LAPPEENRANTA UNIVERSITY OF TECHNOLOGY
Faculty of Technology
Degree Programme in Technomathematics and Technical Physics

Alisa Zeleva

FORECASTING FINANCIAL WEATHER - CAN WE FORESEE MARKET SENTIMENT? SPECTRUM OF STOCK PRICE BEHAVIOR - NYSE STUDY.

Examiners: Professor Tuomo Kauranne
D.Sc. (Tech.) Matylda Jabłońska-Sabuka

ABSTRACT

Lappeenranta University of Technology

Faculty of Technology

Degree Programme in Technomathematics and Technical Physics

Alisa Zeleva

Forecasting financial weather - can we foresee market sentiment? Spectrum of stock price behavior - NYSE case study.

Master's thesis

2015

55 pages, 40 figures, 1 table

Examiners: Professor Tuomo Kauranne

D.Sc. (Tech.) Matylda Jabłońska-Sabuka

Keywords: stock market, NYSE, time series, singular values, singular vectors, correlation, ergodicity, co-occurrence matrix.

The desire to create a statistical or mathematical model, which would allow predicting the future changes in stock prices, was born many years ago. Economists and mathematicians are trying to solve this task by applying statistical analysis and physical laws, but there are still no satisfactory results. The main reason for this is that a stock exchange is a non-stationary, unstable and complex system, which is influenced by many factors.

In this thesis the New York Stock Exchange was considered as the system to be explored. A topological analysis, basic statistical tools and singular value decomposition were conducted for understanding the behavior of the market. Two methods for normalization of initial daily closure prices by Dow Jones and S&P500 were introduced and applied for further analysis.

As a result, some unexpected features were identified, such as a shape of distribution of correlation matrix, a bulk of which is shifted to the right hand side with respect to zero. Also non-ergodicity of NYSE was confirmed graphically. It was shown, that singular vectors differ from each other by a constant factor. There are for certain results no clear conclusions from this work, but it creates a good basis for the further analysis of market topology.

Acknowledgements

First of all, I would like to express my deep gratitude to my supervisor Assoc. Prof. Tuomo Kauranne, this thesis would not have been without his guidance.

My thanks go to D.Sc. Matylda Jabłońska-Sabuka for her ongoing support, essential consultations and valuable advice.

Finally, I want to say special thank to my parents, who always believe in me, and my lovely husband Vladimir for his support and care.

Lappeenranta, January, 2015.

Alisa Zeleva

Contents

List of Symbols and Abbreviations **5**

1 Introduction **6**

2 Literature review **8**

3 Basic concepts of stock market **11**

4 Research methodology **13**

4.1 Correlation and anticorrelation between stock prices 13

4.2 Normalization of initial data by the Dow Jones and S&P500 indexes . 18

4.3 Market ergodicity 21

4.4 Singular values and co-occurrence matrix. 25

4.5 Parameter estimation. 36

4.6 Moving average analysis of singular values. 41

5 Conclusion **49**

References **50**

List of Tables **52**

List of Figures **53**

List of Symbols and Abbreviations

NYSE	New York Stock Exchange
DJI	Dow Jones Index
S&P500	Standard and Poor's stock market index
NASDAQ	National Association of Securities Dealers Automated Quotation
CSE	Chinese stock exchange
ARCH	Autoregressive Conditional Heteroskedasticity
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
HCT	hierarchical cluster tree
ln	Natural logarithm
pdf	Probability density function
C	Correlation matrix
D	Distance matrix

1 Introduction

Countless mathematicians and economists are engaged in solving the problem of forecasting of stock prices, but nobody has yet succeeded. The main difficulty for the accurate prediction is that the data collected over a quite long period of time are non-stationary and often form a chaotic system, the behavior of which is influenced by many factors, for instance, inflation, political events and military operations, the level of bank interest rates and many more other criteria. The main objective of this research is to understand the topological behavior of the exchange, because on the basis of these laws that affect the change in stock prices, one could build the future model of stock market.

Rosario N. Mantegna [5; 6] is one of the firsts who suggested to apply a statistical and topological tools to stock prices, considering this system as a physical measures. Bernd Rosenow continued and extended this theory in his articles [8; 9; 10]. Chinese researchers, Chi K.Tse [11] and Ying Yuan [12], studied a stock exchange as a complex network, basing on Mantegna's work.

So far, chaotic systems, such as stock prices, can not be successfully predicted. There is no good statistical model which is able to forecast future changes in stock sales with minimal risks. A stock exchange is a complicated and non-stationary system, which depends on many factors and undergoes changes during a financial crisis.

The aim of this work is to examine, using various statistical approaches, the daily closure prices taken from NYSE in period of time from June of 2006 to February of 2013. For better understanding the behavior of NYSE we have applied basic statistical tools. In particular, correlation matrices C are found by calculating correlation coefficients for all pairs of stocks. Furthermore, we calculated a metric between two stocks, called distance [5], based on which a hierarchical cluster tree can be constructed. Two types of normalization of initial daily closure prices by Dow Jones and S&P500 indexes are provided and used for further calculations. For opportunity to detect industry clusters and companies belonging to it, singular value decomposition was performed.

This Master's thesis has the following structure. The next Section 2 covers the theoretical background and review of earlier research into the problem. Section 3 includes a basic information about a stock market and a stock exchange, short history of New York Stock Exchange and summary about Dow Jones and Standard and Poor's 500 indexes. The main Section 4 is devoted to different statistical approaches

for studying of NYSE. Section 5 is a short conclusion, which highlights results and gives proposals for future work.

2 Literature review

A lot of research papers are devoted to the study of correlation between stock prices in different stock exchanges, statistical and topological analysis are conducted for such stock markets as NYSE, NASDAQ and CSE. However, overwhelming majority of research works take for survey either individual industries or a narrow range of companies that all can depend on the price of oil, gas, gold and other resources. Therefore, a quite narrow specification of research tasks, assigned in this thesis, complicates search of the literature corresponding to the topic of this thesis.

"Introduction to econophysics. Correlations and Complexity in Finance" by Rosario N. Mantegna is a basic work that underpins this master's thesis. In Chapter 12, methods for research the correlation and anticorrelation between a pair of stocks are described in detail and useful formulas are given. In Chapter 13, the concept of distance between stocks is introduced. These statistical tools, necessary for the empirical research on market, will be used in this thesis.

In the article "Random magnets and correlations of stock price fluctuations" by Bernd Rosenow a framework of magnetic systems is used to describe and explain cross-correlations between stock price fluctuations. Using methods of random matrix theory, it was shown that the methods of physics can be advantageously applied to the scrutiny of financial market.

An article "A network analysis of the Chinese stock market" by Wei-Qiang Huang states that a complex network describes a massive dataset such as a stock market, which is a complex system. A comprehensive statistical analysis and topological studies of the properties of this network are given in the article. Finally, it was shown that the financial network follows a power-law model. Clustering coefficient, connected components, clique and independent set are also presented in this work.

An idea of creating a network constructed from stock prices is supported in "The structure and resilience of financial market networks" by Thomas Kauê Dal'Maso Peron. It is verified that the network organization becomes less robust when it undergoes great changes during crashes. A new regression model, proposed by the author, detects a collective behavior during financial crisis. Network construction is based on correlation matrix and distances between two stocks. It was calculated, that a time-window equal 28 days with 1 day lag is the best for capturing important financial crises.

Another important article is "A network perspective of the stock market" by Chi K. Tse. Using a "winner-take-all" approach and the following condition: the pair of stocks is connected if the cross correlation of the daily prices of these stocks is bigger than a threshold (e.g., 0.9), the author creates a Minimal Spanning Tree to simplify the form of a graph, which alleviate the further analysis. In conclusion, author suggests that the majority of stocks is affected by a relatively small number of stocks. It was shown that stocks in the financial sector highly influence the entire market.

"Price-volume multifractal analysis and its application in Chinese stock markets" by Ying Yuan are devoted to the study of price returns. Using a Mandelbrot theory and his multifractal method, it was shown that a cross-correlation of stock price returns and trading volume variation is multifractal.

Another article related to the topic of the thesis, called "Cluster Analysis and Stock Price Comovement" by Robert D. Arnott, states about analyzing market behavior by dividing stocks into five clusters "corresponding to major extra-market factors". A multifractal model is applied for examine extra-market price behavior. The provided model can be applied for stock classification or portfolio optimization.

Methods of random matrix theory were suggested for analyzing a cross-correlation matrix in "Application of random matrix theory to study cross-correlation of stock prices", written by Bernd Rosenow. The analysis was based on examining the properties of the largest eigenvalues and corresponding to eigenvectors. There are two main results. Firstly, universal predictions of random matrix theory and the most eigenvalues in the spectrum of the cross correlation matrix are well consistent. Secondly, cross-correlation matrix "satisfies the universal properties of the Gaussian orthogonal ensemble of real symmetric random matrices".

The following article, called "Dynamics of cross-correlations in the stock market" by Bernd Rosenow extends previous work by applying random matrix theory for forecasting cross-correlation. The most part of the paper is assigned to study the eigenvalues and eigenvectors of cross-correlation matrices. As a result the large eigenvalues and corresponding eigenvectors describe the economically meaningful information in the empirical cross-correlation matrix.

Relationship between stock price and trading activity are studied in "Impact of Stock Market Structure on Intertrade Time and Price Dynamics" by Plamen Ch. Ivanov. In article two markets, namely NYSE and NASDAQ, are considered and

compared. As a research methodology was chosen a detrended fluctuation analysis method, "which has been shown to detect and accurately quantify long-range power-law correlations embedded in noisy non-stationary time series with polynomial trends". This approach is better than traditional techniques such as power spectral, autocorrelation and Hurst analysis, because it is suited to non-stationary data. This work is unique because there is presented a study about changes "in the trading dynamics of stocks of companies that moved from one market to the other".

In the paper "Stochastic GARCH dynamics describing correlations between stocks" by G. Prat-Ortega non-linear GARCH models are proposed like stochastic process which properly fits the mean, the variance and the memory of the correlations between stocks, and develop a procedure to find an optimal set of parameters for this process.

There are a lot of approaches for solving one difficult problem: how to create a forecast for stock prices with minimal risks? Moreover, many articles and research papers are devoted to the study of the behavior of stock exchange. Despite the variety of methods, common in all research works is that statistical tools and topological analysis are widely used for the study of such complex, multifactorial, non-stationary system, as a stock exchange. Different exchanges, various companies, industry clusters and time periods are considered in research works. However, the ideal model that would allow to predict the behavior of prices and to reduce the risk, has not yet been created.

3 Basic concepts of stock market

A stock market (or equity market) is an integral part of the financial market, where the turnover of securities takes place. A stock exchange "provides services for stock brokers and traders to buy or sell stocks (shares), bonds, and other securities" [13]. The major stock exchanges are New York Stock Exchange, NASDAQ, Japan Exchange Group, Euronext and London Stock Exchange Group.

The main objective of stock exchange is to ensure the market (real or virtual) by facilitating the exchange of securities between buyers and sellers. The stock exchange performs the following functions: organization of trading, preparation and implementation of exchange contracts, quotation of stock exchange prices, execution of stock exchange transactions. The trading information on the listed securities to facilitate the detection of prices available in real-time.

The most famous and largest stock exchange in the world is New York Stock Exchange (NYSE), "located at 11 Wall Street, Lower Manhattan, New York City, New York, United States" [13]. The history of NYSE starts on May 17, 1792, when the "Buttonwood Agreement", devoted to creation of the New York Stock Exchange, was signed by twenty four brokers. Stocks of "The Bank of New York" became the first quoted on this stock exchange.

NYSE is sometimes called as "the Big Board". Trading time starts every day (except Saturdays, Sundays and official holidays) from 9:30 a.m to 4:00. The trading happens in the form of an auction, during which one may perform transactions in shares and other securities. Securities of 3504 companies have been listed on NYSE by the end of 2013.

The commonly known stock market index Dow Jones (or Dow Jones Industrial Average) for shares of industrial companies is determined on NYSE. This is the oldest index of all American market indexes, which covers the 30 "large publicly owned companies in the United States" [13]. Earlier the index was calculated as the arithmetical average price of the stocks covered by the companies, but nowadays the scaled average is used. It means that the sum price is divided by the divisor, which changes whenever stocks included in the index are split or consolidated.

Another significant American stock market index is Standard and Poor's 500 (or S&P500), which includes exactly 500 selected stocks of the USA having the largest capitalization. These companies are not the same as a list of the largest companies

in USA, since a private ownership and companies, which have insufficient liquidity, are not included in a list of 500. The value of S&P500 reflects the total capitalization of companies, because the weight of each company in the index is proportional to its capitalization.

4 Research methodology

4.1 Correlation and anticorrelation between stock prices

In an exchange market all stocks are interrelated. Topological analysis is used to find out some similarities and differences in stock price behavior. The simplest way for research is to study a correlation coefficient between daily closure prices. The time series of a stock price during the whole period of observation time is presented in Fig.1.

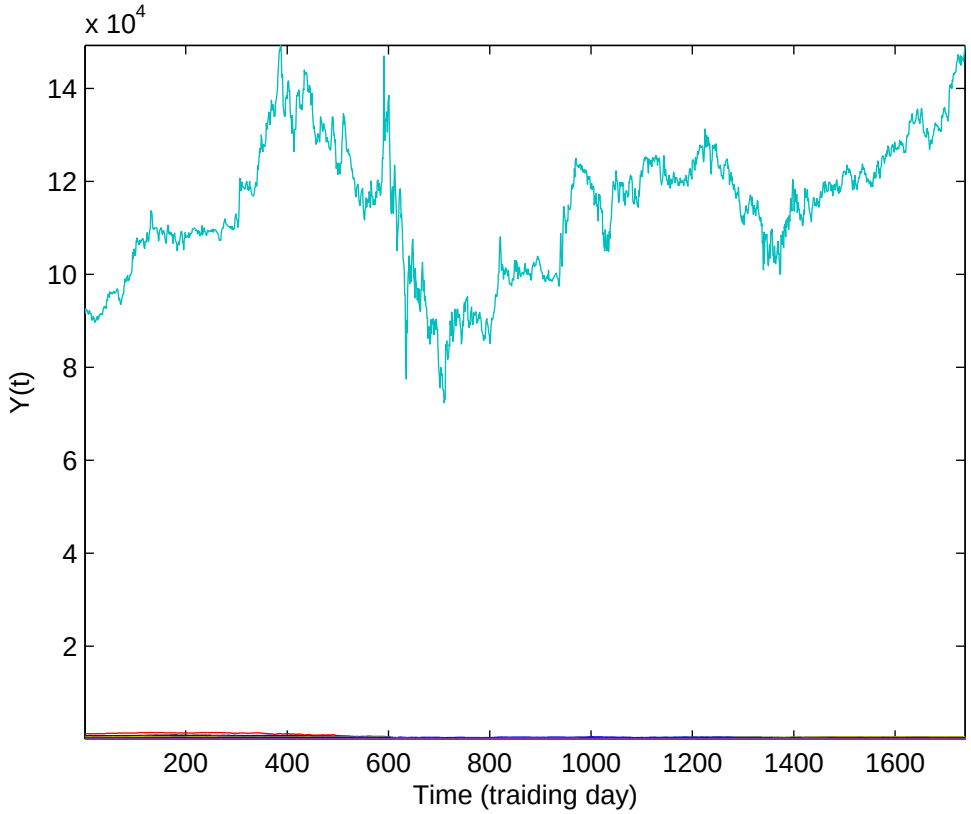


Figure 1: Time series of daily closure NYSE stock prices.

In further analysis a logarithm of daily closure prices is used to avoid large values, which are visible in Fig.1. For this purpose one can apply the following formula, defined for ith stock

$$S_i = \ln Y_i(t),$$

where Y_i is a daily closure price, t - time. The time evolution of a logarithm of daily closure prices is presented in Fig.2.

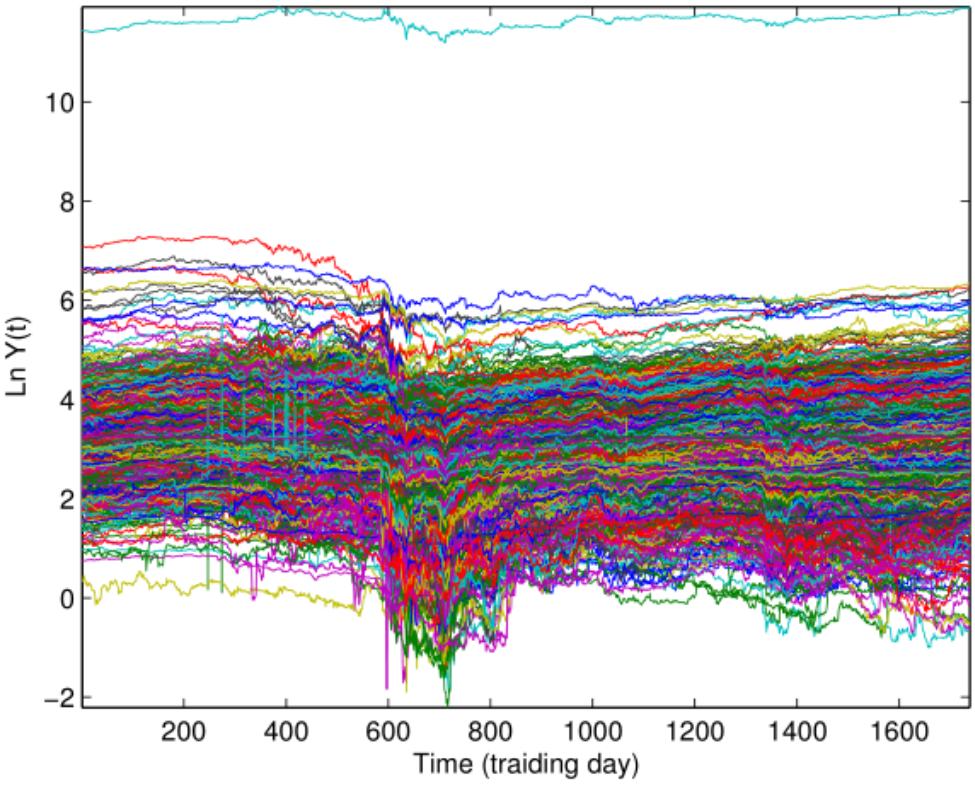


Figure 2: Time evolution of a logarithm daily closure prices.

