International Journal of Academic Research in Business and Social Sciences
2017, Vol. 7, No. 5
ISSN: 2222-6990

The Classification of Stocks with Basic Financial Indicators: An Application of Cluster Analysis on the BIST 100 Index

Bilgehan Tekin
Department of Business Administration, Faculty of Economics and Administrative Sciences, Cankiri Karatekin University, Turkey

Fatih Burak Gümüş
Department of Business Administration, Faculty of Business, Sakarya University Turkey

DOI: 10.6007/IJARBSS/v7-i5/2881 URL: http://dx.doi.org/10.6007/IJARBSS/v7-i5/2881

Abstract: *In the literature, it is seen that has been used the data mining methods frequently for the analysis of the stock market and stocks. The aim in here is to provide making the most rational choice and increasing return by reducing human intervention to a minimum level with the creating the algorithmic process structure. In this study, stocks are classified basis of financial indicators derived from the financial statements of the companies. For this purpose, cluster analysis which is one of the data mining and multivariate statistical methods is used. In this method the aim is to collect most similar the stocks in the same cluster in terms of related variables. The variables used in the study; Price / earnings ratio, market value / book value ratio, dividend yield, return on assets, return on equity, change in sales and equity, return on average, return and risk. As the result of the analysis, 88 stocks in Borsa Istanbul 100 Index are divided into 12 clusters. Among these stocks, the ones that are most suitable to form a portfolio have been tried to be determined based on financial indicators and last one and three years' stock performances.*

Keywords: Portfolio Selection, Stocks, Cluster Analysis, Financial Ratios, Financial Markets

Introduction

Investors, one of the actors of financial or stock markets, are primarily aimed at making profits as a result of their transactions and maximizing their profit percentage. Investors must have a certain level of knowledge of economics, finance and financial analysis to earn income by investing in every kind of company stocks. At the same time they must have sufficient information about the company, has basic knowledge to create a profitable stock portfolio, and must be able to use capital effectively.

As seen in developed markets, Turkey also publishes financial statements of publicly traded companies on certain platforms at certain intervals. Internal and external stakeholders of companies can benefit from these tables and have knowledge about the company through various financial analyzes and statistical techniques.

Due to the growing and developing global economic system, financial markets are constantly developing and growing. The increase in the number of publicly traded companies traded on stock exchanges, the offering of different and new opportunities and the observation that stocks are the most appropriate investment instruments when healthy analyzes are made increase the attractiveness of stock markets. Nevertheless, the issue of how investors will make the right decision under the frequently and unexpectedly changing political and economic conjuncture and in the highly volatile market conditions that emerge as a consequence of this situation is increasing day by day. In this context, new statistical and algorithmic analyzes are continuously performed and new methods are brought to the agenda. One of these methods is the cluster analysis which is a multivariate statistical method.

In this study, the companies included in the Borsa Istanbul (BIST 100) index are classified based on the data obtained from the annual financial statements for the year 2015 and the financial indicators calculated on the basis of the closing prices on April 5, 2016. The aim of the study is to classify the best stocks that can be invested in the context of the financial indicators obtained from the financial statements, and to create the most appropriate stock portfolio using cluster analysis. In this context, another aim of this study is to test the usability of the cluster analysis method to create the most appropriate stock portfolio. At the same time, it will be determined which variables are meaningful for the separation of companies into clusters. Another result of this study is directed towards investors who want to diversify their portfolios with companies in different sectors. The portfolio diversification will be ensured as the companies to be included in the emerging clusters are expected to have companies with different financial indicators and operate in different sectors. With this study, it is aimed to make a contribution to the literature by giving detailed usage of the clustering analysis technique in terms of stock selection and portfolio creation.

Literature Review

One of the studies about the classification of companies and stocks by cluster analysis in Turkey belongs to Karabayır and Doğanay (2010). In this study, the companies listed in the Borsa İstanbul 100 index were divided into 10 groups and compared in two different time periods. Accordingly, has reached the conclusion that an investor when he holds his stock set in his first period of time in his portfolio during his second period of time, he will earn money.

Looking at the work done by Kalfa and Bekçioğlu (2013), it is seen that the 42 selected companies from three different sectors of food, textile and cement were subjected to cluster analysis using financial ratios and then these results were tested by discriminant analysis. As a result of the study, three clusters were obtained and it was seen that the sectors of the companies was a factor for formation of the clusters. In the study, it was confirmed that investing in different sectors that traditional portfolio diversification put forward.

Topak (2010) presented an alternative risk premium determination approach to the CAPM and arbitrage pricing model with cluster analysis. According to the results of the study, the total risk of companies with high business and financial risks will be high and the total risks of companies with low business and financial risks will be low. In terms of total risk, it is seen that the most risky sector is the textile sector and the least risky sector is the stone-land sector.

Aktaş and Doğanay (2007) grouped the stock market in emerging markets according to market data. As a result, they obtained 3 different groups. As a result of the study, the main variables that differentiate emerging markets are total market value, transaction volume and turnover rate. However, based on market data, investable emerging stock markets and advanced stock markets have also been grouped to determine whether the developed and emerging market segmentation is still valid. It is stated that this distinction is lost when market data is taken as basis.

In the international arena, different cluster methods are more common with the groupings of companies and stocks. One such researcher, Arnott (1980), examined the movements in stock prices through cluster analysis. In his study, he formed a set of five stocks corresponding to important foreign market factors. In the cluster process, we observe that the value of a cluster is explanatory for the peak of external market share price movements. Accordingly, the cluster process is terminated when it is started to be diluted by unrelated stocks (for example, at the point where food companies have joined the public services cluster). The resulting clusters have resulted in more than 30 percent more than the single-indexed model, which explains the importance of evaluating foreign market risk.

Tola et al. (2008) argue that cluster algorithms can increase the reliability of portfolios in terms of the ratio between expected and actual risk. They performed portfolio optimization using filtered correlation coefficient matrices in their studies. The matrices are obtained by applying different filtering methods to the original correlation coefficient matrix. Researchers have proposed two filtering methods based on average link and single link cluster procedures. The optimal portfolio obtained by these two new methods was compared with the model proposed by Laloux et al. (2000) and Rosenow et al. (2002). This model applies under ideal conditions and under more realistic conditions.

Da Costa et al. (2005), categorized 816 companies which have a daily average transaction volume higher than $ 100,000, from a total of 1959 publicly traded companies which are listed in the Economata data base according to their risk, return, price-earnings ratio, market value-book value, price-to-sales ratio, number of stocks-sales ratio and dividend yields obtained from 2 different time intervals by cluster analysis. As a result of the study, if an investor chooses according to the stocks listed in the clusters in the first time interval, it is seen that the stocks will be profitable to the investor in the second time interval.

Another work on this subject belongs to Momeni, Mohseni and Soofi (2015). In the study they conducted, 87 companies in three different sectors traded on the Teheran Stock Exchange have set aside clusters by prioritizing the variables of ROA, ROE, net profit / sales, earnings per share and operating profit margins according to the analytical hierarchy process (AHP). As a result, all companies were assembled in 2 clusters using the K-means cluster method.

Fodor, Jorgensen and Stowe (2015) classified the various companies in the United States according to their financial and operational characteristics. In their study they used cluster analysis, 1,641 companies classified clusters by K-Average methods which is one of the non-hierarchical cluster methods. 21 variables were used and 25 clusters were formed as a result of the study. As a result of the study, there were significant differences between the clusters in terms of their financial characteristics and the sector they are in. According to another finding of the study, membership of the cluster reveals a significant difference between the returns of different stocks. The clusters and sectors were compared to explain the relationship between the returns, and it was seen that there was a strong relationship in both.

The companies in the CSI 300 index are cluster with the Isomap (Isometric Specification Mapping) process based on their closing prices by Liu, Cai and Luo (2012). Isomap is one of the nonlinear size reduction algorithms. According to the graph which is formed as a result of the analysis carried out by the researchers in the Matlap program, 8 clusters have emerged. When Isomap bundles stocks according to their trends, stocks in the same group have similar trends. In short, using Isomap, they have shown that stocks can be grouped according to trends they exhibit. In addition, Isomap has shown to be more effective when compared to LLE (Local Linear Placement), another nonlinear size reduction algorithm.

Some other studies on the classification of companies include Zhou et al. (2002), Doherty et al. (2005), Basolto et al. (2005), Xu et al. (2008), Yu and Wang (2009) and Nanda et al. (2010).

The hypothesis generated as a result of the literature survey is as follows.

> ***H_1:*** *It is possible to classify the stocks traded in the Stock Exchange Istanbul by the cluster analysis and to determine the shares with the highest capital gains (difference between purchase and sale).*

Research Methodology

Cluster Analysis

One of the problems faced by researchers working in different areas is how to make the observed data meaningful, in other words. One of the methods proposed to answer this question is the cluster analysis. Cluster analysis was first used by Tyron (1939). Tyron (1939) and Cattell (1944) were the first to introduce mathematical procedures that classify objects according to observed similarities (Gore 2000). Cluster analysis refers to the classification of a set of objects or subjects based on a set of specific variables, such that those with similar characteristics are included in the same clusters. According to another definition, cluster analysis is an uncontrolled classification process. This technique has been frequently used in different fields such as marketing, health, mathematics, geology, finance, production and meteorology (Gan and Wu 2007). This method does not make a distinction according to whether the variables used in the research are relevant or unrelated. For this reason, variables to be subjected to cluster analysis need to be conceptually supported. The significance of this point is increased if the probability that the clusters to be formed will depend on unrelated variables (Cornish 2007).

