MACROECONOMIC DEVELOPMENTS BASED ON THE MAIN COMPONENTS' METHOD

PhD. Mariana BĂLAN
Institute for Economic Forecasting – NIER, Romanian Academy, Romania,
mariana_prognoza@yahoo.com
Prof. univ. PhD. Valentina VASILE
Institute for National Economy – NIER, Romanian Academy, Romania
valentinavasile2009@gmail.com

Abstract:

The main components' method allows for identifying the fundamental structures of some complex data basis, the highlighting of significant relationships of interdependence and the selection of a low number of factors the features of which may be used as basis for estimates or decisions.

The paper intends to investigate some variables with direct and significantly propagated impact on the evolution of macroeconomic indicators – oil price, human capital and educational stock. The suggested methodology permits the identification of measures and instruments of policies' adjustment, including some prevention mechanisms.

The outcomes of using the model are presented based on: a) determining the influences of oil price changes (including shocks) on some macroeconomic indicators for the same period of time in which these occur; b) determining the employability on industries of the national economy with major impact on economic development, particularly for transition countries exposed to deep structural changes; c) influence of diminishing rate of non-participation to education and/or early school-leaving on economic performances.

Key-words: main components' method, employment, educational stock, labour market

JEL Classification: C15, C32, I21, J21

INTRODUCTION

The main components' method is one of the most used descriptive methods for data analysis.

The underlying principles of the method are old, but the development and diversity of these techniques are recent. The substantiation of the classic factorial theory was made by Spermann Thurstone and Burt. The analysis by the main components' method in the economic field was introduced by Karl Pearson in 1901 who studied the issue in the case of non-stochastic variables under different circumstances. The technique was generalised by Hotelling (1933) for the case of stochastic variables.

By means of this method an assembly of data may be reduced to a compact form which still can highlight some fundamental structures of the data. The method allows for underpinning some significant interdependency relationships which might remain unknown at a simple examination of the data. The purpose of this analysis is to reduce complexity, by identifying a small number of factors the characteristics of which can be the basis of some assessments or decisions.

The suggested methodology allows for the study of the influences of quarterly changes of oil price on some macroeconomic indicators, of possible oil price shocks on each of the considered variables. The analysis of the labour market dynamics in the Central and Eastern European countries by this methodology allows highlighting the sectors with the highest share with respect to

the number of civil employed, but also a hierarchy of the countries depending on unemployment rate for various age groups.

The main components' method used in the study of the educational level in countries from Central and Eastern Europe makes possible to underscore the differences between various education segments for the examined countries, but also the enrolment rate to the educational process of children and teens between 6 and 18 years, as well as the participation to lifelong learning (persons aged between 25-64 years).

The principles underlying the method are old, but the development and diversity of these techniques are quite recent. The basics of the classic factorial theory were laid by S. Ledermann and their improvement was realised by Spermann Thurstone and Burt. The analysis based on main components' method (Pearson and Hotelling) allows for a descriptive synthesis of an assembly of *n* observations for *p* variables.

Like all other descriptive methods, also the main components' method resorts to some general outcomes from mathematical statistics.

1. FORMULATING THE PROBLEM

An important factor in the proposed analysis is also the data set underlying the study.

By means of this method a set of data can be reduced to a compact form, but which still can highlight certain fundamental structures of the data. The method permits underscoring significant interdependency relationships which could not be known by simple data examination. The purpose of this analysis is to diminish complexity, by identifying a small number of factors the characteristics of which may be the basis for some assessments or decisions [1]-[2].

The core idea of the main components' method resides in determining the share (percentage) in total variation (sum of variances for those initial *p* variables) of each new variable.

