Ministry of Science and Higher Education of Russian Federation
FGAEU HE “Samara State University of Economics”
Institute of National and International Economics

Bachelor’s Thesis:

«Dynamic portfolio optimisation based on stock market network clustering»

Student: Kuchko Anna Iurievna
Field of Study: 38.03.01 Economics
Program: Business-Analytics and Statistics
Academic Advisor: Prof. Persteneva Natalya Pavlovna

- **Research Goal and Objectives**
- **Theoretical Foundations of Developing the Dynamic Strategy for Portfolio Management**
- **Applied Research of Portfolio Strategy Optimisation**
- **Attachments**

The goal of this research is to conduct the mathematical and statistical optimisation of the dynamic portfolio investment strategy with the means of financial networks clustering

Objectives

1. Conduct a literature revies of the recent developments in the area of dynamic portfolio strategy optimisation using the methods of clustering analysis of financial market networks

2. Analyse the evolution of topological characteristics of Russian and American financial markets networks during different stages of financial market cycle

3. Develop a Python module for efficient optimal portfolio strategy selection based on the methods studied

4. Conduct a modelling of optimal portfolio investment strategy for Russian and American markets using the developed framework

The relevance of the work is determined by the limitations of available methods and tools for choosing an optimal investment portfolio in a timely manner (considering changeability of market conditions)

Research object

investment portfolio strategy

Research subject

mathematical and statistical methods of optimization of dynamic investment strategy

Scientific novelty and practical significance:

- Comprehensive assessment of using financial network modelling tools to select an optimal investment strategy effectiveness
- Development of a specialized software package to increase the efficiency and convenience of developing an investment strategy for interested investors

Data Resources

Investing.com

МОСКОВСКАЯ БИРЖА

STANDARD & POOR'S

- Research Goal and Objectives
- Theoretical Foundations of Developing the Dynamic Strategy for Portfolio Management
- Applied Research of Portfolio Strategy Optimisation
- Attachments

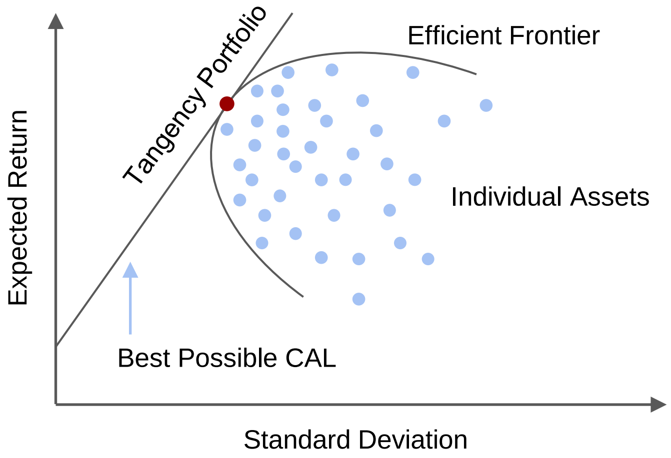
The task of optimizing the investment strategy is to maximize the expected return at a given level of risk

Two perspectives on the optimisation function

1. $E(R_{t+\varphi+\Delta t}) \longrightarrow max$ given $\sigma^2_{t+\varphi+\Delta t} \leq \sigma'^2_{t+\varphi+\Delta t}$,
2. $\sigma^2_{t+\varphi+\Delta t} \longrightarrow min$ given $R_{t+\varphi+\Delta t} \geq R'_{t+\varphi+\Delta t}$,

where $E(R_{t+\varphi+\Delta t})$ – mathematical expectation, $\sigma^2_{t+\varphi+\Delta t}$ – variation of portfolio returns given the maximum level of anticipated risk $\sigma'^2_{t+\varphi+\Delta t}$ or minimum level of anticipated return $R'_{t+\varphi+\Delta t}$, for the investment horizon $(t+\varphi, t+\varphi+\Delta t]$.

Classical solution - Modern Portfolio Theory (MPT)



Approach Drawbacks

- theoretical prerequisites of the approach are not observed in reality
- the assessment of the expected return and variation of individual stocks is not a sufficient argument in favour of making an informed decision on the selection of stocks for the portfolio
- computational complexity when introducing additional conditions and constraints

In a financial network, a vertex is a representation of a specific stock; an edge with a weighting factor is a measure of the distance between a pair of stocks

Distance measures

- Ultrametric distance based on Pearson correlation
- Mean Manhattan distance based on Theil index
- Mean Manhattan distance based on Atkinson index
- Dynamic Time Warping

Let $d_i = \sum s_{ij}$ be the i-th vertex degree, s_{ij} – weights of the adjacent edges, and D – diagonal matrix with diagonal elements d_i. In unweighted undirected graph vertex degree is equal to the number of adjacent edges

Topological characteristics of stock market network

- Degree (or valency) of the graph vertex
- Betweenness centrality
- Degree centrality
- Eigenvalue centrality

To allow a cluster analysis and identification of “central” and “peripheral” stocks in the network one needs to calculate relevant topological parameter (characteristics) of the network

Based on the network model of the market, it is possible to select stocks in a portfolio using a certain set of presumably optimal rules

