USING PCA IN FINANCIAL MARKETS FIELD[1]

Stefan-Alexandru IONESCU
PhD, Assistant Professor, Romanian-American University, Bucharest, Romania
Postdoctoral Fellow, The Romanian Academy - POSDRU/159/1.5/S/137926

E-mail: ionescusa@gmail.com

Abstract:
Principal Component Analysis is a multivariate technique which aims the reduction of the initially causal space, with a minimal loss of the variance. Using this technique I identified a very interesting aspect, subsequently confirmed too by specific methods of discriminant analysis. If before the financial crisis, companies differentiate themselves primarily by rates of profitability, now it's obvious the increasing discriminatory power of liquidity and efficiency rates. We can also see that most companies register values under the standard parameters considered, even for a period of crisis.

Key words: PCA, multivariate, financial market, quantitative

1. Introduction

In the economic-social field, any problem for analysis and prediction entails, as a necessary step, the detailed investigation of the functional ties between explicative variables (independent variables). These represent symbols that express various quantitative or qualitative aspects of phenomena that constitute *influencing factors* or *causes* for other phenomena and processes.

Hence, the sum of all explicative variables involved in a certain multidimensional analysis defines a certain numerical space, called ***initial causal space***. The number of causal variables involved in the analysis defines the dimension of the initial causal space. Every causal variable is represented on one axis of this space, and a certain point on a certain axis represents a possible value that the variable associated with the respective axis can take. The points of the causal space are represented by the objects under investigation and the projections of these objects on the axes of the space are the values registered by these objects in relation with the characteristics associated with the axes.

The most important *characteristic* for all the techniques of multivariate analysis of data revealed by the *causal space* is the *variability* of this space. The variability contained in the causal space determined by explicative variables may be expressed under various more or less efficient forms. From this vantage point, the principal component analysis can be regarded as a *technique to decompose the total variability of the initial causal space* into *a smaller number of components* and without any *informational redundancies*.

In some analyses, finding out the principal components represents a goal in itself, everything being finalized with the actual interpretation of data. However, in other cases, the analysis of the principal components is just an intermediary step, the principal components constituting the input for other analyses. Rancher (1998) shows that a special case is represented by regression analysis. If the number of initial variables is relatively great in comparison with the number of observations, or if the initial variables are strongly correlated among themselves, the testing can be inefficient and the regression coefficients can be unstable.

2. The Principal Components

The principal component analysis technique consists in the calculation of the projections of every point from the initial space, determined by the original variables under analysis, onto the axes of a new space, the dimension of which is significantly reduced. This method of multivariate analysis has as a purpose the determination of new variables called ***principal components***. These are in fact abstract vector variables, in the form of linear combinations of maximum variance of the original variables, and practically define a new space of objects. Ruxanda (2009) shows that the determination of linear combinations used in the construction of principal components takes into consideration the following aspects:

- The number of principal components is equal to the number of original variables.
- The sum of the variances of the principal components coincides with the sum of the variances of the original variables, so that the principal components take the variability contained in the original variables.
- *The first principal component* is a normalized linear combination, the *variance* of which *is maximum*, and the *second principal component* is a linear combination that is *uncorrelated* with the first principal component, having a variance that is as great as possible, but smaller than that of the first component etc. The principal components are scaled according to the magnitude of its variance, the first having the maximum variance and the last the minimum variance.
- The axes of the new space are orthogonal two by two and define the new variables, that is the *principal components* that are obviously uncorrelated two by two.
- The sum of the squares of the coefficients that define the linear combination that corresponds to a principal component equals one;

The coordinates of objects in the new space, i.e. the projections of the objects onto its axes, are evaluations of objects in relation with the new variables and are called scores of the principal components or ***principal scores***.

2.1. The mathematical model of principal components

We consider that the initial causal space under investigation is determined by an *n* number of explicative variables, noted $x_1, x_2, \dots, x_n$, which symbolize characteristics of the objects under analysis. What results is that every object is characterized by *n* variables.
The determination of the principal components is described through a transformation of the following type:

$$\Psi:\ \Re^n \rightarrow \Re^k \qquad (1)$$

where $\Re^n$ and $\Re^k$ are two real vector spaces, and the dimension of the second space is smaller than the dimension of the first space, respectively $k < n$.

Through the $\boldsymbol{\Psi}$ transformation, an object **x,** belonging to the *n*-dimensional $\Re^n$ space, is transformed into an object **w** belonging to the *k*-dimensional $\Re^k$ space. Through this transformation the coordinates of the object change and the number of coordinates is reduced.

