



## Characterizing the level of economic development of countries

**Alexandra NASTU**

*The Bucharest University of Economic Studies, Bucharest, Romania*

[alenastu@yahoo.com](mailto:alenastu@yahoo.com)

**Stelian STANCU**

*The Bucharest University of Economic Studies, Bucharest, Romania*

[stelian.stancu@csie.ase.ro](mailto:stelian.stancu@csie.ase.ro)

**Andreea DUMITRACHE**

*The Bucharest University of Economic Studies, Bucharest, Romania*

[dumitrache.andreea03@gmail.com](mailto:dumitrache.andreea03@gmail.com)

**Abstract.** *The main purpose of this paper is to provide an objective analysis of the economic development level of countries. This is done by measuring it through a new index and by classifying the countries in an optimal number of clusters, each group characterizing different levels of economic development. The proposed methodology is based on three steps: creating a composite index (by applying the principal component analysis), establishing the optimal number of development groups (based on the number of principal components and on the hierarchical clustering) and clustering countries into them (with the help of k-means analysis). Therefore, this approach solves the difficulty of classifying the countries, complication that is mentioned in the specialized literature. Also, the paper creates a better understanding on the economic development level of countries, as, usually, the papers examine the economic growth level of countries. The analysis is conducted at the level of 60 countries for year 2015, using 12 indicators from categories that influence economic development (income, inequality, health, education and living conditions). The empirical results revealed that the countries can be grouped in two groups: economical developed countries (approximately 2/3) and economic developing countries (approximately 1/3). The countries that are most developed from an economic point of view are: Singapore, Luxemburg and Finland.*

**Keywords:** economic development, composite index, principal component analysis, k-means, SPSS.

### Introduction

In the context of the current global economic development, it is necessary to study the economic growth and economic development of the countries. As there are many studies on economic growth, the main purpose of this paper is to characterize and group countries according to their level of economic development. Therefore, we create an objective methodology based on indicators that cover all categories that define the economic development of a country (such as health, education, income, inequality and living conditions). By accomplishing this goal, we cover two points in the literature which are not sufficiently analyzed. The first point is that most studies confuse the analysis of economic development with the analysis of economic growth. Therefore, the result of classifications of countries in groups shows the level of economic growth of the countries and not that of the economic development. The second point is that the division of countries into groups is not

done in an objective way, but by manually setting certain thresholds.

Economic development is the process that focuses on the qualitative and quantitative growth of the economy. It measures all aspects that lead the population of a country to become richer, healthier and to have access to good quality housing. Economic development is different from economic growth, the latter being the quantitative component that leads to economic development (Scutaru, 2013). The qualitative part of economic development is represented by the quality of life of the population.

Economic growth is the increase of existing goods and services in a nation, over a period. This is resulted due to an improvement in the quality of education, technology or in any way possible of obtaining an additional value of the goods and services produced by each sector of the economy. Most often, economic growth is measured by analyzing the value of real GDP.

The importance of the analysis of quality of life, which is the qualitative component of economic development, is highlighted in the OECD book "Beyond GDP" (2018). The authors provide a detailed explanation of why GDP should not be the only indicator to characterize the economic and social performance of a country. No individual variable is a good measure to show the state of a country. Political actions need to be guided by a dashboard of indicators that can inform about the material condition of people and the quality of their lives, their inequalities and sustainability. This dashboard must include indicators that allow assessment of the population condition throughout the economic cycle.

Regarding the second point: the classification of countries, the analysis can be either very clear, without the need for delimitation thresholds or very complicated. The complexity is given by the values of the country's indicators. An example of an easy classification of two countries in the renowned groups: developed or developing countries, is shown the following example. In 2009, a Burkina Faso citizen earns an average of \$ 510 compared to \$ 37,870 earned by a Japanese citizen. In Burkina Faso, only 29% of the adult population was educated, and the life expectancy of a newborn was 53 years. In Japan, almost all adults were educated, and life expectancy was equal to 83 years. For a better understanding of these social and economic differences, countries can be placed in groups. For this example, the classification is clear, Japan is a developed country, and Burkina Faso is a developing country. However, classification is not so obvious for other countries (Nielsen, 2011).

