At sum calculation the following combinatorial identity was used. $\sum_{k=0}^{n} C_{r+k}^{k} = C_{r+n+1}^{n}$. The right hand side of (2) is the total official's deposit to grand coalition (since all officials have the same rights, then each one get the amount $\frac{n-k}{k+1}$). So, the sum, which is remaining for businessmen, equals to $(n-k) - \frac{k}{k+1}(n-k) = \frac{n-k}{k+1}$, so, each businessman get $\frac{1}{k+1}$.

Note, that each official's income equals to the total one of all of the businessmen.

It seems at first sight, that at fixed number of officials the businessmen income must increase as their number increase, cause bribes may be collected by shares among businessmen. But indeed it's not the case, and official's claim to bribe increase as number of businessmen increase at the rate, than each businessman income $\frac{1}{k+1}$ remains constant.

Conclusion. It seems that fortunately in real business bribes fraction is less, than at considered above "ideal" model. It stipulated by the fact, that corruption is still illegal and criminal punishment fear works as restriction factor. If the corruption were legitimate (or at least actually unpunishable), then it cause the situation closer to the model considered above.

1. Branzei, Dimitrov, Tijs. Models in cooperative game theory. Springer, 2005.

Надійшла до редколегії 05.05.12

JEL classification C38

I. Fedorenko, PhD, Associate Professor,
V. Rybalko,
Taras Shevchenko National University of Kyiv

APPLYING PRINCIPLE COMPONENTS ANALYSIS FOR MODELING INVESTMENT ACCEPTANCE OF COMPANIES

Дана стаття присвячена підходу до оцінки доцільності інвестицій за допомогою факторного аналізу. Інвестори стикаються з проблемою, як систематизувати дані, обрати основні чинники та їх конфігурації, що впливають на ціну акцій компаній. В цьому випадку метод аналізу головних компонент як один з методів факторного аналізу дозволяє вирішувати таке завдання. За результатами аналізу головних компонент були складені групи український компаній з найвищою пріоритетністю з точки зору вартості інвестицій.

Ключові слова: факторний аналіз, аналіз головних компонент, доцільність інвестицій компаній, дисперсія, факторна вага, кореляція.

Данная статья посвящена подходу к оценке целесообразности инвестиций с помощью факторного анализа. Инвесторы сталкиваются с проблемой, как систематизировать данные, выбрать основные факторы и их конфигурации, влияющие на цену акций компаний. В этом случае метод анализа главных компонент как один из методов факторного анализа позволяет решать такую задачу. По результатам анализа главных компонент были составлены группы украинских компаний с наивысшей приоритетностью с точки зрения стоимости инвестиций.

Ключевые слова: факторный анализ, анализ главных компонент, целесообразность инвестиций компаний, дисперсия, факторній вес, корреляция.

This article deals with an approach to estimation of investment acceptance by factor analysis. Investors face the problem how to systematize data, select basic factors and their configurations that influence shares price of companies. In this case Principle Components Analysis (PCA) method as one of factor analysis methods helps to solve such a task. The highest-priority groups of Ukrainian companies in terms of investment value were made according to the results of Principle Components Analysis.

Keywords: factor analysis, Principle Components Analysis, investment acceptance of companies, variance, factor's weight, correlation.

Lately investors more frequently face the task of system consideration results of Fundamental analysis, Technical analysis and Liquidity analysis in the process of making investment decision. According to every type of analysis investment object is described by the plural performances: financial ratios, performances that describe prices fluctuations on the stock market exchange, macroeconomic indicators, expert's estimations and ext. Moreover, a lot of those performances are interdependent. The professionalism of investors is expressed exactly in ability to correctly select priority factors and identify dominant configurations. In such conditions, applying factor analysis enables to determine the structure of these interconnections and provide the clench of information, explaining the plurality of indicators through a small, as a rule, number of factors. It is assumed that these factors not only provide the concentration of information but also are the most significant characteristics of investigated object. Principle Components Analysis is the most appropriate for solving such tasks.