If $\boldsymbol{x} \in \Re^n$ and $\boldsymbol{w} \in \Re^k$, then the transformation $\boldsymbol{\Psi}$ is a linear application of the following type:

$$\boldsymbol{w} = A^t \cdot \boldsymbol{x} \qquad (2)$$

where A is a matrix of real numbers, of a $n \times k$ dimension.

The solution to this problem consists in the determination of matrix A, so that an object **w** should constitute the best possible representation for object **x**.

The *n* principal components, corresponding to the analyzed causal space, present themselves in the form of an *n*-dimensional vector, noted **w**:

$$\boldsymbol{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \qquad (3)$$

Every w_i component of this vector represents a principal component defined in relation with the original variables, through the following linear combination:

$$w_i = \alpha_1^{(i)} \cdot x_1 + \alpha_2^{(i)} \cdot x_2 + \cdots + \alpha_n^{(i)} \cdot x_n \qquad i = 1,2,\ldots,n \qquad (4)$$

In order to determine the w_i principal component, the determination of the $\alpha_j^{(i)}$ coefficients which define the corresponding linear combination of this principal component is necessary. The $\alpha_j^{(i)}$ coefficients are precisely the coordinates of the eigenvectors corresponding to the covariance matrix of the original variables $x_1, x_2, \ldots, x_n$, and the variances of the principal components are precisely the eigenvalues of this matrix. According to the definition of the principal components, the determination of these coefficients is performed so that the w_i principal component should have the maximum variance.

If we consider that the *n* coefficients $\alpha_1^{(i)}, \alpha_2^{(i)}, \ldots, \alpha_n^{(i)}$ of the above-mentioned linear combination are the coordinates of the *n-dimensional* $\propto^{(i)}$ vector, namely:

$$\boldsymbol{\alpha}^{(i)} = \begin{pmatrix} \alpha_1^{(i)} \\ \alpha_2^{(i)} \\ \vdots \\ \alpha_n^{(i)} \end{pmatrix} \qquad i = 1,2,\ldots,n \qquad (5)$$

we can define the principal component w_i under the form:

$$w_i = \left(\boldsymbol{\alpha}^{(i)}\right)^t \cdot \boldsymbol{x} \qquad i = 1,2,\ldots,n \qquad (6)$$

where the coordinates of the $\boldsymbol{\alpha}^{(i)}$ vector are chosen so that the maximization of the variance of the principal component w_i be ensured.

Since the *i* principal component, w_i is a linear transformation of the elements of vector **x**, supposedly normally distributed, of a $\boldsymbol{\mu}$ average and covariance matrix Σ, what results is that this principal component is a random, normally distributed variable. Based on the relationship in (6), the average and the variance of this principal component can be inferred, as follows:

$$E(w_i) = E\left[\left(\boldsymbol{\alpha}^{(i)}\right)^t \cdot \boldsymbol{x}\right] = \left(\boldsymbol{\alpha}^{(i)}\right)^t \cdot \boldsymbol{\mu} \qquad (7)$$

$$Var(w_i) = \left(\alpha^{(i)}\right)^t \cdot \Sigma \cdot \alpha^{(i)} \tag{8}$$

The result is:

$$w_i \sim N\left[\left(\alpha^{(i)}\right)^t \cdot \mu, \left(\alpha^{(i)}\right)^t \cdot \Sigma \cdot \alpha^{(i)}\right] \qquad i = 1,2,\dots,n \tag{9}$$

The solution to this principal component analysis is equivalent to the solution of the following extreme value problem:

$$\begin{cases} \underset{A \in M_{n \times k}}{\text{opt}} \ \Phi(\boldsymbol{x}, \boldsymbol{w}) \\ SR: \quad \boldsymbol{w} = A^t \cdot \boldsymbol{x} \end{cases} \tag{10}$$

where the optimum criterion can be either maximum or minimum, according to the nature of the Φ function:

- If the Φ function is a distance function, then the optimum criterion will be represented by the minimization of the Φ function.
- If the Φ function is a measure of the quantity of information brought by the new means of object representation, the optimum criterion will be represented by the maximization of the Φ function.[2]

In order to define the mathematical model of the principal component analysis, we will consider that the $\alpha^{(i)}$ vectors represent the columns of an A matrix of $n \times n$ dimension of the form:

$$A = \begin{pmatrix} \alpha_1^{(1)} & \alpha_1^{(2)} & \dots & \alpha_1^{(n)} \\ \alpha_2^{(1)} & \alpha_2^{(2)} & \dots & \alpha_2^{(n)} \\ \vdots & \vdots & \dots & \vdots \\ \alpha_n^{(1)} & \alpha_n^{(2)} & & \alpha_n^{(n)} \end{pmatrix} \tag{11}$$