Therefore, the present paper proposes to measure the level of economic development based on a set of indicators covering both the quantitative part of the notion of economic development and the qualitative part. Thus, income, inequality, health, education and living conditions are the branches covered by the selection of indicators. Another mentioned difficulty is to provide a classification of the countries according to the level of economic development. This classification is complex due to the multiple indicators that should be considered, so creating a composite index based on the principal components, facilitates this process.

The second section of the paper contains the existing literature of the analysis and the classification of countries according to the economic level. Section three proposes a methodology for analyzing the economic development of countries and grouping them according to their specific level. Also, in this section is mentioned the composition of the database formed for this study. The fourth section shows the empirical results, and the last section presents the conclusions of the paper.

## Literature review

Over the years, the analysis of economic development of countries has been addressed by quantification of several macroeconomic indicators, such as GDP per capita, which limits the study to that of economic growth. Even though economic growth has beneficial effects on the population by increasing the employment rate, income and consumption of goods and services, the prospect is incomplete because it does not take into account the real problems of the population. So, variables such as poverty, health, education etc. must be integrated into the analysis.

Because economic development is a multidimensional process and it cannot be represented as a one-dimensional economic indicator, there is no criterion that is generally accepted when it comes to classifying countries according to their level of development (Nielsen, 2013). Classification of countries is a complicated exercise because it groups together countries that can be very heterogeneous when there are considered different aspects of the development process, such as product and export composition, health and education indexes (Vaggi, 2017). Over the years, the UN General Assembly has debated the issue of country classification, but it has never established a classification taxonomy. Instead, international organizations have set up such methodologies, UNDP, the World Bank and the IMF are three examples that periodically publish comparative development reports (economic or not) and country classifications. On the one hand, the World Bank makes the countries classification according to the per capita income level. Although the World Bank recognizes that development is not a matter of income, it considers that this indicator is the best signal of economic capacity and progress (Vazquez & Sumner, 2012). The United Nations Development Program (UNDP) classification system is built around the Human Development Index (HDI). To capture the multifaceted nature of development, HDI is a composite index created based on education, income and longevity indicators. Other aspects of development, such as political freedom and personal security, have been recognized as important, but the lack of data has prevented their inclusion in the HDI. The International Monetary Fund classifies the countries in its own database into two large groups: developed economies and emerging markets and developing economies. The threshold used for differentiation is not clear.

So, international organizations are addressing the development taxonomies differently. An explanation for this is that the institutions have different mandates and can therefore address the different issue both from an operational point of view and from an analytical point of view. At the same time, an overview suggests that systems have similar outcomes in classifying countries in the groups: developed or developing. All three classification methods designate approximately 20% - 25% of countries as developed (Nielson, 2011). In addition to Nielson's works (2011, 2013), the differences between the country classification taxonomies are also presented by Vaggi (2017). He also provides an explanation of why organizations do not change the very unbalanced threshold between developed and developing countries and the reason is that retrograding countries in the lower group affects their national pride.

Vázquez and Sumner (2012) propose clustering as a way of dividing countries into groups that show the level of development. The chosen method is the hierarchical clustering using the Ward method based on the Euclidean square distance. The technique is applied on a database made up of 12 variables that affect development in general, available for 101 low and low to medium income countries. The chosen period is distinct depending on the

availability of the indicator data. Thus, most variables are calculated as the average of the values for the period 2005-2010, and for the indicators for which this was not possible, it was used the most recent year with available data. The result of the analysis divides the observations into five clusters: the high poverty rate countries, with mostly traditional economies, those of countries dependent on natural resources and little political freedom, those of countries with high inequality, countries with emerging economies with serious environmental protection challenges and limited political freedoms, and the last cluster is made up of emerging economies with low dependency on foreign funding. Two-thirds of the world's poorest countries are in the first cluster, about a quarter are found in clusters 3 and 4, and the rest of the countries are split between clusters 2 and 5. As a conclusion, there is no linear representation according to the level Development. There is no cluster that has only good or negative indicators of development. The authors of the paper believe that is better to build taxonomies in this way than to classify and group countries according to the per capita income level. Even though the study did not provide a classification of countries, the advantage is to define a more nuanced picture of the diversity of development challenges. Thus, the right policies to correct the development difficulties can be established in a clearer way.