By means of the main components' method the set of correlated variables $(x_1, x_{2,} ...x_{\mathrm{p}})$ is changed into a set of uncorrelated variables $(y_1, \mathrm{y}_{2,} ...\mathrm{y}_{\mathrm{p}})$ called *main components* by making use of the relationship:

$$\begin{cases} y_1 = a_{11}x_1 + a_{21}x_2 + ... + a_{p1}x_p \\ y_2 = a_{12}x_1 + a_{22}x_2 + ... + a_{p2}x_p \\ : \\ y_p = a_{1p}x_1 + a_{2p}x_2 + ... + a_{pp}x_p \end{cases} \quad (1)$$

Each component is a weighed sum of the variable x, and a_{ij} are weights or coefficients for which certain restrictions are imposed. The analysis by main components allows also for geometrical representations of individuals and their characteristics. Therefore is necessary also the algebraic formulation equivalent to the orthogonal rotation. As result, the coefficients a_{ij} must also satisfy the conditions:

$$\begin{cases} \sum_{i=1}^{p} a_{ij}^2 = 1, \ \mathrm{j} = 1,...\mathrm{p} \\ \sum_{i=1}^{p} a_{ij}a_{ik} = 0, \ \mathrm{j} \neq \mathrm{k}; \mathrm{j} = 1,...\mathrm{p}; \mathrm{k} = 1,...\mathrm{p} \end{cases} \quad (2)$$

An important consequence of the orthogonal condition is that the total variance of the components *y* is equal to the one of the variables *x*, that is:

$$\sum_{j=1}^{p} \text{var}(y_j) = \sum_{i=1}^{p} \text{var}(x_i) \quad (3)$$

In *choosing the number of components* due consideration must be given to the fact that, for the suggested analysis it is recommendable to retain the smallest set of components, but at the same time to have sufficient values that deliver a good representation of initial data. The variance of the *j* component is the own value λ_j. The components are selected in decreasing order of their own values: $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$. If the variables x are standardised (normalized), then the sum of the *x* variances shall be equal to *p*. Under these conditions, the sum of own values of the total variance y shall be also *p* .

The share (percentage) in total variance explained by the *j* component is:

$$\frac{\lambda_j}{\lambda_1 + \lambda_2 + ... + \lambda_p} \quad (4)$$

And the share (percentage) of the first k components, cumulated, is:

$$\frac{\lambda_1 + \lambda_2 + ... + \lambda_k}{\lambda_1 + \lambda_2 + ... + \lambda_p} \quad (5)$$

Among the most used criteria in taking a decision about choosing the main component that may be retained for the analysis are counted:

- retaining the first *k* components when these represent a higher share of the variance (70-80%);

- if the correlation matrix is analysed, then only those components are retained which have own values higher than 1;

- examining graphs: if the dependencies between own values and the number of main components are represented, then those components are selected for which own values decrease very quickly;

- the graph method, proposed by Kaiser and Cattel [6].

The weight given by the variable i in the j component is a_{ij}. The size of a_{ij} reflects the relative contribution given by each variable in the component. Very often, the coefficients are recalculated as coefficients for the most important components. These new coefficients called also *loaded components* are the coefficients used for reconstructing the variables *x* and *y* and they are computed after the relationship:

$$a_{ij}^{*} = \sqrt{\lambda_j}\, a_{ij} \quad \text{i} = 1,..., \text{p}; \ \text{j} = 1,..., \text{p} \quad (6)$$

When the correlation matrix of the variables x is analysed, the a_{ij}^{*} coefficients must be interpreted as correlation coefficients between the variable i and the component j .

2. APPLYING THE METHOD OF MAIN COMPONENTS ON STUDYING THE IMPACT OF THE OIL PRICE CHANGE ON GROSS DOMESTIC PRODUCT

In order to determine the impact of oil price changes on some economic indicators and on the gross domestic product from Romania, a set of 10 economic indicators were taken into account (Table no. 1) for which the monthly and quarterly evolution was realised for the period 1994 - 2008.

Using the main components' analysis to study the impact of oil price changes on gross domestic product requires undergoing several stages:

- realising a monthly data bank for the considered indicators ;

- building the monthly variation indexes for each indicator (against the same month of the previous year);
- making the logarithms of the data bank and building logarithmic differences;
- applying the algorithm of the main components' method for the built data banks.

Table no. 1 Economic indicators used for studying the impact of oil price changes on GDP

	Indicator
X1	Electric energy
X2	Total employees in economy
X3	Oil (production)
X4	Export (FOB)
X5	Import (FOB)
X6	Extracted natural gas
X7	Industrial production
X8	Nominal exchange rate (Lei/Dollar)
X9	M2 (at period's end)
X10	Average of the monthly nominal wage/IPC

Using the main components' method on the set of indicators considered resulted in obtaining the correlation matrix with the help of which were determined the positive or negative correlations between the examined variables.