We assume that $\boldsymbol{x}$ is the vector whose coordinates are the original variables $x_1, x_2, \dots, x_n$ and that $\boldsymbol{w}$ is the vector whose coordinates are the principal components $w_1, w_2, \dots, w_n$. Taking into account relationship number (6) as well, the linear combinations that define the principal components can be written under the following matrix form:

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} \alpha_1^{(1)} & \alpha_2^{(1)} & \dots & \alpha_n^{(1)} \\ \alpha_2^{(1)} & \alpha_2^{(2)} & \dots & \alpha_n^{(2)} \\ \vdots & \vdots & \dots & \vdots \\ \alpha_1^{(n)} & \alpha_2^{(n)} & & \alpha_n^{(n)} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{12}$$

The mathematical model of the principal component analysis is defined as follows:

$$\begin{cases} \underset{A \in M_{n \times n}}{\max} \ Var(\boldsymbol{w}) \\ \boldsymbol{w} = A^t \cdot \boldsymbol{x} \end{cases} \tag{13}$$

The n columns of the A matrix represent in fact the normalized eigenvectors of the Σ covariance matrix, and the variance of every w_i principal component, which is a maximum variance in relation with the variances of the previous principal components, is represented precisely by the λ_i eigenvalue of the same covariance matrix. This method to determine the elements of the A matrix is equivalent to the calculation of the object projections of the $x \in \mathfrak{R}^n$ type onto the linear subspace generated by the vectors of the A matrix columns.

3. Data used in the analysis

For this analysis we have selected 101 Romanian firms of various sizes which have existed for more than one year on the market. All these firms have unfolded their activity and have submitted their balance sheets for the 31st of December to the Registry of Commerce.

An analysis of the size of these firms can be made according to the total assets. The smallest value is of 15,821 lei, and the biggest is over 30 million lei. More than half of these firms are medium-sized, the small and big ones being in similar proportions.

For all the enterprises we have included, data from both the yearly balance sheet and the balance sheets on December has been collected. The most important values extracted have taken into account:

- assets, as well as their classification;
- debts, divided as well into categories, including those to banks and leasing companies;
- capitals and equity capitals;
- data connected with the turnover, profit, taxes and duties.

Although the information contained in the financial statements of the companies offers important elements referring to company performance, it is advisable that the statements should be accompanied by an analysis of financial rates as well. The objective is precisely to analyze the financial situation focusing on four aspects: liquidity, solvability (risk), activity and profitability.

In this analysis we refer to the following eight rates:

- Profitability Ratios**:** Return on Assets ***(ROA),*** Return on Equity ***(ROE),*** Return on Capital Employed ***(ROCE).***
- Efficiency ratios**:** Total Assets Turnover ***(RAT).***
- Liquidity Ratios**:** Current ratio ***(CR),*** Quick ratio ***(QR),*** Cash Ratio ***(CashR).***
- Solvability Ratios**:** General Solvability ***(SP).***

Data processing has been done with the *STATISTICA 8.0* program package.

4. Results.

We have shown above how the dimensionality of initial causal space can be reduced while preserving the least informational loss. Obviously, informational redundancy from the initial observations will be eliminated. We have performed the principal component analysis on the covariance matrix.

Table 1. Eigenvalues of the covariance matrix

	Eigenvalue	% - Total variance	Cumulative Eigenvalue	Cumulative %
W1	1.681320	49.22606	1.681320	49.2261
W2	1.015914	29.74415	2.697234	78.9702
W3	0.258015	7.55421	2.955248	86.5244
W4	0.197061	5.76958	3.152309	92.2940
W5	0.129885	10281	3.282194	96.0968
W6	0.065222	1.90958	3.347416	98.0064
W7	0.059987	1.75631	3.407403	99.7627
W8	0.008105	0.23731	3.415508	100.0000

In the second column we have the eigenvalues of the covariance matrix of the new factors that have been extracted successively. In the third column is presented the variance values in percentages. As we can see, the contribution of the first factor is approximately half of the total variance, whereas the second contributes with approximately 30% to the total variance. The sum of the 8 eigenvalues, i.e. the variance of the 8 principal components, equals the sum of the variances of the original variables. The third column contains the cumulative variances, the last line showing the actual total variance. The last column reveals the cumulative variances in percentages.

As we can see, the first two principal components account for approximately 79% of the variance of the 8 initial variables, and, if we were to take into account the third component as well, 86.5% of the initial variance would be represented.