An important observation is that these classifications of countries show either the economic growth or the economic development or the sustainable development. Although these elements are easily confused in literature, the relationships between them are not of equivalence. Sustainable development includes economic development, which contains the economic growth factor. The World Bank's classification of income countries indicates a measure of economic growth that influences economic development and, therefore, general development of the country. The present paper proposes to measure the level of economic development, and from the methodology described above, only the United Nations Development Program performs classification according to the level of economic development. The components of the Human Development Index (income, education and longevity) influence the economic development.

In this paper we propose to create a composite index to characterize the level of economic development of countries. The countries classification into specific groups is achieved by clustering the value of the composite index.

## Methodology

The main purpose of the paper is to analyze in an objective manner, the level of economic development of countries. This is done based on a set of indicators from several categories that influence economic development. For this purpose, 12 indicators, covering information on income, inequality, health, education and living conditions categories, are extracted from the World Bank database. 60 countries are selected from three continents (Asia, Europe and Africa) and the period review is 2015, the most recent year with available data for all countries.

Therefore, the proposed methodology has three stages, in addition to the data preparation. The first step is to create a composite index showing the level of economic development, the second is to find the optimal number of clusters, each one showing a different level of economic development, and the third step is to group the countries into the clusters.

The creation of the composite index is done by using principal components analyzes

(PCA). This is the most popular multivariate statistical technique and it is used by almost all scientific disciplines. PCA is used to reduce the number of variables from a given data set into a smaller number of dimensions. In mathematical terms, from an initial set of  $n$  correlated variables, the CPA creates non-correlated components, each of which is a linear weighted combination of the initial variables. The property of non-correlation of the components is evidenced by the fact that they are perpendicular, i.e. they have right angles between them, which means that the indices measure different data dimensions. The weights for each major component are given by the correlation matrix own vectors or, if data is standardized, by the co-variance matrix. The components are ordered so that the first one explains the greatest possible amount of variation from the original data. The second component is completely uncorrelated to the former and explains an additional variation, but less than the first component. Subsequent components are not correlated with the previous ones, so each one captures an additional dimension in the data, while explaining the smaller proportions of the variation of the initial variables. The higher is the degree of correlation between the initial variables in the data, the less components are needed to capture common information (Vyas and Kumaranayake, 2006).

We use Davidescu's proposal (2015) of composite index creation as a guide. In the mentioned study, the author proposes the creation of an index measuring environmental innovation at European level. The paper presents the outcome of the twelve variables in five thematic areas of eco-innovation for 27 countries in Europe during 2003-2013. The composite index is created based on the analysis of the main components for the selected indicators. Thus, the six components are multiplied by their weight ratios, and by summing up the composite index they measure ecological innovation. The index is used to measure the phenomenon and help to create the hierarchy of countries.

The second proposed step for this study is the selection of the number of clusters. This is done through the CPA, but also by observing the resulted dendrogram from hierarchical clustering using Ward method.

Hierarchical clustering is designed to produce multiple cluster solutions, called cluster hierarchies. The main feature is that the number of clusters is not known in advance, nor it is suggested by the user. Hierarchical clustering methods are considered heuristic methods. The Ward distance is chosen because it considers the minimization of the intracluster variability or, in other words, the maximization of the intercluster variability, i.e. the degree of homogeneity of the clusters (Ionescu, 2015).

The country classification represents the last step and it is performed using the k-means clustering method. This technique involves supervised learning, so it is needed to have set the number of clusters in which the observations are divided according to the distance of the nearest centroid.