Principal component analysis is central to the study of multivariate data. Although one of the earliest multivariate techniques it continues to be the subject of much research, ranging from new model- based approaches to algorithmic ideas from neural networks. It is extremely versatile with applications in many disciplines.

Practical application of factor analysis and directly Principle Components Analysis were researched by many prominent scientists (such as Tomashevich, 1999; Pearson, 1901; Silvester, 1889; Ayvazyan, 1989). This method was invented by Pearson (1901) and used as one of methods on diminishing data losing the least of information.

Silvester (1889) was the first who created mathematical foundation for PCA in his paper "On the reduction of a bilinear quantic of the *n-th* order to the form of a sum of *n* products by a double orthogonal substitution". Than in twelve years later Pearson (1901) proposed PCA. In many cases the "independent" variables is subject to just as much deviation or error as the "dependent" variable. Pearson (1901) observed x and y and sought the unique functional relation between them. In case he was about to deal with he supposed that the observed variables – all subject to error – to be plotted in plane, three-dimensioned or high space, and he endeavored to take a line (or plane) which will be the "best fit" to such a system of points. The method that was investigated by K. Pearson can be easily applied to numerical problems.

Nowadays the most powerful generalization of PCA was made by Mikhail Belkin (2003) and Partha Niyogi (2003) investigating laplacian eigenmaps for dimensionality reduction and data representation. One of the central problems in machine learning and pattern recognition is to develop appropriate representations for complex data. They considered the problem of constructing a representation for data lying on a low-dimensional manifold embedded in a high-dimensional space. Drawing on the correspondence between the graph Laplacian, the Laplace Beltrami operator on the manifold, and the connections to the heat equation, they proposed a geometrically motivated algorithm for representing the high-dimensional data. The algorithm provides a computationally efficient approach to nonlinear dimensionality reduction that has locality-preserving properties and a natural connection to clustering.

The purpose of this article is investigation of PCA application for the exposure of main factors that influence the investment acceptance of companies.

1. Theoretical framework

Studying the investment acceptance of companies, analysts understand that some of factors have correlation. In this case their consideration at the same time will result in some output of information but on leaving out one of attributes, the loss of important information can take place that finally will also negatively affect quality of the accepted investment decision. Application of PCA allows defining a small number of independent factors with the minimum loss of influence of others.

This method is based on assumption, that description of all attributes equals zero and the number of general factors equals the number of output attributes. In this case the transformation to a linear combination of factors means transformation to the new system of co-ordinates.

Analysis consists of a few stages, at each of them one principal component is found that the most significantly describes the phenomena and every next component holds less information than previous. The first principal component accounts for as much of the variability in the data as possible, and every succeeding component accounts for as much of the remaining variability as possible. It is limited to a few first components in practice because they are enough for complete description in the compressed type of all initial information. The percentage of the explained variance serves as criteria – the relation of summary found components to general variance of initial attributes.

Practically, if number of already found principal components is not more than half of attributes and explanatory by them variance is no less than 70% and next component adds to total dispersion not more than 5%, the factor model is considered acceptable. The value of factor at one company is named factor weight of this company. Factor weights allow ranking companies by every factor.

In practice calculation is limited to a few factor's weights by the most significant components. If factor weight is higher at one object of investment it means this factor shows up in it more strongly.

2. Estimation of shares acceptance

Our investigation is based on implementation of PCA for liquidity indicators indicators of profitability and riskiness of shares issued by Ukrainian companies listed in Table 1.