Although the number of factors that can be taken into account is generally an arbitrary decision, there are nevertheless certain criteria that practitioners use. The first was proposed by Kaiser (1960)[3] and referred to keeping those factors the eigenvalues of which are bigger than 1.

The second criterion is called the "scree test" and was proposed by Cattell (1966)[4]. The eigenvalues from the second column of table 1 are represented and the points are united. In figure 1 we can see that, starting with the third variable, the chart slope is close to zero.

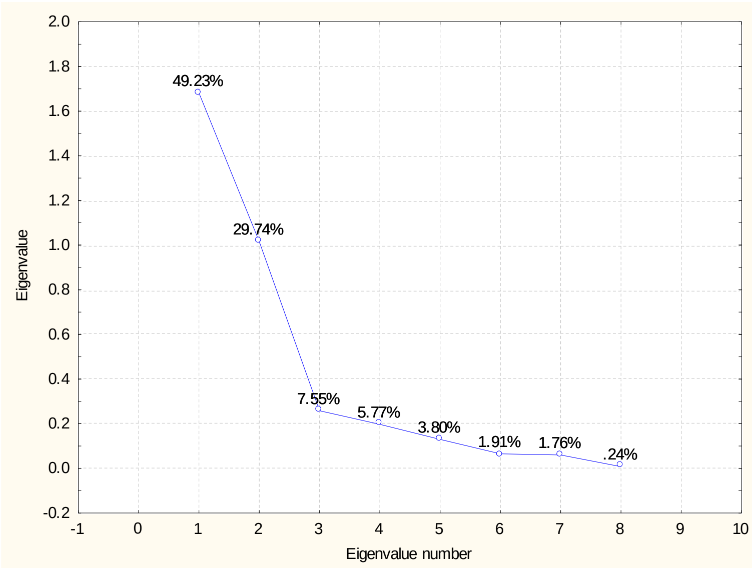

Figure 1. Eigenvalues

Applying both criteria, keeping the first two factors seems the most plausible. Also, keeping more than two factors can make the drawing of future charts more difficult. Hence, the dimension of the new representation space for objects will be 2.

In table 2 we have the eigenvectors of the covariance matrix:

Table 2. Eigenvectors of the covariance matrix

	W 1	W 2	W 3	W 4	W 5	W 6	W 7	W 8
ROA	-0.040117	-0.037710	-0.026787	-0.073412	-0.219236	-0.082868	-0.046354	**0.966324**
ROE	-0.054749	-0.072769	**-0.961826**	-0.014836	0.115778	-0.141111	-0.179759	-0.027359
ROCE	-0.067856	-0.077984	-0.204885	-0.158220	**-0.781037**	0.305138	0.442325	-0.153373
RAT	-0.223403	**-0.960808**	0.096669	-0.039278	0.115039	-0.026270	0.040940	-0.021263
CR	**-0.660927**	0.148632	-0.022416	0.636065	0.051507	-0.131084	0.338144	0.042728
QR	**-0.599534**	0.105551	0.071886	-0.215882	-0.109665	0.421674	**-0.620428**	-0.053660
CashR	-0.369520	0.175291	0.016442	**-0.718330**	0.280798	-0.269826	0.405954	-0.002575
SP	-0.089828	-0.009507	0.130041	-0.025133	-0.470905	**-0.782064**	-0.322708	-0.191789

Every principal component can be defined in relation with the original variables with the help of linear combinations. The coefficients are practically the coordinates of the eigenvectors corresponding to the covariance matrix of the original variables, and the variances of the principal components are precisely the eigenvalues of this matrix and have been presented in table 1. We may notice the relatively small importance of the ROA, SP and ROCE indicators which are represented mainly through the last 4 components. As we can see from table 1, these components preserve less than 8 percent from the total variance. The principal components of the causal space will be determined by the following equations:[5]