The cluster process is carried out in two ways. The first of these is the hierarchical cluster analysis. This is the most commonly used method in practice (Kalaycı 2009). Hierarchical cluster methods are divided into two as hierarchical and divisive. The more preferred of these two is the divisive hierarchical method because it is easy to read and interpret (Kalayci 2009). In this method, the observations obtained in the first place are collected in one cluster and then the observations which are most contrary to this cluster are removed one by one from the cluster

to form the other clusters. The other way in which the cluster analysis is carried out is the non-hierarchical cluster analysis. In this method, firstly the number of clusters is determined on the basis of the researcher's prior knowledge and experience. Then clusters are formed by creating similar observations around a certain observation of each cluster (Kalaycı 2009).
In the cluster analysis, it may be desirable to divide the objects by the predetermined number of clusters or to reveal the number of clusters by the analysis. In both cases, tree diagrams which are called dendrogram are formed (Da Costa, Cunha and Da Silva 2005).
In this study, the hierarchical cluster method of cluster methods, the Ward method of separating companies into clusters and the squared Euclidean distance as distance measure were used. The Ward method is often regarded as the best method in hierarchical cluster methods (Hands and Everitt 1987; Ferreira and Hitchcock 2009). The Ward method is the only method among the agglomerative cluster methods that allows cluster to occur by minimizing the intra-group distribution in each binary fuzzy based on the classical square sum criterion (Murtagh and Legendre 2014). For this reason, the Ward method has a more complex structure than the other hierarchical methods. In this method, the goal is to place the objects within the cluster in such a way that the variance between the objects is minimal (Nakip 2006).
In the cluster analysis, the clusters based on the distance measures compare the similar ones from the objects. However, the samples may be different from each other. Euclidean and square Euclidean distances are the most frequently used as distance measures (Kalayci 2009). Euclidean distance, in a n * p dimensional data matrix, determines distances between i. and j. units (observations, objects) as directly measure unit (Euclidean distance) or in the form of quadratic distances (Squared Euclidean distance). The squared Euclidean distances need to be calculated when the Ward method is applied. Euclidean distance is calculated by the following formula (Özdamar 2010);

$$d(i,j) = \sqrt{\sum_{k=1}^{p} \left(X_{ik} - X_{jk}\right)^2}$$

Where i = 1, 2, ..., n; j = 1, 2, ..., n and k = 1, 2, ..., p. N is the number of units and p is the number of variables. The squared Euclidean distance is also calculated as the Euclidean distance. However, according to the variables, the total distance is not taken as the square root.

$$d(i,j)^2 = \sum_{k=1}^{p} \left(X_{ik} - X_{jk}\right)^2$$

Data and Variables

The study is based on the financial indicators of companies, which are traded on the BIST 100 index as of April 5, 2016, excluding sports clubs and financial companies. The financial statements of the companies were obtained from the website of the Public Disclosure Platform (KAP) and from the web pages of the companies. The variables used in the study are the price / earning ratio, market value / book value ratio, dividend yield, return on assets, return on equity, return, return on average and risk. These variables are the most considered variables in company and stock valuation and portfolio creation studies. Within the scope of this study, the reasons why these variables are chosen are tried to be expressed below.

Price / Earnings (P / E) and Market Value / Book Value (MV / BV) Rates

The P / E ratio shows how much investors are willing to pay per 1 TRY post-tax gain. Calculated by dividing the price per share by gross earnings per share. The P / E ratio should be compared with the growth rate of the profit, and attention should be paid to whether it is proportional to the future growth expectation. This ratio can also be compared with other companies in the sector.

Investors' level of assessment of a company can be understood by looking at MV / BV ratio. The higher MV / BV value of a company, the more likely it is that investors like the company. This ratio also compares the present value of the investments made to the company with its costs. This ratio is the ratio of market value of shares to book value per share. In the calculation of the book value per share, the book value (paid-in capital + non-distributed income) of the company's equity is divided by the number of shares in the market. Researchers such as Demir (2001), Ege and Bayrakdaroğlu (2012) and Korkmaz and Karaca (2013) are some of the researchers who have considered these two ratios in terms of stock turnover and the effect on company performance.

Risk and Return

The risk is, as you know, the possibility of things going undesirable. In the Webster dictionary, risk refers to the concept as "danger", "luck", and "exposure to damage". When considered in terms of financial markets and investments, it is generally accepted that risk is defined as the variance of the probability distribution of the incomes (Mazıbaş 2008). In this study, risk is the uncertainty about the probability of realizing expected returns from stocks. In general, the standard deviation or variance of the expected return is taken into account in the risk calculation. As is known, the standard deviation is the square root of the variance. For this reason, in fact, both alternatives mean the same thing. The formulas for variance and standard deviation are as follows;

$$\sigma^2 = \sum_{i=1}^{N} (R_i - R)^2 P_i \quad \text{(variance)}$$

$$\sigma = \sqrt{\sum_{i=1}^{N} (R_i - R)^2 P_i} \quad \text{(standard deviation)}$$

In these formulas R_i: the return of i. probability, R is the expected rate of return, and P_i expresses the possibility of the realization of the return.

The return expressed in this study is the expected return and is taken as the geometric mean of the daily returns of the stocks. The average return is the arithmetic mean of daily returns of stocks. The arithmetic average is calculated by the following formula (The arithmetic mean of N numbers is calculated by dividing the sum of these numbers by N);

$$X_A = \frac{1}{N} \sum_{i=1}^{N} x_i$$

The geometric mean is calculated by the following formula,

$$X_G = \sqrt[N]{x_1.x_2......x_N}$$

Since the use of this formula for the expected return could lead to deviations with negative consequences, the following formula was used to calculate the expected return;

$$1+\text{Geo}=[(1+X_1)x(1+X_2)x\ldots(1+X_n)]^{1/n}$$

Risk and return ratios are also taken into account in classifying stocks or in stock preferences. Markowitz (1952) has introduced a new approach to portfolio diversification by means of the mean variance model that he has taken into account of these two facts. The modern portfolio theory describes the expected return on the investor's exposure to risk (Demirtaş and Güngör 2004). Turanli, Özden and Demirhan (2002), Momcilovic, Njegic and Jovin (2012), Karabayir and Doğanay (2010) are some of the researchers that take these two financial indicators into account.

Dividend Yield

The factors that encourage investors to invest in stocks are requests to benefit from profit increases, capital gains, free share increase and stock price increases due to repurchase of shares (Ang and Liu 2007). Investors' interest in stocks is largely shaped by the profit share the company has made and its sales revenues (Ünlü, Bayrakdaroğlu and Ege 2009). Throughout the retention period of the stock, investors benefit from the dividend income earned. For this reason, the amount of dividends paid must be taken into consideration in the valuation of stocks. Dividend yield is also known as "yield rate" and is calculated as;

$$\boldsymbol{Dividend\,Yield = (Dividend\ per\ Share)\,/\,(Stock\ Price)}$$

Kurtaran, Kurtaran, Çelik and Temizer (2015) have shown that the dividend yield is a significant positive effect on the company value along with some other financial ratios. In the study done by Aydoğan and Güney (1997), it is stated that the price / earnings ratio and the dividend yield can be important tools in estimating stock returns.

Return on Asset (ROA)

The ROA, a financial ratio taken into account in analyzes by researchers such as Dehuan and Jin (2008), Moderes et al. (2008), Aydemir, Ogel and Demirtas (2012) and Büyükşalvarcı (2011), is a measure of financial performance that shows how much profit the companies have made from the investments they made. In other words, it shows the extent to which an entity effectively uses its assets. The ROA is calculated as follows;

$$ROA = (Net\ Income)\,/\,(Total\ Asset)$$

Return on Equity (ROE)

ROE is another commonly accepted category of financial ratios that is frequently used in domestic and foreign studies in assessing stock returns and performance. Researchers such as Donaldson and Davis (1991), Omran and Ragab (2004), Dehuan and Jin (2008), Siqueira, Otuki and Newton da Costa (2012) are some of the researchers who have studied the relationship between ROE and corporate performance and stock returns. The ROE of a company is calculated by the ratio of the income after the tax to the equity. This ratio represents the amount of money each company earns for each TRY they invest in the company and calculated by the formula as follow;

$$ROE = (Net\ Income) / Equity$$

Change in Sales

Change in sales is often used as one of the growth indicators of companies. For example, Kallapur and Trombley (1999) were considered the growth variables as the growth in sales, assets and equity capital. Other researchers who consider the amount of change in sales are Robinson (1995), Chowdhury and Chowdhury (2010), Lin and Chang (2011), Erdogan (2014), Chen (2015) and Fisher, Shah and Titman (2016). However, Lakonishok, Shleifer and Vishny (1994) point out that there is a negative relationship between the growth direction of sales and future returns of stocks.