The estimation of shares acceptance is carried out on the basis of analysis of daily exchange statistics at periods (periods can be, for example, months, quarters, years). Incoming data for analysis was taken according to results of stock exchange closing range: price of buyer (Bid) and price of offer (*Ask*) of every share, number of request for demand, number of offers for supply and prices of deals. We will use the followings statistical descriptions at the analysis of samples:

- means of shares profitability;
- standard deviation;
- skewness;
- kurtosis;
- spred;
- means of Volume trade;
- means of transaction number.
- means of shares profitability:

– Means of shares profitability:

$$\overline{R(a_i)} = \frac{1}{m}\sum_{j=1}^{m} R_j(a_i), \qquad (1)$$

where $R_j(a_i)$ – profitability i-share for the period j – month; m – number of months.

$$R(a_i, j) = \frac{(Bid(a_i, j+1) + Ask(a_i, j+1)) - (Bid(a_i, j) + Ask(a_i, j))}{Bid(a_i, j) + Ask(a_i, j)} \qquad (2)$$

Standard deviation:

$$\sigma(R(a_i)) = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(R_j(a_i) - \overline{R_j(a_i)}\right)^2} \qquad (3)$$

Useful measure of risk must take into account probability of possible bad results and their size. The risk measure must estimate the degree of possible deviation of actual result from expected instead of measusuring probabilities of different results. Standard deviation is allows to do this. It is lately offered number of approaches based on analysis of higher moments distribution of shares profitabliness. The special attention is devoted to the analysis of asymmetry which is characterized by the third moment and kurtosis which is characterized by the fourth moment.

– Skewness:

An analysis only of standard deviation as risk measures can be insufficient. Especially,

when these values are equal for a few alternative objects (projects). In this case, it is necessary to analyze coefficient of asymmetry as a risk parameter. It is calculated on a formula:

$$Sk = \frac{\mu^3}{\sigma^3}, \qquad (4)$$

where $\mu^3 = \frac{1}{m}\sum_{j=1}^{m}\left(R_j(a_i) - \overline{R(a_i)}\right)^3 \qquad (5)$

If $Sk = 0$, the curve of random variable is located symmetric in relation to mean of distribution. If $Sk > 0$, the random variable has a spatial slant – a "tail" of distribution comes forward on the right. When $Sk < 0$, random variable has a left-side slant – a "tail" of distribution comes forward to the left.

Clearly, that among *n* different alternative objects (projects, strategies) is sense to choose that for which takes place:

$$Sk_k = \max_{k=1\ldots n}(Sk_k)$$

Since unfavorable deviations from expected value located on the left the nearest to the mean of distribution (deviate less from it in an unfavorable side), and the proper (favourable) values are considerably remote from the mean of distribution (these values are located on the right).

The system of risk assessment can be built farther, using such a characteristic as kurtosis.

– Kurtosis:

$$Kurt = \frac{\mu^4}{\sigma^4} \quad (5)$$

The kurtosis gives understanding of the oblongness of random variable curve. If $Kurt > 3$, there is oblongness of curve upward. If $Kurt < 3$, the branches of parabola are pressed to the axis Ox. The kurtosis characterizes the so-called "thick tails" of distribution. What a distributing tail is thicker, the greater probability of taking on extreme values which substantially deviate from mean. Investors negatively refer to possibility of extreme values and want to minimize the kurtosis.

We will add an index, that represents probability of that value of profitableness is higher than mean.

– Frequency of shares profitability:

$$P(a_i) = 1 - P(R(a_i) < 0) \quad (6)$$

The most substantial index of share's liquidity is spred between prices $Ask(a_i)$ and $Bid(a_i)$. This index is calculated to every share according to results of market session:

– Spred:

$$\overline{Spred(a_i)} = \frac{1}{m}\sum_{i=1}^{m} Spred(a_i), \quad (9)$$

$$\text{where } Spred(a_i) = \frac{Bid(a_i) - Ask(a_i)}{\frac{1}{2}(Bid(a_i) + Ask(a_i))} \quad (10)$$

The spred characterizes liquidity not fully from the practical point of view. The sperd can be not so big but at the same time the number of deals that taking place is little or a volume of deals is insignificant. Therefore it is necessary to consider other indexes of liquidity. In particular it is important to consider the indexes of deal's volume for certain period. Such index can be presented as a volume of the real deals that were made for certain period and can be presented as an average between the volume of demand on share at price $Bid(a_i)$ and volume of offer at price $Ask(a_i)$.