Correlation coefficient between a pair of stocks i and j is calculated by the formula

$$\rho(ij) = \frac{cov(i,j)}{\sqrt{cov(i,i)cov(j,j)}},$$

where *cov* means the covariance matrix. By definition, the correlation coefficient takes values in an interval from - 1 to 1. In this instance, three special cases exist

1. $\rho(ij) = 1$ (strong correlation between pair of stocks)
2. $\rho(ij) = 0$ (uncorrelated pair of stocks)
3. $\rho(ij) = -1$ (strong anticorrelation between pair of stocks).

By finding all correlation coefficients a square correlation matrix C is obtained, and it has the following important properties

1. $\rho(ij) = \rho(ji)$

2. $\rho(ii) = 1$

For the data of n stocks the number of unique correlation coefficients is $(n \times (n-1))/2$. In the case of 1930 stocks, this value is equal to 1861485. Distribution $P(\rho)$ of correlation coefficient matrix, presented in Fig.3, is obtained by plotting a normalized histogram of the full set of correlation coefficients. The formula for normalization of ith value is

$$normalized\ x_i = \frac{x_i}{\sum_{i=1}^{n} x_i}.$$

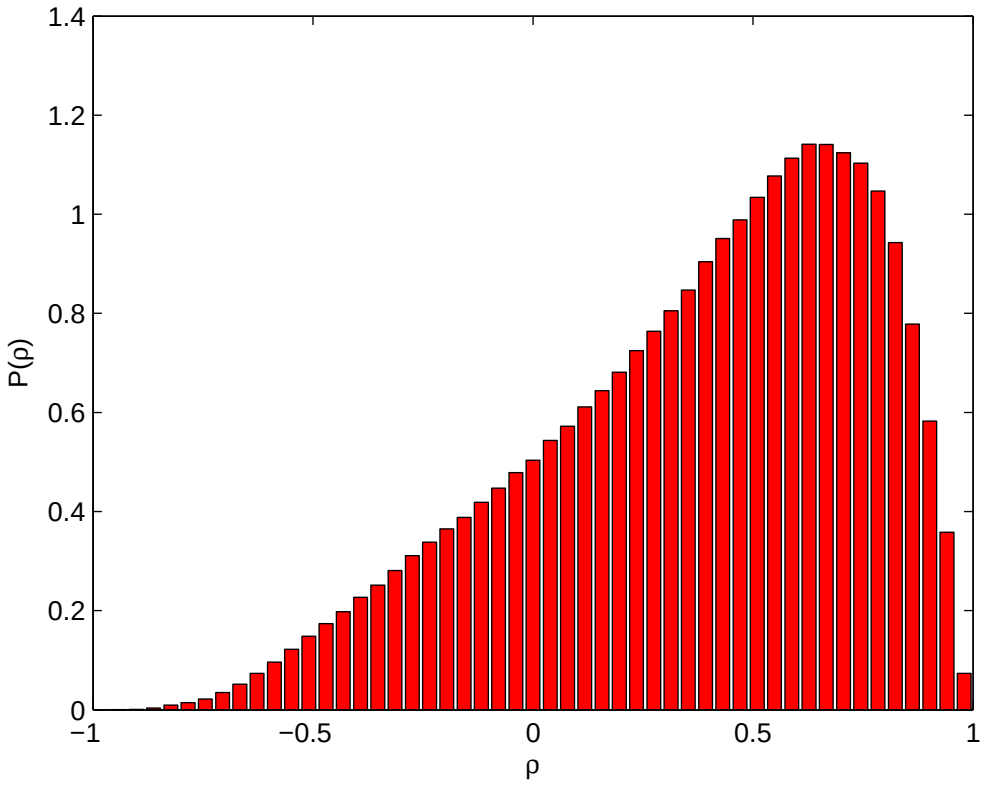


Figure 3: Distribution of correlation coefficients for whole period of time.

The mean of correlation coefficients is equal to 0.3699. The shape of distribution $P(\rho)$ (the bigger mass are shifted in positive side) shows that companies mostly correlate between each other. The $P(\rho)$ reaches a maximum value of 1.1407 at $\rho = 0.6275$.

The shape of distribution is very similar to Maxwell-Boltzmann distribution reflected

about a vertical axis. Its probability density function (pdf) is

$$pdf = \sqrt{\frac{2}{\pi}} \frac{x^2 \exp^{-x^2/2a^2}}{a^3},$$

constant $a = \sqrt{\frac{kT}{m}}$ is a scale parameter, where m is the particle mass, k is the Boltzmann's constant, T is thermodynamic temperature. Parameter a affects the height and width of curve. An example of Maxwell-Boltzmann distribution is presented in Fig.4.

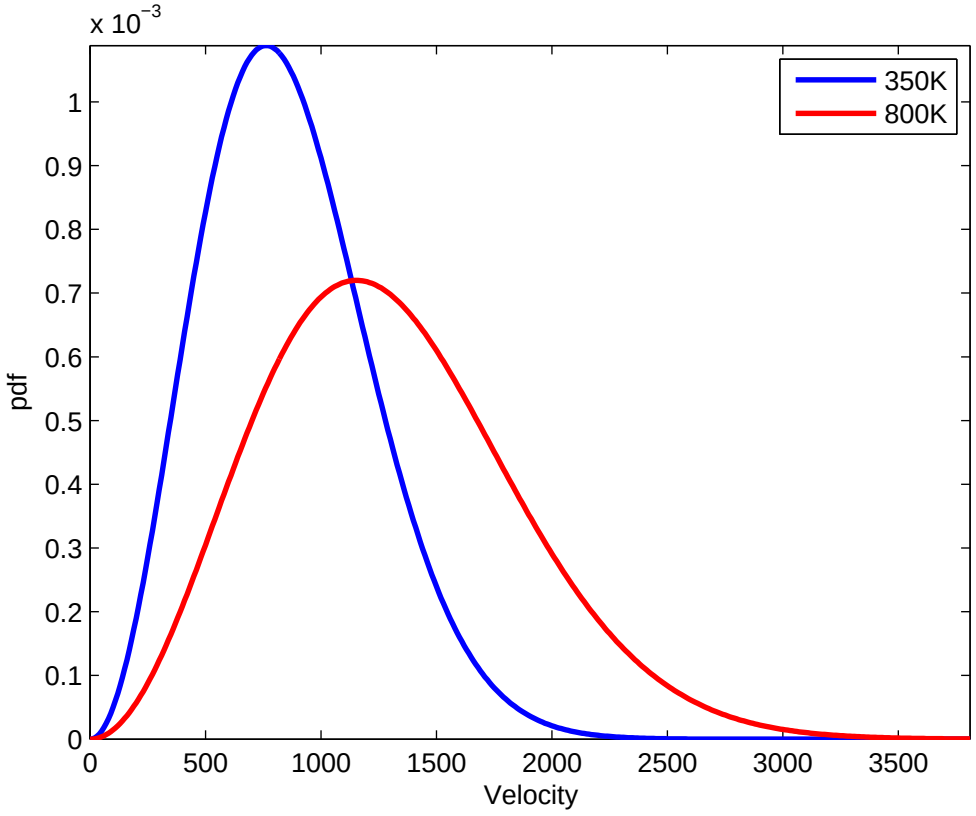
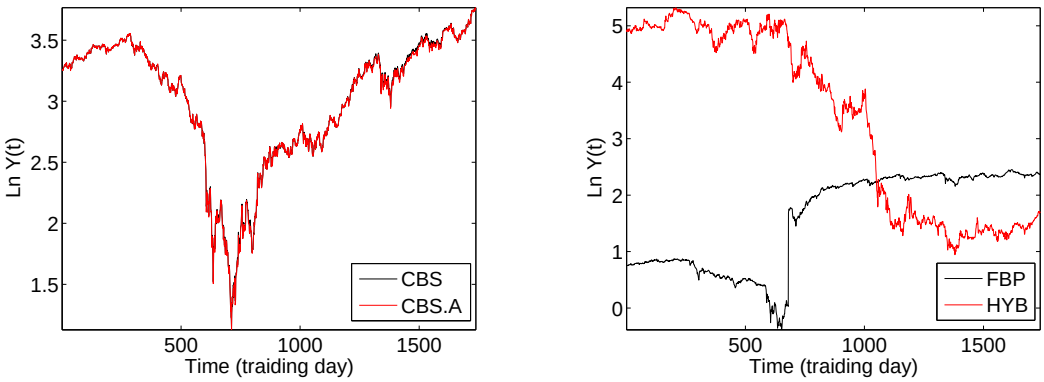


Figure 4: Maxwell-Boltzmann distribution. Red line matches to temperature equals to 350° K, blue - 800° K.

The largest value $\rho(ij) = 1$ corresponds to the pair of stocks "CBS" and "CBS.A", it means one common company, namely "CBS Corporation". The time series of $\ln(Y(t))$ for both stocks is presented in Fig.5a. By analogy, the smallest value $\rho(ij) = -0.960$ corresponds to the pair of stocks "FBP" and "HYB", full names are "First Bancorp" and "New America High Income Fund Inc." respectively. In Fig.5b the time evolution of $\ln(Y(t))$ is presented for these stocks.

In Fig.5a it can be clearly seen that behavior of daily prices for both cases is absolutely identical, whereas Fig.5b depicts similar behavior only on some short time



(a) The most correlated pair of stocks. (b) The most anticorrelated pair of stocks.

Figure 5: Observation pairs of stocks that correspond to the maximum and the minimum correlation coefficient, respectively.

intervals.

The same analysis is made for time intervals equal to one year, namely for periods of 260 working days. In Table 1, a minimum and maximum correlation coefficient for different time intervals are listed. In Fig.6 correlation coefficients are shown for each of the seven years from 16.06.2006 to 14.02.2013.

Table 1: The observed minimum and maximum values $\rho(ij)$ for different years.

Year	Minimum	Maximum
1	-0.95808	0.99969
2	-0.96112	0.99988
3	-0.93597	0.99992
4	-0.92260	0.99979
5	-0.97807	0.99991
6	-0.97761	0.99914
7	-0.96586	0.99939

It was expected that a distribution of correlation coefficients will have a similar bell shape as a probability density function of normal distribution. However, Fig.3 and Fig.6 refute this assumption. Splitting data into one-year-intervals shows, how interdependence between stock prices changes from year to year, and detects a transformation in the shape of $P(\rho)$. Some distributions behave as an exponential function (1, 3, 4 years), while other like Maxwell-Boltzmann distributions.

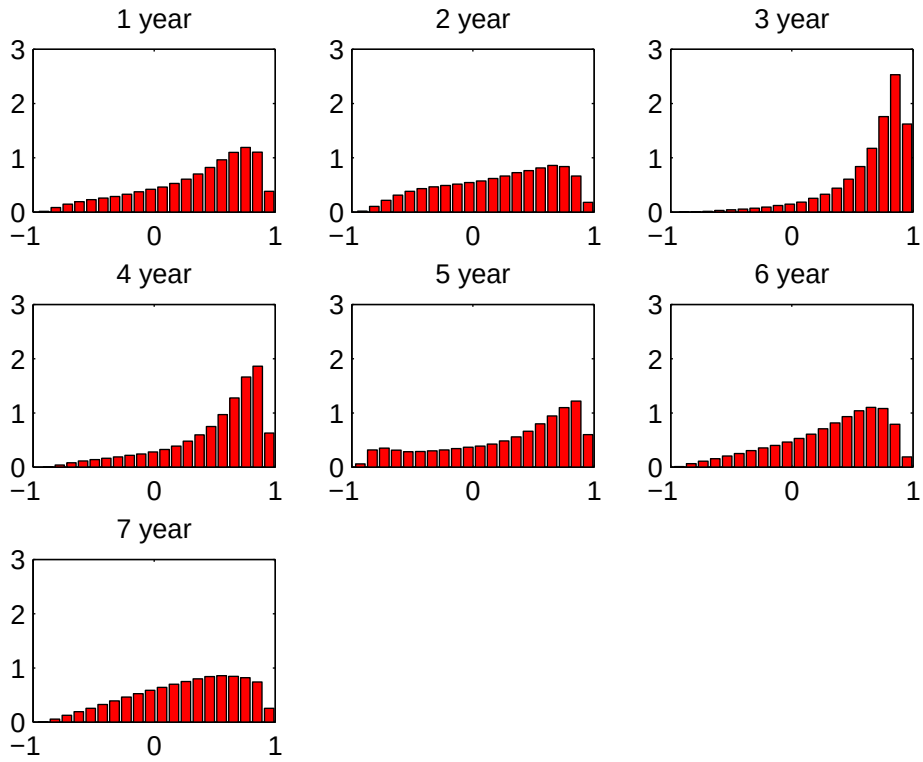


Figure 6: Distribution of correlation coefficients for different years.

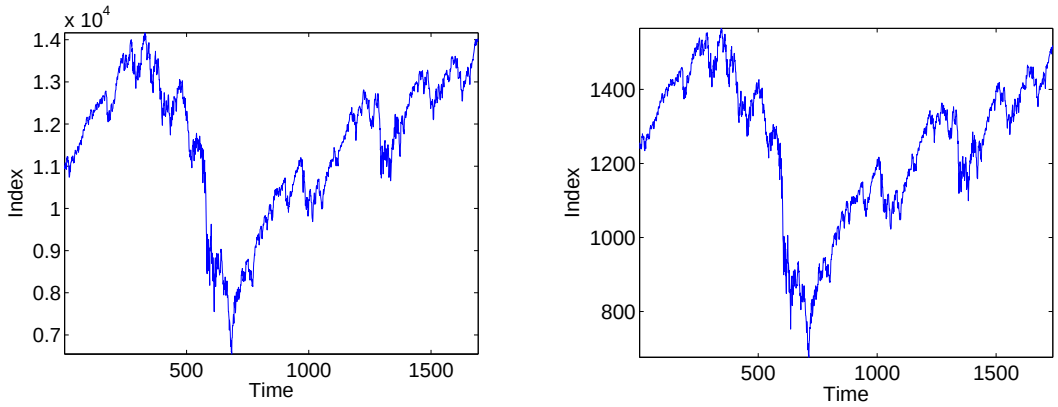
4.2 Normalization of initial data by the Dow Jones and S&P500 indexes

The Dow Jones Index (DJI) and Standard and Poor's 500 (S&P500) are two of several stock market indexes, measuring the value of a section of the stock market. The time series of these indexes are present in Fig.7a and Fig.7b accordingly. These indexes behave almost identically, but their scales are different.

There are two types of normalization applied to the initial daily closure prices by market indexes:

1. **By regression model.** The following functions are taken as a model

$$NormPrice = (Price + \theta_1) * \theta_2.$$



(a) Dow Jones Index.

(b) S&P500 Index.

Figure 7: Stock market indexes.

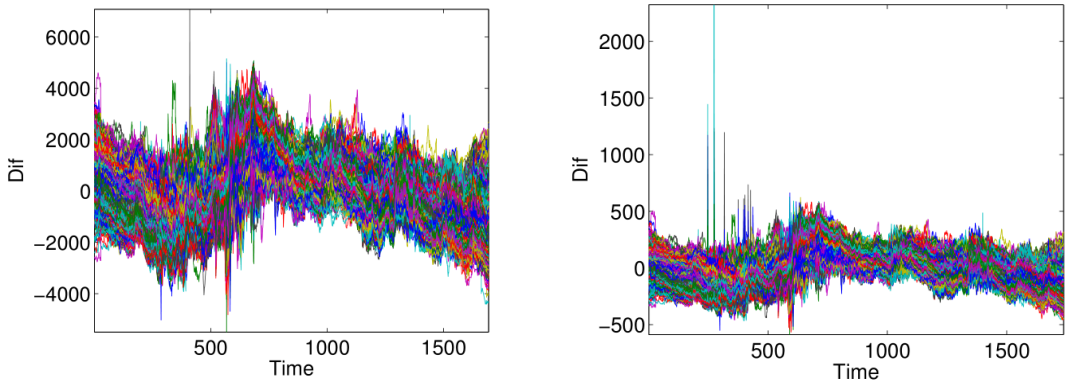
Unknown parameters θ should be estimated from the cost function

$$ss = \sum((\theta_1 * (Price + \theta_2) - Index)^2).$$

After that the differences between normalized daily closure prices and index are calculated as

$$Dif = NormPrice - Index.$$

The time series of Dif for both cases (DJI and S&P500) are shown in Fig.8a and Fig.8b, accordingly.



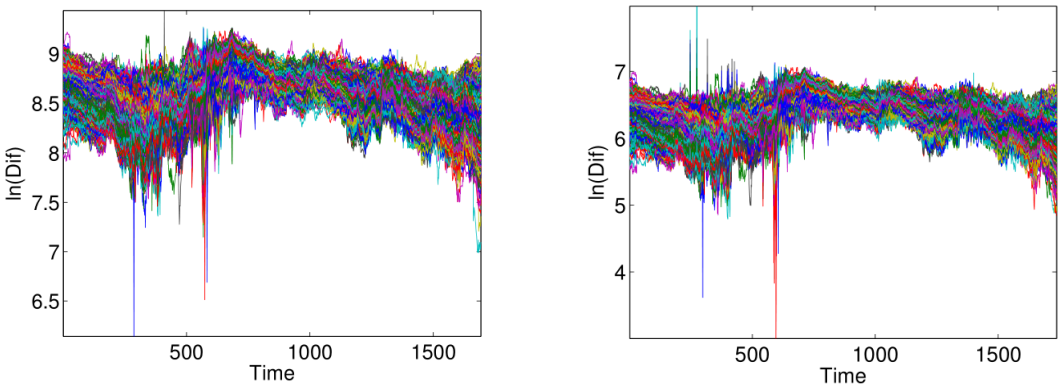
(a) Difference from Dow Jones Index.

(b) Difference from S&P500 Index.

Figure 8: Difference between daily closure prices normalized by regression model and market index.

For better visualization and reducing large scaling, a natural logarithm of these differences is taken and shown in Fig.9.

2. **By mean and standard deviation.** The following formula is used for nor-



(a) Logarithm of difference from Dow Jones Index.

