



Multiscale Partial Correlation Analysis of Tehran Stock Market Indices: Clustering and Inter-Index Relationships



Mohsen Imeni^{1*}, Zongke Bao², Victoria Nozick³

¹ Department of Accounting, Ayandegan Institute of Higher Education, 4681853617 Tonekabon, Iran

² School of Accounting, Zhejiang University of Finance and Economics, 310018 Hangzhou, China

³ Operations and Information Management Group, Aston Business School, Aston University, B4 7ET Birmingham, United Kingdom

* Correspondence: Mohsen Imeni (imeni@aihe.ac.ir)

Received: 12-26-2023

Revised: 02-07-2024

Accepted: 02-24-2024

Citation: M. Imeni, Z. K. Bao, and V. Nozick, "Multiscale partial correlation analysis of Tehran stock market indices: Clustering and inter-index relationships," *J. Oper. Strateg. Anal.*, vol. 2, no. 1, pp. 1–10, 2024. <https://doi.org/10.56578/josa020101>.



© 2024 by the authors. Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: This study delves into the intricate relationships among returns of diverse indices within the Tehran stock market, employing both Pearson and partial correlation coefficients as analytical tools. Utilizing monthly data from fourteen capital market indices, the investigation applies the k-means method for clustering based on four critical attributes: risk, efficiency, average industry index, and the number of companies within each industry. The findings reveal that when the total index is considered as a controlling variable, the partial correlation analysis yields distinct insights into the interconnections among market indices, thereby highlighting the significant influence of the total index on these relationships. Moreover, the clustering analysis categorizes the indices into three distinct groups: the first cluster exclusively comprises the total index; the second cluster includes indices from the automobile, pharmaceutical, metal, cement, chemical, and food sectors; whereas the remaining indices are allocated to the third cluster. This multifaceted approach not only elucidates the dynamic interplay between different stock market indices but also underscores the variability in their interrelations when viewed through the lens of a controlled variable. The study's methodological rigor and its innovative use of multiscale partial correlation analysis contribute to a deeper understanding of the factors shaping the Tehran stock market's behavior, offering valuable insights for investors, policymakers, and scholars alike.

Keywords: Correlation; Clustering; Market index; Tehran stock market; Partial correlation; k-means clustering

1 Introduction

Understanding the complex nature of financial markets, especially after the recent crisis in 2008, left a big challenge for everyone. Recent studies have analyzed a set of big data on financial markets and modeled the static and dynamic behaviors of this very complex system [1]. One of the prominent features in financial markets is the correlation (positive or negative) between the price movements of different financial assets. The existence of a high degree of cross-correlation between the simultaneity of a set of stock returns is a well-known empirical fact [2, 3]. Pearson's correlation coefficient provides information about the similarity in the behavior of specific stock price changes. Much effort has been made to extract information from observed correlations to gain insights into financial markets' underlying structure and dynamics [4]. Most of these cases have been used in finance, especially in portfolio analysis, where an understanding of the linear relationship between different stock market returns has been proposed. For example, the average variance portfolio theory has been used to calculate portfolio variance [5].

However, despite its popularity and ease of calculation, Kenett et al. [4] emphasized that the Pearson correlation coefficient also had a critical limitation when being used in financial applications. A high correlation value for stock market returns does not necessarily mean a direct relationship between two stocks but a common underlying influence by macroeconomic or psychological factors related to investors [6, 7].

To properly understand the direct correlation structure between stocks in such a situation, it is essential first to consider the common factors that influence stock movements, represented mainly by the stock market index in the framework of the Capital Asset Pricing Model (CAPM) [8]. The partial correlation coefficient enables such

calculations by quantifying the correlation between two stock returns while considering the effect of the third variable that affects both returns [9]. On the other hand, clustering is an essential topic in financial markets. Cluster analysis is used to group sets of objects with similar characteristics. Investors use cluster analysis in financial markets to develop a cluster trading approach [10, 11]. This analysis helps them to build a diversified portfolio. Stocks with a high correlation are based on returns and placed in a basket. In the other basket, those with less correlation will be placed [12].

In the same way, clustering will be done until every share is in one category. If done correctly, different clusters will show minimal correlation with each other. In this way, investors get all the virtues of diversification, including loss reduction, capital preservation, and taking riskier trades without adding to the total risk [6, 13].