– Mean of Volume trade:

$$\overline{V(a_i)} = \frac{1}{m}\sum_{i=1}^{m} V(a_i), \quad (11)$$

where $V(a_i)$ – is the summary volume in UAH of the transactions during the month.

An index does not take into account diversification of deals in time.During a month one large deal can take place that is the indicator of small level of liquidity. So it is expedient to enter the mean number of deal's amount $\overline{W(a_i)}$ as the next index of liquidity. The more deals took place the share is considered to be more liquid.

– Mean of transaction number:

$$\overline{W(a_i)} = \frac{1}{m}\sum_{i=1}^{m} W(a_i), \quad (12)$$

where $W(a_i)$ – the number of transactions during the month.

Moreover an amount of quotations is important indicator. In relation to the amount of quotations it is necessary to distinguish bilateral quotations (quotations on demand and on sale). The most considerable are bilateral quotations then quotation on a purchase and finally the least considerable are quotations on a sale. This fact is represented in "Rules of making ratings of securities in the PFTS" where they are differentiated with scales 10, 7 and 5 accordingly. In our model taking into account the structure of database it is possible to use such index:

– Quantity of shares quotations:

$$\overline{Q(a_i)} = \frac{1}{m}\sum_{i=1}^{m} Q(a_i), \quad (7)$$

$$\text{where } Q\ (a_i) = 10 \cdot Q_D(a_i) + 7 \cdot Q_S(a_i), \quad (8)$$

$Q_D(a_i)$ – number of request for demand, $Q_S(a_i)$ - number of offers for supply.

Also performances of assets profitability are included to the analysis:

- EPS (Earnings per Share)
- P/E (Price per Share/Earnings per Share)
- P/S (Market Cap/Revenues)
- Div/N (dividends per share)
- Div/P (dividends per share's price).

EPS indicator performs how good shares are protected by profits of companies.

The coefficient that estimates efficiency of investments – so-called coefficient of price – earning ratio (p/e) is founded on the basis of net income index. It shows how market estimates the results of company activity and it's prospect. The large value of this ratio specifies that these shares are quick-growing and comparatively the low value characterizes stable shares. A comparative analysis of index of P/E is a simple and effective tool for determination of the wrong appraised shares (comparison can be made a relatively middle-market or middle-branch indexes). For the developed economy it is possible to consider the average value of coefficient P/E in the range 7-9 in the period of recession and 15-18 in the period of economic growth.

The coefficient P/E can also characterize the expectations of income growth of companies. This is important especially for investors which consider expedience of long-term investments of capital.

Very often calculation of P/E index is to difficult or it does not represent the reality as a result of errors at the calculation of net income, related to the accounting features and other objective and subjective factors.

In addition different companies are characterized by the considerable degree of recurrence and by the significant variations of income. Many analysts consider that in such conditiions it is useful to use other index – so called coefficient of price – sales (P/S).

In the conditions of economic stability it is appropriate value of P/S from 0.4 to 0.8 for large companies. The value of P/S below this level indicates on underestimation of share's value (if, certainly, financial position of firm is stable enough).

Table 1. **Companies of PFTS-index**

Company (Issuer)	PFTS Ticker
KYIVENERGO	KIEN
UKRNAFTA	UNAF
INTERPIPE NYZHNYODNIPROVSKY PIPE-ROLLING PLANT	NITR
DNIPROENERGO	DNEN
DONBASENERGO	DOEN
STIROL	STIR
ZAKHIDENERGO	ZAEN
CENTRENERGO	CEEN

Applying the PCA we will select principal components in every group of factors that reflect information for all of these indexes. Basing on principal components the most significant attributes will be determined; it will give opportunity for decision making person to form his own rating.

We took first eight companies from the PFTS index and collected all data monthly aggregated for 3 years (Table 2).

