

THE SELECTION OF STOCKS USING PRINCIPAL COMPONENT ANALYSIS

Camelia Catalina Joldes¹⁸³

<https://doi.org/10.31410/itema.2018.454>

Abstract: *The purpose of this paper is to show how principal component analysis (PCA) can be used to construct the most efficient portfolios. Due to the large number of existing variables, it is difficult to perform statistical analysis of data, but PCA will help us in choosing the best performing shares. Through the PCA we will show the utility of this statistical method in the selection phase of the data. For this study we decided to select a number of representative companies for four central European countries: Czech Republic, Poland, Germany and Austria. The main objective of this analysis is to identify the most important factors indicating the most profitable shares, so we have considered a number of five financial indicators, representative for the selected firms.*

Keywords: *Stock selection, principal component analysis, diversification, data reduction, variance*

1. INTRODUCTION

The principal component analysis will help us in choosing the best performing shares, by reducing the original space to a smaller one.

In the first phase, we selected a group of shares traded on the stock exchanges in the Czech Republic, Poland, Germany and Austria. Due to the large number of existing variables it is difficult to perform statistical analysis of data, but principal component analysis will help us in choosing the best performing companies traded on the stock exchange. With this statistical tool we will reduce the initial space to a smaller one. The main purpose of this analysis is to identify the most important factors indicating the most profitable companies.

Subsequently, by applying a statistical size reduction method, we will obtain a number of shares to be as the most performing ones, which could be used to build optimal portfolios.

2. LITERATURE REVIEW

Fulga, Dedu and Serban [1] have developed an algorithm that takes into account the pre-selection of the financial titles and their allocation phase, collecting information for a total of 40 assets traded on the BSE. The procedure approached by the authors allowed them to build a diversified portfolio and improve the optimization algorithm.

The method of Principal Component Analysis was applied to three subgroups of stocks of the American index Down Jones Industrial. The authors showed the benefit of the PCA model in optimizing portfolio investments to achieve the best results and financial control [2].

¹⁸³ Bucharest University of Economic Studies, 6 Piata Romana, 1st district, Bucharest, 010374 Romania

Ngai and Cheng [3] described and demonstrated the applicability of two multivariate statistical techniques, principal component analysis and correspondence analysis, as analysis tools for quality professionals. Authors showed that these simple but powerful techniques can facilitate the tasks of quality practitioners, who are frequently concerned with the detection and interpretation of relationships among a given set of quality data.

Meriç, Ding and Meriç [4] use the principal component analysis method to study the global portfolio diversification opportunities for investors of seven developed and twenty emerging stock markets.

The authors selected a total of 12 listed companies and calculate their various financial indicators using the principal component analysis of the SPSS software for the financial performance, evaluation and analysis method [5].

Yap [6] investigates the application of principal component analysis in the selection of financial ratios that are significant and representative for two industry sectors in Malaysia. The aim of Yap's study is to use principal component analysis in order to obtain a more parsimonious and useful smaller set of financial ratios from the many available ratios that are found in the literature.

Another principal component analysis was applied on Australian market to show the value of the method. The study show that highly correlated stocks tend to remain highly correlated over long periods of times. Therefore, the results of this type of analysis can be used by investors to manage their portfolio [7].

By applying three data analysis techniques such as principal components analysis, cluster analysis and discriminant analysis to the data regarding the petitions that were registered by the insurance companies which activated in the Romanian market, Armeanu et. all obtained a classification that assesses the insurance market transparency [8].

3. DATA AND METHODOLOGY

The application we are proposing to illustrate - PCA - the size reduction, will be done on a sample of four Central European countries for a number of 5-8 variables.

The financial indicators selected for the principal component analysis are for the end of 2017 and are taken from the Thomas Reuters database.

Depending on the availability of the financial indicators (at the end of 2017) for the selected companies, we decided to take into account the following:

- for the Czech Republic we considered the following financial indicators: total return (TR), dividends per share (DPS), earnings per share (EPS), market capitalization (MC), total debt (TD), dividend yield (DY);
- for Poland we considered the following: total return (TR), dividends per share (DPS), earnings per share (EPS), total debt (TD), BETA, dividend yield (DY), market capitalization (MC), earnings before interest & taxes (EBIT);
- for Germany & Austria we considered: total return (TR), dividends per share (DPS), earnings per share (EPS), total debt (TD), BETA, dividend yield (DY), total assets (TA), earnings before interest & taxes (EBIT);