Therefore, the current study aims to cluster and analyze the relationship between stock market returns with a partial correlation approach. This study can contribute in two ways: it strengthens the knowledge of using the correlation coefficient regarding diversification and financial contagion. In portfolio diversification programs, the primary consideration is the correct estimation of correlation matrices between asset returns. Correlation coefficients have also been used in the financial contagion literature, where the main objective is to measure intermarket relationships in financial markets after a shock to one or more countries.

2 Research Literature

2.1 Stock Correlation

While people may be focused on how individual stocks in the portfolio react to the market independently of other stocks, understanding how they move alongside other stocks can provide a more cohesive view of the portfolio. "Stock correlation is important because it can show an investor that they may not be as diversified as they think," says Landsberg. "As for stocks in different sectors, if their returns depend on one thing (e.g., the economy in a particular state), the portfolio has almost no protection against diversification." Diversification is a strategy for risk management. It means don't put all eggs in one basket. Having a mix of different types of stocks, mutual funds, bonds, and other investments allows people to insulate their portfolio against the inevitable fluctuations in the market. Portfolios that "overweight" a particular stock or sector are much more sensitive to market fluctuations. Understanding stock correlation can help avoid it. Landsberg: "Investors may be surprised to learn that only a few fundamental factors may be driving their portfolio" [14].

2.2 Index Market

In finance, a stock index is an index that measures the performance of a stock market or a subset of the stock market. It helps investors compare current stock price levels with past prices to calculate market performance. A market index is a hypothetical set of investments that represents a part of the financial market. The calculation of the index value is obtained from the prices of the underlying assets. Some indices have values based on market value weighting, income weighting, floating weight, and base weight. Weighting is adjusting the individual influence of items in an index. The stock market index, also known as the stock index, is a statistical measure that reflects the changes in the market. It is created by grouping several similar stocks among securities listed on the stock exchange, and the selection criteria can be the size of a company, its market capitalization, or the type of industry. The underlying stock price changes affect the index's overall value. If the prices increase, the index increases; and if prices decrease, it index also decreases [15–17].

2.3 Previous Researches

Michis [6] researched multiscale partial correlation clustering of stock market returns. The results indicate cluster formations that vary by time scale, which require different stock selection strategies for investors with different investment horizon orientations. Ren et al. [18] thought that they had evidence of solid risk communication among global stock markets in general and multiscale networks. They have found that their topological properties are different across time-frequency horizons. The U.S. and Eurozone stock markets play a significant role in risk transfer. Most developing markets remain inactive in multiscale risk communication networks.

Millington and Niranjana [19] concluded that the financial sector was considered the most crucial sector for most of the dataset. Finally, they showed a slight negative correlation between the centrality of a firm and its out-of-sample risk. Kenett et al. [4] concluded that this method provided new insights into financial market dynamics. This method reveals implicit effects in the game between stocks.

2.4 Research Questions

Based on the literature, two questions have been raised in this research:

Q1: Is it possible to use the Pearson and partial correlation index for the returns of stock market indices in different sectors of economic activity?

Q2: How can the clustering of stock market indices affect the decisions of analysts and investors?

3 Research Methodology

3.1 Research Population

In the current research, according to the topic and its application, the research community includes all the indices of Tehran's capital market for each industry, whose data is available from 2013 to 2022. In this study, the selection of sample indicators was based on those considered most vital, encompassing industries with a greater number of companies. After the review, 14 indices were randomly selected, including the total index, 30 companies, mass production, automobile, pharmaceutical, agriculture, metals, cement, chemical, oil, food, sugarcane, banks, and insurance.

3.2 Data Analysis Method

This research used Pearson and partial correlation to check the correlation between indices. Also, the k-means clustering method was employed using four characteristics: (1) risk, (2) the number of companies in the industry, (3) the average monthly return, and (4) the average degree of the index.

3.2.1 Pearson correlation

Pearson's correlation is the most widely used correlation statistic to measure the degree of relationship between linearly related variables. For example, Pearson's correlation is used to measure the relationship between two stocks in the stock market. Point-binary correlation is done with Pearson's correlation formula, with the difference that one of the variables is dichotomous. The following formula is used to calculate Pearson correlation r [20]:

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

where, r_{xy} is the Pearson correlation coefficient r between x and y ; n denotes the number of observations; x_i is the value of x (for the first observation); y_i is the value of y (for the i -th observation).