$$\begin{cases} W_1 = -.04 \cdot ROA - .05 \cdot ROE - .07 \cdot ROCE - .22 \cdot RAT - .66 \cdot CR - .60 \cdot QR - .37 \cdot CashR - .09 \cdot SP \\ W_2 = -.04 \cdot ROA - .07 \cdot ROE - .08 \cdot ROCE - .96 \cdot RAT + .15 \cdot CR + .11 \cdot QR + .18 \cdot CashR - .01 \cdot SP \\ W_3 = -.04 \cdot ROA - .96 \cdot ROE - .20 \cdot ROCE + .10 \cdot RAT - .02 \cdot CR + .07 \cdot QR + .02 \cdot CashR + .13 \cdot SP \\ W_4 = -.07 \cdot ROA - .01 \cdot ROE - .16 \cdot ROCE - .04 \cdot RAT + .64 \cdot CR - .22 \cdot QR - .72 \cdot CashR - .03 \cdot SP \\ W_5 = -.22 \cdot ROA + .12 \cdot ROE - .78 \cdot ROCE + .12 \cdot RAT + .05 \cdot CR - .11 \cdot QR + .28 \cdot CashR - .47 \cdot SP \\ W_6 = -.08 \cdot ROA - .14 \cdot ROE + .31 \cdot ROCE - .03 \cdot RAT - .13 \cdot CR + .42 \cdot QR - .27 \cdot CashR - .78 \cdot SP \\ W_7 = -.05 \cdot ROA - .18 \cdot ROE + .44 \cdot ROCE + .04 \cdot RAT + .34 \cdot CR - .62 \cdot QR + .41 \cdot CashR - .32 \cdot SP \\ W_8 = +.97 \cdot ROA - .03 \cdot ROE - .15 \cdot ROCE - .02 \cdot RAT + .04 \cdot CR - .05 \cdot QR - .003 \cdot CashR - .19 \cdot SP \end{cases} \quad (14)$$

As we have shown previously, the first two principal components ensure the preservation of 78.9702% of the variance contained in the initial causal space, determined by the 8 variables. Within the range of a loss of 21.0298%, we can express the 8 initial variables through the intermediary of the first two principal components. Table 3 represents the first two columns of the factor matrix, with the help of which the correlations between the variables and every factor can be interpreted.

Table 3. Matrix of the intensity of factors for the first two principal components

	W 1	W 2
ROA	-0.052018	-0.038009
ROE	-0.070991	-0.073346
ROCE	-0.087986	-0.078602
RAT	-0.289677	**-0.968422**
CR	**-0.856995**	0.149810
QR	**-0.777390**	0.106387
CashR	**-0.479141**	0.176681
SP	-0.116476	-0.009583

Every element may be interpreted as a measure in which the original variable participates in the formation of principal components. The first component describes very well the liquidity variables CR, QR and CashR. To a smaller extent, the variable of efficiency in the use of RAT assets and the variable of patrimonial solvability SP contribute to the significance of this principal component. The second principal component is strongly and in a negative sense correlated with the RAT efficiency variable, displaying weak correlations with the liquidity indicators. All the other correlations, among the first two principal components and the rest of the variables have values of less than 10 percent. The profitability indicators display correlations of very similar intensities with the first two principal components. RAT is the only indicator from the composition of the *DuPont model* that displays important correlations with the two principal components.

Generally speaking, we can say that the first factor, the one that covers 49.23% of the variance, reflects ***liquidity and solvability***, whereas the second factor, which covers 29.74% of the variance, synthesizes, from an informational point of view, company ***efficiency***. Thus, we have an indicator of general solvability, ***W1***, and an indicator of general efficiency, **W2**.

Each line in table 3 defines the coordinates of an original variable, seen as a point in the bi-dimensional reduced space. Figure 2 presents in an extremely suggestive manner the interpretation of these factors. What needs to be mentioned is that the 8 variables appear only in quadrants II and III, which shows correlations that are rather negative.