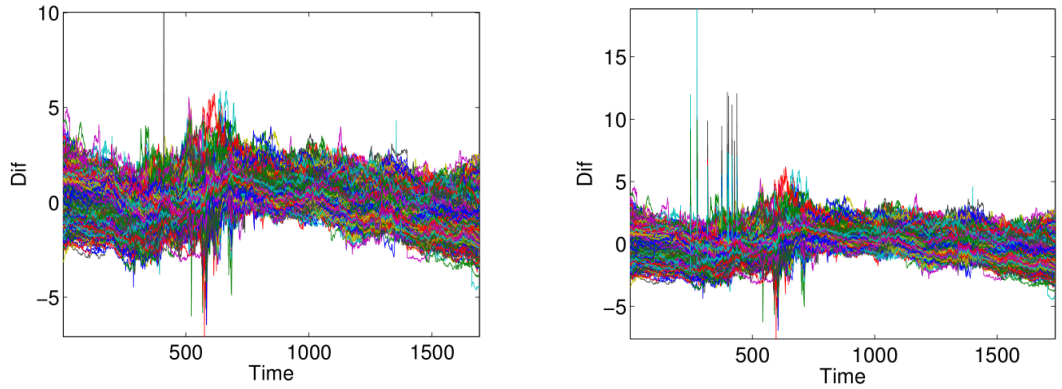
(b) Logarithm of difference from S&P500 Index.

Figure 9: Natural logarithm of difference between daily closure prices normalized by regression model and market index.

malization

$$normalized\ s = \frac{s - \overline{s}}{\sigma(s)},$$

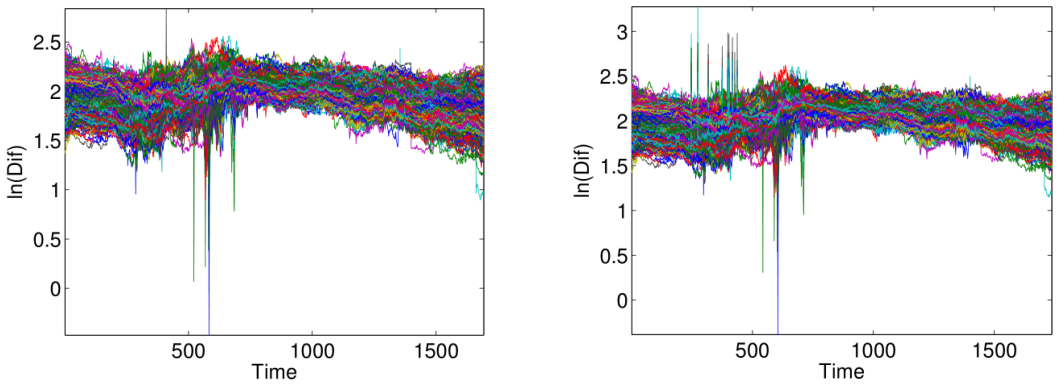
where s is initial values, $\overline{s}$ - mean of s, $\sigma(s)$ - standard deviation of s, are applied for initial closure prices and market index. After that a difference between normalized prices and normalized indexes are calculated. As it was done earlier, the time series of difference and logarithm of it are presented in Fig.10 and Fig.11.



(a) Difference from Dow Jones Index.

(b) Difference from S&P500 Index.

Figure 10: Difference between daily closure prices and index normalized by mean and standard deviation.



(a) Logarithm of difference from Dow Jones Index.

(b) Logarithm of difference from S&P500 Index.

Figure 11: Natural logarithm of difference between daily closure prices and index normalized by mean and standard deviation.

4.3 Market ergodicity

Ergodicity means that with sufficient observation time a systems can be described by statistical methods. The stock market is not an ergodic system, i.e a long average is not the same as a stable average in a short period. In other words, an ergodic system behaves statistically in the same manner across space at one time part and across time.

For the original stock prices a distribution of skewness and kurtosis is found to investigate possible ergodicity of the stock market.

Skewness s is a quantity characterizing the asymmetry of the probability distribution of a random variable around its mean, which is calculated by the following formula:

$$s = \frac{E(x-\mu)^3}{\sigma^3},$$

where μ - the mean of x, σ - the standard deviation of x, $E(x)$ is the expected value of the quantity x.

The NYSE market has a negative skew that is confirmed by Fig.12a, which illustrates that the left tail is longer and a significant mass is concentrated on the right hand side of the figure.

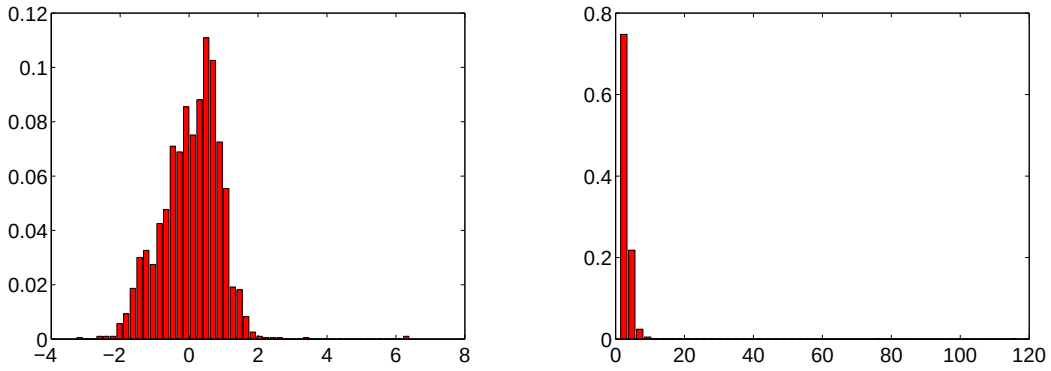
Kurtosis k is a measure of the peakedness of the distribution of a random variable,

which is calculated by the next formula:

$$k = \frac{E(x-\mu)^4}{\sigma^4},$$

where μ, σ, $E(x)$ are the same as earlier.

Fig.12b depicts the distribution of NYSE kurtosis. The shape behaves as an exponential distribution $f(x,\lambda) = \lambda \exp^{-\lambda x} H(x)$, where $\lambda = 0.75$ and $H(x)$ is the heaviside step function.



(a) Skewness. (b) Kurtosis.

Figure 12: A descriptors of the shape of a probability distribution.

The following steps are performed for analysis of possible ergodicity of NYSE.

1. ***Normalized stock prices.***

   A normalization by regression model with S&P500 index is used for the following calculations. Logarithmic returns are defined by the following expression:

   $$r_i = \ln\left(\frac{Y_i}{Y_{i-1}}\right),$$

   where r_i is return at ith day, Y_i and Y_{i-1} are prices (indexes) at day i and $i-1$, respectively.

   - **Sample of individual stock price histogram over time.**

     Five stocks are randomly chosen from among 1930 companies. A price histogram and a histogram of logarithmic returns is plotted in Fig.13 for each of these companies.

   - **Total price histogram on a single day of normalized prices.**

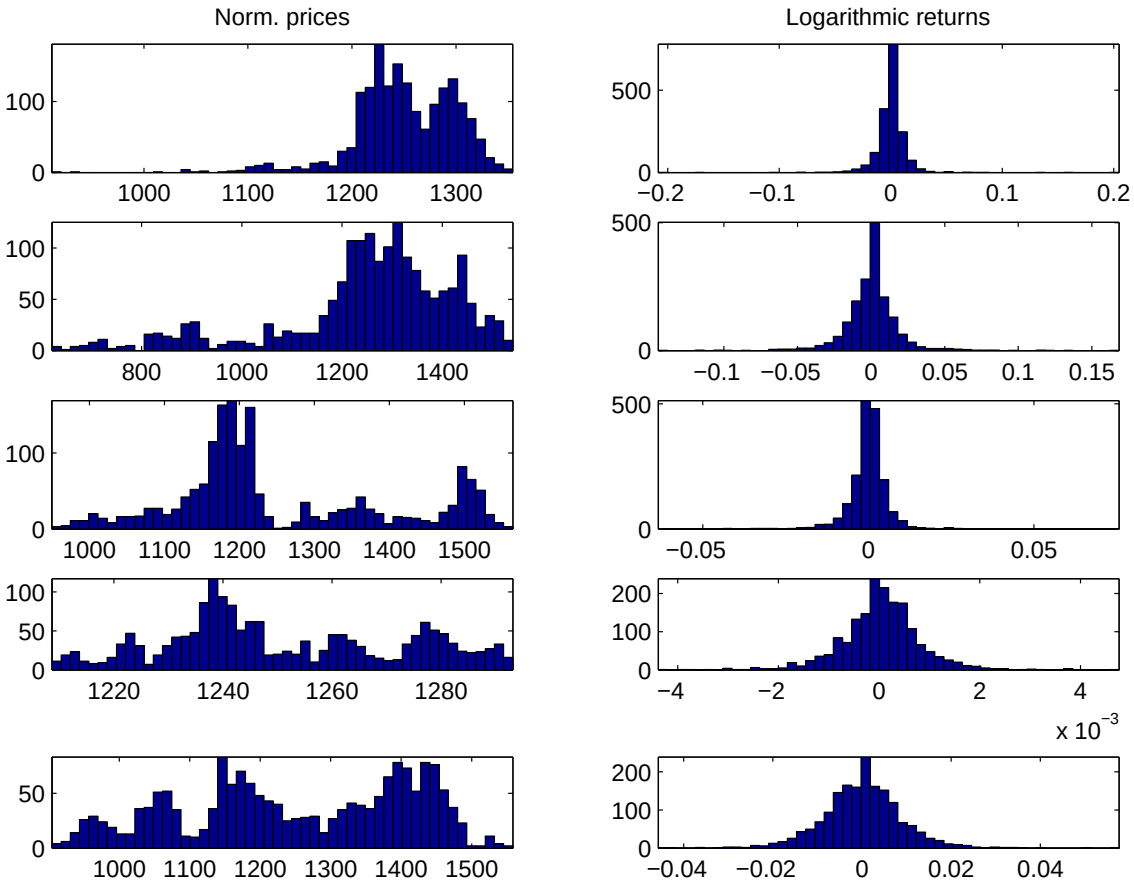


Figure 13: Histogram of individual stock price over time. Left picture - normalized prices, right - logarithmic returns.

Also five days are chosen randomly from June of 2006 to February of 2013 and histograms for all stocks on these days are plotted in Fig.14.

2. ***S&P500 index.***

   Since normalization is created by regression model, the index was not changed. Moreover, the index data is a vector with size 1×1738. In Fig.15 histograms are plotted for both indexes, DJI and S&P500, for their original and logarithm form.

3. ***Difference between normalized prices and S&P500 index.***

   In this case a difference between normalized prices and index is considered instead of normalized prices. Histograms are presented in Fig.16 and Fig.17.

4. **Logarithmic returns.**

   Logarithmic returns are calculated from normalized prices and of difference between these prices and S&P500 index.

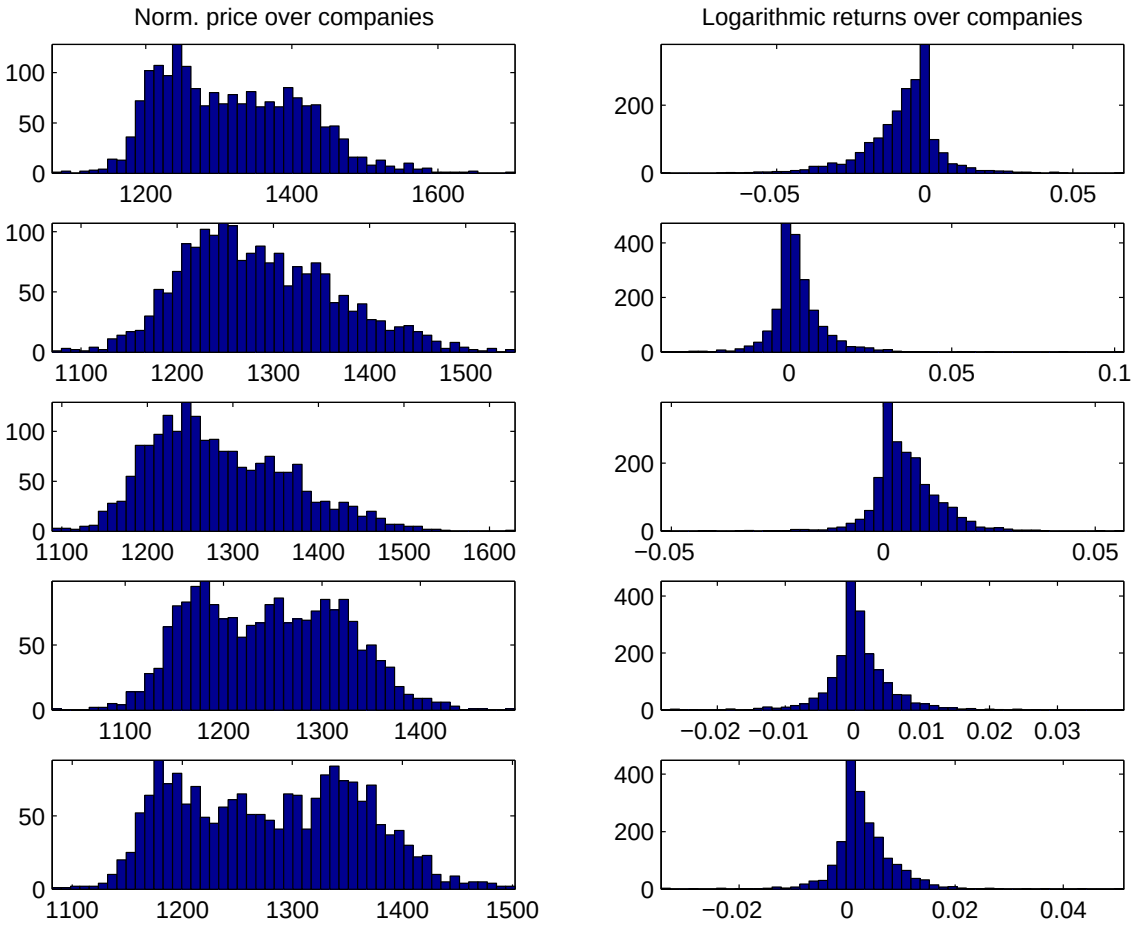


Figure 14: Histogram of all stocks on a single day. Left picture - normalized prices, right - logarithmic returns.

Based on the different shapes of the histograms of individual study and daily snapshot one may conclude that prices form non-ergodic system. A mean, variance, skewness and kurtosis of logarithm returns, presented in Fig.18–21, show that logarithmic returns are not ergodic. A logarithmic returns of difference at ith day, presented in the middle picture in Fig.18, are calculated by one of these equal expressions:

$$\ln \frac{Dif_i}{Dif_{i-1}} = \ln(Dif_i) - \ln(Dif_{i-1}),$$

whereas a difference between logarithmic returns are calculated as follows:

$$\ln \frac{NormPrice_i}{NormPrice_{i-1}} - \ln \frac{Index_i}{Index_{i-1}} = \ln \frac{NormPrice_i * Index_{i-1}}{NormPrice_{i-1} * Index_i}.$$

Moreover, from the fact that the mean of logarithmic returns of the S&P500 index is positive on average, but the mean of logarithmic returns of the difference of all companies from the index is negative on average, we get a conclusion: the companies

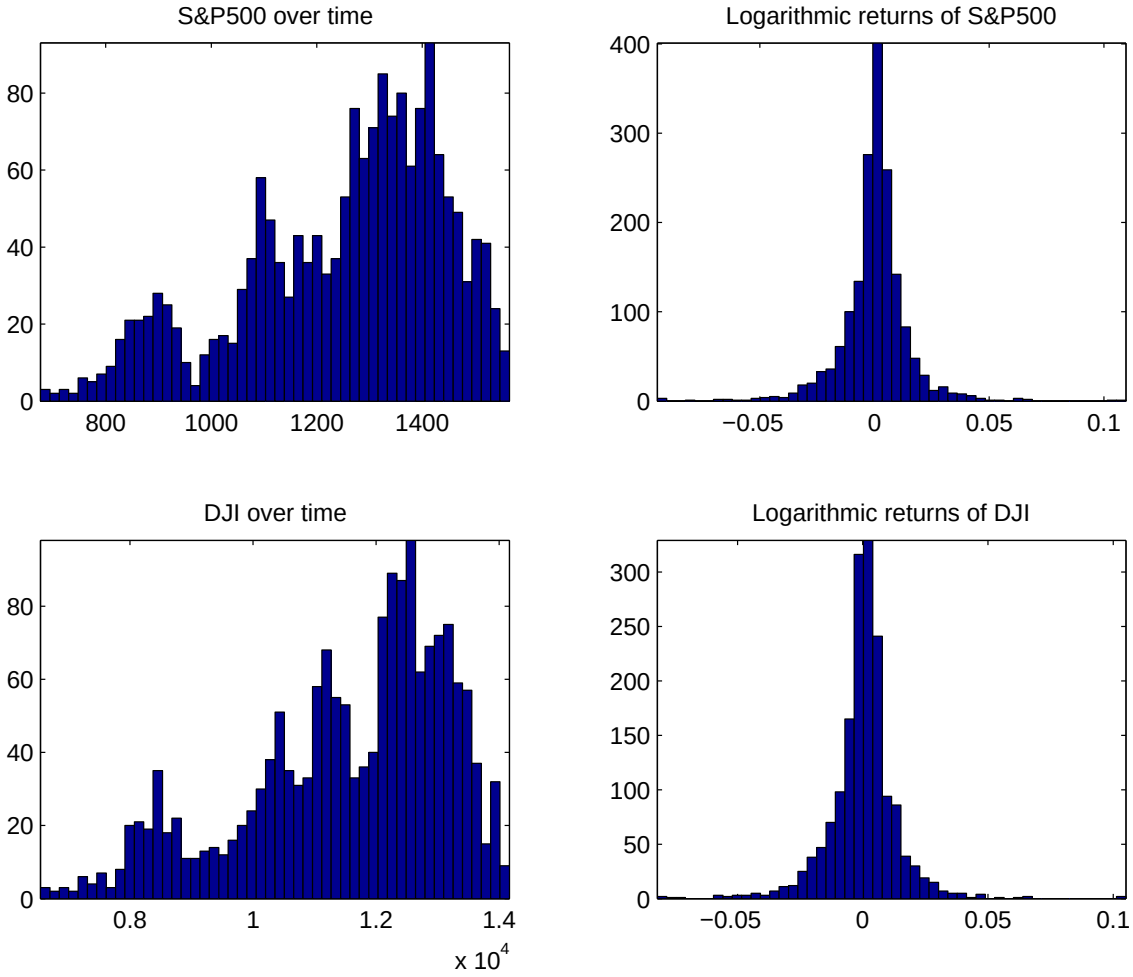


Figure 15: Histogram of the index and logarithmic returns over time.

that are part of the S&P500 index systematically outperform other companies as investments. A third graph in Fig.18 indeed confirms the following statement: an index fund on S&P500 outperforms clearly companies on average. The percentage of companies that outperform the index is less than 18 per cent with both ways of calculating the return on investment.