3.2.2 Cluster analysis

Cluster analysis is a type of data classification done by separating data into groups. The purpose of cluster analysis is to classify n objects into k groups ($k > 1$) called clusters using p variables ($p > 0$). Like many other types of statistics, cluster analysis has different types, each with its own clustering method. The main subsection of clustering methods is that the number of clusters is predefined in the first method. This method is known as the k-means clustering method. When the number of clusters is not defined in advance, hierarchical cluster analysis is used.

In centroid-based clustering, each cluster is represented by a centroid vector, which is not necessarily a dataset member. When the number of clusters is fixed at k , k-means clustering provides a formal definition as an optimization problem. After k cluster centers are found and the objects are assigned to the nearest cluster center, the squared distances from the cluster are minimized. Most k-means type algorithms require that the number of clusters k be determined in advance, which is one of the biggest drawbacks of these algorithms. In addition, the algorithms prefer clusters of approximately the same size since they always assign an object to the closest center. It often leads to incorrect clipping of cluster boundaries (which is not surprising since the algorithm optimizes cluster centers instead of cluster boundaries) [21].

4 Data Analysis and Research Findings

4.1 Descriptive Statistics

In this research, descriptive statistics for index performance and, once again, for the index numbers have been extracted as the standard deviation from the dispersion index group and the average from the central index group. It should be noted that these results were removed based on monthly data. The results are shown in Table 1.

According to Table 1, it can be seen that the highest and lowest returns in 10 years were related to the oil and automobile index, which experienced even more than 100% in terms of returns. Also, regarding average efficiency, the automobile and agriculture index had the highest monthly efficiency. Also, most of the indices created a higher yield compared to the total index, which in financial discussions indicates the creation of excess yield. Also, in terms of risk, measured by the standard deviation of efficiency, the highest risk is related to the car index, and the lowest risk is the overall index. But for all indicators, the standard deviation is higher than the average return, which indicates the high fluctuations of the indicators. Also, Figure 1 shows the performance trends of the indicators during the period under review. As can be seen, the automotive and oil industry achieved the highest efficiency.

4.2 Correlation

Correlation is a bivariate analysis that measures the strength of the relationship between two variables and the direction of the relationship. Regarding the relationship's strength, the correlation coefficient's value varies between +1 and -1. The value of ± 1 indicates the degree of relationship between two variables. As the correlation coefficient value goes toward 0, the relationship between the two variables will be weaker. The coefficient sign indicates the direction of the relationship. The (+) sign indicates a positive relationship, and the (-) sign indicates a negative relationship. In this research, the correlation has been tested once by considering the two-by-two relationship of the indices and once again by controlling the total index. Therefore, the correlation table is presented in two parts. The Pearson correlation coefficient was obtained between the quantities in a section called none. The other section of Table 2 is related to the state included in correlation analyses under the total index controller title for years.

Table 1. Descriptive statistics of the efficiency of each index

	No.	Mean	Max	Min	S.D.
Total index	119	.038	.508	-.2	.104
30 firms	119	.041	.617	-.277	.122
Mass production	119	.033	.483	-.181	.138
Car	119	.051	1.218	-.308	.2
Medicinal	119	.046	.489	-.179	.114
Agriculture	119	.052	.735	-.305	.181
Metals	119	.042	.666	-.287	.132
Cement	119	.042	.495	-.257	.134
Chemical	119	.041	.545	-.233	.111
Petroleum	119	.046	1.017	-.451	.171
Dietary	119	.041	.716	-.261	.135
Sugar	119	.038	.443	-.193	.123
Banks	119	.033	.577	-.213	.117
Insurance	119	.041	.477	-.254	.128