Table 2. **Indicators of shares**

PFTS Ticker	$\overline{R_j(a_i)}$	$\overline{\sigma(R(a_i))}$	Sk	$Kurt$	$P(a_i)$	$\overline{\text{Spred}}$	$\overline{V(a_i)}$	$\overline{W(a_i)}$	$\overline{Q(a_i)}$
KIEN	0.02	0.14	-0.71	1.86	0.62	0.43	1 258	4.22	150
UNAF	0.05	0.15	2.82	9.88	0.52	0.08	1 149	12.71	217
NITR	0.17	0.68	3.97	21.29	0.54	0.70	281	3.96	89
DNEN	0.07	0.34	3.08	12.99	0.50	0.34	475	6.65	187
DOEN	0.04	0.26	1.13	2.70	0.44	0.68	144	3.12	146
STIR	0.06	0.19	1.00	0.88	0.56	0.33	367	2.86	72
ZAEN	0.05	0.18	2.50	9.08	0.50	0.28	515	6.14	179
CEEN	0.07	0.28	2.00	5.56	0.52	0.20	403	6.90	197

Realization of PCA was done in the package of Statistica 6.0. On the first step we'll apply the PCA to information of table 2 and we have got matrix of factor's weights (Table 3).

Table 3. **Matrix of factor's weights**

	Factor 1	***Factor 2***
$\overline{R_j(a_i)}$	*0.92*	*0.18*
$\overline{\sigma(R(a_i))}$	*0.96*	*0.11*
Sk	*0.65*	*0.73*
Kur	*0.73*	*0.57*
$P(a_i)$	*-0.26*	*-0.29*
$\overline{\text{Spred}}$	*0.67*	*-0.56*
$\overline{V(a_i)}$	*-0.69*	*0.22*
$\overline{W(a_i)}$	*-0.39*	*0.89*
$\overline{Q(a_i)}$	*-0.52*	*0.72*

The factor's weights of attributes stand in the columns of this matrix. Factors are always ranked by size of their contribution to total variance of attributes – the first factor is the most significant in this understanding. There are values of contribution in absolute and relative figures in two last lines of the Table 4.

Table 4. **Results of PCA**

No. of active vars: 9	*No. of supplementary vars: 0*
No. of active cases: 8	*No. of supplementary cases: 0*
Eigenvalues: 4.16759 2.68247 1.46124 0.438753 0.157293 …	
Number of factors: 2	*Quality of representation: 76.1%*

So, figure 4.16 (explained variance) means that the first factor explains 4.16 from total variance that equals 8 (8 is a number of attributes in the initial table), it is almost 46.31%.

Thus, already one first component considerably describes initial data. The second component explains almost 29.81% of total variance and together with the first one – more than 76.1%.

Let's pay attention to the most significant factor's weights (for example, those which exceed 0.75 or 0.8), exactly they specify attributes the most closely correlated with this factor. In our case the first factor is the most closely correlated with such attributes that allows us to accept the name "profitability- risk" as a working name for this component.

Risk and Liquidity attributes are the most closely connected to the second component. It is useful to consider factor's weights of objects for more detailed analysis (Table 5).

Table 5. **Matrix of object factor's weights**

	Factor 1	*Factor 2*
KIEN	*-2.3277*	*-1.8009*
UNAF	*-1.8914*	*2.6479*
NITR	*4.3732*	*0.1880*
DNEN	*0.5805*	*1.1545*
DOEN	*0.4656*	*-1.4473*
STIR	*-0.2894*	*-2.0497*
ZAEN	*-0.4558*	*0.61106*
CEEN	*-0.4529*	*0.69625*

Let's look at the matrixes of factor weights in the first column. The positive values of weights testify the showing up of factor higher middle level. As we see, the first factor has the highest weight in relation to NITR shares and some less – to DNEN and DOEN.

NITR shares are characterized by comparatively higher profitability and at the same time high risk. Negative values testify the showing up of factor below middle level.