4.4 Singular values and co-occurrence matrix.

The names of the stocks can be sorted alphabetically or by distance. Distance between a pair of stocks i and j is calculated by following formula

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

This equation defines an Euclidean distance with the following properties:

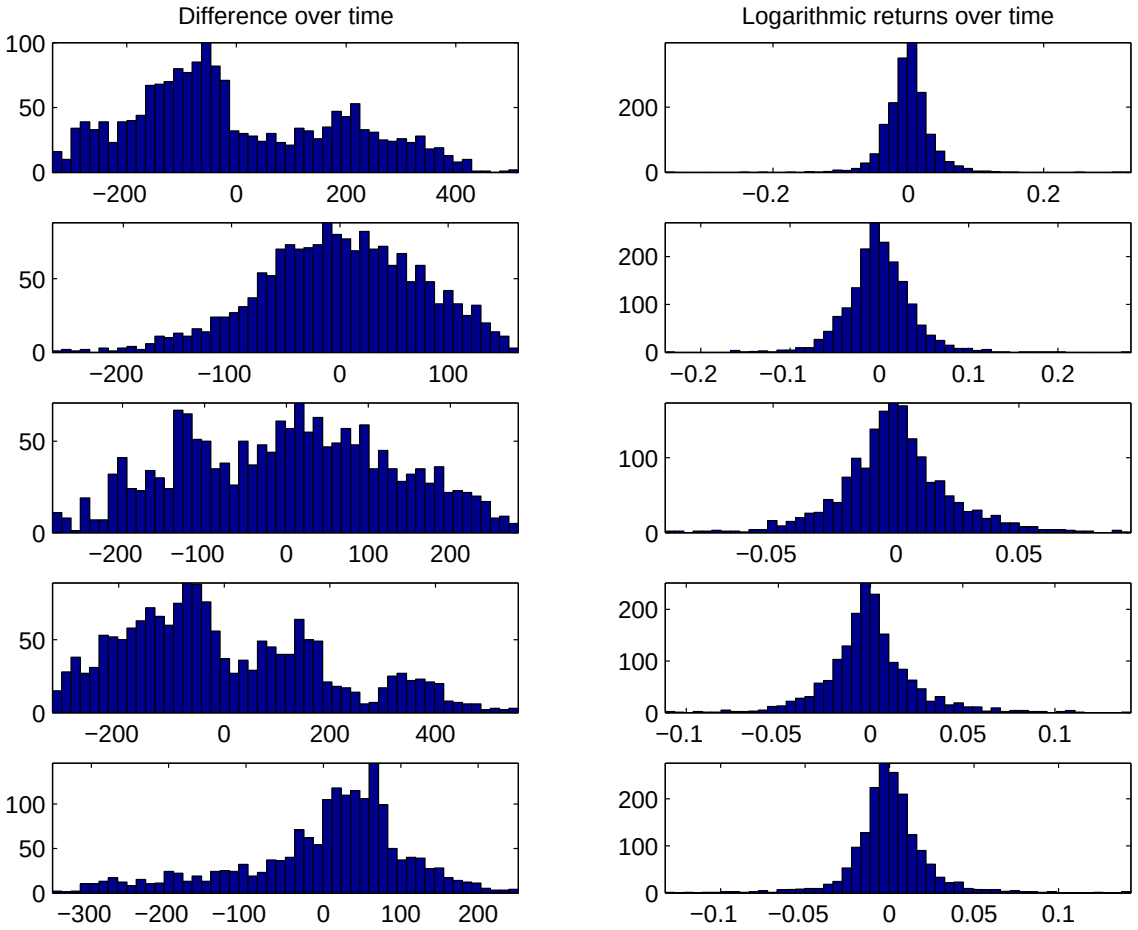


Figure 16: Histogram of difference and logarithmic returns over time.

1. $d_{ij} = d_{ji}$ (symmetric)
2. $d_{ii} = 0$
3. $d_{ij} \leq d_{ik} + d_{kj}$ (triangular inequality)

Notice, if $\rho_{ij} = -1$, then d_{ij} takes a maximal value equal 2. Opposite, distance is close to zero, when stock i and j are positively correlated, it means that the time series for these stocks has a similar behavior over the entire time horizon.

After calculation of d for all pair of stocks, a symmetric matrix of distances D is obtained. Using values from this matrix a hierarchical cluster tree (HCT) can be produced. A new order of stock names (from the most correlated companies to the most anticorrelated) are taken from this tree.

The data used is the difference between daily closure prices and S&P500 index, which are normalized by the first method using the regression model. The same analysis can be done for the other three cases. The correlation matrix of the differences is

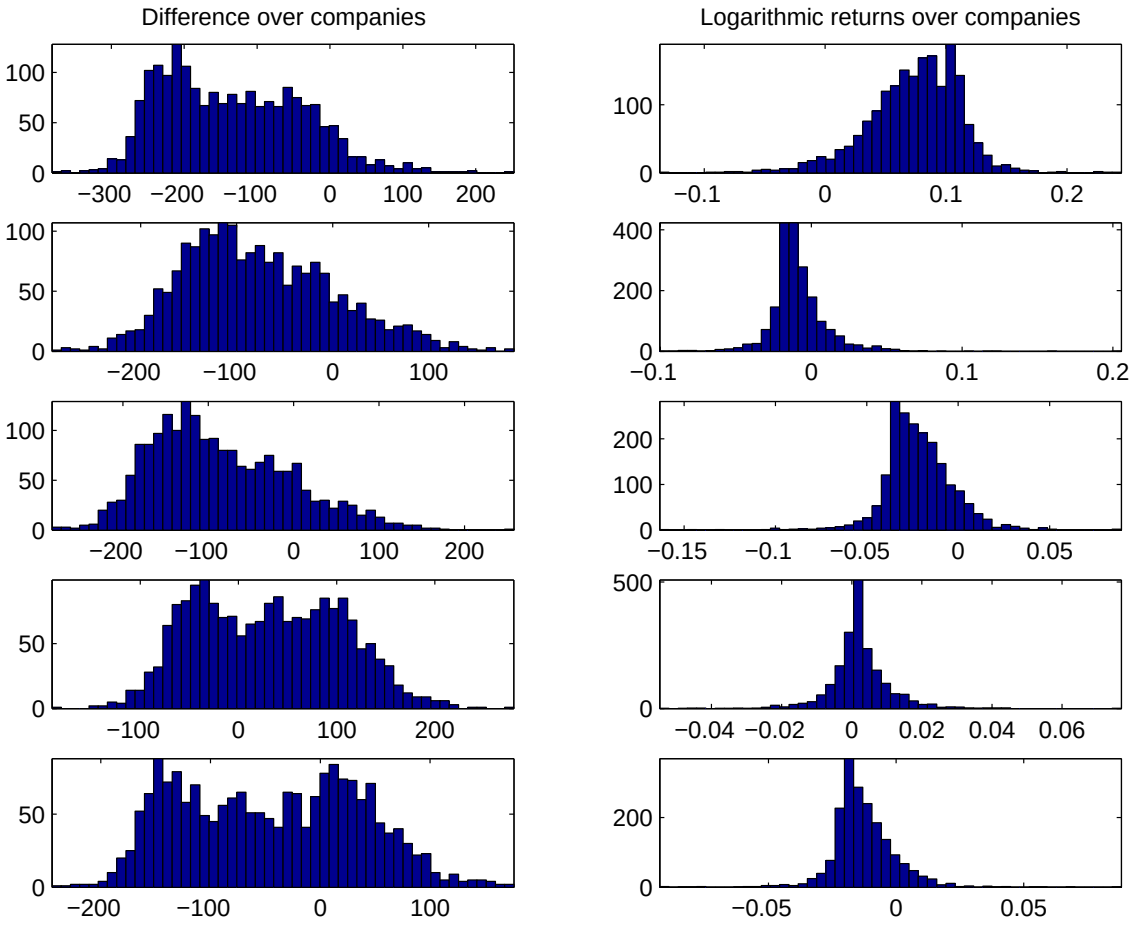


Figure 17: Histogram of the difference and logarithmic returns over companies at one single day.

calculated the same way as it was done for daily closure prices. The distribution of its values is presented in Fig.23.

The fact that the correlation coefficients of deviations of individual stock prices from an index are still biased towards positive correlations indicates that there is a herding or momentum effect at play between companies even after we deduct the impact of the index. Deviations from an index are therefore definitely not normally distributed.

Using singular value decomposition of the covariance matrix, one may find 10 singular vectors S, which correspond to 10 largest singular values. These vectors S are plotted separately in Fig.24. Companies in this case are sorted according to order obtained from HCT.

Further, values of the obtained singular vectors are sorted in descending order and plotted together in Fig.25. Also, the specific companies respective to all singular

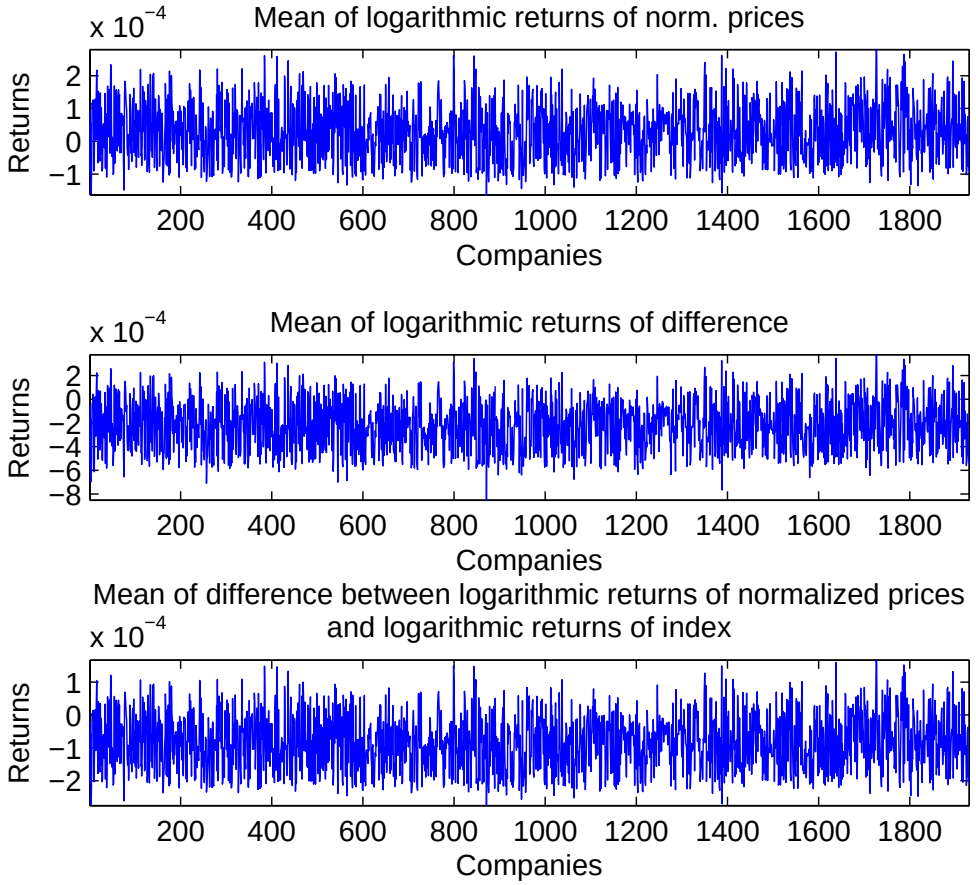


Figure 18: Mean of logarithmic returns.

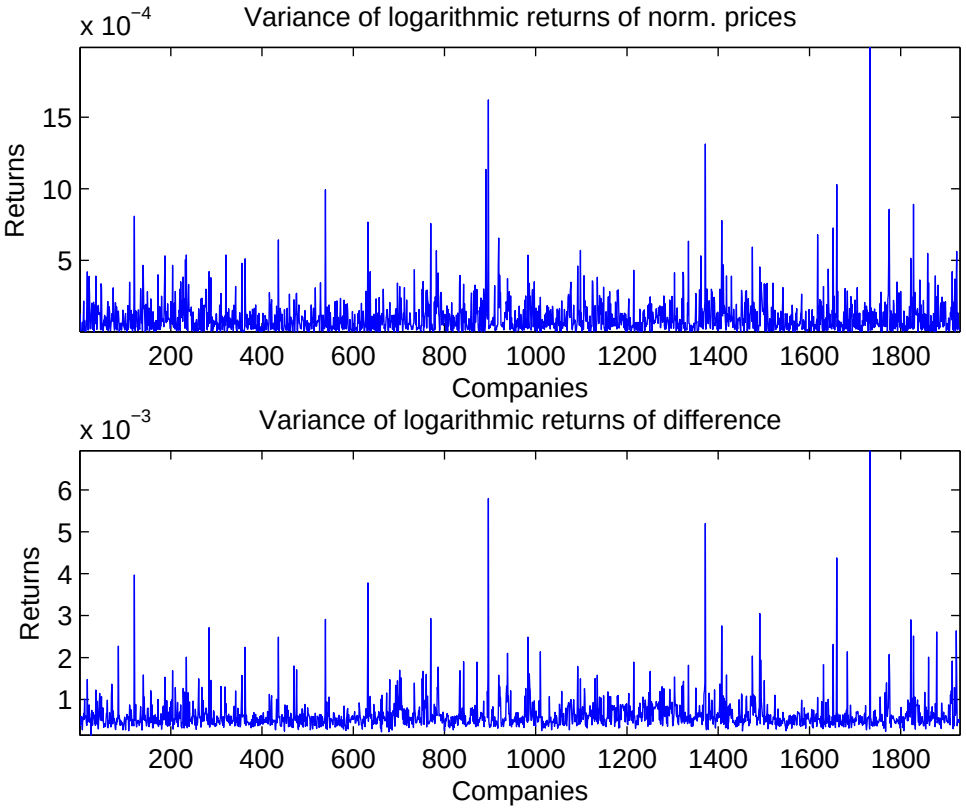


Figure 19: Variance of logarithmic returns.

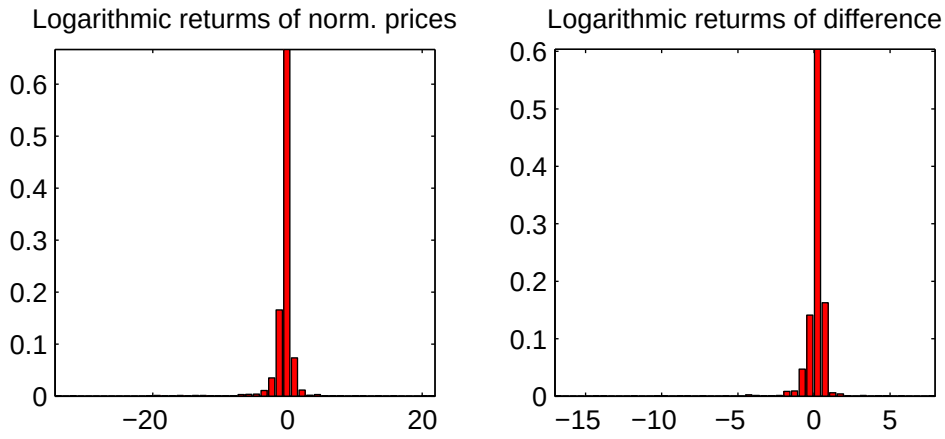


Figure 20: Skewness of logarithmic returns.

values can be identified.

The next step is to calculate a 20×20 co-occurrence matrix for these companies on the basis of how many same companies appear in the top 100 and bottom 100 companies in any of the four combinations, namely "Top-Top", "Top-Bottom", "Bottom-Top" and "Bottom-Bottom" of any given pair of SV's, ordered by descending values. If the permutations of 1930 companies are purely random, it can be expected that

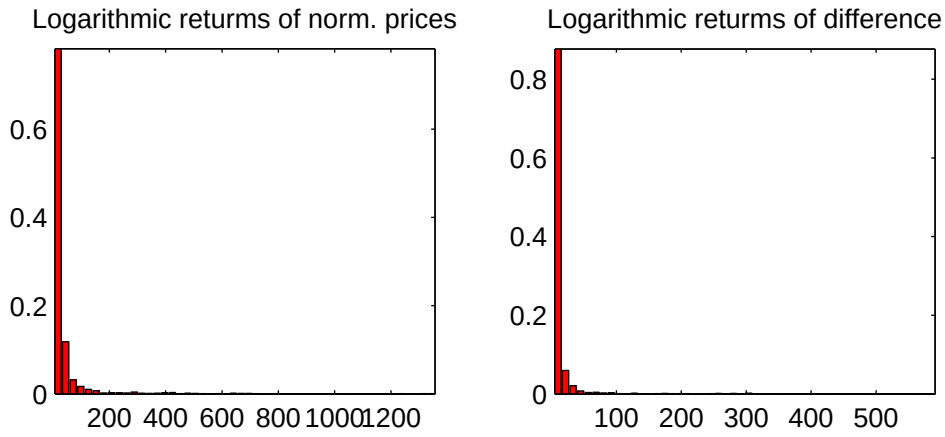


Figure 21: Kurtosis of logarithmic returns.