These fundamental indicators are used in the analysis of securities on the capital market:

- ✓ The **BETA coefficient**- is a measure of the volatility of a financial instrument or of a portfolio as a whole;
- ✓ **EBIT**- represents the gross surplus before tax deduction and interest expense deductions. After the payment of interest and profit tax is resulting the net profit used for the creation of own funds and shareholder remuneration results;
- ✓ **EPS** - one of the most used indicators is earnings per share, that is, a way to report the market value of a share with the returns to a common share;
- ✓ **DPS** - this indicator expresses how many cash units of cash dividends are paid per share in one year;
- ✓ **DY**- is an important indicator for investors who are interested in dividend shares;
- ✓ **MC**- is the market value of all company shares and a very good indicator of the size of a business;
- ✓ **TD** - a set of goods possessed in money and in nature or all the money resources invested in a business;
- ✓ **TR** - the ability of a firm to achieve revenue or, in other words, the ability of an enterprise to generate a surplus over expenditure.

4. RESULTS

Next we apply the Bartlett test of Sphericity which compares the correlation matrix of a random vector with a normal multivariate distribution with the unit matrix - a matrix with correlations equal to 0, except for the main diagonal, which has all elements equal to 1, since the Pearson linear correlation coefficient of variables with itself is equal to 1.

Table 1: Bartlett's sphericity test

	<i>Czech Republic</i>	<i>Poland</i>	<i>Germany & Austria</i>
Chi-square (Observed value)	21.39	53.85	45.66
Chi-square (Critical value)	3.94	16.93	16.93
p-value	0.02	0.00	0.02
alpha	0.95	0.95	0.95

The test's interpretation is:

- H₀: There is no correlation significantly different from 0 between the variables;
- H_a: At least one of the correlations between the variables is significantly different from 0.

As the computed p-value is lower than the significance level alpha=0.95, one should reject the null hypothesis H₀, and accept the alternative hypothesis H_a.

For all countries analyzed, the p-value of Bartlett's sphericity test is less than 5% and therefore we will reject the null hypothesis.

Czech Republic

First, for each variable we will calculate the main descriptive statistical selection indicators, such as: mean and standard deviation.

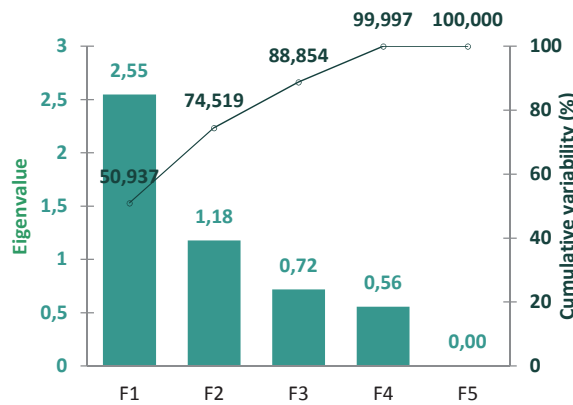
Table 2: Descriptive statistics of indicators

<i>Variable</i>	<i>Mean</i>	<i>Std. deviation</i>
TR	836.53	779.30
DPS	197.50	432.67
EPS	234.88	507.32
TD	49,055,612.33	61,191,792.76
DY	5.85	3.84

In the table 2 we have the standard deviation and the average. The standard deviation indicator records high values about in all cases, which could mean that that there is a high amount of information in the initial space. The higher the standard deviation, the more companies analyzed are very different from each other. Also, high levels of standard deviation can be explained by strong correlations between the original variables.

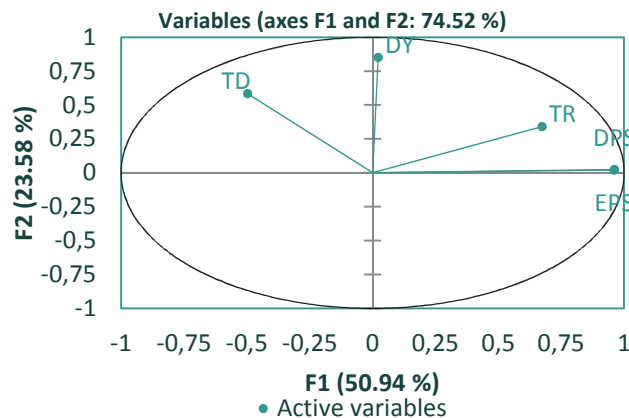
According to Kaiser's criterion, the number of principal components required to be taken into study is given by the number of eigenvalues strictly larger than 1.