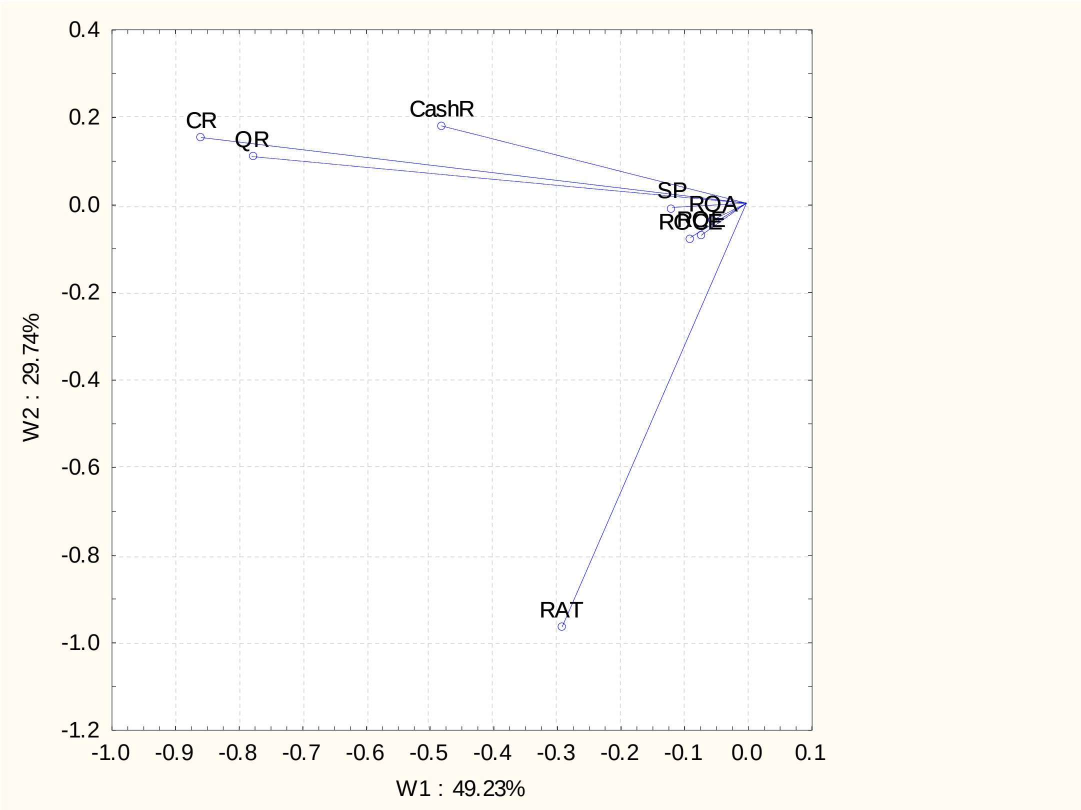

Figure 2. Projection of variables in a Cartesian system of W1 and W2

The variables that are strongly correlated with the first factor, namely liquidity rates, are farther from the *Oy* axis and appear represented in quadrant II. The RAT variable, being correlated in the first place with the second factor, appears even farther apart from the *Ox* axis and is represented in quadrant III.

Table 4 contains the coordinates of each and every one of the 101 firms under analysis, for the bi-dimensional finite space defined by the eigenvectors of the two principal components that we have kept.

Object coordinates in the reduced space are also called the principal scores of objects. Principal scores can be used in the analysis as a substitute of original observations, thus simplifying the initial information base. Ruxanda (2009) considers that principal scores are better suited to be used in analyses, as they are less likely to be affected by errors in comparison with original measurements. The fact that principal scores are more robust in relation with perturbations introduced by errors, that they have a certain invariance in relation with errors, makes them more salient from an informational point of view than original observations.

Due to the fact that the new reduced space has only two axes, the 10 objects can be represented graphically in this space. The graphical representation from figure 3 shows the positioning of the 101 firms in relation with the axes of the new space.