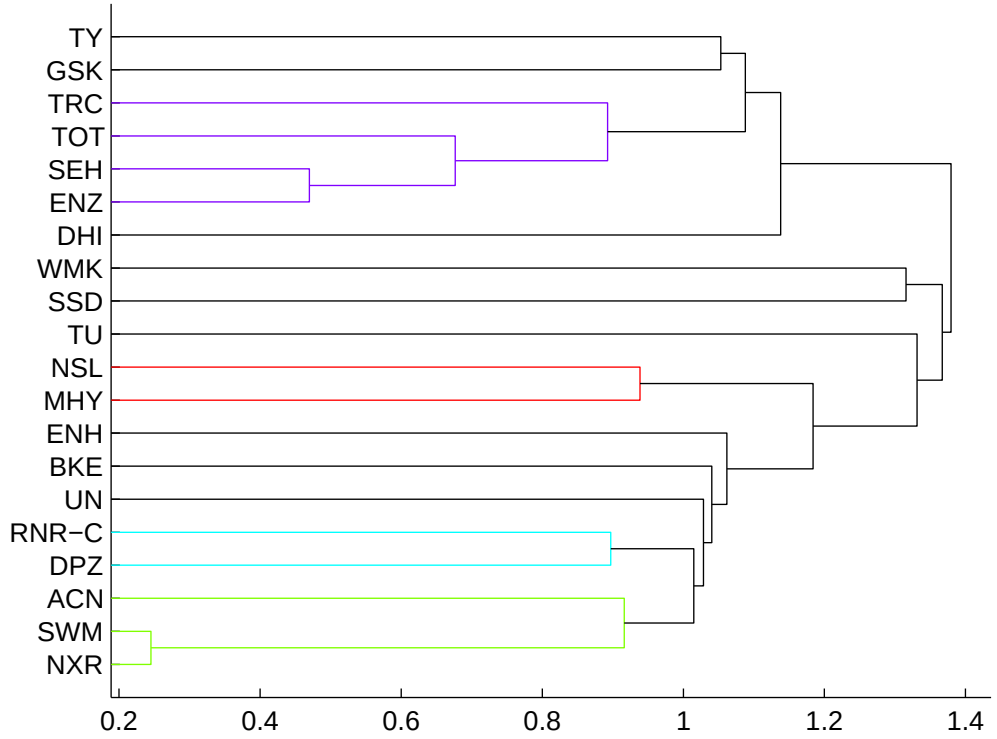


Figure 22: An example of the hierarchical cluster tree for the case of 20 randomly chosen stocks.

the co-occurrence ratio takes value of roughly 0.05 everywhere. An initial hypothesis is that type structure exists in the covariance matrix, and values both higher and lower than 0.05 would indicate some industry sector.

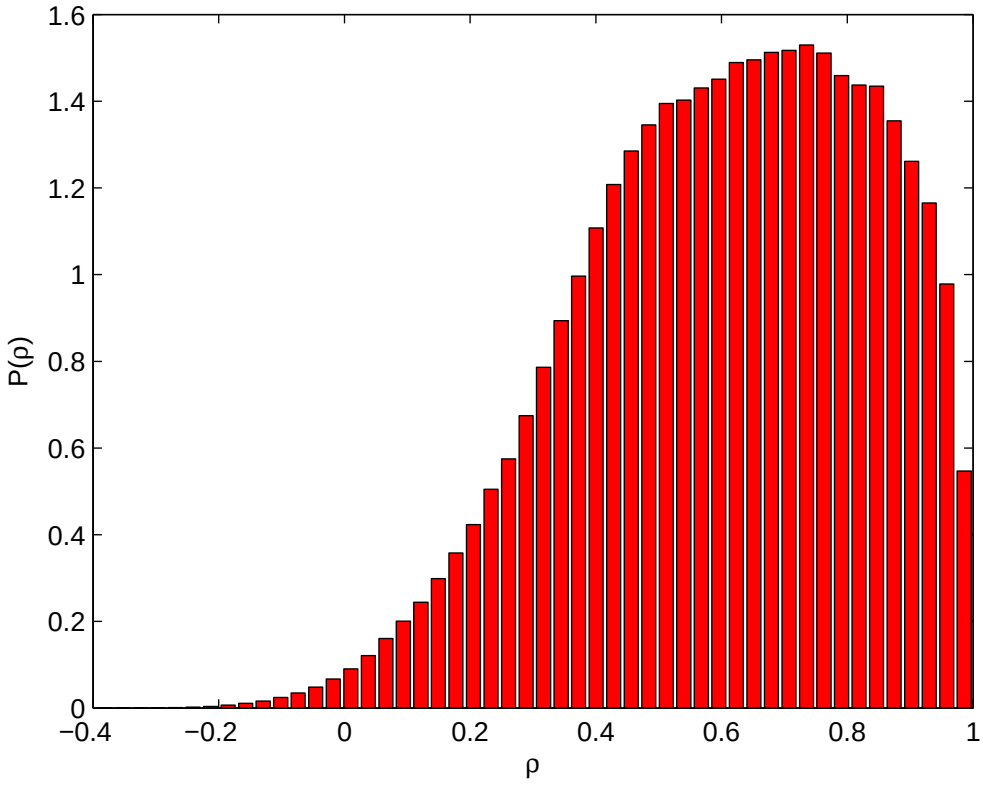


Figure 23: Distribution of correlation coefficients.

A co-occurrence matrix has the following form

$$\begin{bmatrix} Top-Top & Top-Bot \\ Bot-Top & Bot-Bot \end{bmatrix},$$

where "Top-Top", "Top-Bot", "Bot-Top" and "Bot-Bot" are the following 10×10 matrices

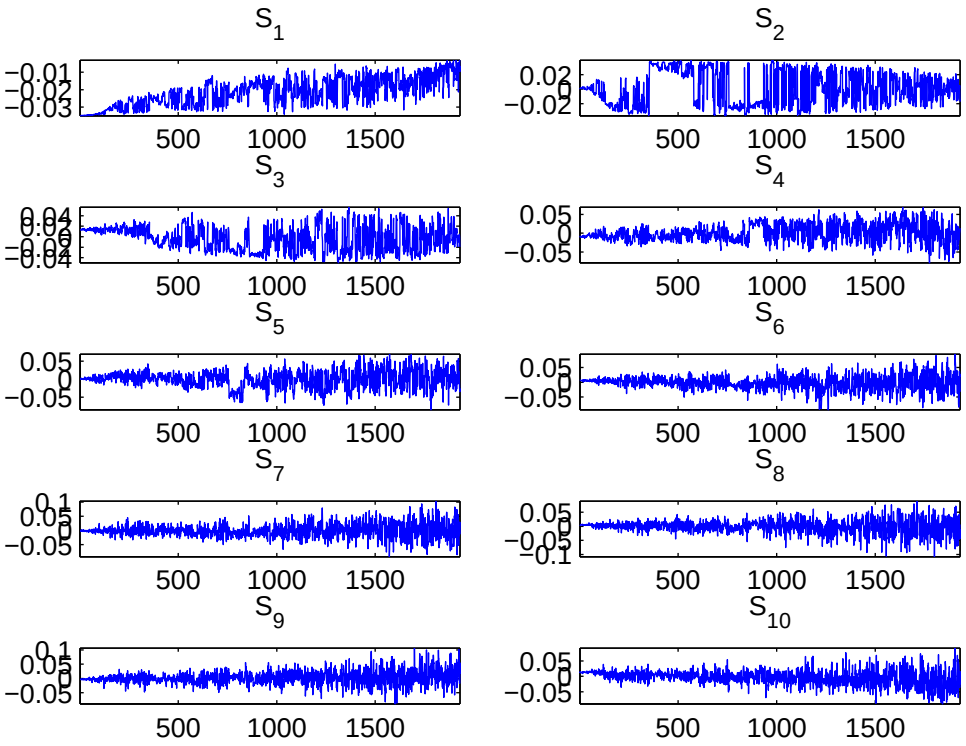


Figure 24: The 10 singular vectors, corresponded to 10 largest singular values.

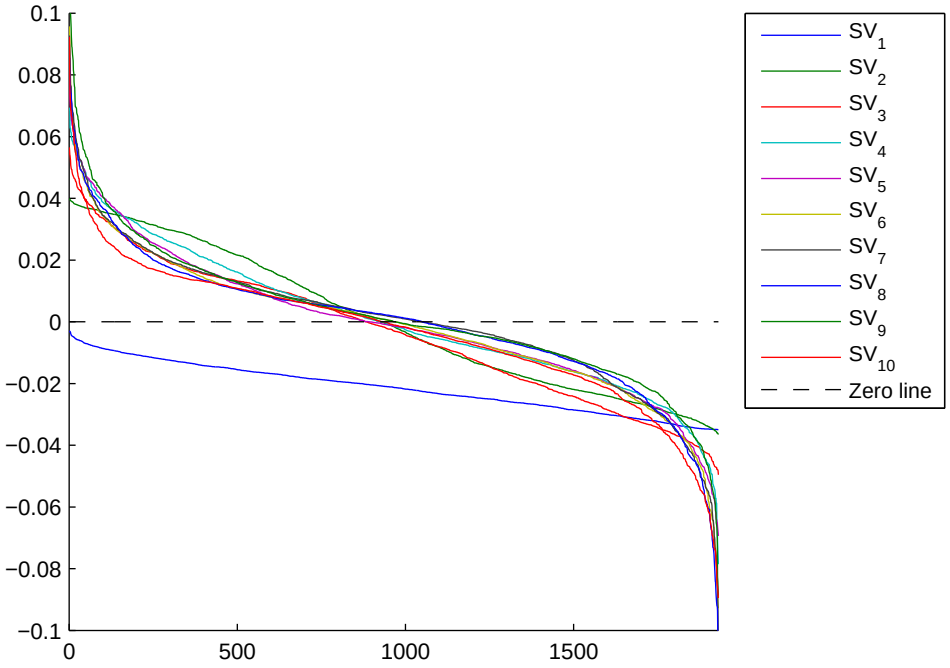


Figure 25: The 10 singular vectors. Values in vectors are sorted in descending order.

$$Top - Top = \begin{bmatrix} 1 & 0 & 0.04 & 0.01 & 0.07 & 0.07 & 0.02 & 0.13 & 0.19 & 0.06 \\ 0 & 1 & 0 & 0 & 0 & 0.01 & 0 & 0.01 & 0 & 0 \\ 0.04 & 0 & 1 & 0.07 & 0.02 & 0.05 & 0.05 & 0.09 & 0.16 & 0.05 \\ 0.01 & 0 & 0.07 & 1 & 0.04 & 0.11 & 0.05 & 0.07 & 0.07 & 0.07 \\ 0.07 & 0 & 0.02 & 0.04 & 1 & 0.04 & 0.04 & 0.07 & 0.12 & 0.09 \\ 0.07 & 0.01 & 0.05 & 0.11 & 0.04 & 1 & 0.1 & 0.13 & 0.14 & 0.09 \\ 0.02 & 0 & 0.05 & 0.05 & 0.04 & 0.1 & 1 & 0.09 & 0.15 & 0.19 \\ 0.13 & 0.01 & 0.09 & 0.07 & 0.07 & 0.13 & 0.09 & 1 & 0.13 & 0.08 \\ 0.19 & 0 & 0.16 & 0.07 & 0.12 & 0.14 & 0.15 & 0.13 & 1 & 0.15 \\ 0.06 & 0 & 0.05 & 0.07 & 0.09 & 0.09 & 0.19 & 0.08 & 0.15 & 1 \end{bmatrix},$$

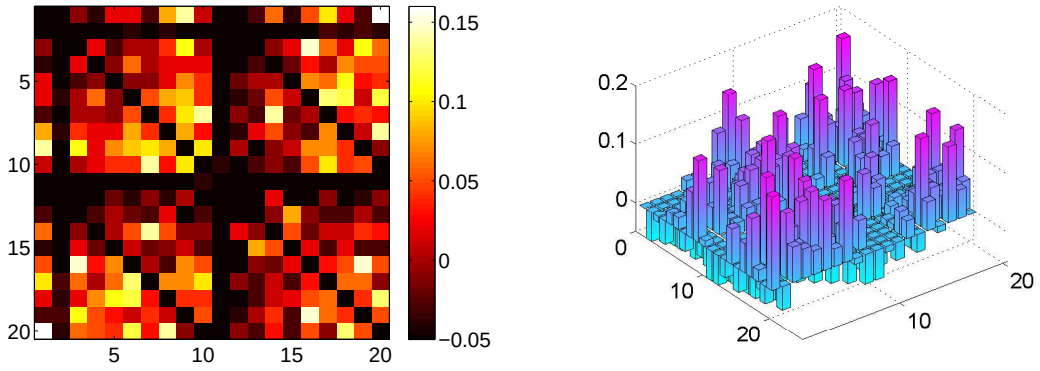
$$Top - Bot = \begin{bmatrix} 0 & 0 & 0.02 & 0.11 & 0.01 & 0.1 & 0.15 & 0.07 & 0.02 & 0.21 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.02 & 0.01 & 0.02 & 0.01 \\ 0 & 0 & 0 & 0.04 & 0.07 & 0.2 & 0.11 & 0.07 & 0.16 & 0.11 \\ 0 & 0 & 0.02 & 0 & 0.03 & 0.08 & 0.05 & 0.12 & 0.1 & 0.1 \\ 0 & 0.03 & 0.05 & 0.05 & 0 & 0.12 & 0.11 & 0.16 & 0.08 & 0.09 \\ 0 & 0.01 & 0.03 & 0.1 & 0.06 & 0 & 0.18 & 0.17 & 0.06 & 0.17 \\ 0 & 0.04 & 0.02 & 0.19 & 0.03 & 0.05 & 0 & 0.08 & 0.09 & 0.08 \\ 0 & 0.01 & 0.07 & 0.1 & 0.04 & 0.02 & 0.12 & 0 & 0.06 & 0.19 \\ 0 & 0.05 & 0 & 0.04 & 0.06 & 0.12 & 0.12 & 0.09 & 0 & 0.04 \\ 0.01 & 0 & 0.02 & 0.04 & 0.02 & 0.1 & 0.15 & 0.09 & 0.1 & 0 \end{bmatrix},$$

$$Bot - Top = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 \\ 0 & 0 & 0 & 0 & 0.03 & 0.01 & 0.04 & 0.01 & 0.05 & 0 \\ 0.02 & 0 & 0 & 0.02 & 0.05 & 0.03 & 0.02 & 0.07 & 0 & 0.02 \\ 0.11 & 0 & 0.04 & 0 & 0.05 & 0.1 & 0.19 & 0.1 & 0.04 & 0.04 \\ 0.01 & 0 & 0.07 & 0.03 & 0 & 0.06 & 0.03 & 0.04 & 0.06 & 0.02 \\ 0.1 & 0 & 0.2 & 0.08 & 0.12 & 0 & 0.05 & 0.02 & 0.12 & 0.1 \\ 0.15 & 0.02 & 0.11 & 0.05 & 0.11 & 0.18 & 0 & 0.12 & 0.12 & 0.15 \\ 0.07 & 0.01 & 0.07 & 0.12 & 0.16 & 0.17 & 0.08 & 0 & 0.09 & 0.09 \\ 0.02 & 0.02 & 0.16 & 0.1 & 0.08 & 0.06 & 0.09 & 0.06 & 0 & 0.1 \\ 0.21 & 0.01 & 0.11 & 0.1 & 0.09 & 0.17 & 0.08 & 0.19 & 0.04 & 0 \end{bmatrix},$$

$$Bot - Bot = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.07 & 0 & 0 & 0.04 & 0 & 0.01 & 0.04 \\ 0 & 0 & 1 & 0.04 & 0.13 & 0.04 & 0.02 & 0.02 & 0.05 & 0.02 \\ 0 & 0.07 & 0.04 & 1 & 0.1 & 0.03 & 0.06 & 0.06 & 0.09 & 0.08 \\ 0 & 0 & 0.13 & 0.1 & 1 & 0.07 & 0.02 & 0.07 & 0.02 & 0.02 \\ 0 & 0 & 0.04 & 0.03 & 0.07 & 1 & 0.08 & 0.1 & 0.2 & 0.1 \\ 0 & 0.04 & 0.02 & 0.06 & 0.02 & 0.08 & 1 & 0.11 & 0.11 & 0.04 \\ 0 & 0 & 0.02 & 0.06 & 0.07 & 0.1 & 0.11 & 1 & 0.09 & 0.18 \\ 0 & 0.01 & 0.05 & 0.09 & 0.02 & 0.2 & 0.11 & 0.09 & 1 & 0.1 \\ 0 & 0.04 & 0.02 & 0.08 & 0.02 & 0.1 & 0.04 & 0.18 & 0.1 & 1 \end{bmatrix}.$$

Notice, that a co-occurrence matrix is a symmetric matrix with ones on the main diagonal. Moreover, a matrix "Top-Bot" is same as a transposed matrix "Bot-Top", but neither of these matrices above is symmetric, unlike "Top-Top" and "Bot-Bot" that are.

This matrix can be presented as a heat map, in which the darker and the lighter colors correspond to the lowest and highest values, and as a heat bar, where the height of bars depends on the magnitude of value. Heat bar and a heat map are shown in Fig.26.



(a) Heat map. (b) Heat bar.

Figure 26: The visualization of a co-occurrence matrix.

Analysis of company names, which correspond to Top-20 and Bottom-20 values taken from S, show that these companies belong to different industry sectors and there is no strong cluster separation.

The time series of singular vectors can be calculated by the formula

$$SV_j = \sum_{i=1}^{1930} S_i^j * Dif_i,$$

where S_i is a matrix of singular vectors with size 1738×10, Dif_i - difference between normalized daily closure prices of stock i and S&P500 index, j means a day. This formula is a matrix multiplication. The matrix SV has a 10×1738 size. The time series of transposed SV is plotted in Fig.27.

For more detailed analysis the time interval was split into:

1. **First half.** It means the first 869 out of 1738 days.
2. **Second half.** It means the last 869 out of full set of days.
3. **Every second day.** Only odd days were taken into account.
4. **Random 30%.** That is a 521 days (this number corresponds to 30% of 1738) are randomly chosen days.

For each of these cases the same calculation was made as it was done for the full period of time. The singular vectors S were found, sorted according to distances and plotted in Fig.28. Vectors S, which values sorted by descending order, are presented

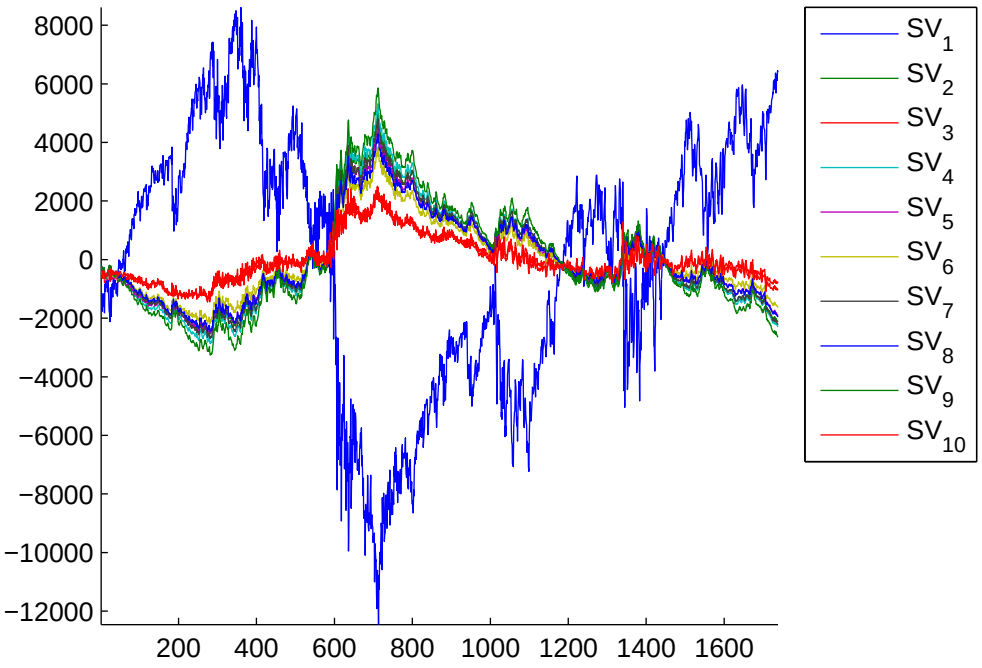


Figure 27: Time series of singular vectors.

in Fig.29. Heat maps, visually describing the co-occurrence matrices, are shown in Fig.30.The time series of singular vectors are plotted in Fig.31.