By analysing the correlation matrix of own values is highlighted the fact that the cumulated percentage in total variance of the first seven components is of 81% (table no.2).

Table no.2. Initial statistics

	Own values	**Dispersion percentage explained by each factor**	**Cumulated percentage**
Comp. 1	1.6912	0.1691	0.1691
Comp. 2	1.4869	0.1487	0.3178
Comp. 3	1.1293	0.1129	0.4307
Comp. 4	1.1117	0.1112	0.5419
Comp. 5	0.9602	0.0960	0.6379
Comp. 6	0.8878	0.0888	0.7267
Comp. 7	0.8756	0.0876	0.8143
Comp. 8	0.7931	0.0793	0.8936
Comp. 9	0.6086	0.0609	0.9544
Comp. 10	0.4556	0.0456	1.0000

The analysis of the variation graphic of own vectors allows for determining the number of main factors (components) (Fig. no. 1) [4]. The fact is underpinned that the first 2 factors are dominant (Fig. no. 1), by having a strong evolution towards the axis ox, thereafter to 4 and 5 and, respectively, to 7 and 8 where a more significant change of the angle coefficient appears of the tangent to the evolution curve of own values. This evolution suggests the fact that for this set of indictors, the first 2 components and the components 4, 5, 7 and 8 are dominant.

In order to determine the correlations between the main components and the initial characteristics, the correlation matrix of the *"loaded" components* was used.

The correlations obtained within the correlation matrix of the *"loaded" components* allow for a first analysis of the impact of the changes in the oil price on each examined indicator, that is, depending on the sign of the values from the tables, positive or negative correlations may be determined between the variables subject to the study and, hence, the way in which these can influence the evolution of the respective indicator.

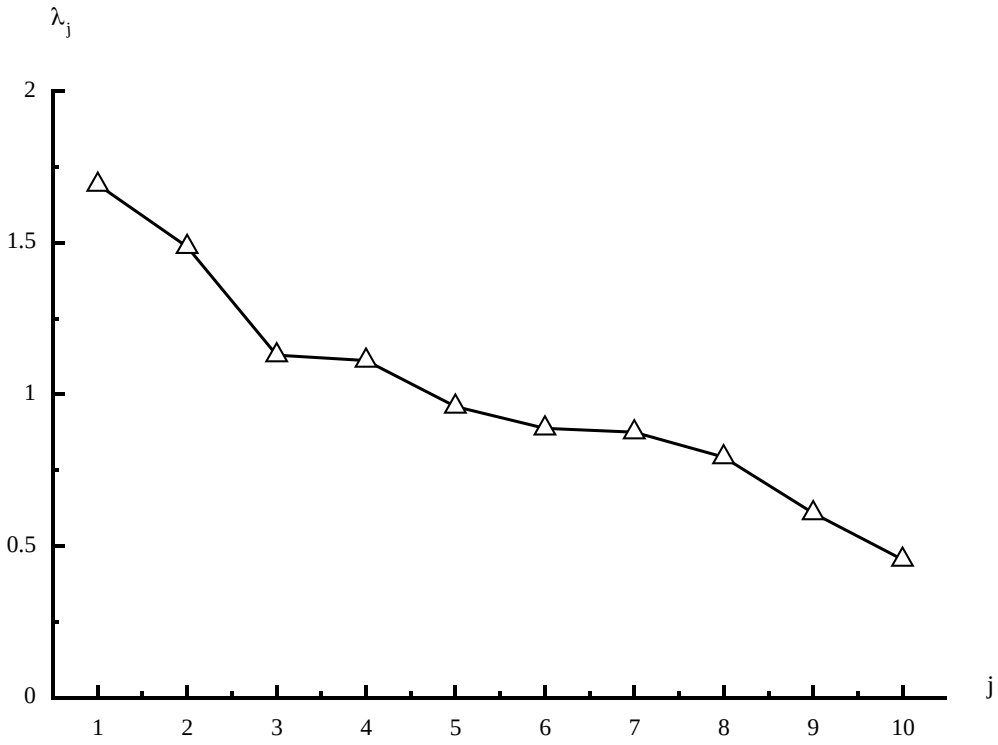

Figure 1. Evolution of own values for 10 economic indicators

By using the relationships (6) and the "loading" diagrams for two main components, is possible to highlight the contribution of each indicator to the development of the gross domestic product (Fig. no. 2).