The change in sales is calculated as follows;

$$\delta_S = \frac{S_T - S_{T-3}}{S_{T-3}}$$

δ : Three-Year Change Rate in Sales

S_T : 2015 Year Sales

S_{T-3}: 2012 Year Sales

Change in Equity

In addition to the above variables, it is considered that the change in the shareholders' equity is also an important financial indicator. As is known, especially during periods of high volatility and crisis periods, companies that have strong equity and continue to grow, offer a stronger image with the power they receive from their own equity. This variable is an indication of how the companies' own capital has changed over the past 3 years. In the case of an increase, shareholders' share of the company's profits is increased, and in the case of a decrease, this share is reduced. In general, investors prefer stocks of firms that have relatively stronger equity. The change in equity is calculated as shown below;

$$\delta_E = \frac{E_T - E_{T-3}}{E_{T-3}}$$

δ : Three-Year Change Rate in Equity

E_T : 2015 Year Equity Value

E_{T-3}: 2012 Year Equity Value

APPLICATION

Determination of Number of Cluster

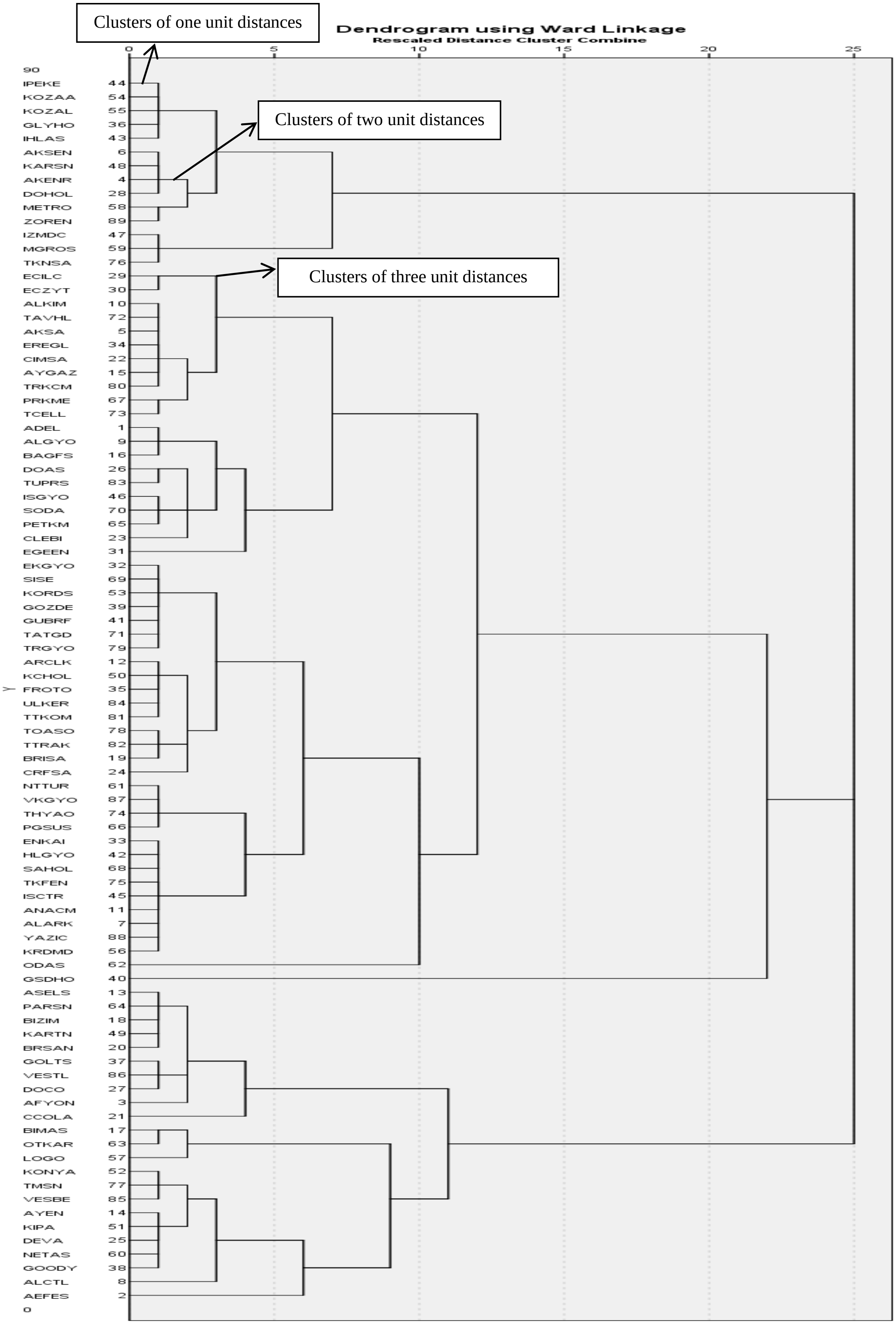
One of the biggest problems of the cluster analysis is determining the number of clusters. The goal is to achieve the optimal number of clusters by determining the cluster number. In this context, there are a number of methods that test the validity of formed clusters (Milligan and Cooper, 1985), determining the number of clusters that occur or will occur in non-hierarchical methods (Marriot 1971; Calinsky and Harabasz 1974). These methods are called validity indices. These indices are used to measure the validity of clusters formed as a result of cluster analysis (Gan and Wu 2007). In the hierarchical cluster analysis, this process can be done by considering the distances between the units and the connections with the help of the dendrogram graphics. Milligan and Cooper (1987) stated that a seven-step structure was used to determine the clusters to be formed in cluster analysis. According to the nature of the research and implementation, these steps are listed as follows (Çakmak 1999);

1. Firstly the units / elements to be clustered should be selected to represent the general structure of the cluster.
2. Then select variables with enough information to allow for the cluster of the individuals.
3. It should be decided whether the data will be standardized or not.
4. Distance or similarity criteria should be determined.
5. The appropriate cluster method should be chosen for the purpose of the research. Different results can be achieved with different methods.
6. The number of clusters should be determined.
7. It is the last and most important step in the cluster analysis. Includes interpretation, testing and applicability. Interpretation is possible only if the researcher has specific knowledge of the field of application. The test involves determining whether the clusters formed as a result of the analysis are meaningful. Applicability is to determine whether the results obtained can be applied to other samples or to the environment. These steps have been taken into account in the analysis carried out within the scope of this study.

Dendrogram Graph and Generated Clusters

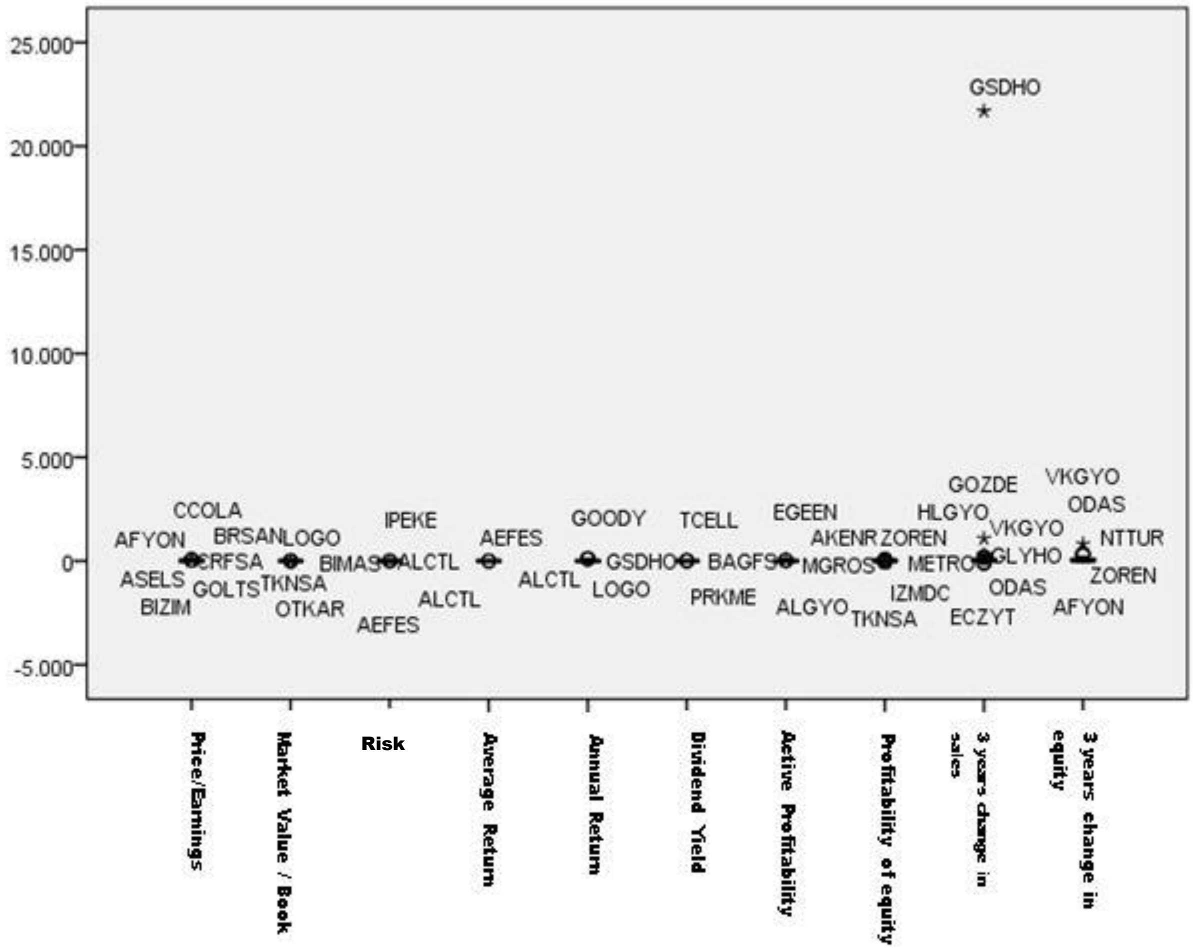
In the hierarchical cluster analysis, used the Ward method as the hierarchical method and the square Euclidean distance as the distance measure. The resulting dendrogram graphic is as follows (Graph 1). On the horizontal axis of the graph is the stocks, on the vertical axis is the distances of the stocks to each other and the connections between them. According to the graph, there are 30 clusters at the "one" unit distance. And 19 "two" unit distance, 15 "three" unit distance, 12 "four" unit distance and 8 "seven" unit distance. It should be noted here that GSDHO shares constitute a cluster of all the distances up to a distance of 22 units.