1. Random selection
2. Industry based
3. Clusters based
4. Based on industry groups inside clusters

Clustering Algorithms

- Kmeans
- Spectral
- SpectralKmeans
- Kmedoids
- GaussianMixture
- SpectralGaussianMixture
- Hierarchical

- Research Goal and Objectives
- Theoretical Foundations of Developing the Dynamic Strategy for Portfolio Management
- **Applied Research of Portfolio Strategy Optimisation**
- Attachments

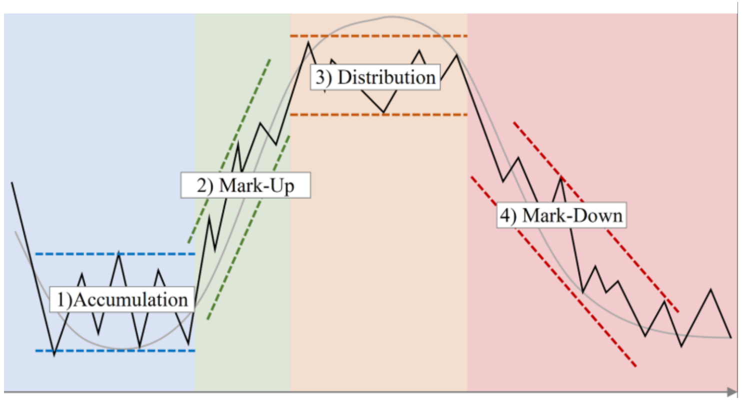
The FiNetwork software package was developed for the purposes of the research

Basic Functionality

- Automatic data collection from the Investing.com website for the required period of time
- Data pre-processing
- Building a weighted network graph of the stock market based on selected metrics
- Identification, analysis and visualization of the market cycle stages
- Cluster graph analysis
- Selection of the optimal portfolio according to one of the criteria
- Validation of results for different phases of the market cycle
- Graph filtering and visualization
- Calculation of related statistics

The Wyckoff theory was used to distinguish the four stages of the stock market for accounting to the market environment (1/3)

Wyckoff Market Cycle Stages



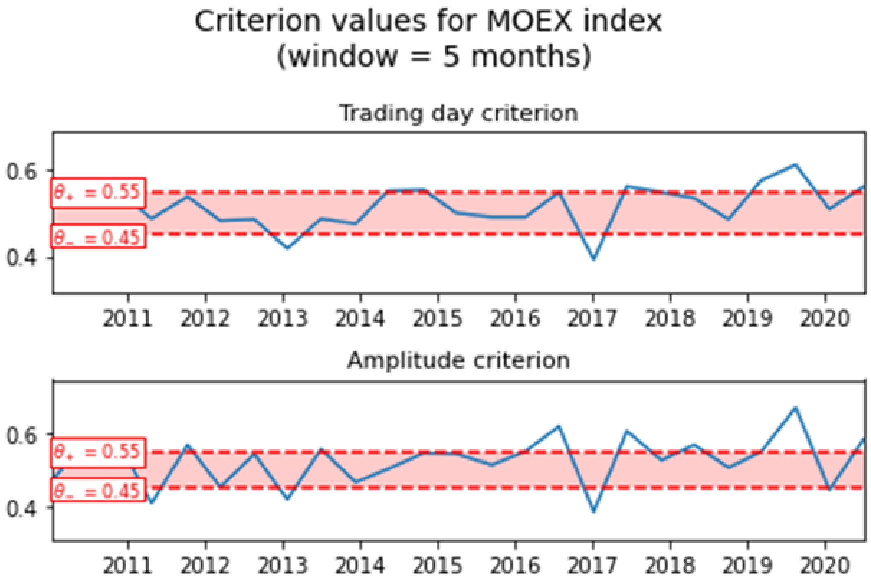
To "mark" the stages of the market cycle, the values of related market indices and coefficients based on those were used

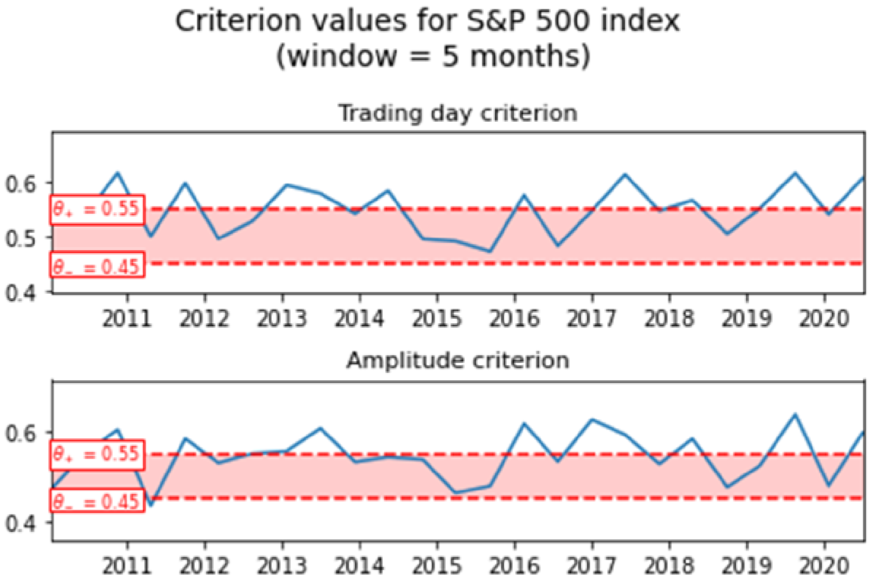
- Trading day coefficient r_d
- Amplitude coefficient r_f