Hence, the components in the first quadrant are those with positive correlations, hence they are not the ones influencing in the strongest way the evolution of GDP according to the changes of the oil price. In the diagram from Figure 2, each "bolded number" represents the order number of the indicator taken into consideration.

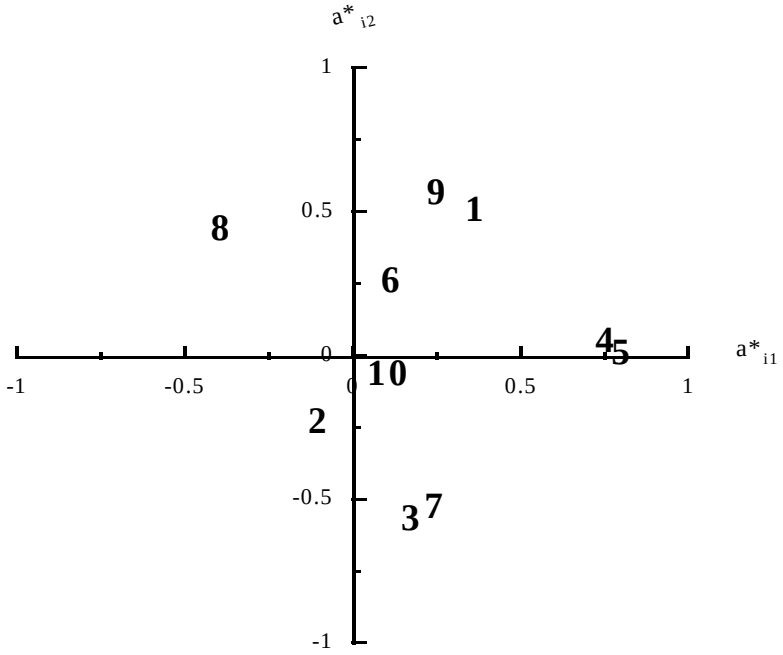

Figure no. 2. Contribution of each indicator to the evolution of Gross Domestic Product under the conditions of changes to the oil price

With the help of the monthly main components were determined the main quarterly components. These were used for estimating the evolution equation of the Gross Domestic Product for each of the considered data sets [6].

For the set of 10 data indicators the following equation was obtained:

$$DLOGGDP = C(1) + C(2)\cdot F_1 + C(3)\cdot F_2 + C(4)\cdot F_4 + C(5)\cdot F_5 + C(6)\cdot F_7 + C(7)\cdot F_8 + AR(6) \quad (7)$$

where (F_i) are the coefficients of the examined main components..

Using the coefficients resulting from equation (7) and the matrices of own vectors were determined the partial derivates. These together with the coefficients from the oil price regression allowed for determining the impact of changes in the oil price on each examined indicator.

The total impact of oil price change on the examined indicators computed based on the methodology of the main components' method is of -0,075.

The calculation way presented above for the gross domestic product for the economic indicators and the price of oil (as differences of the logarithms of quarterly values against the same period of the previous year), allows for quantifying the GDP change whenever a change by a certain percentage of the oil price occurs. Thus, an increase of 10% of the oil price in a certain quarter leads to a diminishment of the GDP by 0.686 % in the same quarter.

Hence, by this method is possible to forecast also the impact of the oil price on the macroeconomic indicators, hence on the GDP together with the monthly forecasts about the GDP and the evolution of the oil price.

3. THE ANALYSIS OF THE LABOUR MARKET DYNAMICS IN THE CENTRAL AND EAST EUROPEAN COUNTRIES BY THE METHOD OF THE MAIN COMPONENTS

The transition to the market economy influenced strongly on the employment rate of the population from the perspective of volume, and structure triggering the adjustment of the labour market policies in parallel with the process of institutional reform. Under the conditions of accelerated restructuring, increasing job uncertainty and of the current economic and financial crisis, on the labour market were accumulated a large number of social issues. Employment became one of the most tension-filled areas of the economies.