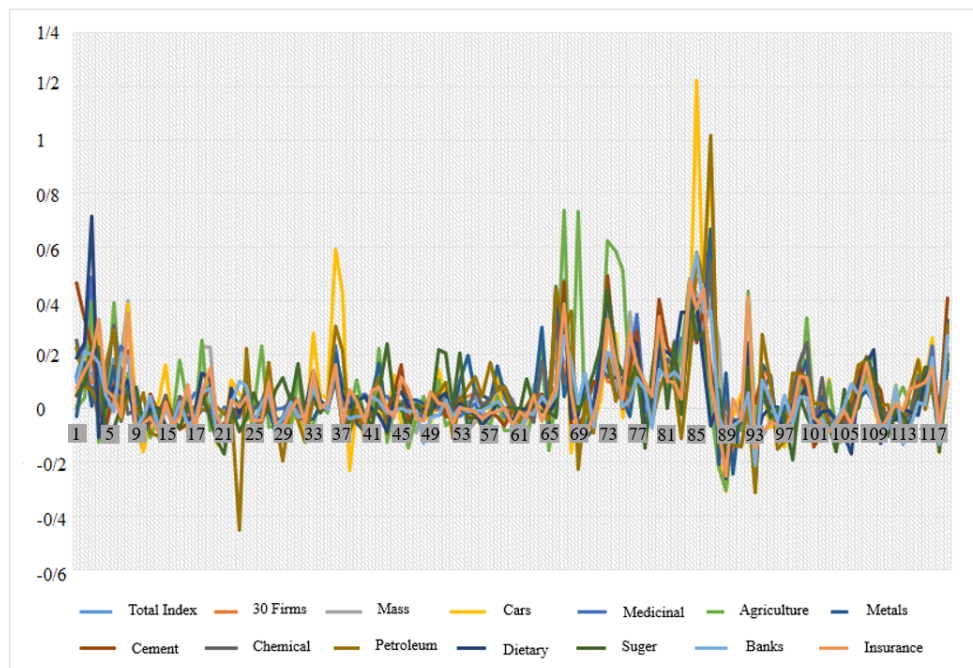


Figure 1. Return trend of selected indices

The results of Table 2 show that all indices have a positive correlation with each other at the 99% confidence level. In addition, the indices of 30 firms, chemical, petroleum, and banking have the most robust correlations with the total index. Conversely, the indices representing the food, agriculture, and sugar industries displayed relatively lower correlation coefficients. In the following table, the study explored partial correlation coefficients, employing the total index as a controlling variable.

Table 2. Pearson's correlation coefficient of index efficiency

Industry No.		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1- Total Index	Pearson	1	.965	.705	.763	.687	.473	.887	.690	.922	.843	.581	.565	.854	.680
	Sig.		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
2- 30 Firms	Pearson	.965	1	.597	.710	.553	.374	.932	.574	9	.839	.434	.467	.818	.588
	Sig.	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
3- Mass	Pearson	.705	.597	1	.710	.706	.517	.487	.747	.507	.481	.722	.673	.737	.762
	Sig.	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
4- Car	Pearson	.763	.710	.710	1	.648	.443	.610	.565	.563	.669	.543	.488	.759	.609
	Sig.	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
5- Medicinal	Pearson	.687	.553	.706	.648	1	.636	.457	.722	.533	.466	.849	.667	.733	.717
	Sig.	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
6- Agriculture	Pearson	.473	.374	.517	.443	.636	1	.317	.559	.374	.250	.628	.536	.461	.550
	Sig.	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000
7-Metals	Pearson	.88	.932	.487	.610	.457	.317	1	.493	.825	.77	.327	.4	.680	.500
	Sig.	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.00	.000	.000
8- Cement	Pearson	.690	.574	.747	.565	.772	.559	.493	1	.561	.475	.745	.696	.691	.726
	Sig.	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000
9- Chemical	Pearson	.922	.90	.57	.563	.533	.374	.825	.561	1	.784	.43	.3	.694	.505
	Sig.	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000
10- Petroleum	Pearson	.843	.839	.481	.669	.466	.250	.770	.475	.784	1	.329	.3	.626	.466
	Sig.	.000	.000	.000	.000	.000	.006	.000	.000	.000		.000	.000	.000	.000
11- Dietary	Pearson	.581	.434	.722	.543	.849	.628	.327	.745	.431	.329	1	.629	.628	.693
	Sig.	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000
12- Sugar	Pearson	.565	.467	.673	.488	.667	.536	.402	.696	.391	.348	.629	1	.619	.654
	Sig.	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000
13-Banks	Pearson	.854	.818	.737	.759	.733	.461	.680	.691	.694	.626	.628	.619	1	.720
	Sig.	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000
14- Insurance	Pearson	.680	.588	.762	.609	.717	.550	.50	.726	.505	.466	.693	.654	.720	1
	Sig.	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
	N	119	119	119	119	119	119	119	119	119	119	119	119	119	119

Table 3. Partial correlation coefficients of the index performance considering the control variable