Table 4. Object coordinates in the bi-dimensional principal space

Obj.	W 1	W 2	Obj.	W 1	W 2	Obj.	W 1	W 2
1	0.92246	0.17083	**34**	0.11369	-0.75858	**68**	-2.25986	-1.83606
2	-2.30047	1.62656	**35**	-0.25561	-0.54107	**69**	1.11784	1.02697
3	0.54505	0.58010	**36**	1.44912	0.39254	**70**	1.16220	0.98494
4	0.94290	0.34056	**37**	-1.01628	-2.55144	**71**	1.31252	0.29129
5	-0.41592	-1.50400	**38**	-0.31971	-0.68276	**72**	0.09453	-1.71136
6	-0.92440	1.15265	**39**	1.06464	0.40374	**73**	0.75357	0.79035
7	-2.27346	1.75198	**40**	0.29465	-1.24653	**74**	-1.83639	1.59073
8	0.30099	-0.31820	**41**	0.60322	0.07182	**75**	1.53481	0.63960
9	-0.16085	1.18733	**42**	0.75149	-0.31289	**76**	0.30684	-0.18174
10	-1.15324	-2.15315	**43**	-0.10130	0.80636	**77**	0.07209	0.71098
11	1.18478	0.05475	**44**	1.67841	0.76009	**78**	0.59218	-0.03836
12	0.26761	-0.02226	**45**	1.55056	-1.51286	**79**	0.00806	0.47547
13	-1.35034	-4.05489	**46**	1.64630	0.72357	**80**	-2.03139	1.54840
14	-0.26368	0.69912	**47**	1.73311	0.45565	**81**	0.21322	-0.08724
15	-1.48029	0.56571	**48**	1.53607	0.82753	**82**	1.01396	-0.54419
16	0.20445	0.54344	**49**	-0.06948	-1.00137	**83**	-0.04110	-0.80137
17	0.11657	-1.01666	**50**	-5.06088	0.88351	**84**	-4.99077	1.16111
18	-1.77799	-0.43436	**51**	1.44254	0.79235	**85**	0.58313	-0.62409
19	-1.38050	0.34749	**52**	-0.85130	1.04399	**86**	0.51682	-0.12252
20	-2.35150	0.23165	**53**	-1.52406	0.62633	**87**	0.30317	-0.42902
21	-0.26505	0.34568	**54**	0.57404	0.37003	**88**	-0.41400	-0.29846
22	0.13788	0.91824	**55**	0.07909	-1.19050	**89**	-2.03296	-0.87358
23	-0.33733	-2.48703	**56**	0.27728	-0.36044	**90**	-0.57988	-0.62456
24	0.94363	0.54725	**57**	0.30565	-1.21259	**91**	0.19400	-1.16720
25	0.94021	-0.70235	**58**	0.23057	-0.43096	**92**	0.55482	-0.53463
26	-1.50101	1.24276	**59**	-0.34549	0.46831	**93**	1.69031	0.51705
27	1.68698	0.67703	**60**	0.37427	-0.24178	**94**	0.48793	-0.44790
28	1.78411	0.59879	**61**	1.42302	0.55391	**95**	-0.95689	-1.43593
29	0.98932	0.21634	**62**	-0.40780	-0.92208	**96**	0.26765	-1.03124
30	0.15451	0.94375	**63**	1.20108	-0.73368	**97**	-2.47667	-0.31404
31	1.54794	0.83926	**64**	0.98130	0.01579	**98**	-0.52374	1.20294
32	1.09693	0.56957	**65**	-0.59865	-0.72998	**99**	-0.10314	0.51226
33	0.57210	0.66365	**66**	1.23019	0.14112	**100**	0.37494	0.58453
			67	-1.51410	1.63574	**101**	0.21414	0.40241

The positioning of objects in relation with the two axes renders a first image regarding the similarities or dissimilarities among the firms. The best distinction seems to be offered by the first factor. We notice a more numerous group of objects gathered on the right side of the *Oy* axis, on the one hand, and a more rarefied, less numerous group on the left side of the *Oy* axis. Analyzing the composition of the first factor, we can say that the firms in quadrants II and III are characterized by liquidity indicators bearing reassuring values. Practically speaking, the farther on the left side of the *Oy* axis is an object placed, the better indicators for liquidity it will have. The second factor separates objects from under the *Ox* axis, characterized primarily by a more intense asset turnover than those above the *Ox* axis which do not use their resources in a satisfactory manner. We notice in the first place the big group in quadrant I. These are firms that register mediocre values for all the indicators and, automatically, for the two principal components as well.

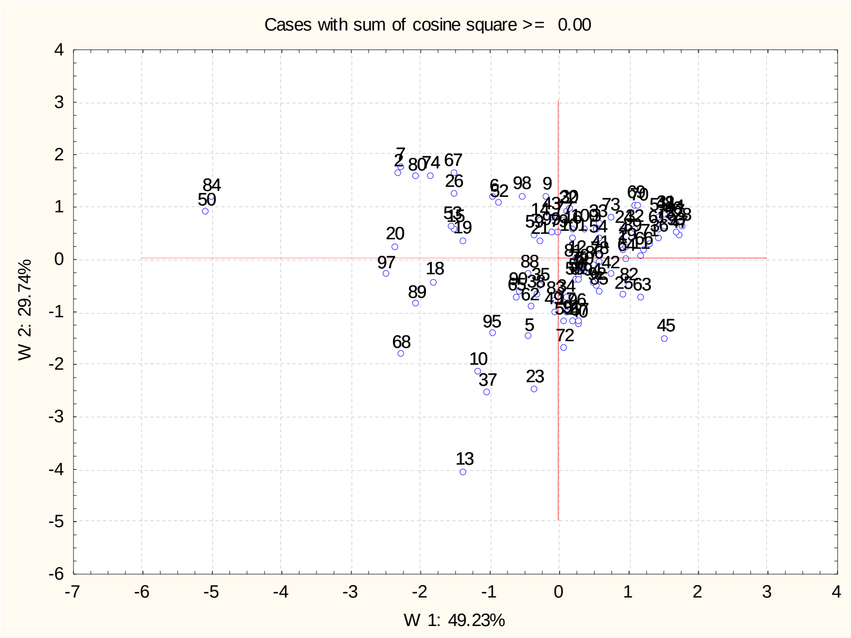

Figure 3. Object representation in the bi-dimensional principal space

5. Conclusions

To conclude, we may assert that principal component analysis is useful in solving two categories of problems: the ***simplification of the structure of causal dependency*** and the ***reduction of space dimensionality***.