According to the first factor KIEN, UNAF, ZAEN, CEEN have the lowest value. Other companies are approximately at the middle level, because their factor weights insignificantly deviate from zero. In accordance with this method the working name of the first factor ("a profitableness-risk") gets indirect confirmation.

Ranking of objects by the second component is quite different. UNAF has the greatest value STIR and KIEN – the lowest ones. It explains that there is factor of liquidity before us in accordance with which the highest-liquidity shares belong to company UNAF and relatively low values belong to shares of STIR and KIEN.

For generalization it is possible to classify considered companies on chart in dimension of two components (Fig. 1).

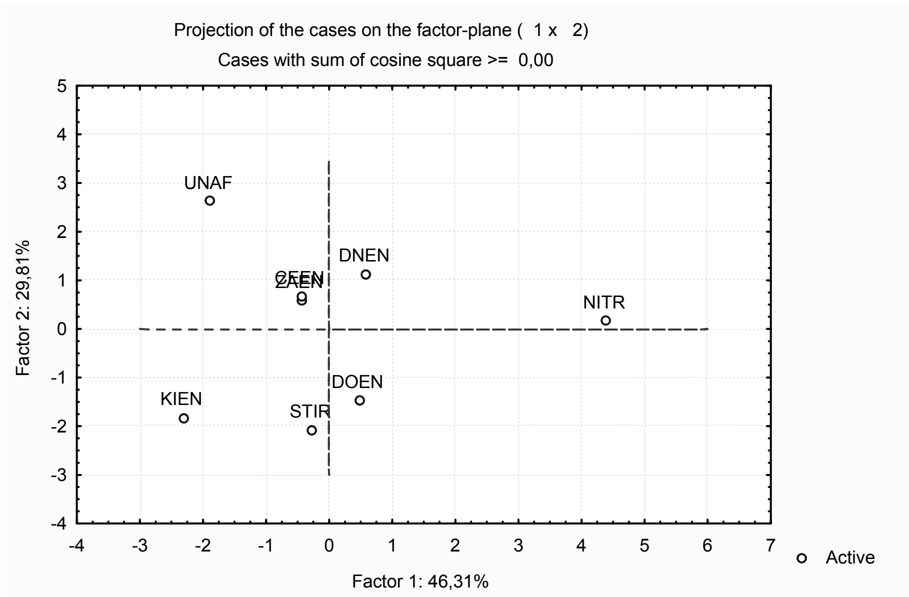

Fig. 1. Projection of the cases on the factor-plane

We'll apply the PCA to the indicators of equity (Table 6).

Table 6. **Performances of assets profitability**

	EPS	P/E	P/S	Div/N	Div/P
KIEN	0.011	122.73	0.48	0.0077	0.57
UNAF	3.08	6.67	2.01	0.35	1.7
NITR	0.063	44.39	0.48	0.01	0.37
DNEN	18.41	3.56	0.77	-	-
DOEN	-1.1	-4.55	0.58	0	0
STIR	1.31	9.73	1.4	0.035	0.27
ZAEN	0.092	214.4	0.82	0.038	0.19
CEEN	-0.042	-19.1	853.38	0	0

Let's select two components that explain 67.4% of total variance. As we see from the table 6 the first factor is the most closely related to the indicator of dividend payments (Div/N) and to indicator of dividend profitability of shares. The second factor reflects information of indexes related to indicators of company earnings and current price of shares.

Table 7. **Matrix of factor's weights**

	Factor 1	*Factor 2*
EPS	*-0.049*	*0.361*
P/E	*0.015*	*-0.916*
P/S	*-0.422*	*0.559*
Div/n	*0.953*	*0.212*
Div/p	*-0.976*	*0.065*

It is possible to set from the matrix of factor's weights (Table 7) that the first component "dividend profitability" has the biggest weight at companies "Ukrnafta" and "Kyivenergo" the second one – in "Tsentrenergo", "Donbasenergo" and "Ukrnafta".

