networks: Models and Analysis*, online course, Stanford University, 2018, available online (1.10.2018): https://www.coursera.org/learn/social-economic-networks/home/welcome (exists also book Jackson, M.O., 2010. Social and Economic Networks, Princetion University Press,)

[5] Ognyanova, K.: *Static and dynamic network visualization with R*, online notes from workshop, Rutgers University, 2018, available online (7.1.2019): http://kateto.net/network-visualization

[6] Namaki, A., et al., 2011. Network analysis of a financial market based on genuine correlation and threshold method. In *Physica A: Statistical Mechanics and its Applications, 390(21-22), p. 3835-3841.*

[7] Newman, M.E.J. 2010. Networks: An Introduction. New York: Oxford University Press Inc., 2010. 772 s. ISBN 978-0-19-920665

[8] Melicherčík, I., Olšárová, L., Úradníček V. 2005. Vybrané kapitoly z finančnej matematiky. Bratislava: EPOS, 2005. 242 s. ISBN 80-8057-651-3

[9] Bollobas, B. 2001. Modern Graph Theory. Beijing: Science China Press, 2001. 394 s. ISBN 978-0-387-98488-9

Appendix A - Tables

Datasets

Table 1: Explanation of industry portfolio abbreviations

No.	Node	Meaning	No.	Node	Meaning
1	Food	Food Products	16	Carry	Aircraft, ships, and railroad equipment
2	Beer	Beer and Liquor	17	Mines	Precious Metals, Non-Metallic, and Industrial Metal Mining
3	Smoke	Tobacco Products	18	Coal	Coal
4	Games	Recreation	19	Oil	Petroleum and Natural Gas
5	Books	Printing and Publishing	20	Util	Utilities
6	Hshld	Consumer Goods	21	Telcm	Communication
7	Clths	Apparel	22	Servs	Personal and Business Services
8	Hlth	Healthcare, Medical Equipment, Pharmaceutical Products	23	BusEq	Business Equipment
9	Chems	Chemicals	24	Paper	Business Supplies and Shipping Containers
10	Txtls	Textiles	25	Trans	Transportation
11	Cnstr	Construction and Construction Materials	26	Whlsl	Wholesale
12	Steel	Steel Works Etc	27	Rtail	Retail
13	FabPr	Fabricated Products and Machinery	28	Meals	Restaraunts, Hotels, Motels
14	ElcEq	Electrical Equipment	29	Fin	Banking, Insurance, Real Estate, Trading
15	Autos	Automobiles and Trucks	30	Other	Everything Else

Table 2: Explanation of MSCI Country Indices abbreviations

Node	Dataset	Index Identification	Country	Node	Dataset	Index Identification	Country
WRLD	MSWRLD$	MSCI World U$ - Price Index	World	JPAN	MSJPAN$	MSCI Japan U$ - Price Index	Japan
AUST	MSAUST$	MSCI Australia U$ - Price Index	Australia	NETH	MSNETH$	MSCI Netherlands U$ - Price Index	Netherlands
ASTR	MSASTR$	MSCI Austria U$ - Price Index	Austria	NWAY	MSNWAY$	MSCI Norway U$ - Price Index	Norway
BELG	MSBELG$	MSCI Belgium U$ - Price Index	Belgium	SING	MSSING$	MSCI Singapore U$ - Price Index	Singapore
CNDA	MSCNDA$	MSCI Canada U$ - Price Index	Canada	SPAN	MSSPAN$	MSCI Spain U$ - Price Index	Spain
DNMK	MSDNMK$	MSCI Denmark U$ - Price Index	Denmark	SWDN	MSSWDN$	MSCI Sweden U$ - Price Index	Sweden
FRNC	MSFRNC$	MSCI France U$ - Price Index	France	SWIT	MSSWIT$	MSCI Switzerland U$ - Price Index	Switzerland
GERM	MSGERM$	MSCI Germany U$ - Price Index	Germany	UTDK	MSUTDK$	MSCI UK U$ - Price Index	United Kingdom
HGKG	MSHGKG$	MSCI Hong Kong U$ - Price Index	Hong Kong	USAM	MSUSAM$	MSCI USA U$ - Price Index	United States
ITAL	MSITAL$	MSCI Italy U$ - Price Index	Italy				

Table 3: Explanation and abbreviation of variables in S&P 500 Dataset

Abbreviation	Full Name	Number of missing values	Frequency	Explanation
permno	ID of the Asset	0	Daily	Identification number of the asset
date	Date	0	Daily	Observed date
ret	Return	19 210	Daily	Return of an asset at the observed date
std.ret	Standard Deviation of Returns	NA	Period of time	Standard deviation of returns calculated for specified period of time
bid	Bid Price of Share	19 646	Daily	Bid price of asset share
ask	Ask Price of Share	19 646	Daily	Ask price of asset share at the observed date
spread	Spread of asset	NA	Daily	Spread of the asset calculated as $spread = \frac{ask-bid}{\frac{ask+bid}{2}}$
numtrd	Number of trades	3 415 709	Daily	Number of trades of the asset that has been made
vol	Volume	18 810	Daily	Total quantity of shares of an asset traded
prc	Price	18 804	Daily	Price of a single share of an asset
volpri	Turnover	NA	Daily	Total turnover of an asset calculated as $turnover = volume * price$

Note: Number of missing values are presented from whole raw data matrix which contained 5 617 623 rows.