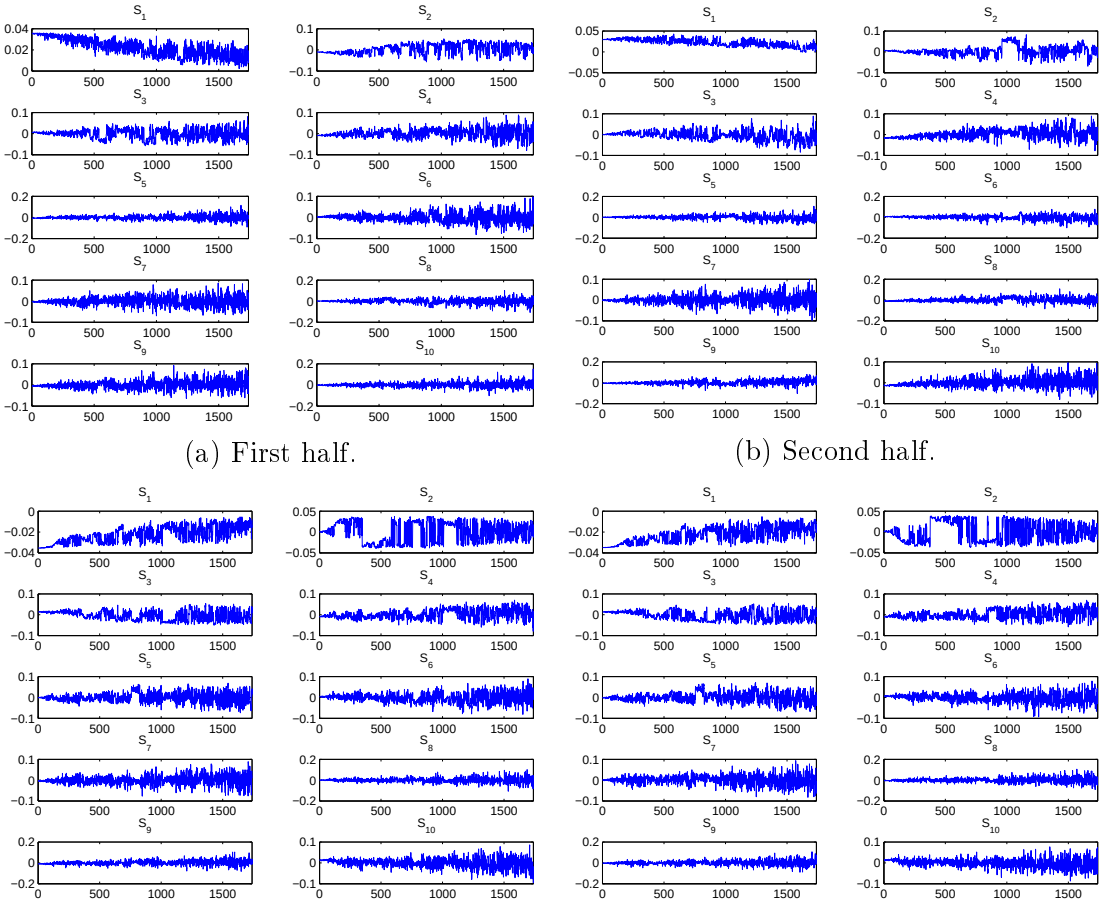
The cases of "First half" and "Second half" are more distinct from the case with full data than cases of "Random 30%" and "Every second day".

4.5 Parameter estimation.

Another hypothesis is that singular vectors are orthogonal over time. However, Fig.27 refutes this theory. Singular vectors do not behave as waves with different amplitude and frequency. However, there is some strong structure: some singular vectors can be obtained from another through multiplication by a constant. For verifying this statement a parameter estimation procedure is performed.

Let us choose from the matrix SV^τ two random vectors with size 1×1738, one of them suggested as a data vector X, second vector as a vector Y. Then a constant b can be found using the following expression

$$b = X \setminus Y.$$



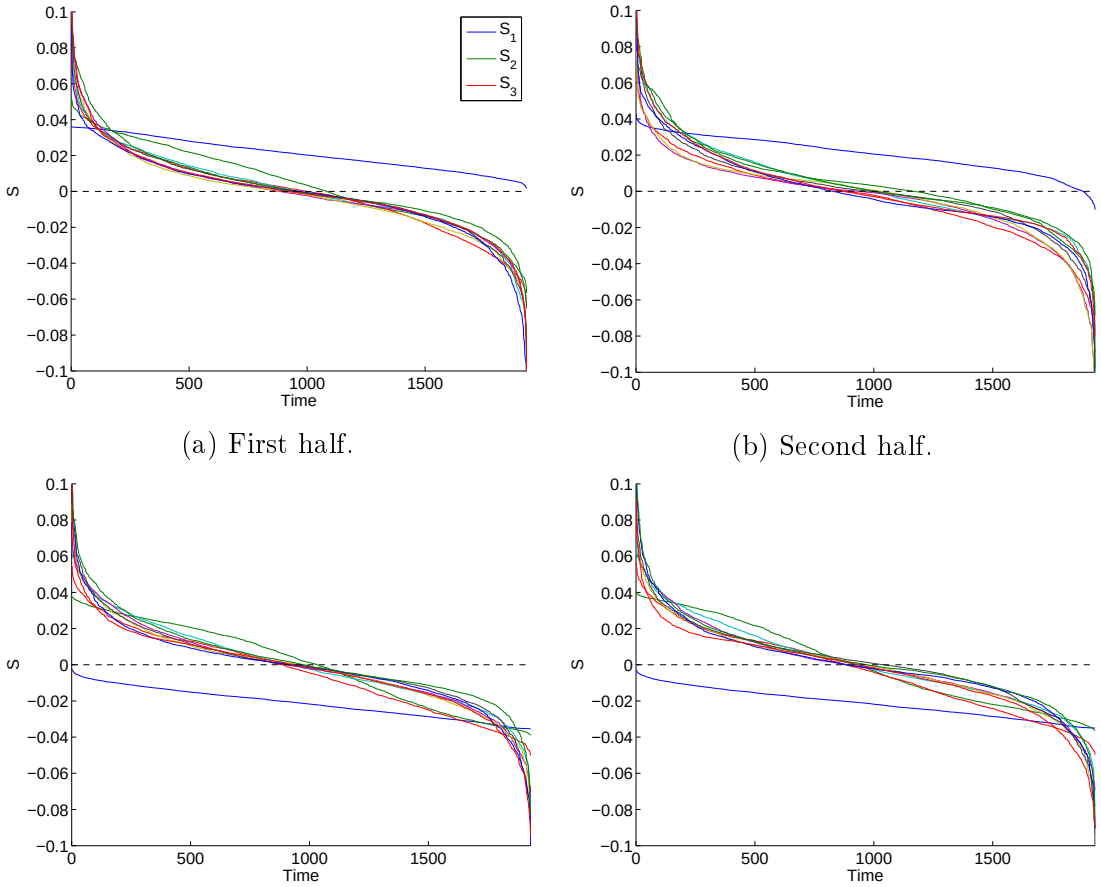
(a) First half. (b) Second half.

(c) Random 30%. (d) Every second day.

Figure 28: 10 singular vectors corresponded to 10 biggest singular values.

Estimated vector $Y_{est} = X * b$ is plotted together with initial vector Y. In Fig.32 several random cases are presented. For each case a constant b is calculated and collected in the following vector

$$B = \begin{bmatrix} 0.98918 \\ 0.37457 \\ 1.9415 \\ -0.26273 \\ 1.9368 \\ -0.30614 \\ 0.87869 \\ 1.1249 \\ 0.60399 \end{bmatrix}.$$



(a) First half.

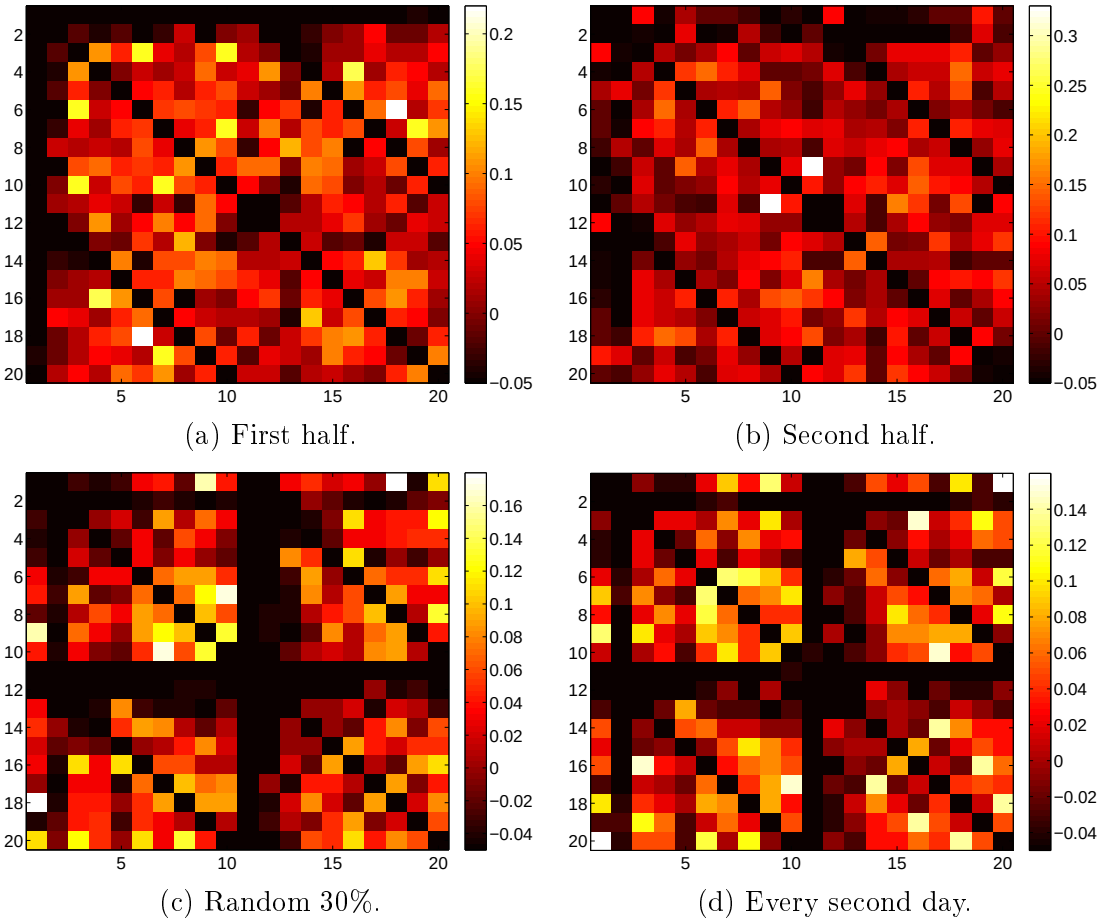
(b) Second half.

(c) Random 30%.

(d) Every second day.

Figure 29: The 10 singular vectors, sorted by descending order.

In addition, the same analysis is made for vectors taken from SV's obtained from 4 different cases described earlier. However, in this instance vectors X and Y are not chosen from the same SV matrix, the cases are also taken randomly. The designation $SV_j(i)$ means ith singular vector taken from jth case, $i = \overline{1,10}, j = \overline{1,4}$. Fig.33 shows that in many cases $SV_j(i)$ vectors differ from each other by a constant.



(a) First half.

(b) Second half.

(c) Random 30%.

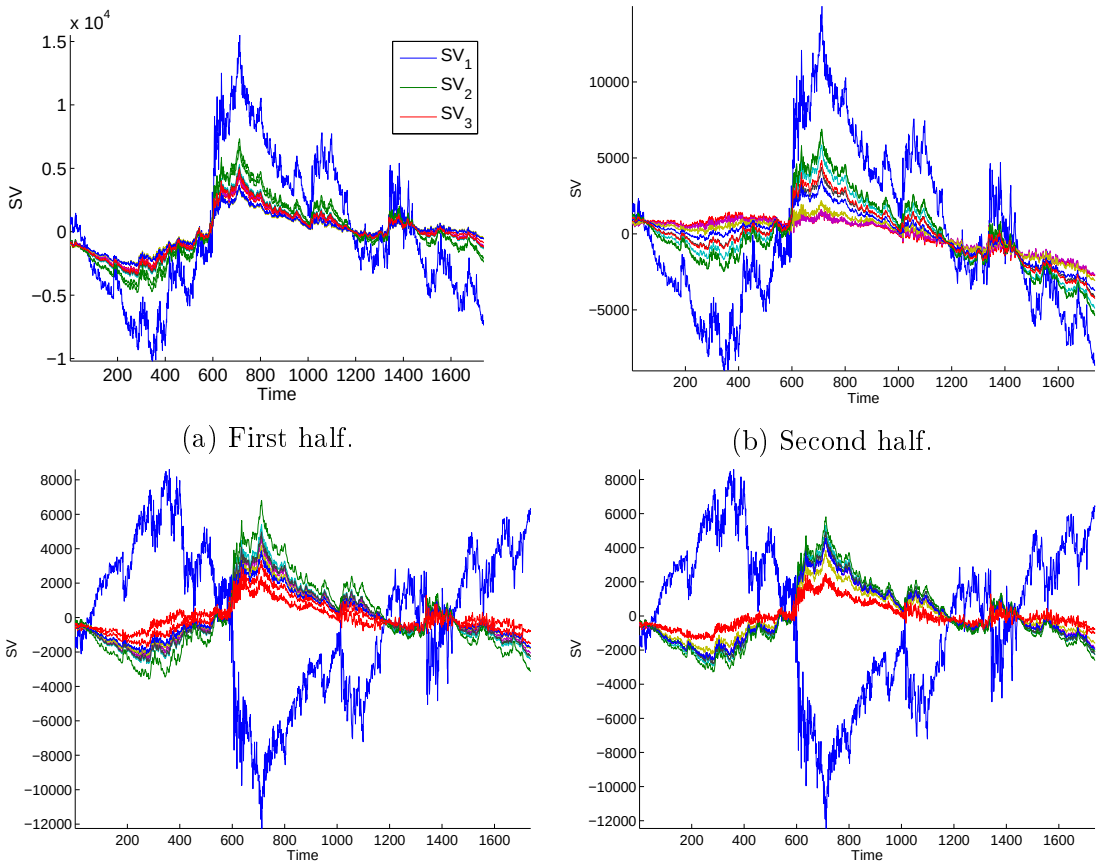
(d) Every second day.

Figure 30: Heat map corresponded to co-occurrence matrix.

Vector of obtained constants for each estimation is

$$B = \begin{bmatrix} 0.21982 \\ 0.64164 \\ 2.3383 \\ -2.6952 \\ 1.4719 \\ 0.79655 \\ 0.44419 \\ 1.0498 \\ 0.27961 \end{bmatrix}.$$

Moreover, vectors $SV_1(i)$ fit very well to vectors $SV_3(i)$ and $SV_4(i)$, but not to $SV_2(i)$.



(a) First half.

(b) Second half.

(c) Random 30%.

(d) Every second day.

Figure 31: Time series of singular vectors.

The set of graphics in Fig.34 support this observation.

$$B = \begin{bmatrix} b_1 & b_2 & b_3 \\ 0.95417 & -0.80695 & -0.801 \\ 0.74549 & 0.60495 & 0.54047 \\ -0.011353 & 0.44782 & 0.35428 \\ 0.77104 & 0.90129 & 0.85336 \\ 0.041168 & 0.88266 & 0.91045 \\ 0.1412 & 0.89387 & 0.95209 \\ 0.55197 & 0.87704 & 0.81965 \\ 0.41676 & 1.0235 & 1.0388 \\ 0.81609 & 0.81203 & 0.8056 \\ 0.66595 & 0.4637 & 0.42032 \end{bmatrix}$$

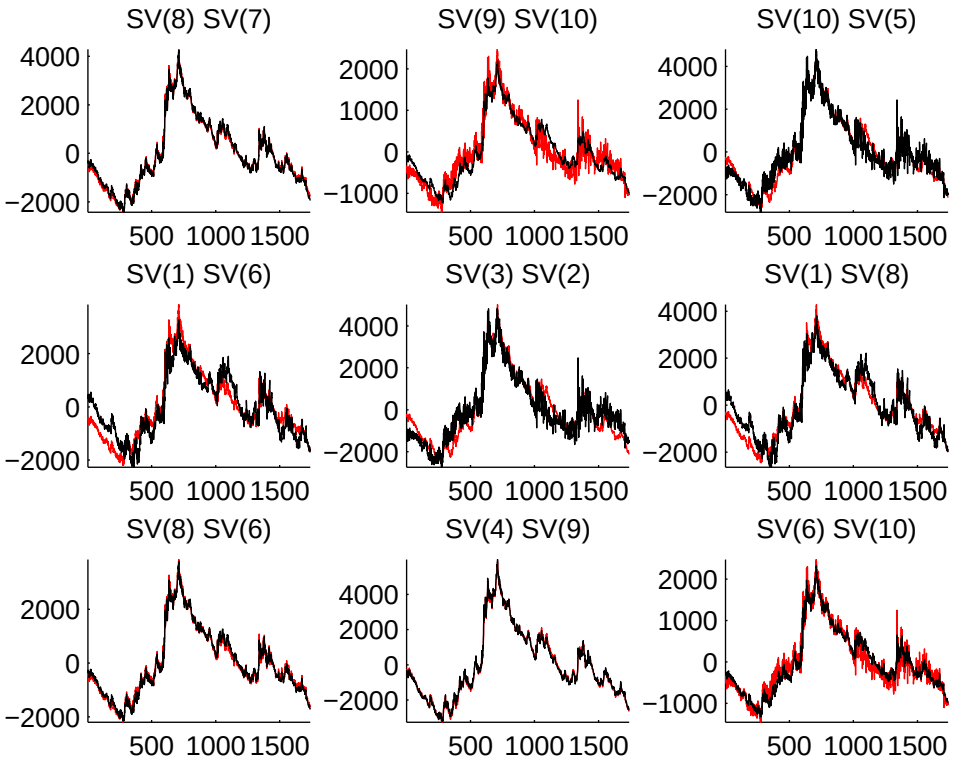


Figure 32: Parameter estimation. Red line is an initial vector Y, black line - vector Y_{est}. Numbers in a round brackets denote a number of vector, chosen from the matrix SV.