Table 8. **Matrix of object factor's weights**

	Factor 1	*Factor 2*
KIEN	*0.085*	*-1.099*
UNAF	*3.312*	*0.829*
NITR	*-0.153*	*-0.344*
DNEN	*-0.755*	*0.895*
DOEN	*-0.653*	*0.028*
STIR	*-0.146*	*0.087*
ZAEN	*-0.193*	*-1.988*
CEEN	*-1.496*	*1.593*

Consequently, in accordance with the indexes of equity, shares of company UNAF have absolute advantages (Fig. 2).

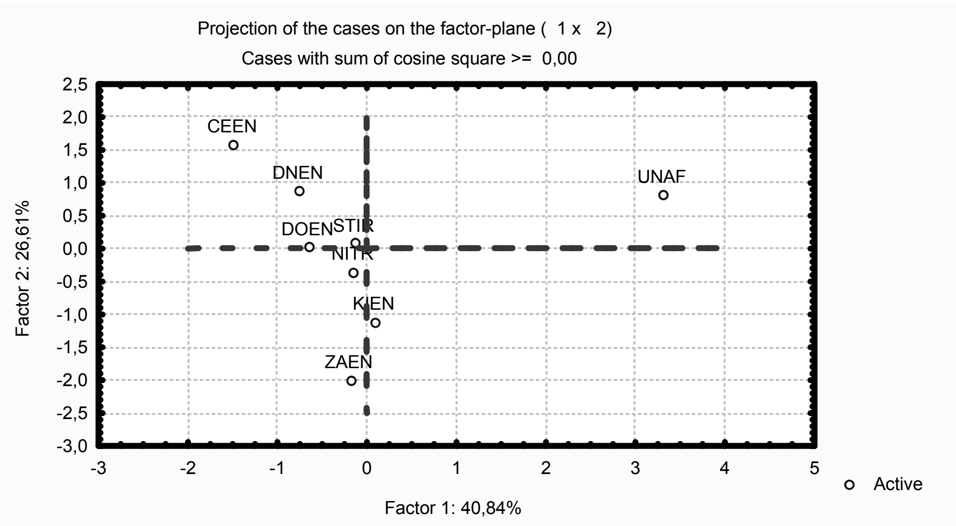

Fig. 2. Progection of the cases on the factor plane

Conclusions

Analysis of shares stock exchange statistics and equity indicators of companies that are included in the PFTS-index, gives opportunity to investor basing on PCA to form the most acceptable portfolio of companies taking into account priorities to some factors.

In accordance to component "profitableness-risk" and "liquidity" advantage must be given to the following companies: "Ukrnafta", "Zakhidenergo", "Centrenergo", in conformity to the component "dividend profitability" – "Ukrnafta", "Dniproenergo", business concern "Stirol" have leading position.

1. Ayvazyan, S., Bukhtaber, V., Enyukov, I., Meshalkin, L. (1989). Applied statistics. Classification and decline of dimension. – M.: Finances and statistics. – 607 p. 2. Belkin, M., Niyogi P. (2003). Laplacian Eigenmaps for Dimensionality Reduction and Data Representation, Neural Computation, Vol. 15, No. 6, Pages 1373-1396. 3. Dubrov, A. (1978). Processing of statistical data by Principal Component Analysis. – M.: of Statistician. 4. Frećhet, M. (1948), Les élements aléatoires de nature quelconque dans un espace distancié. Ann. Inst. H. Poincaré, 10 215–310. 5. Pearson, K. (1901). On lines and planes of closest fit to systems of points in space, Philosophical Magazine. 6. Soshnikova, L., Tomashevich, V., Sheber, M. (1999). A multidimensional statistical analysis in an economy: Studies. Manual for high school/ ed. prof. Tomashevich. – M.: YUNITI – DANA, 598 p. 7. Silvester, J.J. (1889), On the reduction of a bilinear quantic of the n-th order to the form of a sum of n products by a double orthogonal substitution, Messenger of Mathematics, 19 (42–46).

Надійшла до редколегії 05.05.12