For studying the dynamics of labour force in various areas of economic activity in Romania, data were taken into account about civil employed population, on activities of the Romanian activities at CANE section level by the end of the year for the period 1991-2008.

The correlation matrix of the own vectors determined for the examined variables have highlighted that agriculture, industry and thermal electric energy, gas and water are three components which gather 96.3% of the total variation.

The positive or negative correlations of the first three main components for each of the analysed fields allow for a yearly distinction of the sectors with a high share with respect to the number of civil employed.

If only the correlations between the first two *"loaded" components* are taken into account and the analysed economic sectors (Table 3).

The position differences between the sectors are due to the fact that the sign of the scores for these components is changed from one sector to another. The negative values of the scores for agriculture, hunting and forestry, extracting industry, electric and thermal energy, gas and water, transports, warehousing and communications corresponding to the first component indicate that these are sectors with higher employment.

In order to make a comparative analysis of unemployment rate evolution on various categories of age and gender of employees, data were considered about this indicator for the 17 countries from Central and South-Eastern Europe.

Table no.3. The score corresponding to each field of activity depending on the first two main components

Economic activity field	a_{i1}^{*}	a_{i2}^{*}
Agriculture, hunting and forestry	-0.0002	0.0003
Fishing and fishery	0.0180	0.6669
Extracting industry	-0.0092	-0.0024
Processing industry	0.0013	-0.0035
Electric and thermal energy, gas and water	-0.0055	0.0034
Constructions	0.0030	0.0009
Trade	0.0013	-0.0004
Hotels and restaurants	0.0045	-0.0015
Transports, warehousing and communications	-0.0046	0.0535
Financial intermediation	0.0142	0.0325

Real estate transactions and others	0.0018	0.0003
Public administration and defence	0.0139	0.0122
Education	0.0021	-0.0153
Health and social assistance	0.0022	0.0053
Other activities of the national economy	0.0030	0.0006

The analysis of the correlation matrix of the own vectors highlights the fact that the first two components cumulate 96.5% from total variance.

The *scores* corresponding to each segment of the unemployment rate examined for the 17 countries depending on a certain main component may be calculated based on the relationship:

$$y_j^* = -0,49053x_1^* + ... - 0,46858x_5^* \quad (8)$$

where x_j^* is the standardised variance of x_j, which represents the percentage of the *j* component of the unemployment rate from the examined country, and the obtained results for each country are presented in the diagram in Figure 3.