Graph 1: Dendrogram Graph



The reason for the emergence of this situation will be explained more clearly by "extreme value analysis". The graph below shows extreme value analysis. This graph better explains the difference in GSDHO feeling and why it forms a cluster alone. Accordingly, looking at the variables immediately below the GSDHO, it will be seen that the 3-year sales growth equals 20.000-25.000 per cent. As a matter of fact, the value calculated as 109.000 TL in 2012 was realized as 31.583.000 TL in 2015. This is much higher than the BIST average. According to this analysis, we reach the result that GSDHO's value is in contradiction with 3 year sales growth.

Graph 2: Extreme Value Analysis



Observations in the clusters formed at different distances seem to be clustered generally at "four" unit distances. For this reason, the number of suitable clusters is 12. When this result is achieved, it is considered those 30 clusters at a unit distance and a cluster at a distance of 25 units, and it is desirable that the number of observations per cluster be as equal as possible. Also, in cluster analysis, the main purpose is to separate units or objects as much as possible and to aggregate the most similar units in the same cluster. Now, let's show the clusters formed in tabular form.

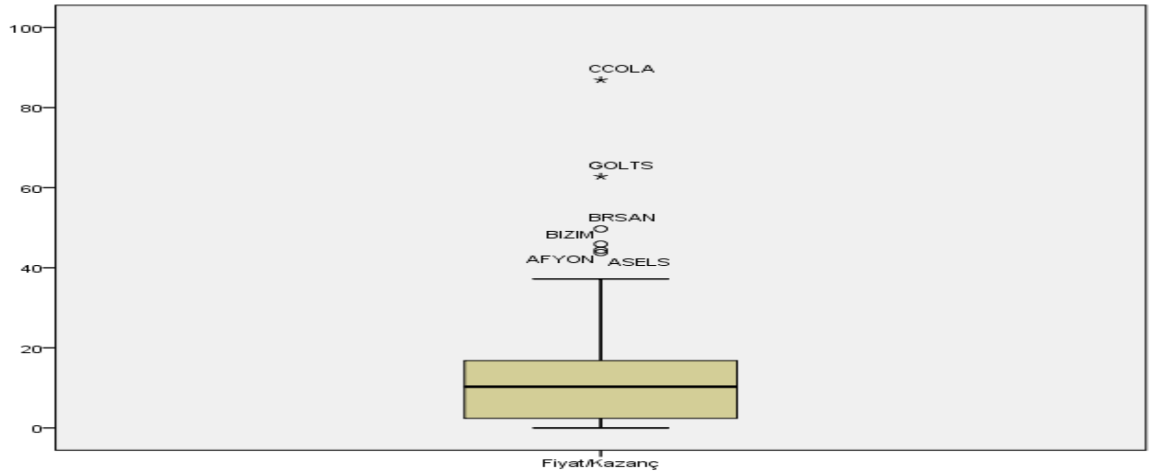
The following Table 1 shows 30 clusters formed according to “one” unit distance. When these clusters are examined in general, the following substances multiply;

Table 1: Clusters based on "one" unit distance

Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**Cluster 7**	**Cluster 8**	**Cluster 9**	**Cluster 10**
IPEKE	AKSEN	METRO	IZMDC	ECILC	ALKIM	PRKME	ADEL	DOAS	ISGYO
KOZAA	KARSN	ZOREN	MGROS	ECZYT	TAVHL	TCELL	ALGYO	TUPRS	SODA
KOZAL	AKENR		TKNSA		AKSA		BAGFS		PETKM
GLYHO	DOHOL				EREGLI				
IHLAS					CIMSA				
					AYGAZ				
					TRKCM				
Cluster 11	**Cluster 12**	**Cluster 13**	**Cluster 14**	**Cluster 15**	**Cluster 16**	**Cluster 17**	**Cluster 18**	**Cluster 19**	**Cluster 20**
CLEBI	EGEEN	EKGYO	ARCLK	TOASO	CRFSA	NTTUR	ENKAI	ODAS	GSDHO
		SISE	KCHOL	TTRAK		VKGYO	HLGYO		
		KORDS	FROTO	BRISA		THYAO	SAHOL		
		GOZDE	ULKER			PGSUS	TKFEN		
		GUBRF	TTKOM				ISCTR		
		TATGD					ANACM		
		TRGYO					ALARK		
							YAZIC		
							KRDMD		
Cluster 21	**Cluster 22**	**Cluster 23**	**Cluster 24**	**Cluster 25**	**Cluster 26**	**Cluster 27**	**Cluster 28**	**Cluster 29**	**Cluster 30**
ASELS	GOLTS	AFYON	CCOLA	BIMAS	LOGO	KONYA	AYEN	ALCTL	AEFES
PARSN	VESTL			OTKAR		TMSN	KIPA		
BIZIM	DOCO					VESBE	DEVA		
KARTN							NETAS		
BRSAN							GOODY		

- Looking at the companies that form the cluster alone, it is seen that they are CLEBI, AEGEN, CRFSA, ODAS, GSDHO, AFYON, CCOLA, LOGO, ALCTL and AEFES. These companies operate in different sectors and are not included in any cluster. That is why they must have their own financial indicators. Again, extreme value analysis was used to identify this situation and to identify stocks with outliers for each variable used in the study. Accordingly, in the analysis chart (Graph 3), which is made in the context of the following P / E variable, it is seen that the CCOLA is the one with the most contrary value although there are other outliers. At that time, the main reason for the cluster formation of the CCOLA feeling alone is that the P / E ratio deviates considerably from the average.

Graph 3: Outlier analysis by Price / Earnings



ratio

- When the same analysis is done for the other stand-alone set of stocks have quite different values in terms of LOGO MV / BV and annual return variables, AEFES risk and average return variables, ALCTL risk, annual and average return and risk variables, ODAS 3 year change in equity variable, CLEBI ROE variable, EGEEN ROA variable, the CRFSA MV / BV variable, AFYON also MV / BV variable and the 3 year change in equity, the GSDHO dividend yield and the 3 year sales growth variables. These values are the main causes of cluster formation on their own. In short, stocks forming a single cluster at a unit distance constitute contradictory observations in terms of relevant financial indicators.
- This is more clearly seen in the table below (Table 2). The "Case Number" column in the table shows the order of the stocks' in data file (for example, CCOLA in the 21st row from top to bottom) and the column of "Financial Indicators Value" shows the value per share for each variable (e.g. CCOLA's P / E ratio is 86.90). The "Highest" and "Lowest" lines represent the first five stocks with the highest relative variable value and the first five with the lowest relative variable value.