All these data analysis techniques are performed with the help of SPSS version 25. It is the most important statistical software and it is used to solve research problems or ad hoc analysis, hypothesis tests and predictive analyses. The software helps to understand better the selected database, to analyze trends, to perform forecasts and to validate hypotheses that lead to precise conclusions.

## Results and discussions

Prior of applying the above methodology, we standardize the data as the multiple units of indicators influence the reliability of the results. We perform this process in SPSS by the Z

score function. Another important aspect is that not all variables have a positive effect on economic development. Therefore, we turn all indicators to measure the same by multiplying with -1 the values of the variables whose bigger values means poor economic development. A high value of the indicator's values must have a positive impact on economic development.

**Table 1. Reliability Statistics**

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.885	.885	12

Source: Authors' own research results /contribution.

The last step before applying the analysis of the principal components is to check how well the chosen variables can measure a single object. For this purpose, we perform Cronbach's Alpha statistics in SPSS. C-alpha is not a statistical test, but a reliability factor based on the correlation between individual indicators, the value of which should exceed 0.6 (for indulgent persons) or 0.7 (OECD, 2008). The result (0.885) is positive (Figure 1), but the "Squared Multiple Correlation" at indicator level (Figure 2) shows that the indicators: "Corruption Perception" and "Unemployment" do not help in our analysis because they are not homogeneous with the rest of the indicators from the database. Therefore, we remove these two variables from the database.

**Table 2. Reliability statistics at variable level**

Item-Total Statistics					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Zscore(Inflation)	.0000000	53.917	.599	.479	.876
Zscore: Life expectancy at birth	.0000000	49.743	.920	.954	.857
Zscore: Access to electricity	.0000000	51.250	.801	.917	.864
Zscore: Cause of death, by communicable diseases and maternal, prenatal and nutrition conditions	.0000000	50.849	.832	.876	.862
Zscore: People using at least basic drinking water services	.0000000	50.624	.850	.903	.861
Zscore: Cost of business start-up procedures	.0000000	53.258	.648	.601	.873
Zscore: Time required to start a business	.0000000	56.085	.443	.458	.884
Zscore: GDP per capita	.0000000	52.527	.703	.662	.870
Zscore: Corruption Perception	.0000000	68.092	-.325	.204	.922
Zscore: HCI: Learning-Adjusted Years of School	.0000000	50.144	.888	.990	.859
Zscore: HCI: Total	.0000000	50.018	.898	.994	.858
Zscore(Unemployment)	.0000000	62.697	.001	.184	.907

Source: Authors' own research results /contribution.

To be vigorous, we also use KMO and Barlett's sphericity tests to verify that the selected indicators can be used to create an index. KMO value must be greater than 0.6 to be able to use the chosen variables. In our case, the KMO value is 0.851 (Figure 3). In addition to KMO, Bartlett's sphericity test is used to test the null hypothesis that the indicators are uncorrelated in a correlation matrix. The results of our analysis show a level of significance of zero, a value that allows rejection of the null hypothesis and the conclusion that the force of the relation between the variables is strong.

**Table 3. KMO and Barlett's Test**

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.851
Bartlett's Test of Sphericity	Approx. Chi-Square	812.204
	df	45
	Sig.	.000

Source: Authors' own research results /contribution.

Therefore, all these diagnostic procedures indicate that the analysis of the main components is appropriate and that the composite index can be created.

The Principal Component Analysis applied to the ten variables reveals two main components that explain 79.926% of the data set variability. The number of main components is given by the condition that the eigenvalues should be greater than one. The number of two principal components is also in accordance with Kaiser's criterion, that the coverage of the extracted principal components should be over 70%.

**Table 4. The component Eigenvalues**

Component	Total Variance Explained								
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.886	68.856	68.856	6.886	68.856	68.856	5.341	53.409	53.409
2	1.107	11.070	79.926	1.107	11.070	79.926	2.652	26.518	79.926
3	.593	5.926	85.852						
4	.515	5.152	91.004						
5	.378	3.781	94.786						
6	.291	2.915	97.701						
7	.106	1.060	98.761						
8	.065	.646	99.407						
9	.056	.556	99.962						
10	.004	.038	100.000						

Extraction Method: Principal Component Analysis.