Industry No.		1	2	3	4	5	6	7	8	9	10	11	12	13
1	Correlation	1	-.45	-.16	-.58	-.36	.63	-.49	.1	.18	-.6	-.36	-.05	-.36
	Sig.		.000	.09	.000	.000	.000	.000	.28	.05	.000	.000	.61	.000
2	Correlation	-.45	1	.38	.43	.29	-.42	.51	-.52	-.3	.54	.47	.37	.54
	Sig.	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
3	Correlation	-.16	.38	1	.26	.14	-.22	.08	-.56	.08	.19	.11	.32	.19
	Sig.	.09	.000		.000	.12	.02	.37	.000	.42	.04	.25	.000	.04
4	Correlation	-.58	.43	.26	1	.49	-.45	.57	-.36	-.29	.76	.47	.39	.47
	Sig.	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
5	Correlation	-.36	.29	.14	.49	1	-.25	.36	-.18	-.31	.49	.37	.13	.35
	Sig.	.000	.000	.12	.000		.01	.000	.05	.000	.000	.000	.17	.000
6	Correlation	.63	-.42	-.22	-.45	-.25	1	-.36	.04	.09	-.5	-.26	-.32	-.3
	Sig.	.000	.000	.02	.000	.01		.000	.65	.34	.000	.000	.000	.000
7	Correlation	-.49	.51	.08	.57	.36	-.36	1	-.27	-.27	.58	.51	.27	.48
	Sig.	.000	.000	.37	.000	.000	.000		.000	.000	.000	.000	.000	.000
8	Correlation	.1	-.52	-.56	-.36	-.18	.04	-.27	1	.03	-.33	-.41	-.47	-.43
	Sig.	.28	.000	.000	.000	.05	.65	.000		.72	.000	.000	.000	.000
9	Correlation	.18	-.3	.08	-.29	-.31	.09	-.27	.03	1	-.37	-.29	-.34	-.27
	Sig.	.05	.000	.42	.000	.000	.34	.000	.72		.000	.000	.000	.000
10	Correlation	-.6	.54	.19	.76	.49	-.5	.58	-.33	-.37	1	.45	.31	.5
	Sig.	.000	.000	.04	.000	.000	.000	.000	.000	.000		.000	.000	.000
11	Correlation	-.36	.47	.11	.47	.37	-.26	.51	-.41	-.29	.45	1	.32	.45
	Sig.	.000	.000	.025	.000	.000	.000	.000	.000	.000	.000		.000	.000
12	Correlation	-.05	.37	.32	.39	.13	-.32	.27	-.47	-.34	.31	.32	1	.37
	Sig.	.61	.000	.000	.000	.17	.000	.000	.000	.000	.000	.000		.000
13	Correlation	-.36	.54	.19	.47	.35	-.3	.48	-.43	-.27	.5	.45	.37	1
	Sig.	.000	.000	.04	.000	.000	.000	.000	.000	.000	.000	.000	.000	
	d.f	116	116	116	116	116	116	116	116	116	116	116	116	116

As can be seen in Table 3, controlling the total index affects all market indices. The correlation between all indices is no longer positive two by two, and it can be seen that the correlation between them is negative at a significance level of 99%. It can also be seen that the correlation coefficient between other indices has also decreased because the effect of the total index has been controlled.

4.3 Cluster Analysis

Cluster analysis is a type of data classification that separates data into groups. The purpose of cluster analysis is to classify n objects into k groups ($k > 1$) called clusters using p variables ($p > 0$). Like many other types of statistics, cluster analysis has different types, each with its own clustering method. The main subsection of clustering methods is that the number of clusters is predefined in the first method. This method is known as the k-means clustering method. Hierarchical cluster analysis is used when the number of clusters is not predefined. This research uses the k-means clustering method based on four risk characteristics: standard deviation of returns, index size (index average in 10 years), number of companies in each index, and average return. The results are as follows:

4.3.1 Initial cluster centers

As seen in Table 4, two clusters are considered for each quantity, and three ones are considered for initial cluster centers.

Table 4. Initial cluster centers

	Cluster		
	1	2	3
Risk	.10	.17	.11
Index size	5.7	6.26	4.54
Number of firms in the industry	970	13	65
Average monthly return	.04	.05	.04

4.3.2 Iteration history

Table 5, with the name of repetition history, shows the change in the center of the clusters from the repetition algorithm until reaching the selected convergence criterion at the determined convergence value, i.e., 0.

Table 5. Iteration history

Convergence	Change in Cluster Centers		
	1	2	3
1	0	12.409	16.5
2	0	0	0

Note: Convergence was achieved due to no or little change in cluster centers. The maximum absolute coordinate change for each center is 0.000. The current iteration is 2. The minimum distance between primary centers is 52.029.

4.3.3 Cluster membership

In Table 6, the cluster number of each row and its distance from the cluster's center are obtained.