Figure 1: Eigenvalue



According to the coverage percentage criterion (Timm) [9], the aggregate information of the main components should retain about 70% - 80% of the total information. Figure 1 highlights the share of information that each principal component takes as a proportion of its cumulative total. Thus, the first principal component takes up 50.94% of the total information. The second principal component explains 23.58% of the total dispersion, and cumulatively, the first two principal components take up 74.52% of the total dispersion.

Figure 2: Correlations between variables and factors



In Figure 2, the factor matrix elements are represented, which are the correlation coefficients between the original variables and the main components.

The first main component is strongly correlated in positive way with the TR, DPS and EPS indicators, and the second component is strongly correlated in positive way with TD and DY. The first principal component will provide us with information about their ability to generate profit and profitability. The second principal component will provide us information about their strategy of financing.

Table 3: Descriptive statistics of indicators

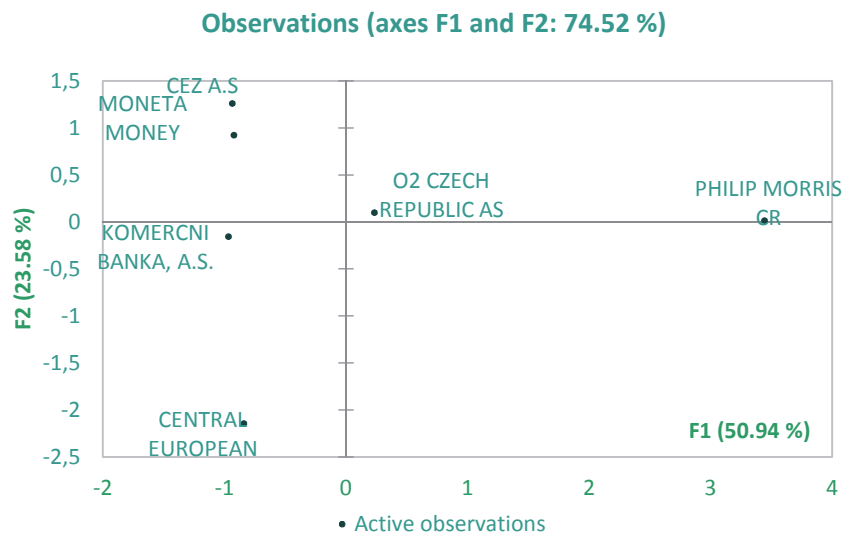
<i>Indicator</i>	<i>Communality</i>	<i>Specificity</i>
TR	0.57	0.43
DPS	0.93	0.07
EPS	0.92	0.08
TD	0.59	0.41
DY	0.72	0.28

A measure of how well the selected principal components explain the variance of each of the variables is given by a statistic called communality. In Table 3 we find the communality of the indicators representing the sum of the squares of the elements on the matrix factor lines, and the specificities represent the decrease between 1 and the communalities.

For DPS and EPS, communality is over 90%, this means that common factors are determined and the influence of specific factors is unrepresentative, representing less than 10% of the variance of indicators. Note that TR has the lowest communality, which indicates that TR is less well explained by the analysis than the other indicators.

In Figure 3 we note that PHILIP MORRIS CR and O2 CZECH REPUBLIC AS have the highest scores after the first main component. KOMERCNI BANKA and CENTRAL EUROPEAN have negative values of factors score.

Figure 3: Factor scores



Poland

In the case study for Poland we considered the following: total return (TR), dividends per share (DPS), earnings per share (EPS), total debt (TD), BETA, dividend yield (DY), market capitalization (MC) and earnings before interest & taxes (EBIT). We highlight the following descriptive statistics for indicators:

Table 4: Descriptive statistics of indicators

<i>Variable</i>	<i>Mean</i>	<i>Std. deviation</i>
TR	919.56	1,500.08
DPS	3.06	5.24
EBIT	2,790,362.73	3,011,643.35
EPS	6.11	7.12
MC	21,128,080.82	18,023,584.08
TD	9,844,252.73	10,379,277.62
BETA	0.92	0.37
DY	2.35	2.73

In Figure 4 we can see the correlation matrix's own values as well as the cumulative variance. Between all indicators only the first two are above 1, the rest being appropriate values of 0. From the above, we will have two principal components, being the only eigenvalues to have over-unit values: 3.431 and 2.324.