Source: Authors' own research results /contribution.

In order to create the composite index, we use PCA in two steps, as proposed by Davidescu (2015). As a first step, the Varimax rotation technique is used to obtain the components on orthogonal axes. Loading results after orthogonal rotation are the correlation coefficients of each variable with the factor, so that they naturally range from -1 to +1. A negative value means that results must be interpreted in the opposite direction to the way it is formulated. The second step involves identifying the initial variables that are most

correlated with each major component. Therefore, each component can be explained by the initial variables for which the correlation coefficients have the highest values. Thus, to explain the meaning of the main components obtained by our analysis, we need to analyze the correlation of the coefficients between the two main components and the original indicators. By applying such a methodology, the process of explaining the main components becomes relatively simple.

**Table 5. Rotated Component Matrix**  
**Rotated Component Matrix<sup>a</sup>**

	Component	
	1	2
Zscore(Inflation)	.375	.648
Zscore: Life expectancy at birth	.792	.536
Zscore: Access to electricity	.916	.161
Zscore: Cause of death, by communicable diseases and maternal, prenatal and nutrition conditions	.884	.268
Zscore: People using at least basic drinking water services	.922	.240
Zscore: Cost of business start-up procedures	.807	.135
Zscore: GDP per capita	.557	.563
Zscore: HCl: Learning-Adjusted Years of School	.763	.550
Zscore: HCl: Total	.748	.585
Zscore: Time required to start a business	.006	.900

Extraction Method: Principal Component Analysis.  
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

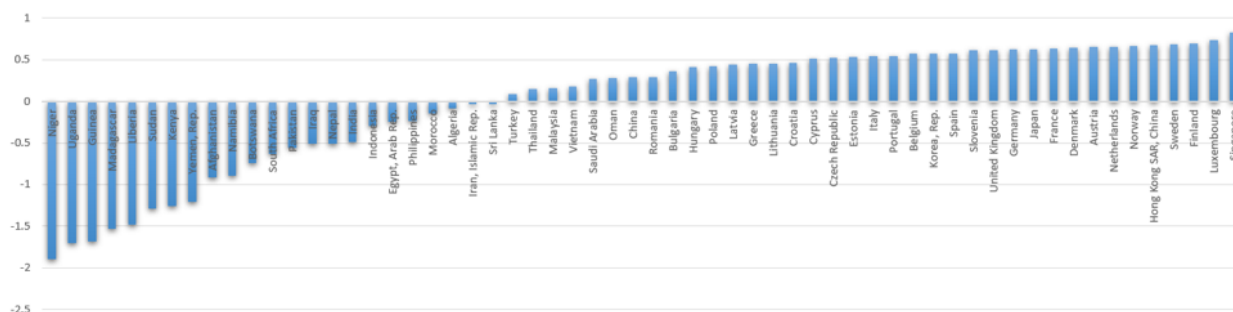
Source: Authors' own research results /contribution.

Therefore, as we can see in Table 5, the first main component is most closely related to social indicators that show the population's access to drinking water and electricity. It is also characterized by other variables that have as a common point the population: life expectancy, the level of studies and the expectation level of a newborn (from the point of view of health, education and so on). The second component has a more economic-financial tinge, being focused mainly on aspects like: time required to start a business, the level of inflation, the GDP level.

We use the factors (principal components) to create a variable that shows the level of economic development of countries. The weights used for capturing the contribution of each component within the index are determined by the proportion of variance recovered by each component in the total variance (Table 4). Thus, we use the below formula to create the composite index:

$$\text{Economic development} = (53,409 / 79,926) * \text{PC 1} + (26,518 / 79,926) * \text{PC 2} \quad (1)$$