Table 2: The first five stocks with extreme value for each variable (Financial Indıcator)

			Case Number	**BİST 100 Stocks**	**Financial Indicator Value**
Price / Earnings	Highest	1	21	CCOLA	86,90
		2	37	GOLTS	62,80
		3	20	BRSAN	49,70
		4	18	BIZIM	45,90
		5	3	AFYON	44,40
	Lowest	1	89	ZOREN	,00
		2	88	YAZIC	,00
		3	76	TKNSA	,00
		4	62	ODAS	,00
		5	59	MGROS	,00[a]
Market Value / Book Value	Highest	1	57	LOGO	12,70

		2	17	BIMAS	11,20
		3	63	OTKAR	10,60
		4	76	TKNSA	9,20
		5	24	CRFSA	7,60
	Lowest	1	58	METRO	,20
		2	54	KOZAA	,30
		3	61	NTTUR	,40
		4	44	IPEKE	,40
		5	43	IHLAS	,40[b]
Risk	Highest	1	2	AEFES	6,41
		2	44	IPEKE	4,22
		3	8	ALCTL	4,22
		4	54	KOZAA	4,03
		5	85	VESBE	3,95
	Lowest	1	24	CRFSA	1,46
		2	18	BIZIM	1,49
		3	15	AYGAZ	1,53
		4	33	ENKAI	1,58
		5	50	KCHOL	1,59
Average Return	Highest	1	2	AEFES	,50690
		2	8	ALCTL	,36910
		3	21	CCOLA	,34340
		4	26	DOAS	,27520
		5	57	LOGO	,25840
	Lowest	1	88	YAZIC	-,11740
		2	66	PGSUS	-,11330
		3	56	KRDMD	-,10450
		4	76	TKNSA	-,07480
		5	47	IZMDC	-,05240
Annual Return	Highest	1	57	LOGO	117,23
		2	8	ALCTL	117,20
		3	38	GOODY	89,08
		4	27	DOCO	70,92
		5	14	AYEN	70,05
	Lowest	1	44	IPEKE	-43,95
		2	58	METRO	-40,00
		3	54	KOZAA	-37,32
		4	55	KOZAL	-32,80
		5	66	PGSUS	-30,04
Dividend Yield	Highest	1	40	GSDHO	16,00
		2	67	PRKME	13,20
		3	73	TCELL	13,20
		4	29	ECILC	11,90
		5	26	DOAS	10,40
	Lowest	1	89	ZOREN	,00
		2	87	VKGYO	,00
		3	86	VESTL	,00
		4	77	TMSN	,00
		5	74	THYAO	,00[a]

ROA	Highest	1	31	EGEEN	47,08
		2	9	ALGYO	25,00
		3	16	BAGFS	24,46
		4	57	LOGO	21,65
		5	1	ADEL	21,61
	Lowest	1	58	METRO	-12,61
		2	76	TKNSA	-9,55
		3	47	IZMDC	-9,03
		4	4	AKENR	-8,19
		5	59	MGROS	-6,53
ROE	Highest	1	23	CLEBI	72,71
		2	31	EGEEN	61,09
		3	16	BAGFS	57,25
		4	17	BIMAS	41,36
		5	57	LOGO	40,98
	Lowest	1	76	TKNSA	-71,40
		2	47	IZMDC	-54,09
		3	59	MGROS	-52,05
		4	89	ZOREN	-41,67
		5	4	AKENR	-32,76
Change in Sales	Highest	1	40	GSDHO	21686,00
		2	39	GOZDE	1078,00
		3	42	HLGYO	271,76
		4	36	GLYHO	262,65
		5	79	TRGYO	179,09
	Lowest	1	58	METRO	-100,00
		2	30	ECZYT	-98,55
		3	87	VKGYO	-65,69
		4	54	KOZAA	-47,00
		5	44	IPEKE	-43,62
Change in Equity	Highest	1	62	ODAS	816,64
		2	66	PGSUS	344,61
		3	3	AFYON	330,39
		4	87	VKGYO	326,12
		5	89	ZOREN	297,77
	Lowest	1	76	TKNSA	-61,01
		2	59	MGROS	-59,15
		3	47	IZMDC	-57,95
		4	43	IHLAS	-43,93
		5	6	AKSEN	-29,00

- When we look at the clusters formed by stocks other than the stand-alone stocks, it seems that even at the smallest distance, 9 shares in the 18th cluster are located together. This is another indication that the relevant variables are effective in separating stocks into clusters.
- Identifying variables that influencing the formation of clusters is another important subject. ANOVA analysis was performed for this purpose. As can be seen in Table 3, all variables are significant in separating clusters ($p < 0.01$).

Table 3: ANOVA analysis of clusters according to one unit distance

		Sum of Squares	F	p
P/E	Between Groups	19186,606	15,202	,000
	Within Groups	2567,785		
MV/BV	Between Groups	471,949	14,875	,000
	Within Groups	64,549		
Risk	Between Groups	48,497	17,815	,000
	Within Groups	5,539		
AR	Between Groups	,874	10,718	,000
	Within Groups	,166		
Return	Between Groups	65326,129	8,114	,000
	Within Groups	16378,833		
Dividend Yield	Between Groups	1025,902	14,038	,000
	Within Groups	148,678		
ROA	Between Groups	5733,162	17,664	,000
	Within Groups	660,316		
ROE	Between Groups	40920,568	21,562	,000
	Within Groups	3861,022		
Change in Sales	Between Groups	462883635,561	852,329	,000
	Within Groups	1104888,356		
Change in Equity	Between Groups	1059854,106	22,793	,000
	Within Groups	94602,115		

The 30 clusters and notable points in a unit distance are summarized above. The reason why the first steps of the clusters are given in detail is to ensure that clusters in the next stages, how the most appropriate cluster is formed, and the other causes underlying these clusters are followed in detail. Now let's continue by examining the clusters that are formed at a distance of 4 units that we have determined as the best number of clusters. Table 4 shows the clusters formed at a distance of four units. When the resulting clusters are examined, the following conclusions are reached;

Table 4: Clusters according to "four" unit distance

Cluster 1		**Cluster 2**	**Cluster 3**		**Cluster 4**		**Cluster 5**		
IPEKE	DOHOL	IZDMC	ECILC	AYGAZ	ADEL	PETKM	EKGYO	ARCLK	BRISA
KOZAA	METRO	MGROS	ECZYT	TRKCM	ALGYO	CLEBI	SISE	KCHOL	CRFSA
GLYHO	ZOREN	TKNSA	ALKIM	PRKME	BAGFS	EGEEN	KORDSA	FROTO	
IHLAS			TAVHL	TCELL	DOAS		GOZDE	ULKER	
AKSEN			AKSA		TUPRS		GUBRF	TTKOM	
KARSN			EREGLI		ISGYO		TATGD	TOASO	
AKENR			CIMSA		SODA		TRGYO	TTRAK	

Cluster 6			**Cluster 7**	**Cluster 8**	**Cluster 9**		**Cluster 10**	**Cluster 11**		**Cluster 12**
NTTUR	SAHOL	KRMD	ODAS	GSDHO	ASELS	VESTL	BIMAS	AYEN	KONYA	AEFES
VKGYO	TKFEN				PARSN	DOCO	OTKAR	KIPA	TMSN	
THYAO	ISCTR				BIZIM	AFYON	LOGO	DEVA	VESBE	
PGSUS	ANACM				KARTN	CCOLA		NETAS		
ENKAI	ALARK				BRSAN			GOODY		
HLGYO	YAZIC				GOLTS			ALCTL		

Let's first examine which of the clusters that occur at one unit distance are joined at a distance of four units.

- Accordingly, the first, second, and third clusters converge to form a cluster 1 at a distance of four units.
- The 4th cluster is not merged with any cluster and is formed the 2nd cluster.
- 5th, 6th and 7th clusters merged to form 3rd cluster.
- 8th, 9th and 10th clusters merged to form 4th cluster, 13th, 14th, 15th and 16th clusters clusters merged to form 5th cluster, 17th and 18th clusters merged to form 6th cluster.
- The 19th and 20th clusters were unchanged and they formed the 7th and 8th.
- 21, 22, 23, 24 clusters formed the cluster 9, clusters 25 and 26 clusters formed the cluster 10 and 27, 28 and 29 clusters formed the cluster 11.
- The 30th cluster is unchanged and formed the 12th cluster.
- Another consequence is that companies in the clusters are involved in different sectors. This is an appropriate result for portfolio diversification.

As a result, 12 clusters came to the four-unit distance and 18 of the clusters based on one unit distance merged with another cluster. When the clusters formed at a distance of four units are examined in general, the following results are achieved by moving from the mean values of the clusters given in Table 5. According to this;

- The first three clusters with the highest P / E ratio are Cluster 9 (46,98), Cluster 10 (33,20) and Cluster 11 (16,63). Since this ratio is meaningless (for companies that disclose the loss in the period when the research is conducted), the clusters for which it is assumed to be zero are Cluster 2, Cluster 7 and Cluster 12.
- The first three clusters with the highest MV / BV ratio are Cluster 10 (11,50), Cluster 2 (6,30) and Cluster 11 (3,19). The cluster with the lowest rate is Cluster 8 (0,40), Cluster 6 (0,80) and Cluster 1 (0,96).
- When the groups are evaluated in terms of risk and return variables, it is seen that the first three groups of the most risky shares are Cluster 12, Cluster 11 and Cluster 1. The lowest clusters are Cluster 8, Cluster 6, and Cluster 3. The highest return clusters are cluster 10. Cluster 11 and Cluster 3. In that case, Cluster 3 is the most appropriate investment in terms of risk and return.
- When the dividend yields of the clusters are examined, it is seen that they are ranked as Cluster 8, Cluster 3 and Cluster 4. At this point again Cluster 3 is on the agenda. It includes stocks that can be preferred in terms of dividend yield as well as risk and return.
- Given their ROA, it is seen that Cluster 4, Cluster 10 and Cluster 8 are in the top three ranks. In terms of ROE, Cluster 4, Cluster and Cluster 8 are the first three clusters.
- Cluster 8, in which the growing sales obviously observe include GSDHO stock. This increase is due to the increase in sales figures in 2015. In 2015, sales increased from TRY 109,000 to TRY 31,583,000.