In the context of principal component analysis, the values of characteristics are coordinates of the points that define the elements of the analyzed population. The most advantageous means of representation from an informational point of view is obtained by considering a new *space of representation*, which defines through its axes new characteristics of objects. These new characteristics are the ***principal components***, and the values registered by objects to these new characteristics are called ***scores***. Representation in a reduced space is known as *dimensionality reduction* and the principal component analysis is known as the *dimensionality reduction technique*.

The representation of objects in the new principal space is simplified and non-redundant. Every firm appears as a bi-dimensional vector, the components of which are represented by the corresponding values of the two principal components. These values are called principal scores. Having in mind the values from table 2, the two principal components can be expressed in the following manner:

$$\begin{cases} W_1 = -.04 \cdot ROA - .05 \cdot ROE - .07 \cdot ROCE - .22 \cdot RAT - .66 \cdot CR - .60 \cdot QR - .37 \cdot CashR - .09 \cdot SP \\ W_2 = -.04 \cdot ROA - .07 \cdot ROE - .08 \cdot ROCE - .96 \cdot RAT + .15 \cdot CR + .11 \cdot QR + .18 \cdot CashR - .01 \cdot SP \end{cases} \quad (15)$$

In broad lines, we may say that the first component, which covers 49.23% of the variance, reflects liquidity and solvability, whereas the second component, which covers 29.74% of the variance, synthesizes firm efficiency in an informational manner.

The advantage of using principal components is given by the simplification of calculations, and the downside consists in a more difficult interpretation of results, if the analysis is corroborated with other techniques of multidimensional analysis.

Bibliography

1. Abdi, H. and Williams, L.J. **Principal Component Analysis,** John Wiley & Sons, Inc. WIREs Comp Stat, vol. 2, 2010, pp. 433–459
2. Back, A.D. and Weigend, A.S. **Discovering Structure in Finance Using Independent Component Analysis,** Advances in Computational Management Science, Vol. 2, 1998, pp 309-322
3. Cattell, R. **The Scree Test for the Number of Factors. Multivariate Behavioral Research,** vol. 1, 1966, pp. 245-276
4. Chen, K.H. and Shimerda, T.A. **An Empirical Analysis of Useful Financial Ratios,** Financial Management, Vol. 10, No. 1, Spring 1981, pp. 51-60
5. Dieckmann, S. and Plank, T. **Default Risk of Advanced Economies: An Empirical Analysis of Credit Default Swaps during the Financial Crisis,** Review of Finance, vol. 16, no. 4, 2012, pp. 903-934
6. Geambasu, C., Jianu, I., Herteliu, C. and Geambasu L. **Macroeconomic Influence on Shares' Return. Study case: Arbitrage Pricing Theory (APT) Applied on Bucharest Stock Exchange,** Economic Computation and Economic Cybernetics Studies and Research, Vol 48, no. 2, 2014, pp 133-150
7. Kaiser, H. **The Application of Electronic Computers to Factor Analysis,** Educational and Psychological Measurement, 1960, pp. 141-151
8. Leon, C. and Murcia, A. **Systemic Importance Index for Financial Institutions: A Principal Component Analysis Approach,** ODEON, no. 7, 2013.
9. Rencher, A. **Methods of Multivariate Analysis,** New York, USA: John Wiley & Sons, 2002
10. Ruxanda, G. **Analiza Datelor,** Bucharest, ASE, 2001
11. Ruxanda, G. **Construirea, estimarea și implantarea software a metodelor matematice,** Cercetarea stiintifica in ASE, 2010

---

[1] **Acknowledgements**
This work was financially supported through the project "Routes of academic excellence in doctoral and post-doctoral research - READ" co-financed through the European Social Fund, by Sectoral Operational Programme Human Resources Development 2007-2013, contract no POSDRU/159/1.5/S/137926.

[2] This situation is specific for the standard variance used to solve the principal component problem, in which the objective is to maximize the variance of the principal components, as a measure for the quantity of information expressed by each of these. Thus, in what follows we will particularize matters for this case.

[3] It is the criterion that bears his name: *the Kaiser criterion*.

[4] "Scree" means rock debris, it is a term used in geology and refers to the bits of rock resulted from mountain fragmentation that gather at the base of mountain slopes. What is important is to find the point where the slope becomes visibly less abrupt, namely the base of the slope.

[5] It is about the equations described in (12).