Table 6. Cluster membership

Industry No.	Indices	Cluster	Interval
1	Total index	1	.000
2	30 firms	2	4.715
3	Mass production	2	8.769
4	Car	3	9.507
5	Medicinal	3	1.501
6	Agriculture	2	4.304
7	Metals	3	1.78
8	Cement	3	4.612
9	Chemical	3	16.5
10	Petroleum	2	12.409
11	Dietary	3	5.508
12	Sugar	2	7.288
13	Banks	2	1.138
14	Insurance	2	10.716

As seen in Figure 2, based on the four characteristics of the index, 30 firms are located in cluster 2, along with mass production, agriculture, oil, sugarcane, banks, and insurance. Automotive, pharmaceutical, metal, cement, chemical, and food indices are placed in cluster 3. Only the total index is placed in cluster 1.

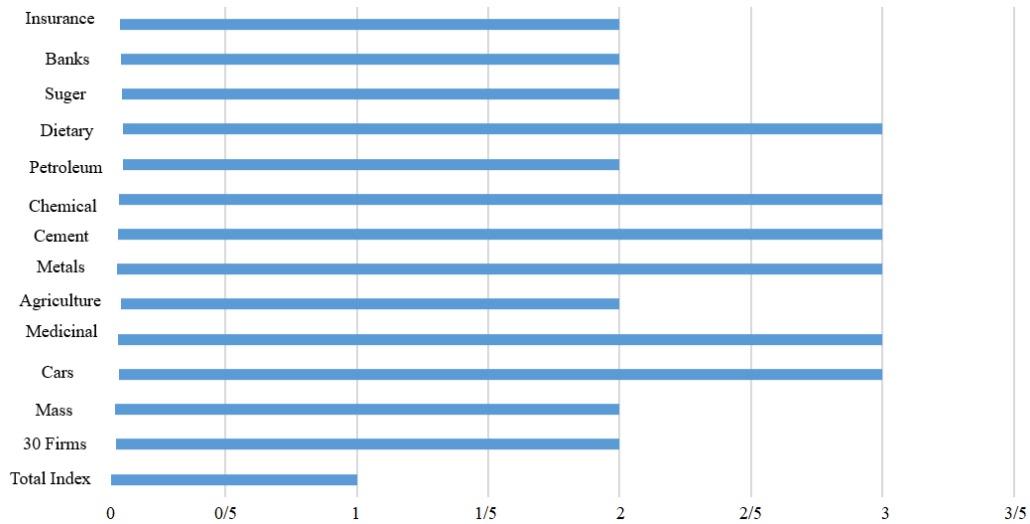


Figure 2. The membership of each index in the cluster

4.3.4 Results of the iterative clustering algorithm

In Table 7 of the final cluster centers, the center of the clusters in the last iteration is displayed.

Table 7. Final cluster centers

	Cluster		
	1	2	3
Risk	.104	.140	.138
Index size	5.699	4.518	4.589
Number of firms in the industry	970	25	49
Average monthly return	.038	.041	.044

From the comparison of Table 7 with Table 4, it can be seen that after the repetition algorithm, the centers of the clusters have become closer to each other. Also, Table 8 shows the distance of the center of each cluster from another cluster after the last step of the iteration algorithm.

Table 8. Distances between the centers of the final cluster

Cluster	1	2	3
1		944.715	921.501
2	944.715		23.214
3	921.501	23.214	

4.3.5 Analysis of Variance (ANOVA) test

Since the members of each cluster can be considered as a sample of a community, it is easy to check the existence of a significant difference between the clusters by using the ANOVA test. The results are shown in Table 9.

Table 9. ANOVA test

	Cluster		Error		F	Sig.
	Average of Squares	df	Average of Squares	df		
Risk	.001	2	.001	11	.671	.531
Index size	.621	2	.691	11	.898	.435
Number of firms in the industry	405892.964	2	77.539	11	5234.697	.000
Average monthly return	0	2	0	11	.697	.519

Note: F-tests should be used for descriptive purposes only, as clusters were chosen to maximize differences between cases in different clusters. Observed significance levels are not corrected for this. Therefore, they cannot be interpreted as hypothesis tests of the equality of cluster means.

4.3.6 Number of items per cluster

Table 10 indicates the number of items in each cluster after the last step of the iteration algorithm.

Table 10. Number of items per cluster

	1	1
Cluster	2	7
	3	6
Validity		14
Missing		0

As seen in Table 10, cluster 1 has one index, cluster 2 has seven indices, and cluster 3 has six indices, which are consistent with the results in Table 6.