The two components explain 71.94% of the total variation of the initial space, which means that the reduction of the eight components to two records an informational loss of about 28.06%.

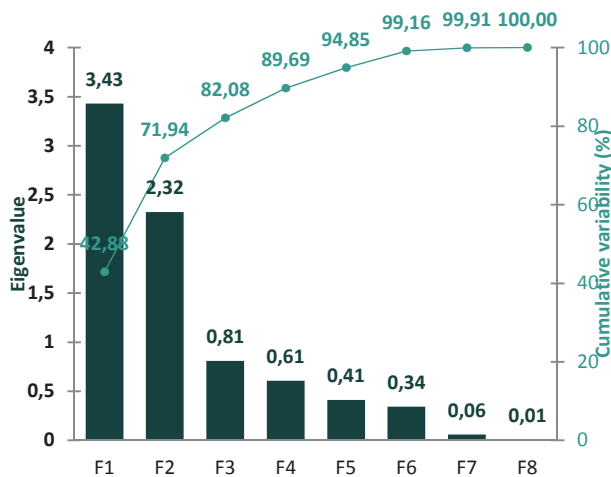


Figure 4: Eigenvalue

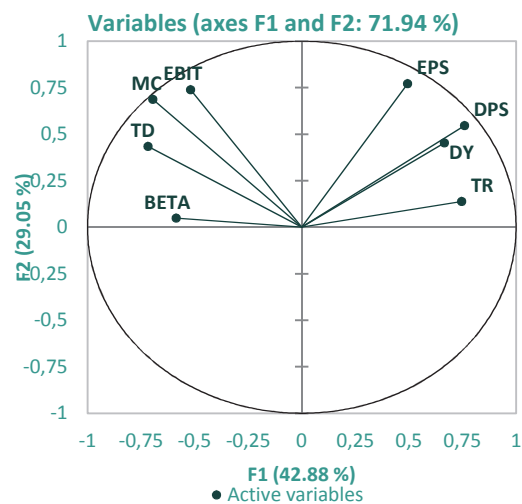


Figure 5: Correlations between variables and factors

The first principal component is strongly correlated in a positive way with the TR, DPS, DY and negatively with MC, TD and BETA indicators, and the second component is strongly correlated in a positive way with EBIT and EPS. Thus, the first principal component will provide us with information about the performance of the company in 2017, the size of the companies studied. The second principal component will provide us with information on operating profit and about the company's profitability.

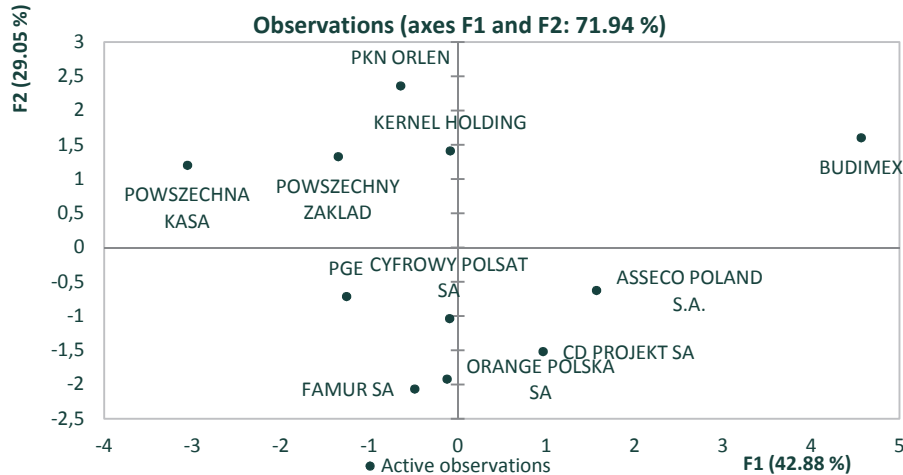
Table 5 Descriptive statistics of indicators

<i>Indicator</i>	<i>Communality</i>	<i>Specificity</i>
TR	0.58	0.42
DPS	0.87	0.13
EBIT	0.82	0.18
EPS	0.84	0.16
MC	0.95	0.05
TD	0.70	0.30
BETA	0.34	0.66
DY	0.65	0.35

For the DPS, EBIT, EPS, MC, TD and DY indicators, communality is over 58%, which means that the specific factors do not have a significant influence, which represents around 5%-40% of the indicator variance. The BETA indicator is more strongly influenced by specific factors (more than 60% of the variance indicators), as this indicator is more sensitive to the specifics of the company.