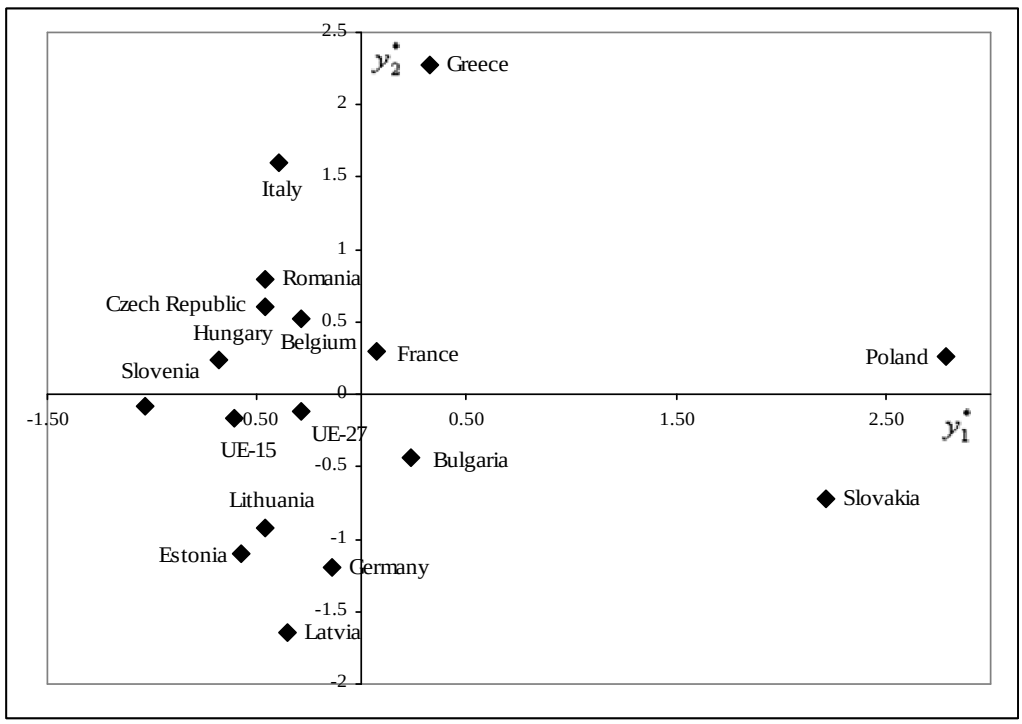

Figure 3. Evolution of the unemployment rate in EU and South-East European countries by using main components

For instance, the strongly negative values of the two components separate Latvia, Lithuania, Germany and Estonia and the other analysed countries. This indicates the fact that in the economy of these countries unemployment and the component corresponding to the unemployed aged under 25 years are preponderant.

The current situation of these economies is characterised by high unemployment rates and output and labour productivity increases.

Despite productivity gains from the majority of economies, the gains in diminishing unemployment were disappointing. The dimensions of labour force and the level of the employment rate of the population were declining in the period 1993-2008 (with brief revigorating periods around the years 2005). In the context of the current economic and financial crisis it should be expected that high values of the unemployment rate occur for individuals up to the age of 25 years and among those older than 50 years of age.

4. THE STUDY OF THE EDUCATIONAL LEVEL IN EU MEMBER COUNTRIES (EU-15 AND EU-25) AND IN CENTRAL AND EAST-EUROPEAN COUNTRIES BY THE METHOD OF MAIN COMPONENTS

The method of main components allows also for highlighting some significant interdependency relationships which could not be known by simple data examination.

For the suggested study were considered data on the pre-school enrolment rate, on education enrolment of children and teens aged 6 to 18 years, as well as on participation to permanent education (persons aged between 25 and 64 years of age).

The data set for these indicators comprised the period 1994-2008 and they were examined for some countries of Central and Eastern Europe.

Based on the correlation matrix for these indicators obtained by applying the main components' method positive or negative correlations were determined between the examined variables. Hence the only negative correlation underpinned is the one between the pre-school enrolment rate and the enrolment rate of children and teens aged between 6-18 years.

Because the aggregated percentage of the first two components is of 87.6%, for the suggested analysis were considered as significant only the first two components.

Using the relationships (6), depending on the first two main components, were calculated the scores of the indicators analysed for each of the countries subject to the study (Fig. 4).

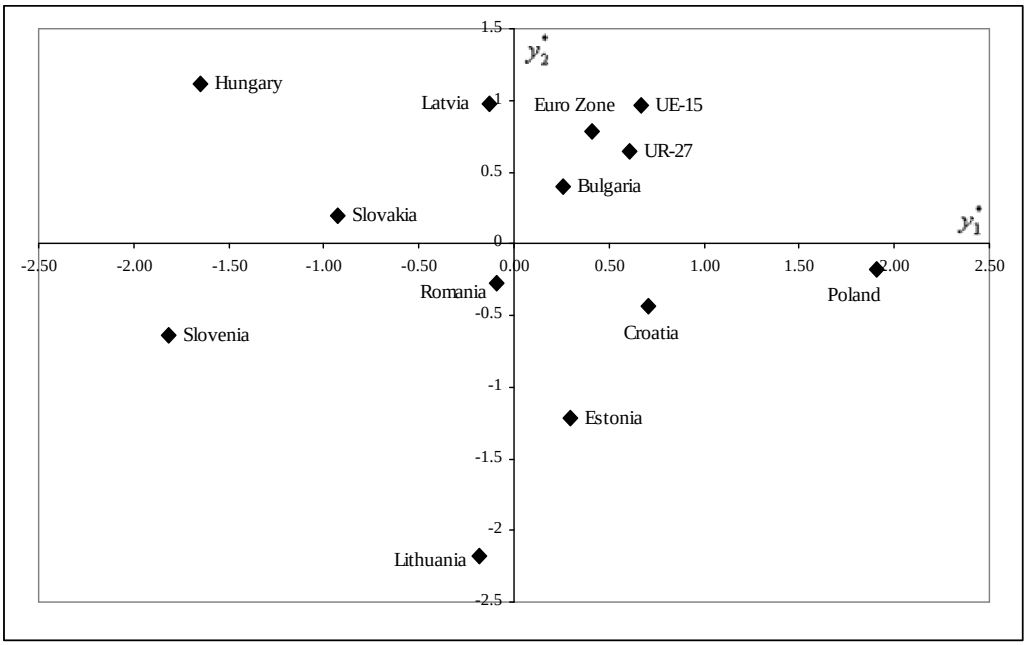

Figure 4. Evolution of the educational level using the scores of the two main components for EU countries, and Central and Eastern-European countries