Criterions for Identifying Market Cycle Stages

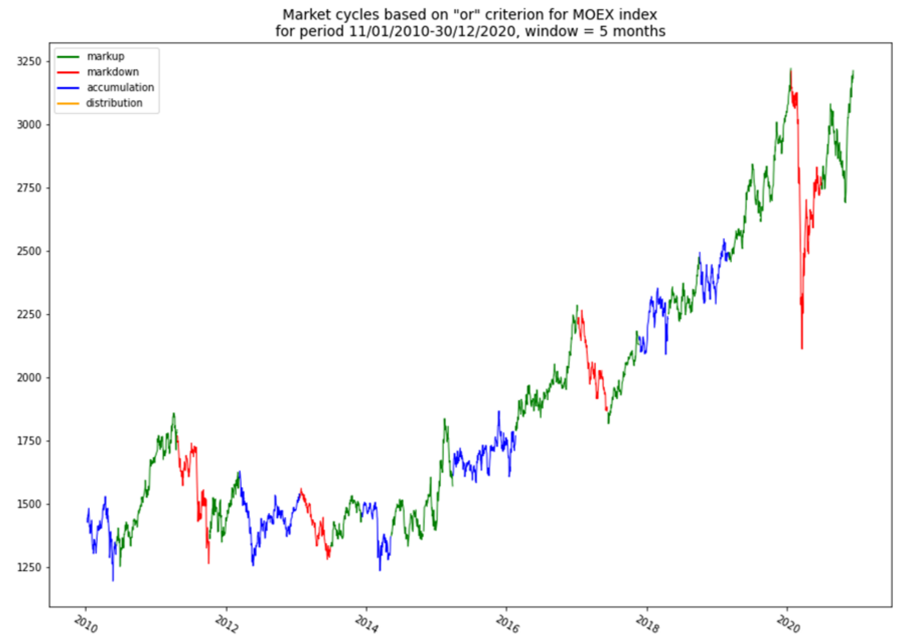
Stage of market cycle	Trading day criterion	Amplitude criterion	"AND" criterion	"OR" criterion
Accumulation	$\theta_- \le r_d \le \theta_+$ and previous stage – Mark-Down	$\theta_- \le r_f \le \theta_+$ and previous stage – Mark-Down	$\theta_- \le r_d \le \theta_+$ *or* $\theta_- \le r_f \le \theta_+$ and the previous stage– Mark-Down	$\theta_- \le r_d \le \theta_+$ *and* $\theta_- \le r_f \le \theta_+$ and previous stage – Mark-Down
Mark-Up	$r_d > \theta_+$	$r_f > \theta_+$	$r_d > \theta_+$ *and* $r_f > \theta_+$	$r_d > \theta_+$ *or* $r_f > \theta_+$
Distribution	$\theta_- \le r_d \le \theta_+$ and previous stage – Mark-Up	$\theta_- \le r_f \le \theta_+$ and previous stage – Mark-Up	$\theta_- \le r_d \le \theta_+$ *or* $\theta_- \le r_f \le \theta_+$ and the previous stage – Mark-Up	$\theta_- \le r_d \le \theta_+$ *and* $\theta_- \le r_f \le \theta_+$ and previous stage – Mark-Up
Mark-Down	$r_d < \theta_-$	$r_f < \theta_-$	$r_d < \theta_-$ *and* $r_f < \theta_-$	$r_d < \theta_-$ *or* $r_f < \theta_-$

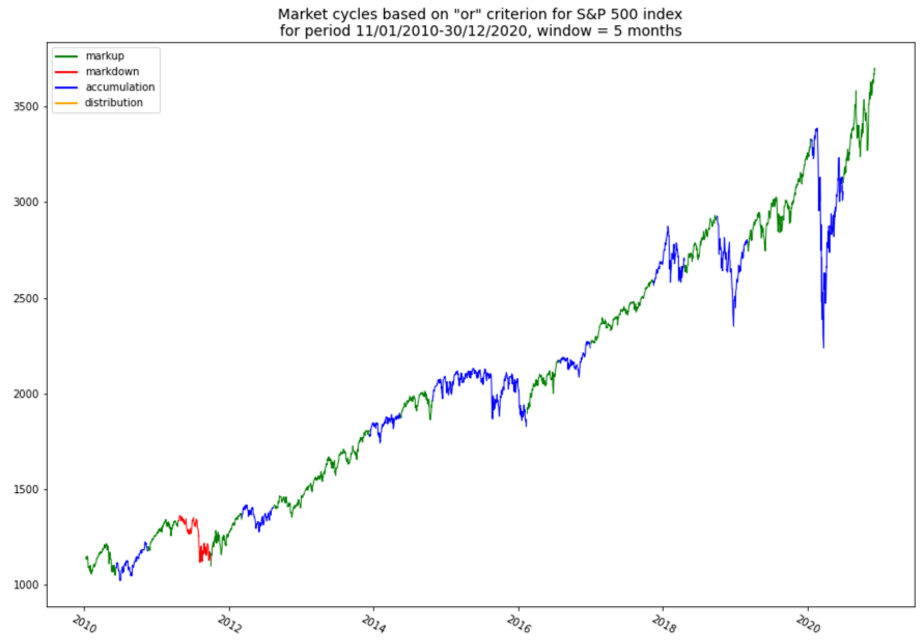
The Wyckoff theory was used to distinguish the four stages of the stock market for accounting to the market environment (2/3)