5 Conclusions and Discussion

This research has answered the central questions of this research by using partial correlation, Pearson methods, and the k-means clustering method. In the following, after the scientific explanation of the results to answer the research questions, suggestions based on the results and recommendations for future research will also be presented. Is it possible to use the Pearson and partial correlation index for the returns of stock market indices in different sectors of economic activity? To answer this research question, first, the relationship between selected indices was investigated using Pearson's correlation coefficient, which was positive and significant at the significance level of 99%. However, despite this method's popularity and ease of calculations, Kenett et al. [4] emphasized that Pearson's correlation coefficient had a critical limitation when being used in financial programs. A high correlation value for stock market returns does not necessarily mean a direct relationship between two stocks but a common underlying influence by macroeconomic or psychological factors related to investors or other factors that can be influential. Therefore, the common factors of the total stock market index were used to understand the real direct correlation structure between the stock market indices. The results showed that the partial correlation coefficient conditions such calculations by quantifying the correlation between two stock returns while conditioning the effect of a third variable that affects both returns. Therefore, the results show that the correlation between returns in the market between indices with the influence of the total index factor is positive and significant. In market analysis and portfolio creation, this case cannot be suitable for analysts and portfolio managers because it questions the principle of diversification [22, 23].

For the second research answer, how does the clustering of stock market indices affect the decisions of analysts and investors? Based on the results obtained in this research, the market indices were examined in three clusters with four characteristics, which were used in a way that included both the risk factor and the return factor, which are of interest to investors in the market. In addition, it is vital for investors in fundamental analysis that the industry's size is essential regarding the number of companies and their index degree. Because companies have more competitors in an industry, usually according to financial and investment theories, the field of competition has increased for them, and they cannot take a large share of the market as before. On the other hand, another analysis is that looking at industries with more companies can open the way for investors to choose their favorite stocks and even create sector funds and other stock funds. Therefore, the clustering of market indices has been able to classify different markets based on the desired characteristics of this research.

Based on the results of this study, the following suggestions can be made:

1. Portfolio managers are suggested to diversify their portfolios by forming clusters with different consequences for investors.
2. Portfolio managers are suggested to use the partial correlation coefficient to select stocks of companies in a specific industry to form partial funds.
3. Investors are suggested to buy and cluster assets with related returns suitable for different market segments with cluster analysis.

Author Contributions

Conceptualization, M.I., and V.N.; Methodology, M.I. and Z.B.; Software, Z.B.; Validation, M.I., and V.N.; Formal analysis, Z.B.; Data maintenance, M.I.; writing-creating the initial design, M.I.; writing-reviewing and editing, V.N.