- The other clusters with the greatest change in sales are Cluster 7, Cluster 12, and Cluster 5. The average of these clusters indicates a sales change of over 100%.
- Growth in equity is most observed in Cluster 7, Cluster 6 and Cluster 4.

Table 5: Financial Indicators of the Clusters

	P/E	*MV/BV*	*RİSK*	*AR**	*R**	*D**	*ROA**	*ROE**	*CS**	*CE**
CLUSTER 1	1,95	0,96	3,11	0,06	-18,21	0,32	-3,89	-15,39	28,43	51,34
CLUSTER 2	0,00	6,30	1,99	-0,05	-9,91	0,60	-8,37	-59,18	35,35	-59,37
CLUSTER 3	15,14	1,40	1,99	0,05	27,27	8,82	7,35	12,64	19,13	29,13
CLUSTER 4	6,95	2,44	2,09	0,14	21,57	5,88	19,80	38,65	50,75	105,09
CLUSTER 5	13,61	3,07	2,04	0,11	5,16	2,93	6,33	18,16	111,67	46,88
CLUSTER 6	7,87	0,80	1,95	-0,01	-2,32	1,78	2,86	6,36	51,17	115,39
CLUSTER 7	0,00	1,90	2,66	0,14	-25,97	1,30	-0,11	-0,43	177,78	816,64
CLUSTER 8	1,30	0,40	1,82	-0,01	-15,33	16,00	8,11	30,52	21686,00	35,31
CLUSTER 9	46,98	2,79	2,41	0,09	26,42	0,70	2,62	5,69	43,10	81,84
CLUSTER 10	33,20	11,50	2,91	0,23	54,91	1,97	14,33	38,27	99,22	92,32
CLUSTER 11	16,63	3,19	3,21	0,20	53,23	1,27	6,08	12,32	30,57	18,92
CLUSTER 12	0,00	1,70	6,41	0,51	4,29	1,10	-0,94	-2,58	136,25	15,00

*AR: Average Return; G: Return; D: Dividend; CS: Changes in Sales; CE: Changes in Equity

In summary, we can say that, in terms of P / E and MV / BV ratios the clusters 10 and 11, in terms of risk, return and dividend yield cluster 3, in terms of ROE and ROA cluster 4 and in terms of increases in sales and equity Cluster 7 contain the best stocks to invest. What are the more favorable clusters among these clusters than the others, in order to build a portfolio, when only one cluster of stocks is excluded? At this point, an evaluation can be made by going through the stock performance of the related stocks within the last one year (06.05.2016-31.03.2017). Accordingly, it is seen that the cluster 3 has the highest performance with 27.18%, when the related groups in Table 6 are analyzed in terms of share performance in the last one year. The performance of cluster 11 is calculated to be -4.5%. In that case, we can say that the most suitable portfolio according to the performance of the last 1 year stocks is the portfolio which will be formed from the stocks in Cluster 3.

Table 6: Last 1 Year Stock Performance of Stocks in the Clusters and the Cluster Cumulative

Cluster 3		Cluster 4		Cluster 10	Cluster 11	
ECILC	**AYGAZ**	**ADEL**	**PETKM**	**BIMAS**	**AYEN**	**KONYA**
+51,86%	*+28,05%*	*-16,28%*	*+36,83%*	*-7,26%*	*+23,71%*	*-13,16%*
ECZYT	**TRKCM**	**ALGYO**	**CLEBI**	**OTKAR**	**KIPA**	**TMSN**
+21,25%	*+63,30%*	*+47,28%*	*-24,64%*	*+9,81%*	*-22,18%*	*-20,69%*
ALKIM	**PRKME**	**BAGFS**	**EGEEN**	**LOGO**	**DEVA**	**VESBE**
+77,89%	*+12,56%*	*-29,49%*	*-15,39%*	*+6,69%*	*+4,03%*	*+12,02%*
TAVHL	**TCELL**	**DOAS**	**ISGYO**	**Cluster Mean**	**NETAS**	**ALCTL**
-14,49%	*+4,72%*	*-26,50%*	*-6,75%*	3,08%	*-12,71%*	*+24,58%*
AKSA	**EREGLI**	**TUPRS**	**SODA**		**GOODY**	
+9,04%	*+40,70%*	*+23,65%*	*+51,46%*		*-36,49%*	
	CIMSA	**Cluster Mean**			**Cluster Mean**	
	+4,08%	4,01%			-4,5%	
Cluster Mean						
27,18%						

At this point, a different assessment can be made by going through the stock performance of the related stocks in the past periods. Accordingly, when the retrospective 3-year (03.04.2013-06.04.2016) stock performance of the relevant clusters is considered, performance of Cluster 10 is 59.54%, performance of Cluster 11 is 94.88%, that of Cluster 3 is 58.80% and that of Cluster 4 is 118%. In that case, we can say that the most suitable portfolio according to the performance of the last 3 years stocks is the portfolio which will be formed from the stocks in Cluster 4.

Another point is that ARCLK, KCHOL, TTRAK, FROTO, TOASO in Koç Holding have to take place together in the 5st cluster. Likewise, BIRSA, CRFSA, KORDSA which are operating while parts of Sabancı Holding were in the 5th cluster together. It is thought that the reason for this situation to occur is the fact that the management forms, activities and results of operations are similar to each other in the sense that they are different companies operating in the same holding company.

According to these results, it can be said that the performance of stocks composed of stocks with high equity and return on assets is higher and more preferable than others. Sorting according to stock performance can be given as Cluster 3, Cluster 4, and Cluster 10.

Interpretation of Agglomerative Table

The merging processes of the clusters described above can also be seen by examining the following agglomerative table (Table 7). In the agglomerative table, the shares most similar to each other are matched according to the coefficients related to the financial variables. Agglomerative table has n-1 stages. According to this, there are 89 stocks and 89 - 1 = 88 phases. Through this table, it can also be seen which stock is clustered with which stock at which stage. For example, IPEKE (44) and KOZAA (54) are merged in the first stage of cluster (stage 1) while EKGYO (32) and SISE (69) are merged in the second stage. Again according to this table, it can be detected the most similar shares. For this, it is necessary to look at the coefficients. The smaller coefficient, the more relevant the stocks are to each other. The most similar feelings come together in the first steps. Accordingly, as seen in Table 7, IPEKE and KOZAA are the most similar shares in terms of the variables involved (coefficient: 0,059). It is noteworthy that both companies are in Koza Holding. This can be interpreted as a sign of the consistency of the analysis performed. It is also present in companies under the same holding roof and in different clusters. However, it appears that these companies are among the most similar clusters to each other (e.g. TTRAK, KCHOL and TUPRS in Cluster 4 and Cluster 5).

Table 7: Most resembling stocks and financial indicators

	P/E	MV/BV	RİSK	AR*	R*	D*	ROA*	ROE*	CS*	CE*
IPEKE	7,9	0,4	4,22	-0,0132	-43,95	0	0,10	0,37	-43,62	22,13
KOZAA	6,7	0,3	4,03	-0,0174	-37,32	0	0,11	0,05	-47	24,68

*AR: Average Return; G: Return; D: Dividend; CS: Changes in Sales; CE: Changes in Equity

The stocks in the last period of the agglomerative table (Table 9) are least similar stocks. Table 8 gives the least similar stocks in terms of related variables. These stocks are ADEL and AEFES (coefficient: 880,000).

Table 8: Shares that are least similar to each other

	P/E	MV/BV	RİSK	AR	R	D	ROA	ROE	CS	CE
ADEL	6,9	2,4	2,69	0,0881	6,63	6,1	21,61	40,39	57,64	96,87
AEFES	0	1,7	6,41	0,5069	4,29	1,1	-0,94	-2,58	136,25	15

Again according to this table, the stages of formation of the clusters can be determined. For example, ECILC (29) and ECZYT (30) are members of Cluster 3, which takes place at the 72st stage (next stage column) with AKSA, at 81th stage with ADEL, at 85th stage with ALARK and continuing in this way.