The Wyckoff theory was used to distinguish the four stages of the stock market for accounting to the market environment (3/3)



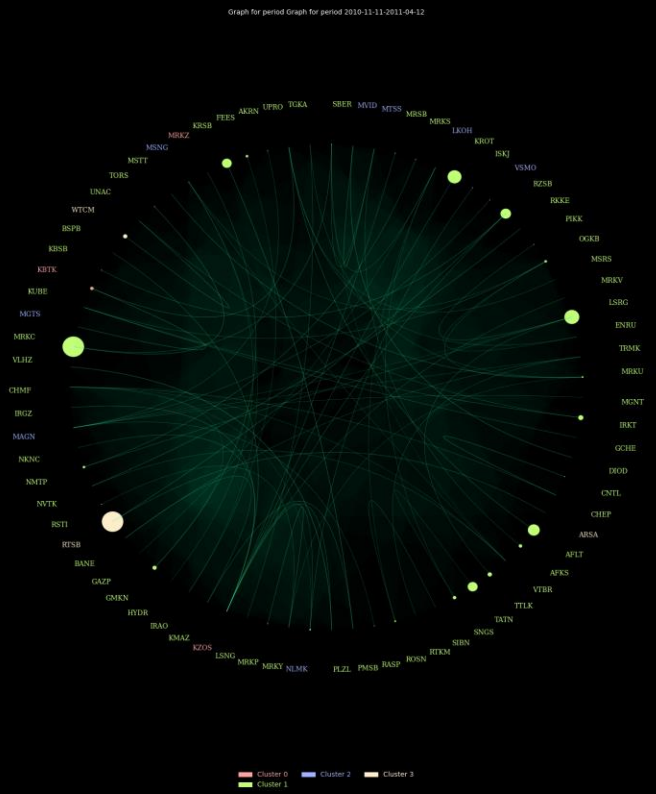


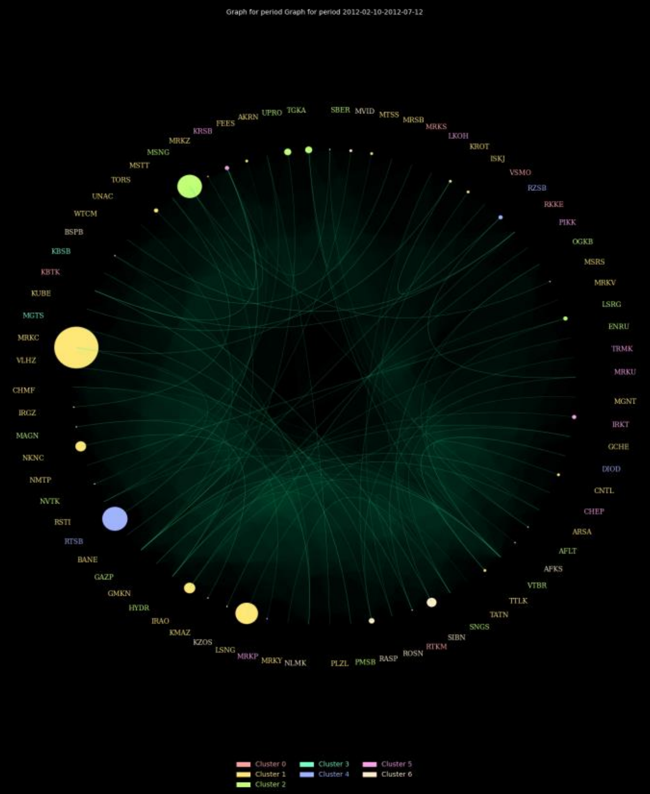
Dynamics of the MOEX stock graph by sectors and clusters (pearson-Spectral), 01/01/2010 - 01/01/2021