From the figure 6 we have three company classifications. A first class is made up of BUDIMEX - has the highest scores after the first main component, the second category being made up of FAMUR SA, ORANGE POLSKA SA, KERNEL HOLDING, CYFROWY POLSAT SA and PGE these recording negative factor scores.

Figure 6: Factor scores



Germany and Austria

In the case study for Germany & Austria we considered the following: total return (TR), dividends per share (DPS), earnings per share (EPS), total debt (TD), BETA, dividend yield (DY), total assets (TA), and earnings before interest & taxes (EBIT). We highlight the following descriptive statistics for indicators:

Table 6: Descriptive statistics of indicators

<i>Variable</i>	<i>Mean</i>	<i>Std. deviation</i>
TR	1,374.36	1,524.74
DPS	0.73	0.81
EBIT	2,099,109.70	1,767,560.69
EPS	1.49	2.94

TD	23,680,921.30	47,734,472.00
TA	89,652,202.30	141,071,118.05
BETA	0.94	0.51
DY	1.98	1.82

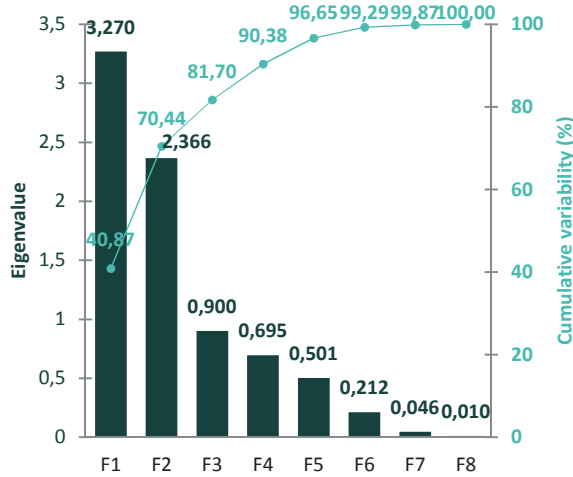


Figure 7: Eigenvalue

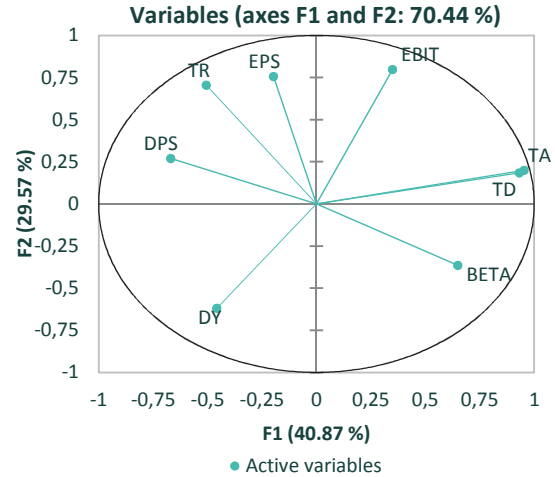


Figure 8: Correlations between variables and factors

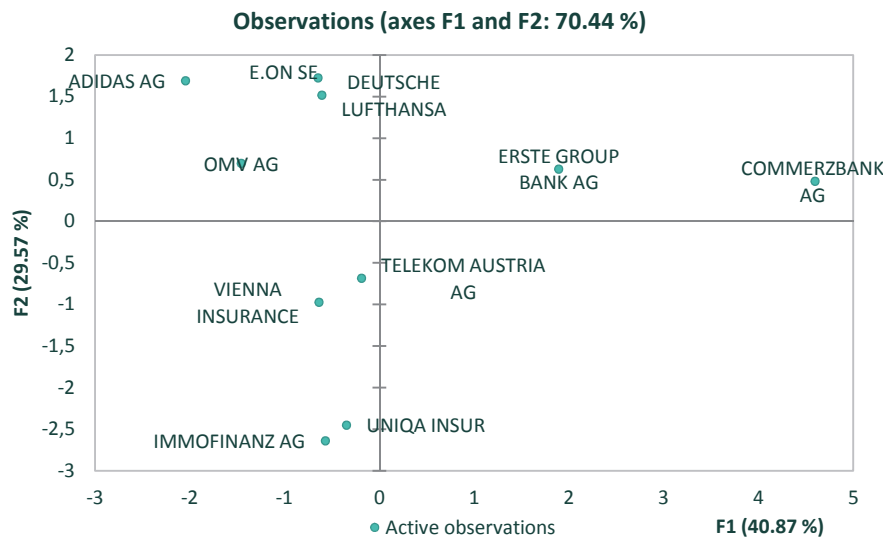
Only values above 1 are of interest because they have a richer information packet than the initial variables. The first two components account for 70.44% of the total variation. Of the eight values, only the first two components are over 1, the remainder having an appropriate value of 0. Out of the eight components, only the first two components are over 1, the rest are appropriate value of 0.