Table 9: Agglomerative Table

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	44	54	,059	0	0	41
2	32	69	,208	0	0	24
3	68	75	,395	0	0	9
4	33	42	,635	0	0	28
5	5	34	,887	0	0	17
6	12	50	1,231	0	0	37
7	36	43	1,704	0	0	52
8	39	41	2,241	0	0	35
9	45	68	2,794	0	3	15
10	6	48	3,357	0	0	25
11	35	84	3,948	0	0	23
12	25	60	4,584	0	0	50
13	10	72	5,268	0	0	34
14	61	87	5,968	0	0	38

15	11	45	6,675	0	9	28
16	46	70	7,430	0	0	43
17	5	22	8,251	5	0	27
18	71	79	9,135	0	0	35
19	7	88	10,045	0	0	45
20	47	59	11,003	0	0	51
21	13	64	11,976	0	0	56
22	14	51	12,968	0	0	57
23	35	81	13,963	11	0	37
24	32	53	14,978	2	0	42
25	4	6	16,138	0	10	32
26	26	83	17,316	0	0	60
27	5	15	18,494	17	0	34
28	11	33	19,684	15	4	58
29	18	49	20,913	0	0	40
30	78	82	22,233	0	0	48
31	29	30	23,611	0	0	72
32	4	28	25,080	25	0	67
33	52	77	26,587	0	0	53
34	5	10	28,174	27	13	46
35	39	71	29,909	8	18	42
36	17	63	31,830	0	0	69
37	12	35	33,754	6	23	61
38	61	74	35,679	14	0	55
39	1	9	37,611	0	0	44
40	18	20	39,561	29	0	56
41	44	55	41,511	1	0	52
42	32	39	43,525	24	35	74
43	46	65	45,620	16	0	60
44	1	16	47,842	39	0	70
45	7	56	50,140	19	0	58
46	5	80	52,582	34	0	62
47	67	73	55,028	0	0	62
48	19	78	57,739	0	30	61
49	37	86	60,519	0	0	59
50	25	38	63,302	12	0	57
51	47	76	66,162	20	0	80
52	36	44	69,280	7	41	73
53	52	85	72,522	33	0	68
54	58	89	76,062	0	0	67
55	61	66	79,847	38	0	75
56	13	18	83,856	21	40	63
57	14	25	88,134	22	50	68
58	7	11	92,432	45	28	75
59	27	37	96,933	0	49	63
60	26	46	101,767	26	43	66

61	12	19	106,683	37	48	65
62	5	67	111,751	46	47	72
63	13	27	117,217	56	59	64
64	3	13	123,758	0	63	76
65	12	24	130,702	61	0	74
66	23	26	138,212	0	60	70
67	4	58	146,297	32	54	73
68	14	52	154,574	57	53	71
69	17	57	163,230	36	0	82
70	1	23	173,225	44	66	77
71	8	14	184,385	0	68	79
72	5	29	196,014	62	31	81
73	4	36	207,958	67	52	80
74	12	32	220,774	65	42	78
75	7	61	235,581	58	55	78
76	3	21	250,550	64	0	84
77	1	31	266,888	70	0	81
78	7	12	289,743	75	74	83
79	2	8	316,371	0	71	82
80	4	47	346,317	73	51	87
81	1	5	376,929	77	72	85
82	2	17	415,697	79	69	84
83	7	62	459,236	78	0	85
84	2	3	507,237	82	76	88
85	1	7	557,652	81	83	86

Anova Analysis

An ANOVA analysis was conducted to determine whether there is any difference in the financial indicators examined among the 12 clusters at the 4-unit distance, which is determined as the optimal number of clusters. Test hypotheses are as follows;

H1: The P / E ratios of stocks differ according to the clusters they are in and it is meaningful that the stocks are divided into clusters.

H2: The MV / BV ratios of stocks differ according to the group they are in and it is meaningful to separate the stocks into clusters.

H3: The RISK of stocks variable differ according to the group they are in and it is meaningful to separate the stocks into clusters.

H4: The RETURN of stocks variable differ according to the group they are in and it is meaningful to separate the stocks into clusters.

H5: The AVERAGE RETURN of stocks variable differ according to the group they are in and it is meaningful to separate the stocks into clusters.

H6: The DIVIDEND YIELD of stocks variable differ according to the group they are in and it is meaningful to separate the stocks into clusters.

H7: The ROA of stocks variable differ according to the group they are in and it is meaningful to separate the stocks into clusters.

H8: The ROE of stocks variable differ according to the group they are in and it is meaningful to separate the stocks into clusters.
H9: CHANGE IN SALES of stocks variable differ according to the group they are in and it is meaningful to separate the stocks into clusters.
H10: CHANGE IN EQUITY of stocks variable differ according to the group they are in and it is meaningful to separate the stocks into clusters.
The analysis results are presented in Table 10 below. Because of the all variables' sig. values are < 0.05, all the variables used in this study show a meaningful difference according to the clusters and all of them are meaningful for the stocks to be divided into the clusters.

Table 10: ANOVA analysis results of 12 Clusters

		F	Sig.
P / E	Between Groups	18,391	,000
	Within Groups		
MV / BV	Between Groups	16,826	,000
	Within Groups		
Risk	Between Groups	14,658	,000
	Within Groups		
Average Return	Between Groups	10,560	,000
	Within Groups		
Return	Between Groups	7,894	,000
	Within Groups		
Dividend Yield	Between Groups	16,800	,000
	Within Groups		
ROA	Between Groups	12,790	,000
	Within Groups		
ROE	Between Groups	22,847	,000
	Within Groups		
Change in Sales	Between Groups	2552,810	,000
	Within Groups		
Change in Equity	Between Groups	11,102	,000
	Within Groups		

Again, looking at the "mean plots" for each variable that is one of the outputs of the ANOVA analysis, it is seen which variable is more determinative in which clusters are formed (Appendix 1).

Conclusion and Evaluation

In this study, the availability of the cluster analysis was tested for the classification of stocks. At the same time, attempts have been made to create a portfolio from stocks traded in Borsa İstanbul with a different approach by means of this method. For this purpose, the stages of the cluster analysis and the results obtained are presented. In doing so, it is desirable to bring together the stocks with the highest possible capital gains, which have a greater return potential than the market return. The companies included in the BIST 100 index are divided into 12 different clusters by 10 financial ratios calculated from the financial statements. All of the variables used in dividing stocks into clusters are found to be meaningful.

According to P/E and MV/BV ratios most suitable stocks or portfolios to be invested as a result of the study, in the clusters of 10 and 11, in terms of risk, return and dividend yields cluster 3 and in terms of ROE and ROA cluster 4 and in terms of sales and equity, stocks in cluster 7. On the basis of last one year (05.04.2016[1]-30.03.2017) performances stocks in Cluster 3, Cluster 4 and Cluster 10, respectively, can be preferred. And on the basis of last 3 year (03.04.2013-05.04.2016) performances Cluster 4, Cluster 11, Cluster 10, and Cluster 3, can be preferred.

With the cluster analysis used in this study, complex and a large number of stocks that prices and performances depend on many variables have been transformed into smaller and more meaningful clusters that will help investors to make decisions more accurately.

Cluster analysis is a simple and at the same time practical method for classifying a set of complex data on the basis of certain variables, making it meaningful, analyzing it, present meaningful results, and using the obtained results as a helpful to decision making. Because of this feature, it is possible and feasible to use when it is desired to create a stock portfolio among a number of stocks on the stock market based on a large number of financial indicators or base factors.

In this study, the cluster analysis is referred to in detail at every step because the cluster process should be observed and a precise understanding of the operation is desired. It is thought that the cluster analysis method can be used as an effective method by the investors in order to create stock preference and portfolio. In subsequent studies it is thought that analytical hierarchy process can be used to determine the most used financial ratios by investors' in investment decisions and on the basis of these ratios stock selection can be made by using cluster or fuzzy cluster analysis.

Corresponding Author

Asst. Prof. Bilgehan Tekin,
Department of Business Administration, Faculty of Economics and Administrative Sciences,
Cankiri Karatekin University,
e-mail: btekin@karatekin.edu.tr.

Acknowledgement

Asst. Prof. Fatih Burak Gümüş
Department of Business Administration, Faculty of Business, Sakarya University
Email: fbgumus@sakarya.edu.tr.

REFERENCES

Aktaş, R. & Doğanay, M. M. (2007). Grouping Emerging Stock Markets Based on Market Data, ***BDDK Bankacılık ve Finansal Piyasalar***, 1(2).

Arnott, R. D. (1980). Cluster analysis and stock price comovement. ***Financial Analysts Journal, 36*** (6), pp.56-62.

[1] The date of when the analysis is done

Aydoğan, K. & Güney, A. (1997). Hisse senedi fiyatlarının tahmininde F/K oranı ve temettü verimi. ***İMKB Dergisi*, *1*(1), pp.83-96.

Aydemir, O., Serdar, Ö. G. E. L., & Demirtaş, G. (2012). The Role of Financial Ratios in Determining the Stock Prices. ***Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi***, *19*(2), pp.277-288.

Büyükşalvarcı, A. (2011). Finansal Analizde Kullanılan Oranlar ve Hisse Senedi Getirileri Arasındaki İlişki: Ekonomik Kriz Dönemleri İçin İmkb İmalat Sanayi Şirketleri Üzerine Ampirik Bir Uygulama. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, *25* (1).

Chen, M. H. (2015). Understanding the impact of changes in consumer confidence on hotel stock performance in Taiwan. ***International Journal of Hospitality Management***, 50, pp.55-65.

Chowdhury, A. & Chowdhury, S. P. (2010). Impact of capital structure on firm's value: Evidence from Bangladesh. *Business and Economic Horizons, 3*(3), pp.111-122.

Cornish, R. (2007). ***Statistics: 3.1 Cluster Analysis*** http://www.statstutor.ac.uk/resources/uploaded/clusteranalysis.pdf (02.03.2016)

Da Costa Jr, N., Cunha, J., & Da Silva, S. (2005). Stock selection based on cluster analysis. ***Economics Bulletin,*** *13*(1), pp.1-9.

Dehuan, J. & Jin, Z. (2008). Firm Performance and Stock Returns: An Empirical Study of the Top Performing Stocks Listed on Shanghai Stock Exchange. ***Academy of Accounting and Financial Studies Journal*, 12**(1), pp.79.