https://github.com/annakuchko/FiNetwork/blob/main/imgs/industries_partition.gif

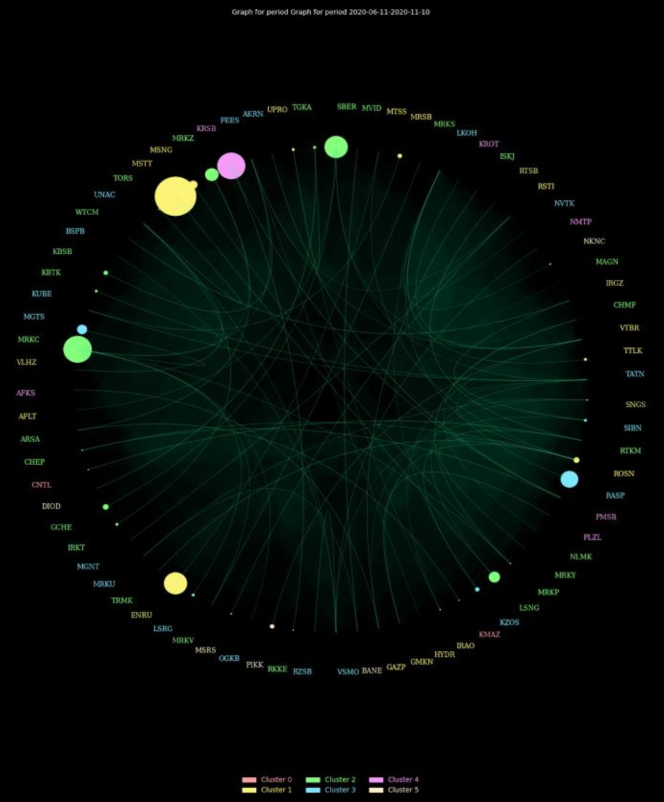
https://github.com/annakuchko/FiNetwork/blob/main/imgs/pearson_Spectral_clustering_partition.gif

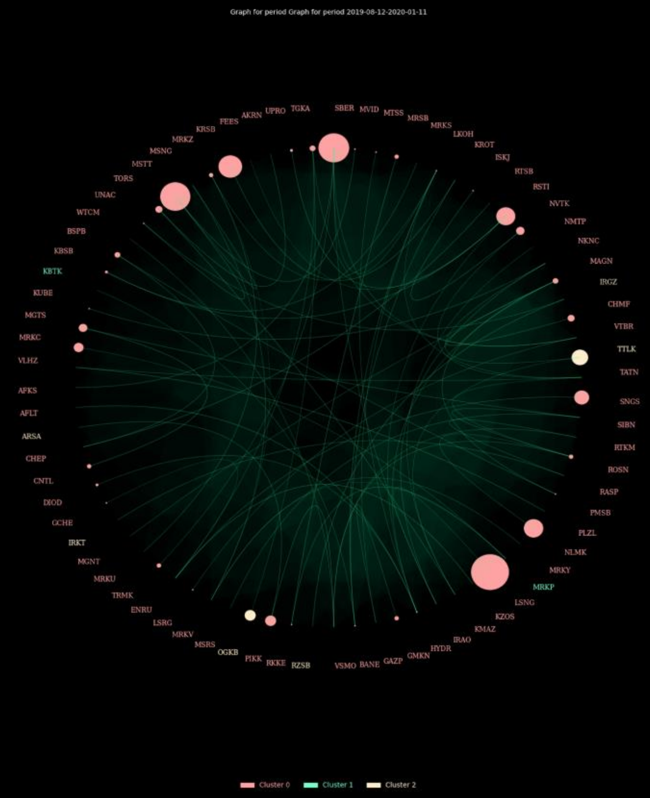
Dynamics of the MOEX stock exchange graph into clusters (pearson-Spectral), 11/11/2010 - 12/04/2010 and 10/02/2012 - 12/07/2012





Dynamics of the MOEX stock exchange graph into clusters (pearson-Spectral), 11/06/2020 - 10/11/2020 and 12/08/2019 - 11/01/2020





A cluster analysis was performed to select the most optimal method of graph partitioning

Internal clustering validation metrics for distances based on Pearson correlation

Clustering Method	Avg. Kalinski-Harabasz score	Avg. Silhouette score	Avg. Davies-Bouldin score
K-means	9.1612	0.1446	1.1585
Spectral	3.3468	0.0973	2.7331
Spectral Gaussian mixture	0.1632	0.3361	2.1845
Hierachical	3.1558	0.0241	3.6237

Metrics of internal clustering validation for distances based on the Atkinson index

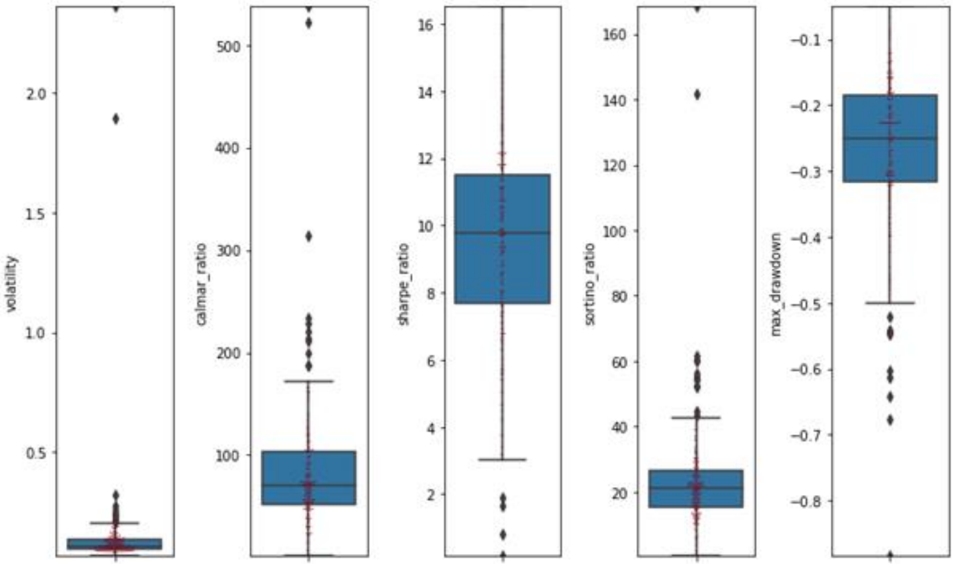
Clustering Method	Avg. Kalinski-Harabasz score	Avg. Silhouette score	Avg. Davies-Bouldin score
K-means	44.2678	0.7817	0.1449
K-medoids	17.7982	0.3147	1.5895
Spectral Gaussian mixture	0.5129	0.0899	1.1598
Gaussian mixture	30.4568	0.6113	0.9459