Table 7: Descriptive statistics of indicators

<i>Indicator</i>	<i>Communality</i>	<i>Specificity</i>
TR	0.75	0.25
DPS	0.52	0.48
EBIT	0.76	0.24
EPS	0.61	0.39
TD	0.90	0.10
TA	0.95	0.05
BETA	0.55	0.45
DY	0.59	0.41

For total assets and total debts, communality is over 90%, this means that common factors are determined and the influence of specific factors is unrepresentative, representing less than 10% of the indicator variance. These variables are influenced to a large extent by economic growth, capital market developments, etc., rather than company-specific factors. DPS and BETA are more strongly affected by specific factors - more than 40% of the variance of indicators, being more influenced by the company's specificity.

Figure 9: Factor scores



The first principal component is positively correlated with total debt, total assets, BETA, which means that companies with the highest scores on this component are the most performing in terms of these indicators. The second component is strongly correlated in a positive way with profitability, operating profit and negative earnings per share.

Taking into account the above, companies are distinguished according to each character. A first category includes COMMERZBANK AG and ERSTE GROUP BANK AG, another one includes IMMOFINANZ AG, UNIQA INSUR, VIENNA INSURANCE and TELEKOM AUSTRIA AG and another E.ON SE, COMMERZBANK AG, DEUTSCHE LUFTHANSA, ADIDAS AG, ERSTE GROUP BANK AG and OMV AG.

4. CONCLUSIONS

In this paper we showed that using a statistical size reduction method, we have succeeded in selecting actions more easily and efficiently. From the multitude of financial indicators that characterize the evolution of a company, we decided to use five indicators in the PCA. Based on the results of the PCA we made a classification of the selected actions. Through the PCA, we showed the usefulness of this statistical method at the data selection stage.

The theme addressed in this paper is of interest to both investors and the academic side. That is why we believe it is necessary to continue researching and discovering the best solutions and methods in efficient portfolios management. We plan in the near future to develop this analysis by addressing other selection methods such as discriminatory or factorial analysis in order to obtain a fair selection of assets.

REFERENCES

- [1] Fulga, C., Dedu, S., Şerban, F. (2013) *Portfolio Optimization with Prior Stock Selection*, Economic Computation and Economic Cybernetics Studies and Research.
- [2] Pasini, G.(2017) *Principal Component Analysis For Stock Portfolio Management*, International Journal of Pure and Applied Mathematics, Volume 115 No. 1 2017, pp.153-167.

- [3] Ngai, E.W.T., Cheng, T.C.E. (1997) *Identifying potential barriers to total quality management using principal component analysis and correspondence analysis*, International Journal of Quality & Reliability Management, Vol. 14 Issue: 4, pp.391-408.
- [4] Meriç, İ., Ding,J., Meriç, G. (2016) *Global Portfolio Diversification with Emerging Stock Markets*, Volume 6 No 1.
- [5] Lia, Y., Zhangb,Q.(2011) *The Application of Principal Component Analysis on Financial Analysis in Real Estate Listed Company*, Procedia Engineering, Volume 15, pp. 4499-4503.
- [6] Yap, B.(2013) *The Application of Principal Component Analysis in the Selection of Industry Specific Financial Ratios*, British Journal of Economics, Management & Trade. 3. pp.242-252.
- [7] Yang,L., Rea, W., Rea, A.(2015) *Identifying Highly Correlated Stocks Using the Last Few Principal Components*. arXiv:1512.03537v1
- [8] Armeanu, D., Istudor, N., Florinel, S. M., Burca,A.-M. (2014) *Analysis of the Romanian Insurance Market Based on Ensuring and Exercising Consumers' Right to Claim*, Amfiteatru Economic, 16(36), pp.550–562.
- [9] Timm, N. H.(2002) *Applied Multivariate Analysis*, Springer-Verlag, New York.