Demir, A. G. D. Y. (2001). Hisse Senedi Fiyatını Etkileyen İşletme Düzeyindeki Faktörler ve Mali Sektör Üzerine İMKB'de Bir Uygulama. ***Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 6**(2).

Demirtaş, Ö. & Güngör, Z. (2004). Portföy yönetimi ve portföy seçimine yönelik uygulama. ***Journal of Aeronautics and Space Technologies*, 1**(4), pp.103-109.

Doherty, K.A.J., Adams, R.G., Davey, N. & Pensuwon, W. (2005). Hierarchical Topological Cluster Learns Stock Market Sectors, ***ICSC Congress on Computational Intelligence Methods and Applications***, Istanbul. http://dx.doi.org/10.1109/CIMA.2005.1662299

Donaldson, L. & Davis, J. H. (1991). Stewardship theory or agency theory: CEO governance and shareholder returns. ***Australian Journal of management*, 16**(1), pp.49-64.

Ege, İ. & Bayrakdaroğlu, A. (2012). İMKB Şirketlerinin Hisse Senedi Getiri Başarılarının Lojistik Regresyon Tekniği İle Analizi. ***Uluslararası Yönetim İktisat ve İşletme Dergisi***, *5*(10), pp.139-158.

Erdoğan, A. İ. (2014). Büyüme İmkanı Göstergelerinin Şirket Büyümesi Tahminindeki Başarısı: BİST Örneği. ***Yönetim ve Ekonomi Araştırmaları Dergisi**, 12*(23), pp.113-125.

Ferreira, L. & Hitchcock, D. B. (2009). A comparison of hierarchical methods for cluster functional data. ***Communications in Statistics-Simulation and Computation**, 38*(9), pp.1925-1949.

Fisher, G. S., Shah, R. & Titman, S. (2016). ***Analyst Long-Term Growth Forecasts, Accounting Fundamentals and Stock Returns***. Available at SSRN: http://ssrn.com/abstract=2759332 or http://dx.doi.org/10.2139/ssrn.2759332, (03.04.2016)

Gan, G., Ma, C. & Wu, J. (2007). ***Data Cluster Theory, Algorithms And Applications (Asa-Sıam Series On Statistics And Applied Probability)***, Canada:SIAM Society for Industrial and Applied Mathematics Publishing

Gore, P. A. (2000). ***Cluster Analysis***, In book: Handbook of Applied Multivariate Statistics and Mathematical Modeling, pp. 297-321.

Hands, S, Everitt, B (1987). A Monte Carlo study of the recovery of cluster structure in binary data by hierarchical cluster techniques. ***Multivar. Behav. Res****. 22*, pp. 235-243.

Kallapur, S. & Trombley, M. A. (1999). The association between investment opportunity set proxies and realized growth. ***Journal of Business Finance & Accounting****, 26*(3-4), pp. 505-519.

Korkmaz, Ö. & Karaca, S. S. (2013). Şirket Performansini Etkileyen Faktörler ve Türkiye Örnegi/The Factors Affecting Firm Performance: The Case of Turkey. ***Ege Akademik Bakis*****,** *13*(2), pp. 169.

Lakonishok, J., Shleifer, A. & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. ***The journal of finance****, 49*(5), pp. 1541-1578.

Lin, F. L. & Chang, T. (2011). Does debt affect firm value in Taiwan? A panel threshold regression analysis. ***Applied Economics,*** *43*(1), pp.117-128.

Liu, R., Cai, H. & Luo, C. (2012). Cluster Analysis of Stocks of **CSI 300 Index Based on Manifold Learning.** ***Journal*** *of Intelligent Learning Systems and Applications, 4,* pp.120-126.

Mazıbaş, M. (2005). İMKB Piyasalarındaki volatilitenin modellenmesi ve öngörülmesi: Asimetrik GARCH modelleri ile bir uygulama. ***VII. Ekonometri ve İstatistik Sempozyumu***, İstanbul Üniversitesi İktisat Fakültesi Ekonometri Bölümü, İstanbul, Çevrimiçi, 28, 2008.

Markowitz, H. (1952). Portfolio selection. *The journal of finance*, *7*(1), pp. 77-91.

Momcilovic, M. I. R. E. L. A., Njegic, J. O. V. A. N. & Jovin, S. L. O. B. O. D. A. N. K. A. (2012). Risk, return and stock performance measures. In *Proc. 1st WSEAS Int. Conf. on Finance, Accounting and Auditing (FAA'12)* pp. 20-22.

Murtagh, F. & Legendre, P. (2014). Ward’s hierarchical agglomerative cluster method: Which algorithms implement ward’s criterion?. ***Journal of Classification*****,** ***31***(3), pp. 274-295.

Nakip, M. (2006), ***Pazarlama Araştırmaları: Teknikler ve (SPSS Destekli) Uygulamalar****,* Ankara: Seçkin Yayıncılık.

Nanda, S.R. Mahanty, B. & Tiwari, M.K. (2010). Cluster Indian stock market data for portfolio management. ***Expert Systems with Applications****, 37* (12) , pp. 8793–8798.

Omran, M. & Ragab, A. (2004). Linear versus non-linear relationships between financial ratios and stock returns: empirical evidence from Egyptian firms. ***Review of Accounting and finance*****,** ***3***(2), 84-102.

Özdamar, K. (2010), ***Paket Programlar ile İstatistiksel Veri Analizi: Çok Değişkenli Analizler***. Eskişehir: Kaan Kitabevi.

Robinson, K. C. (1995). *Measures of entrepreneurial value cre ation: An investigation of the impact of strategy and industry st ructure on the economic performance of independent new ventures*. An unpublished doctoral dissertation in the Department of Management. The University of Georgia, Athens, GA.

Siqueira, E., Otuki, T. & Da Costa, N. (2012). Stock Return and Fundamental Variables: A Discriminant Analysis Approach. ***Applied Mathematical Sciences***, ***6***(115), pp. 5719-5733.

Tola, V., Lillo, F., Gallegati, M., & Mantegna, R. N. (2008). Cluster analysis for portfolio optimization. ***Journal of Economic Dynamics and Control**, 32*(1), pp. 235-258.

Ünlü, U., Bayrakdaroğlu, A. & Ege, İ. (2009). Hisse Senedi Endeks Getirileri ve Temettü Verimi: İMKB 100 ve S&P 500 Endeksleri Üzerine Bir Uygulama. ***Gazi University Journal of Economics & Administrative Sciences***, *11*(1).

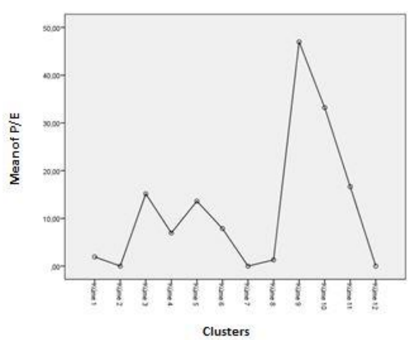
Xu, Z.-C., Liang, Y.-C. & Shi, X.-H. (2008). Analysis Method of SOM-Based of Stock Cluster, ***Computer Engineering and Design**, 29* (9), pp. 2426-2428.

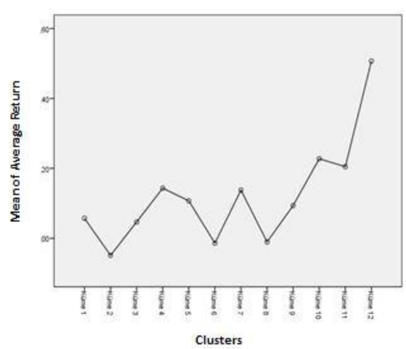
Yu, L.-A. & Wang, S.-Y. (2009). Kernel Principal Component Cluster Methodology for Stock Categorization, ***System Engineering: Theory & Practice***, 29 (12).

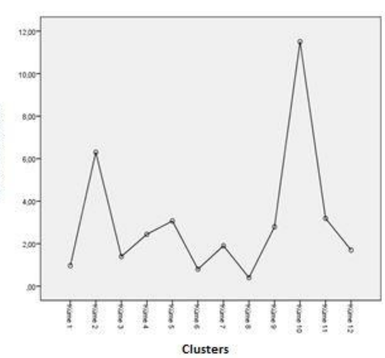
Zhou, Z.-H., Chen, W.-N. & Zhang, Z.-Y. (2002). Application of Cluster Analysis in Stock Investment, ***Journal of Chongqing University**, **25*** (7), pp. 122-126.

Appendix 1: Average Graphs of Variables

It is seen that the P / E ratio in the graphs is more determinant for the cluster 9 than the others. MV / BV value is more determinant for Cluster 10. When looking at the risk variable, the cluster has the highest level with 6.41 in 12 and is a decisive factor. The yield variant is a decisive factor for the cluster 10 and 11. Likewise, the average return, ROA and ROE, as well as changes in sales and equity variables can be examined on a set of charts in the following charts.







International Journal of Academic Research in Business and Social Sciences
2017, Vol. 7, No. 5
ISSN: 2222-6990