The selection of shares in the portfolio was carried out using various combinations of model parameters

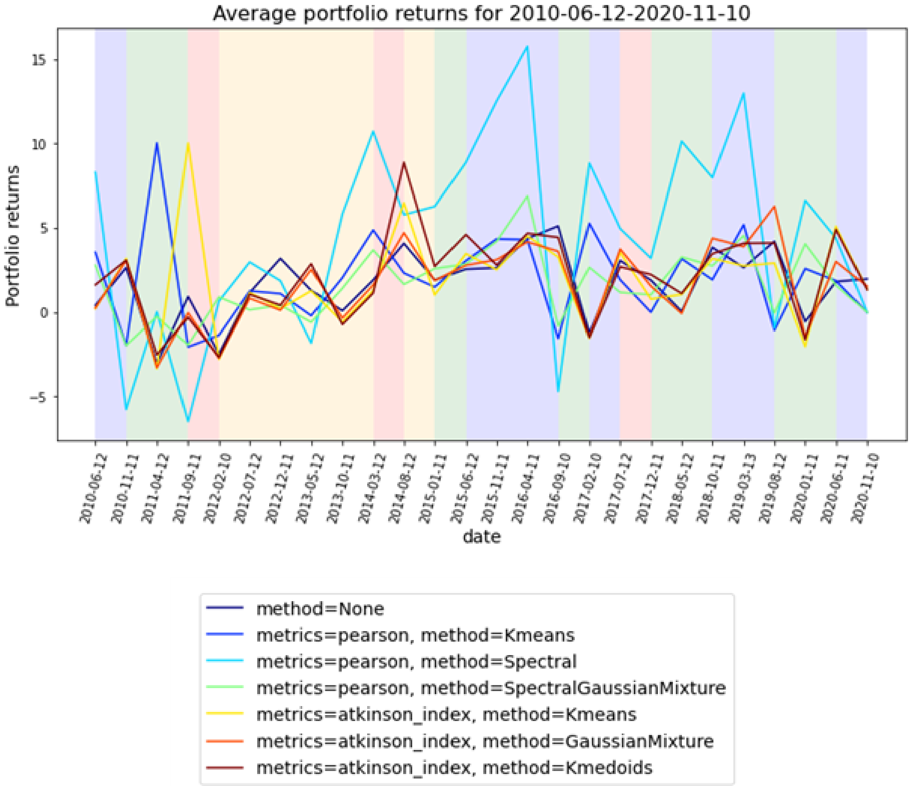
Average values of portfolio performance metrics for various combinations of model parameters

Metrics	Method	Volatility	Calmar ratio	sharpe_ ratio	sortino_ ratio	max_ drawdown
Pearson	K-means	0.2136	96.7784	9.0484	28.0426	-0.2605
	Spectral	0.1086	77.7222	8.9880	21.4607	-0.2415
	Spectral Gaussian mixture	0.1313	99.2924	8.9445	28.8019	-0.2661
Atkinson index	K-means	0.2009	69.0319	7.9287	19.3753	-0.3262
	K-medoids	0.1404	91.7309	9.7559	25.3013	-0.2715
	Gaussian mixture	0.1253	73.6568	9.2336	20.2947	-0.3103
-	None	0.1087	83.2218	10.8492	24.5597	-0.2426

Box-plot diagram of investment portfolio performance metrics for portfolios of various sizes selected using different combinations of model parameters



Dynamics of average returns of financial asset portfolios based on cluster analysis and random selection for the Russian market, 2010-2020



Main modelling results

- ANOVA showed the significance of differences between different models in terms of portfolio performance
- Leven's test showed a significant difference in variation at the level of acceptance of the hypothesis 0.065 (<0.1) only for splitting by portfolio size
- A model with pearson-Kmeans parameters shows greater profitability for the majority of time intervals
- In most market phases, the best quality, on average, is shown by a portfolio with the parameters atkinson_index-Kmedoids

A cluster analysis was performed to select the most optimal method of graph partitioning

Internal clustering validation metrics for distances based on Pearson correlation

Clustering Method	Avg. Kalinski-Harabasz score	Avg. Silhouette score	Avg. Davies-Bouldin score
K-means	109.0422	0.6407	1.2918
Spectral	17.6567	0.0998	3.4774
Spectral k-means	0.1728	0.3720	1.7981
Hierarchical	32.9781	0.0941	2.6156

Metrics of internal clustering validation for distances based on the Atkinson index