4.6 Moving average analysis of singular values.

One important issue is that no fixed subspace stays invariant under nonlinear dynamics. Therefore, no span of a small number of singular vectors can prevail for long, but eventually they all always blend into a random (ergodic) sample of the state space. Calculating SV's over 7 years explains why these vectors seem so random samples of all the stocks.

For this case subspaces do prevail over short time intervals, and then get replaced by new subspaces. Time interval is split into subintervals as follows:

1. **Every 6 month interval started every 3 months**, so that the intervals overlap by 3 months.

2. **Every 3 month interval started every 2 months**, so that the intervals overlap by 2 months.

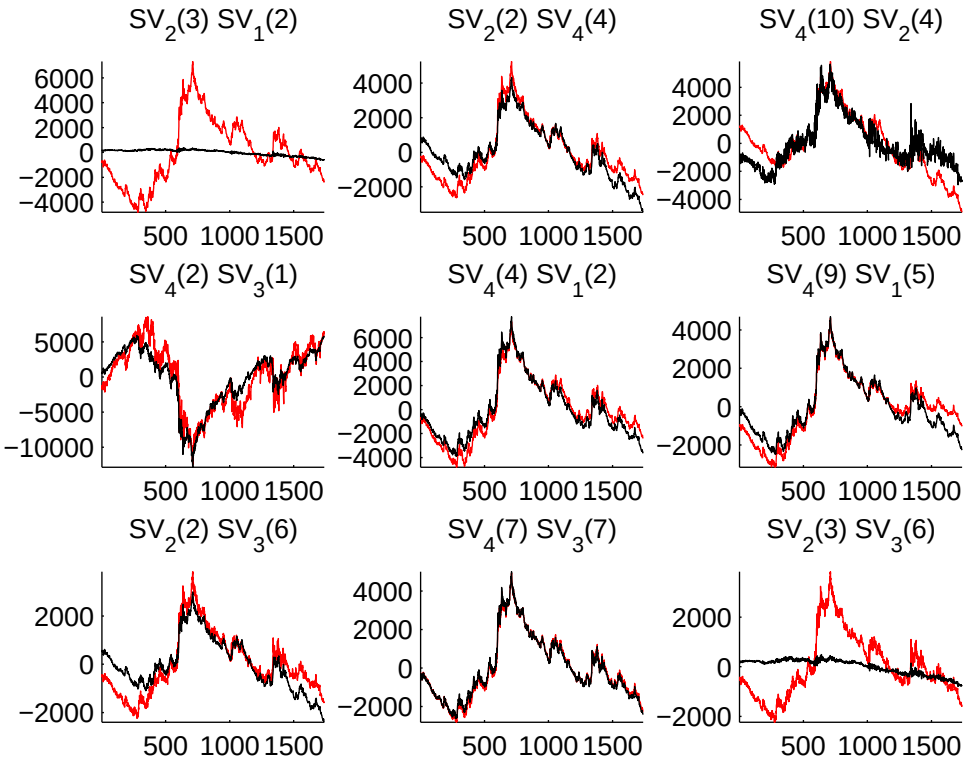


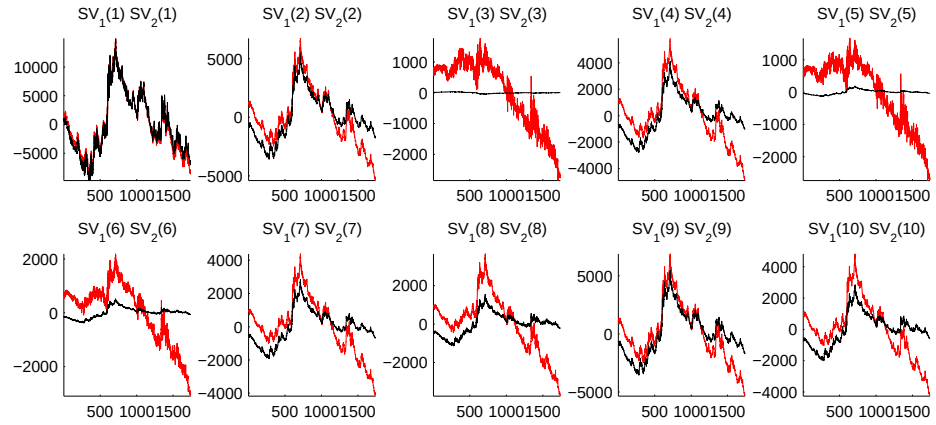
Figure 33: Parameter estimation. Red line is an initial vector Y, black line - vector Y_{est}.

3. **Every 12 month interval started every 6 months**, so that the intervals overlap by 6 months.

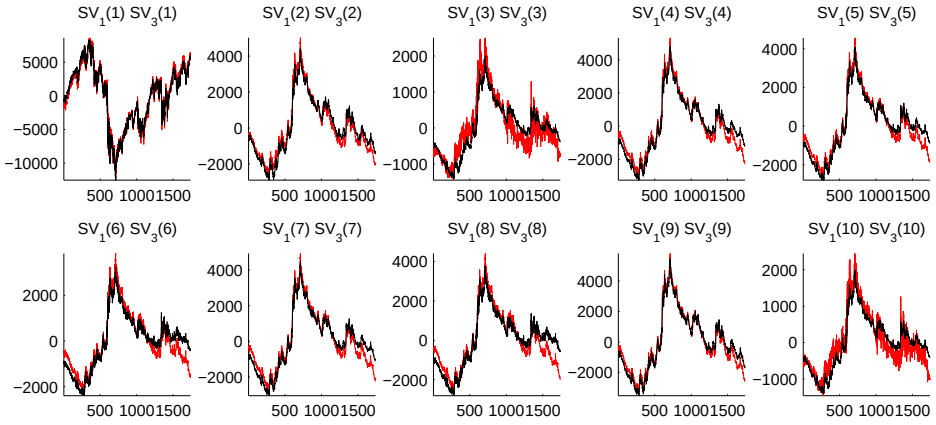
For each of the cases above the Top 10 singular values are calculated and the time series of the 10 singular values are plotted together in a Fig.35a, Fig.36a and Fig.37a. Vectors of singular values contain unnecessary zeroes which can be removed. Reduced vectors decreased their size by a factor of 10. The time series of reduced vectors, which contains only non-zero elements, are shown in Fig.35b, Fig.36b and Fig.37b accordingly to each type of splitting of time interval.

Let us consider only 3 out of the 10 largest singular values and denote the singular vectors, that correspond to them, as SV_1, SV_2 and SV_3, accordingly. The time series of the logarithm of singular values are present in Fig.35c, Fig.36c and Fig.37c. Fig.35c, Fig.36c and Fig.37c depict the time series of a number of common companies between singular vectors SV_1 and SV_2 and between SV_2 and SV_3.

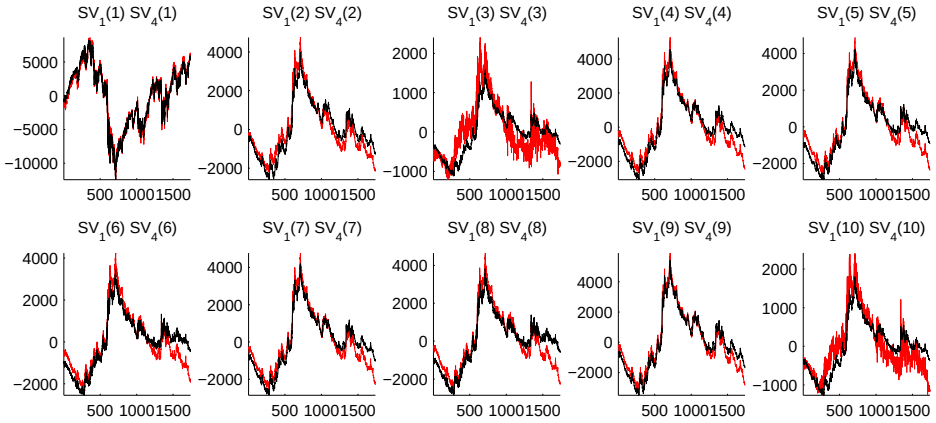
The main surmise is that the number of common companies becomes bigger when vectors SV_i's approach each other, and smaller when vectors diverge from each



(a) SV_1 and SV_2.



(b) SV_1 and SV_3.



(c) SV_1 and SV_4.

Figure 34: Parameter estimation.

other. For proving this statement, let us choose the second type of splitting, namely 3 month time interval with overlap of 2 month. When the number of common companies exceeds 10 (for vectors SV_1 and SV_2) and 5 (for vectors SV_2 and SV_3),

the colors between SV_i's are swapped, i.e whenever the overlap reaches 10 companies, the SV_i's change places: SV_1 turns into SV_2 and vice verse, and the same for SV_2 and SV_3. In Fig.38 one may see that the hypothesis are mostly confirmed: colors are not swapped when SV's are far from each other.

Next, for each 3 month time interval the next procedure is made: the values of $SV_1(i+1)$ and $SV_2(i)$, where i means ith interval, are swapped if a number of common companies between $SV_1(i)$ and $SV_2(i+1)$ is more than number between $SV_1(i)$ and $SV_1(i+1)$. So the basic principle is always that each SV time series follows the set of companies that maximizes the overlap of companies between time steps i and $i+1$, and then between $i+1$ and $1+2$ and so on. Fig.39 depicts the crossing between new time series of SV_1 and SV_2.

Sample autocorrelation with 95 percent confidence, which is provided in Fig.40, shows a time dependence and has some significant phenomena, which will be explored in future works. Each lag corresponds to a time interval of one month. According to obtained figures, for SV_1 the 1st and 33th lags (excluding value 0, where autocorrelation is always equal to 1) seem the most significant, meanwhile for SV_2 the most significant lag is only one and equals 3.

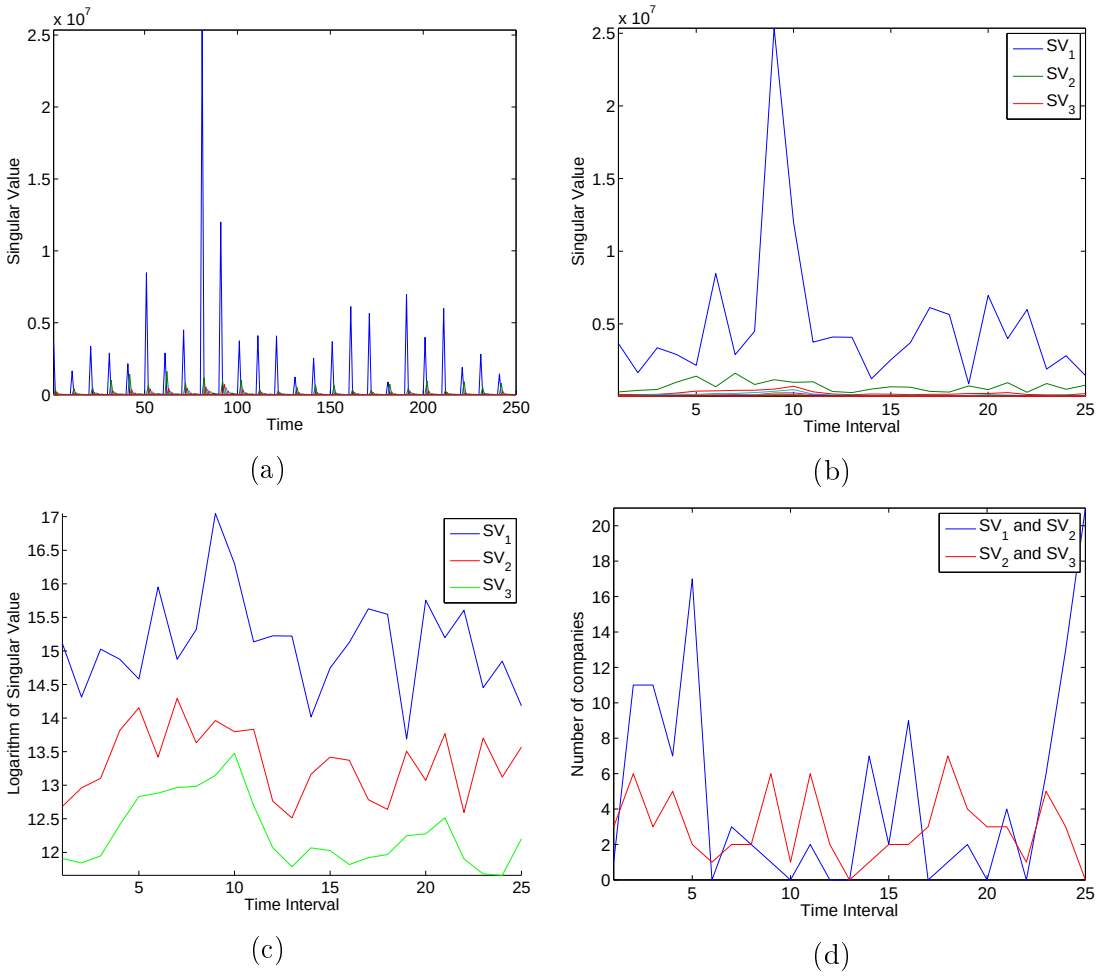


Figure 35: Interval of time 6 month with overlap 3 months. (a) Time series of singular values. (b) Time series of singular values with reducing zeros. (c) Time series of logarithm of 3 largest singular values. (d) Time series of number of common companies in sets with overlap at top-top plus bottom-bottom taken together between: blue line - SV_1 and SV_2, red line - SV_2 and SV_3.

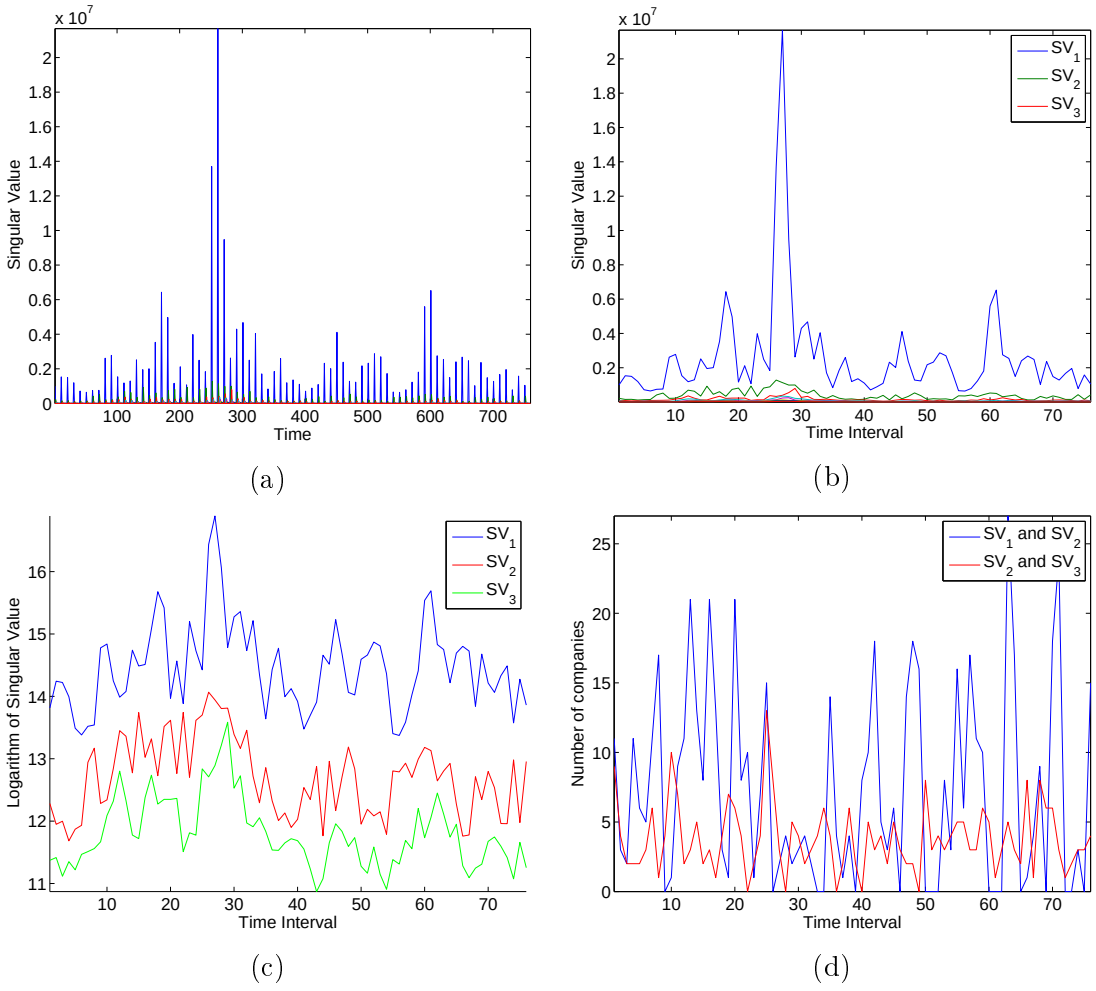


Figure 36: Interval of time 3 month with overlap 2 months. (a) Time series of singular values. (b) Time series of singular values with reducing zeros. (c) Time series of logarithm of 3 largest singular values. (d) Time series of number of common companies in sets with overlap at top-top plus bottom-bottom taken together between: blue line - SV_1 and SV_2, red line - SV_2 and SV_3.