Data Availability

The stock exchange information technology management company and the financial information processing center section at <https://www.fipiran.com/> have been used to collect the data needed to calculate the variables.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] M. M. Hasan, J. Popp, and J. Oláh, “Current landscape and influence of big data on finance,” *J. Big Data*, vol. 7, no. 1, p. 21, 2020. <https://doi.org/10.1186/s40537-020-00291-z>
- [2] E. F. Fama, “Efficient capital markets: A review of theory and empirical work,” *J. Finance*, vol. 25, no. 2, pp. 383–417. <https://doi.org/10.2307/2325486>
- [3] N. Cakici and A. Zaremba, “Salience theory and the cross-section of stock returns: International and further evidence,” *J. Financ. Econ.*, vol. 146, no. 2, pp. 689–725, 2022. <https://doi.org/10.1016/j.jfineco.2021.10.010>
- [4] D. Y. Kenett, X. Huang, I. Vodenska, S. Havlin, and H. E. Stanley, “Partial correlation analysis: Applications for financial markets,” *Quant. Finance*, vol. 15, no. 4, pp. 569–578, 2015. <https://doi.org/10.1080/14697688.2014.946660>
- [5] K. Cuthbertson and D. Nitzsche, *Quantitative Financial Economics: Stocks, Bonds and Foreign Exchange*. John Wiley & Sons, Ltd., 2005.
- [6] A. A. Michis, “Multiscale partial correlation clustering of stock market returns,” *J. Risk Financ. Manag.*, vol. 15, no. 1, p. 24, 2022. <https://doi.org/10.3390/jrfm15010024>
- [7] R. Aljifri, “Investor psychology in the stock market: An empirical study of the impact of overconfidence on firm valuation,” *Borsa Istanbul Rev.*, vol. 23, no. 1, pp. 93–112, 2023. <https://doi.org/10.1016/j.bir.2022.09.010>
- [8] J. M. Chen, “The capital asset pricing model,” *Encyclopedia*, vol. 1, no. 3, pp. 915–933, 2021. <https://www.mdpi.com/2673-8392/1/3/70>
- [9] D. Jing, M. Imeni, S. A. Edalatpanah, A. Alburaikan, and H. A. E. W. Khalifa, “Optimal selection of stock portfolios using multi-criteria decision-making methods,” *Mathematics*, vol. 11, no. 2, p. 415, 2023. <https://doi.org/10.3390/math11020415>
- [10] S. Dzuba and D. Krylov, “Cluster analysis of financial strategies of companies,” *Mathematics*, vol. 9, no. 24, p. 3192, 2021. <https://doi.org/10.3390/math9243192>
- [11] Y. J. Wang and H. S. Lee, “A clustering method to identify representative financial ratios,” *Inf. Sci.*, vol. 178, no. 4, pp. 1087–1097, 2008. <https://doi.org/10.1016/j.ins.2007.09.016>
- [12] V. Sushko and G. Turner, “The implications of passive investing for securities markets,” *BIS Q. Rev.*, pp. 113–131, 2018.
- [13] W. Tang, X. Xu, and X. Y. Zhou, “Asset selection via correlation blockmodel clustering,” *Expert Syst. Appl.*, vol. 195, p. 116558, 2022. <https://doi.org/10.1016/j.eswa.2022.116558>
- [14] S. W. Narayan, M. U. Rehman, Y. S. Ren, and C. Ma, “Is a correlation-based investment strategy beneficial for long-term international portfolio investors?” *Financ. Innov.*, vol. 9, no. 1, pp. 1–26, 2023. <https://doi.org/10.1186/s40854-023-00471-9>
- [15] E. R. González Pozo, “An argument against stock-picking and market-timing: An empirical approach,” *Investig. Desarro.*, vol. 20, no. 2, pp. 93–106, 2021. <https://doi.org/10.23881/idupbo.020.2-6e>
- [16] F. C. Dumiter, F. Turcaş, Ş. A. Nicoară, C. Benţe, and M. Boiţă, “The impact of sentiment indices on the stock exchange—The connections between quantitative sentiment indices, technical analysis, and stock market,” *Mathematics*, vol. 11, no. 14, 2023. <https://doi.org/10.3390/math11143128>
- [17] M. A. Basuony, M. Bouaddi, H. Ali, and R. EmadEldeen, “The effect of COVID-19 pandemic on global stock markets: Return, volatility, and bad state probability dynamics,” *J. Public Aff.*, vol. 22, no. S1, p. e2761, 2022. <https://doi.org/10.1002/pa.2761>
- [18] Y. Ren, W. Zhao, W. You, and K. Zhai, “Multiscale and partial correlation networks analysis of risk connectedness in global equity markets,” *Physica A Stat. Mech. Appl.*, vol. 573, p. 125911, 2021. <https://doi.org/10.1016/j.physa.2021.125911>
- [19] T. Millington and M. Niranjana, “Partial correlation financial networks,” *Appl. Netw. Sci.*, vol. 5, no. 1, pp. 1–19, 2020. <https://doi.org/10.1007/s41109-020-0251-z>
- [20] L. Egghe and L. Leydesdorff, “Relation between pearson’s correlation coefficient r and salton’s cosine measure,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, no. 5, pp. 1027–1036, 2009. <https://doi.org/10.1002/asi.21009>
- [21] L. Dalton, V. Ballarin, and M. Brun, “Clustering algorithms: On learning, validation, performance, and applications to genomics,” *Curr. Genomics*, vol. 10, no. 6, pp. 430–445, 2009. <https://doi.org/10.2174/138920209789177601>
- [22] M. Heidari Haratemeh, “Portfolio optimization and random matrix theory in stock exchange,” *Innov. Manag. Oper. Strateg.*, vol. 2, no. 3, pp. 257–267, 2021. <https://doi.org/10.22105/imos.2021.289758.1109>
- [23] M. Rasoulzadeh, S. A. Edalatpanah, M. Fallah, and E. Najafi, “The combination of markowitz mean-variance

models with trapezoidal fuzzy efficiency and cross-efficiency data envelopment analysis with network structure in order to find the optimal investment portfolio,” *J. Decis. Oper. Res.*, 2023. <https://doi.org/10.22105/dmor.2023.396910.1743>