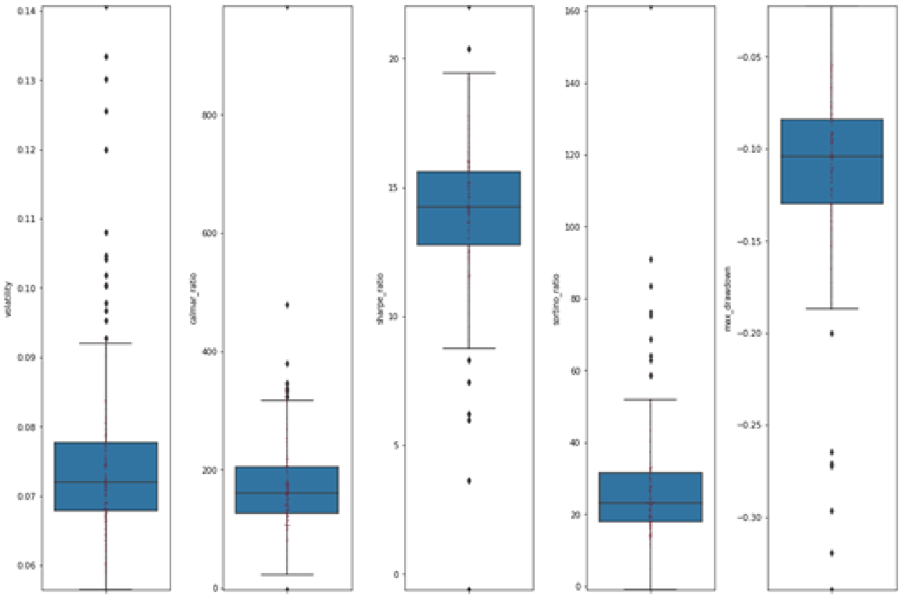
Clustering Method	Avg. Kalinski-Harabasz score	Avg. Silhouette score	Avg. Davies-Bouldin score
K-means	227.3429	0.2352	1.2359
Spectral	67.7112	0.1841	1.2100
Spectral k-means	0.4705	0.1384	3.0637
K-medoids	143.5340	0.0672	2.5837
Spectral Gaussian mixture	0.6219	0.1361	3.3367
Gaussian mixture	224.6047	0.2328	1.2266
Hierarchical	214.3039	0.1906	1.3886

The selection of shares in the portfolio was carried out using various combinations of model parameters

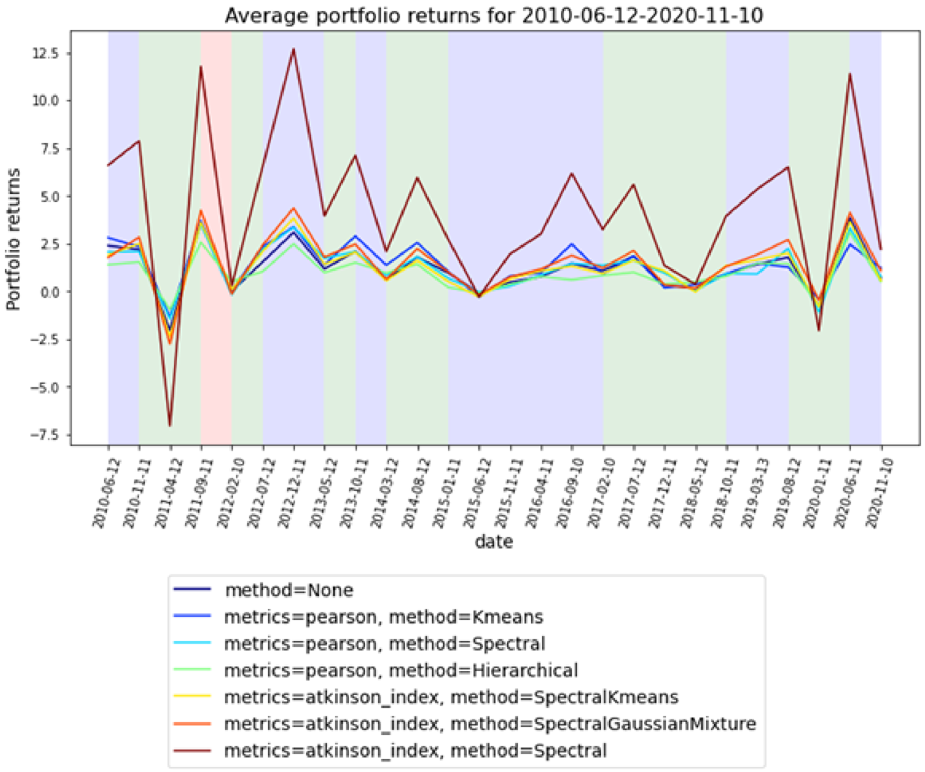
Average values of portfolio performance metrics for various combinations of model parameters

Metrics	Clustering Method	Volatility	Calmar ratio	Sharp ratio	Sortino ratio	Max drawdown
Pearson	К-means	0.0678	226.8679	15.5472	33.0405	-0.0893
	Spectral	0.0761	175.6716	14.3517	29.3157	-0.1084
	Hierarchical	0.0721	209.6962	14.8302	32.1533	-0.0939
Atkinson Index	Spectral k-means	0.0790	132.6965	12.9849	20.5005	-0.1319
	Spectral	0.0770	190.9069	14.0749	34.5049	-0.1018
	Spectral Gaussian mixture	0.0773	145.6613	13.9545	20.7990	-0.1250
-	None	0.0753	199.4947	13.5698	32.3332	-0.1108