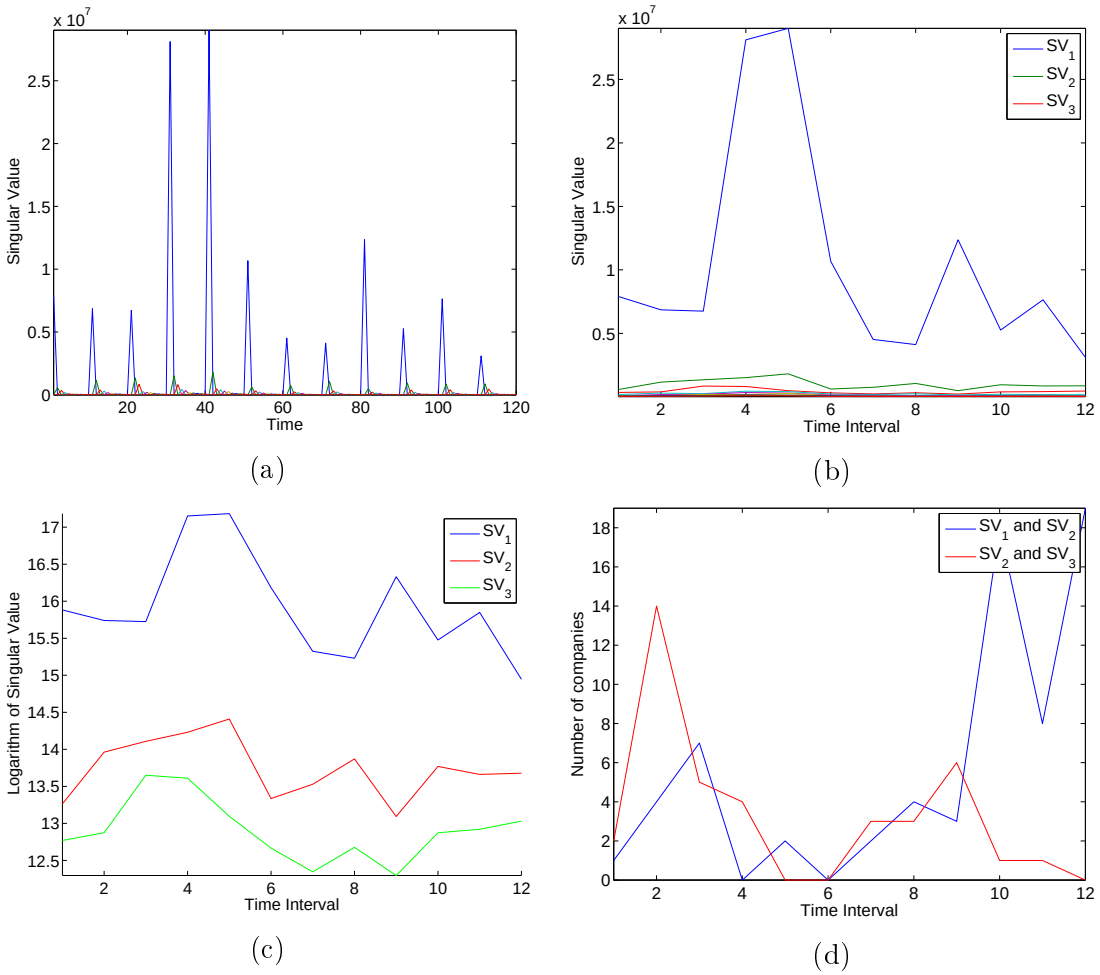
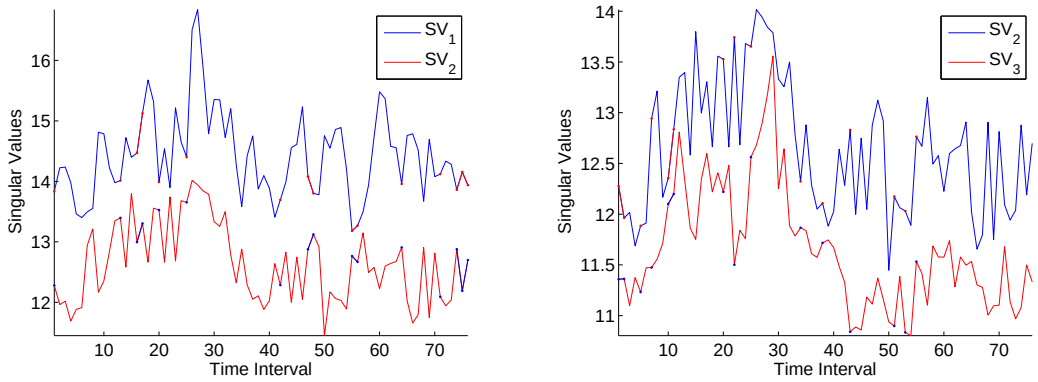


Figure 37: Interval of time 12 month with overlap 6 months. (a) Time series of singular values. (b) Time series of singular values with reducing zeros. (c) Time series of logarithm of 3 largest singular values. (d) Time series of number of common companies in sets with overlap at top-top plus bottom-bottom taken together between: blue line - SV_1 and SV_2, red line - SV_2 and SV_3.



(a) Cross-over between first and second singular values.

(b) Cross-over between second and third singular values.

Figure 38: Swapping the colors between SV_i's means that the overlap exceeds: 10 companies for SV_1 and SV_2, 5 companies for SV_2 and SV_3.

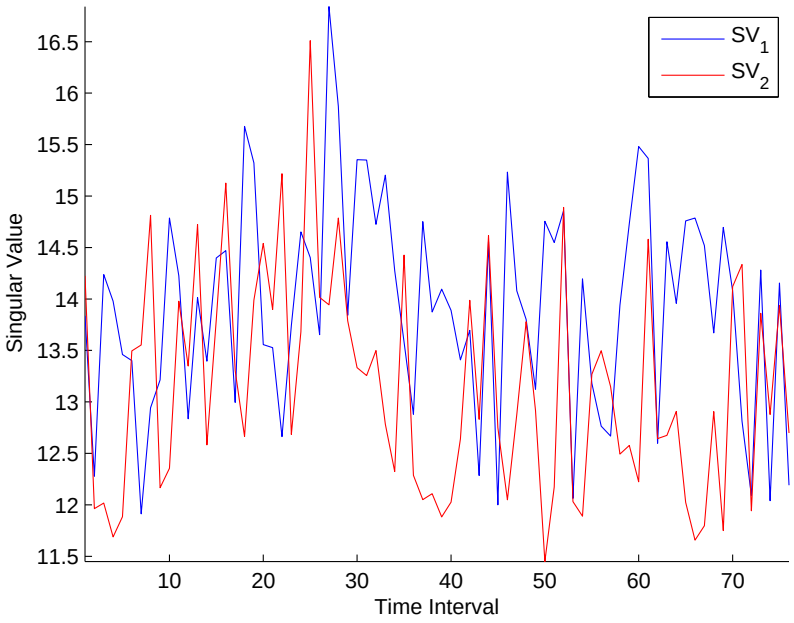


Figure 39: A values of $SV_1(i+1)$ and $SV_2(i)$ are swapped if a number of common companies between $SV_1(i)$ and $SV_2(i+1)$ is more than number between $SV_1(i)$ and $SV1(i+1)$.

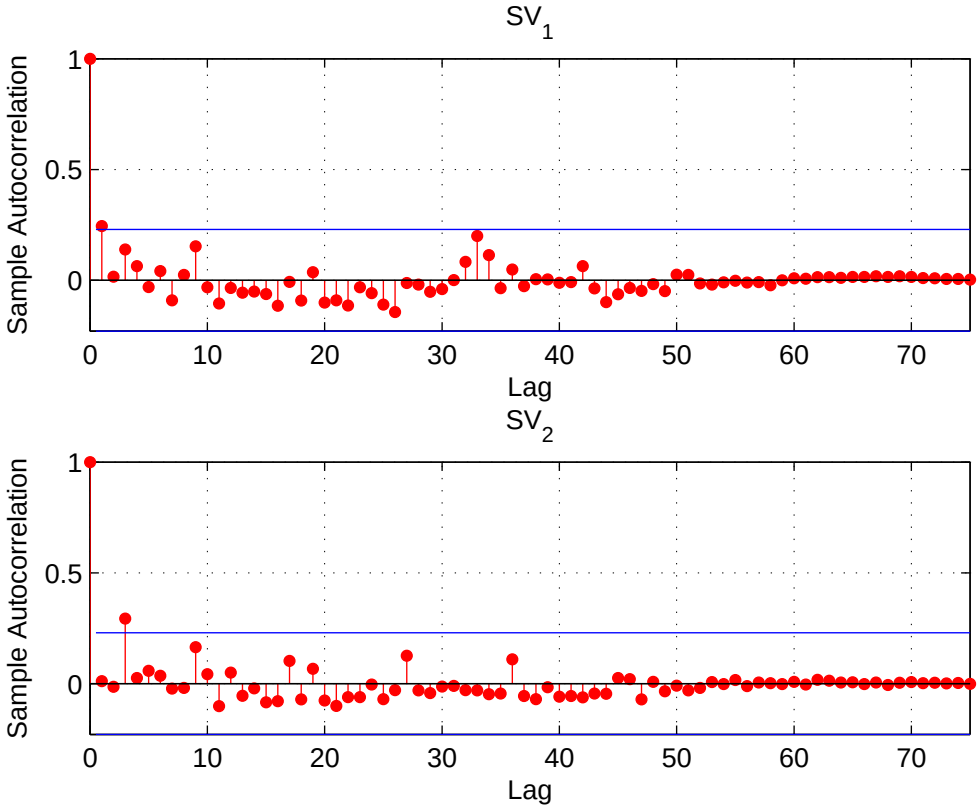


Figure 40: Sample autocorrelation diagram of singular values.

5 Conclusion

A stock exchange has an unstable and non-stationary structure, which greatly complicates the investigation and the possibility of forecasting. It is complicated to create a universal statistical model, which would allow to predict future changes in stock prices for a remote period of time with minimal risks. A statistical analysis helps to understand the nature of the market and its behavior. Many scientists dedicated their works to studying this problem, however no one totally understands a phenomenon of the market.

The aim of this work is to perform a basic statistical research for identifying possible factors, affecting the behavior of the NYSE, or finding the key features of the stock exchange. For these purposes there was calculated a correlation matrix C and a distance matrix D, based on which a hierarchical cluster tree is created, besides, a singular value decomposition is produced and examined.

The most significant findings, obtained during the study, are as follows:

- A distribution of correlation coefficient matrix, calculated for 1930 companies, taken from NYSE, looks more like an reflected Maxwell-Boltzmann distribution, than a Normal distribution, which has a bell-shaped form.
- The NYSE is not an ergodic system. It means, that a stable mean in a short period of time is not the same as a mean in a long period.
- Singular vectors, corresponded to largest singular values, are not orthogonal over time, but they are obtained from each other by a constant factor. This is different for different periods of time.

Nevertheless, this work has no clear findings, which would provide a new look at the behavior of the market, and a question of forecasting the market is still open. However, the results obtained in this thesis can be a good foundation for a future work, because there are still many unexplored tasks, such as more detailed study of singular vectors and companies, that correspond to them, detection of possible clusters and the relationships between industry sectors.

References

[1] Arnott, R.D., (1980) Cluster Analysis and Stock Price Comovement. *Financial analysis journal.* [Online] CFA Institute 36(6). p.56–62. Available from: http://www.cfapubs.org. [Accessed: 12th December 2014].

[2] Huang, W., Zhuang, X., Yao, S., (2009) A network analysis of the Chinese stock market. *Physica A: Statistical Mechanics and its Applications.* [Online] ScienceDirect 388(7). p.2956–2964. Available from: http://www.sciencedirect.com. [Accessed: 12th December 2014].

[3] Ivanov P.Ch., Yuen, A., Perakakis, P., (2014) Impact of Stock Market Structure on Intertrade Time and Price Dynamics. *PLoS One.* [Online] PLoS One 9(4). e92885. Available from: http://dx.doi.org/10.1371/journal.pone.0092885. [Accessed: 12th December 2014].

[4] Kauê, T., Peron, D., 1 Costa, L.F. and Rodrigues, F.A., (2012) The structure and resilience of financial market networks. *Chaos.* [Online] EBSCO 22(3). p013117. Available from: http://www.ebsco.com. [Accessed: 12th December 2014].

[5] Mantegna, R.N. and Stanley, H.E., (2000) Correlation and anticorrelation between stocks. In: Mantegna, R.N. and Stanley, H.E., *Introduction to econophysics. Correlations and Complexity in Finance.*. Cambridge: The Press Syndicate of the University of Cambridge.

[6] Mantegna, R.N. and Stanley, H.E., (2000) Taxonomy of a stock portfolio. In: Mantegna, R.N. and Stanley, H.E., *Introduction to econophysics. Correlations and Complexity in Finance.*. Cambridge: The Press Syndicate of the University of Cambridge.

[7] Prat-Ortega, G., Savel'ev, S.E., (2009) Stochastic GARCH dynamics describing correlations between stocks. *Physica A: Statistical Mechanics and its Applications.* [Online] ScienceDirect 410(9). p.623–627. Available from: http://www.sciencedirect.com. [Accessed: 12th December 2014].

[8] Rosenow, B., Plerou, V., Gopikrishnan, P., Amaral, L.A.N. and Stanley, H.E., (2000) Application of random matrix theory to study cross-correlation of stock prices. *International Journal of Theoretical and Applied Finance.* [Online] World Scientific 3(9). p.399–403. Available from: http://www.worldscientific.com. [Accessed: 12th December 2014].

[9] Rosenow, B., Gopikrishnan, P., Plerou, V., Stanley, H.E., (2002) Random magnets and correlations of stock price fluctuations. *Physica A: Statistical Mechanics and its Applications.* [Online] ScienceDirect 314(1). p.762–767. Available from: http://www.sciencedirect.com. [Accessed: 12th December 2014].

[10] Rosenow, B., Gopikrishnan, P., Plerou, V., Stanley, E.H., (2003) Dynamics of cross-correlations in the stock market. *Physica A: Statistical Mechanics and its Applications.* [Online] ScienceDirect 324(6). p.241–246. Available from: http://www.sciencedirect.com. [Accessed: 12th December 2014].

[11] Tse, C.K., Liu, J., Lau, F.C.M., (2010) A network perspective of the stock market. *Journal of Empirical Finance.* [Online] ScienceDirect 17(9). p.659–667. Available from: http://www.sciencedirect.com. [Accessed: 12th December 2014].

[12] Yuan, Y., Zhuang, X., Liu, Z., (2012) Price-volume multifractal analysis and its application in Chinese stock markets. *Physica A: Statistical Mechanics and its Applications.* [Online] ScienceDirect 391(6). p.3484–3495. Available from: http://www.sciencedirect.com. [Accessed: 12th December 2014].

[13] http://www.wikipedia.org

List of Tables

1 The observed minimum and maximum values $\rho(ij)$ for different years. 17

List of Figures

1 Time series of daily closure NYSE stock prices. 13

2 Time evolution of a logarithm daily closure prices. 14

3 Distribution of correlation coefficients for whole period of time. 15

4 Maxwell-Boltzmann distribution. Red line matches to temperature equals to 350° K, blue - 800° K. 16

5 Observation pairs of stocks that correspond to the maximum and the minimum correlation coefficient, respectively. 17

6 Distribution of correlation coefficients for different years. 18

7 Stock market indexes. 19

8 Difference between daily closure prices normalized by regression model and market index. 19

9 Natural logarithm of difference between daily closure prices normalized by regression model and market index. 20

10 Difference between daily closure prices and index normalized by mean and standard deviation. 20

11 Natural logarithm of difference between daily closure prices and index normalized by mean and standard deviation. 21

12 A descriptors of the shape of a probability distribution. 22

13 Histogram of individual stock price over time. Left picture - normalized prices, right - logarithmic returns. 23

14 Histogram of all stocks on a single day. Left picture - normalized prices, right - logarithmic returns. 24

15 Histogram of the index and logarithmic returns over time. 25

16 Histogram of difference and logarithmic returns over time. 26

17 Histogram of the difference and logarithmic returns over companies at one single day. 27

18 Mean of logarithmic returns. 28

19 Variance of logarithmic returns. 29

20 Skewness of logarithmic returns. 29

21 Kurtosis of logarithmic returns. 30

22 An example of the hierarchical cluster tree for the case of 20 randomly chosen stocks. 30

23 Distribution of correlation coefficients. 31

24 The 10 singular vectors, corresponded to 10 largest singular values. . 32

25 The 10 singular vectors. Values in vectors are sorted in descending order. 32

26 The visualization of a co-occurrence matrix. 35

27 Time series of singular vectors. 36

28 10 singular vectors corresponded to 10 biggest singular values. 37

29 The 10 singular vectors, sorted by descending order. 38

30 Heat map corresponded to co-occurrence matrix. 39

31 Time series of singular vectors. 40

32 Parameter estimation. Red line is an initial vector Y, black line - vector Y_{est}. Numbers in a round brackets denote a number of vector, chosen from the matrix SV. 41

33 Parameter estimation. Red line is an initial vector Y, black line - vector Y_{est}. 42

34 Parameter estimation. 43

35 Interval of time 6 month with overlap 3 months. (a) Time series of singular values. (b) Time series of singular values with reducing zeros. (c) Time series of logarithm of 3 largest singular values. (d) Time series of number of common companies in sets with overlap at top-top plus bottom-bottom taken together between: blue line - SV_1 and SV_2, red line - SV_2 and SV_3. 45

36 Interval of time 3 month with overlap 2 months. (a) Time series of singular values. (b) Time series of singular values with reducing zeros. (c) Time series of logarithm of 3 largest singular values. (d) Time series of number of common companies in sets with overlap at top-top plus bottom-bottom taken together between: blue line - SV_1 and SV_2, red line - SV_2 and SV_3. 46

37 Interval of time 12 month with overlap 6 months. (a) Time series of singular values. (b) Time series of singular values with reducing zeros. (c) Time series of logarithm of 3 largest singular values. (d) Time series of number of common companies in sets with overlap at top-top plus bottom-bottom taken together between: blue line - SV_1 and SV_2, red line - SV_2 and SV_3. 47

38 Swapping the colors between SV_i's means that the overlap exceeds: 10 companies for SV_1 and SV_2, 5 companies for SV_2 and SV_3. 47

39 A values of $SV_1(i+1)$ and $SV_2(i)$ are swapped if a number of common companies between $SV_1(i)$ and $SV_2(i+1)$ is more than number between $SV_1(i)$ and $SV1(i+1)$. 48

40 Sample autocorrelation diagram of singular values. 48