The position differences between the countries are due to the fact that the sign of the scores for these components is changed from one country to another. The negative values of the scores for Slovenia, Estonia and Lithuania indicate that these have a higher enrolment of children at the corresponding level for the ages 6 to 18 years, and lower for the other education forms. Romania, Croatia, Poland are approximately on the same horizontal, with positive values of the component y_1^*. The countries with both of these components positive indicate that the differences between them are due to the different sign of the scores for the enrolment rates of pre-school education and of the education level corresponding to the ages between 6 and 18 years.

CONCLUSIONS

Using the method of main components and of the regression equations for the Gross Domestic Product, as well as for the other variables examined have allowed for determining the impact of the oil price changes on each considered economic indicator. This made possible to evaluate the GDP changes to the percentage changes of the oil price.

Applying the main components method in analysing the influence of the oil price changes on Gross Domestic Product allowed also for highlighting the variable most affected by these changes, as well as for quantifying their changes according to the shocks underwent by the oil price.

. Also, this methodology permits, that based on supplied data, to realise forecasts for determining the impact of oil price changes on some economic indicators.

In analysing the dynamics of the labour force with the help of this methodology, was highlighted which of the studied variables have a wider or lesser expansion in each of the countries subjected to the analysis. At the same time it allowed also for a comparative analysis of the evolution stage of the considered variables.

Applying the method of main components for the comparative analysis of the various education segments in different European countries highlighted which is the most developed segment.

This information together with similar analyses on other groups of indicators can become very useful in analysing the economic evolution of a country at a given moment, as well as in taking decisions on the economic-social development of a society.

Applying the method of main components on some economic phenomena, and for the issues of some economic sectors, to analysing the correlation degree of the main macroeconomic indicators in view of developing macroeconomic forecasts may contribute to economic growth turnaround which became a priority of the macroeconomic policy starting with its importance.

SELECTED BIBLIOGRAPHY

1. Alvin, C., (1998), *Multivariante Statistical Interface and Application*, Ed. A. Wiley Interscience Publication John Wiley&Sons. Inc.;
2. Bartholomew, D., Steele, F. Moustaki, I., (2002), *The Analysis and Interpretation of Multivariante Data for Socila Scientes*, Ed. Chapman&Hall/CRC;
4. Jonston, J., (1997), *Econometric Methods*, Ed. The McGraw-Hill Companies, Inc.;
5. L. Klein, Mariana Nicoale, A. Roudoi, V. Eskin, (2004), *Principal components models of the Romanian economy. Study of the oil price impact upon GDP*, Romanian Journal of Economic Forecasting, Year V, Supplement / 2004;
10. M. Nicolae, (2003), *Utilizarea metodei componentelor principale la analiza stării de „sănătate" a societăţilor bancare*, revista Oeconomica, Nr. 2/2003, pag. 117-133;

13. St. Stefănescu, (2000), *Numerical Analysis*, University of Bucharest Publishing House, Bucharest;

14. Tomasz Michiewicz, (2001), *Convergence in Employment Structure: Transition Countries versus the EU-Reforms, Income Level or Specialisation Patterns, Ten years of Economic Transormation*, Volume III-Societies and Institutions in Transition, Ed. By K. Liuhto, no 16/2001 ;

*** *Anuar Statistic* 1990-2008 - Institutul Naţional de Statistică

*** *Buletin lunar* –Banca Naţională a României, 1994-2008

*** *Buletin statistic lunar* –Institutul Naţional de Statistică, 1994-2008

*** Date Eurostat, www. Eurostat.com

*** World Development Indicators (WDI) Washington, DC, www.worldbank.org/