Box-plot diagram of investment portfolio performance metrics for portfolios of various sizes selected using different combinations of model parameters



Dynamics of average returns of financial asset portfolios based on cluster analysis and random selection for the American stock market, 2010-2020



Main modelling results

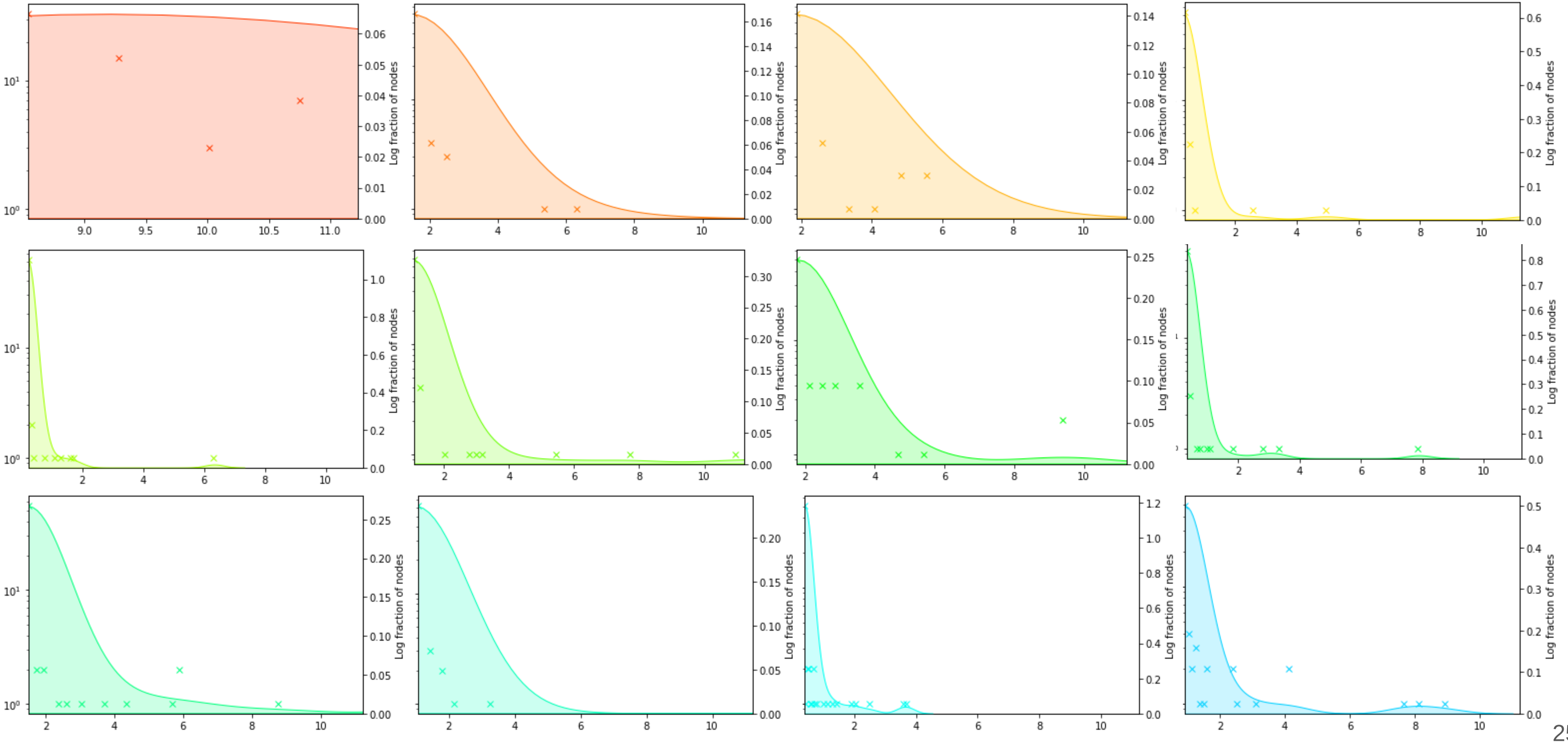
- ANOVA showed the significance of differences in the variation of metrics for different model specifications
- Leven's test showed a significant difference in portfolio efficiency depending on the distance metric used in modelling and the clustering algorithm
- Consistently the best results are shown by a portfolio selected using atkinson_index-Spectral model parameters with high volatility of results

- **Research Goal and Objectives**
- **Theoretical Foundations of Developing the Dynamic Strategy for Portfolio Management**
- **Applied Research of Portfolio Strategy Optimisation**
- **Attachments**

Dynamics of the adjacency matrix of an unfiltered graph. MOEX stock exchange graph, 01/01/2010 - 01/01/2021, Atkinson index, without scale

https://github.com/annakuchko/FiNetwork/blob/main/imgs/Unfiltered%20adjacency%20matrix_Atkinson%20index_MOEX.gif

The log-distribution of the logarithm of the degree of the graph corresponds to the power law of distribution. MOEX Stock Exchange graph, 01/01/2010 - 01/01/2021



Project Github repository

github.com/annakuchko/FiNetwork