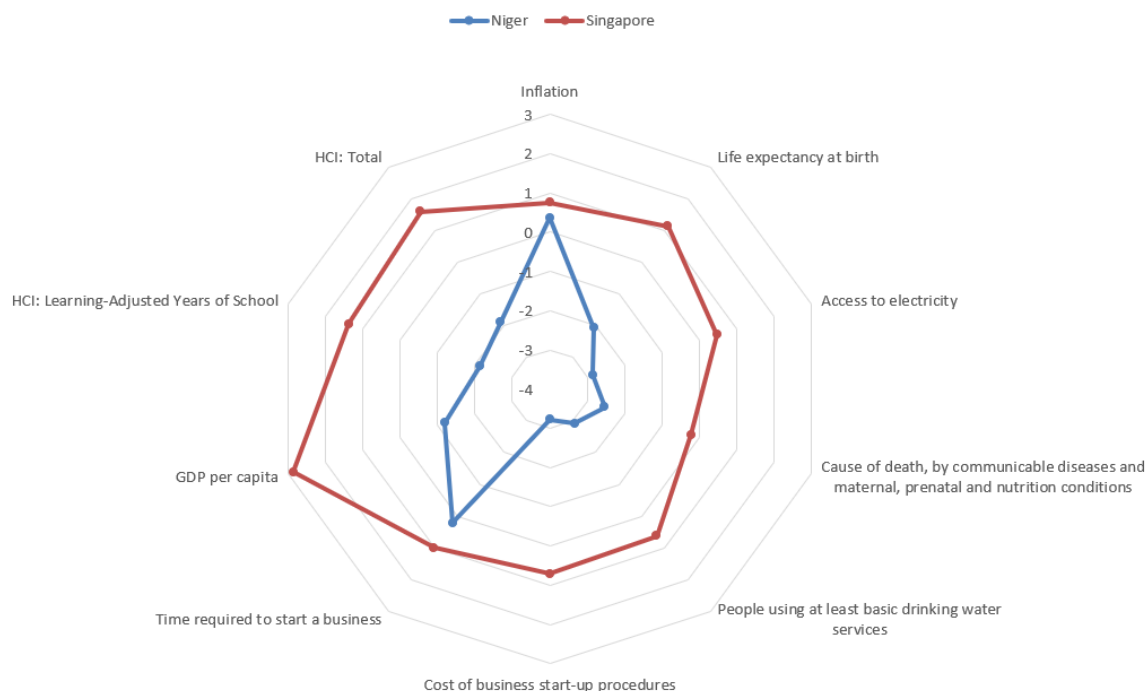




**Figure 1. The level of economic development of countries**

Source: Authors' own research results /contribution.

The values of the composite index that measures the level of economic development of countries, is shown in Figure 1. A first observation is that the values are positive for some countries, and negative for others. As the index measures the economic development, we can affirm that the countries with negative values do not have a good level of economic development. From our 60 countries data set, Niger is the one with the poorest level, followed by Uganda and Guinea. At the other pole, Singapore has the greatest economic development level, followed by Luxemburg and Finland. The reason why Singapore and Niger are in diametrically opposed positions is visible in the figure below. A mention is that all indicators are standardized and transformed so their high values influence in better the economic development (for example, the original data of "Time required to start a business" was multiplied to -1 and then standardized).



**Figure 2. Comparative analysis of variables values for Niger and Singapore**

Source: Authors' own research results /contribution.

Because the composite index contains both positive and negative values, we can easily

split countries into two groups. This is an advantage of the chosen methodology because the classification of countries is a great challenge in the literature. However, the disadvantage is that the objective division can be done only in two groups (countries with the index value greater than zero and opposite).

Another possibility of grouping is by applying a k-means clustering technique. We choose the optimal number of clusters based on the number of principal components and based on the resulting dendrogram from hierarchical clustering analysis. In both cases, the optimal number of clusters is two. So, we have one group for economically developed countries and one for developing countries. We apply k-means on the composite index and then on full the data base. In this way we can check the index's veracity. The composite variable explains about 80% of the information variability from the entire database. Thus, the results from the index clustering should have almost the same output as the results from the clustering of the whole database. The outcome shows that only 3 countries out of 60 are not clustered in the same way: Egypt, Philippines and Iran. So, the results are encouraging.

We also compare the clustering results with the grouping done based on the index sign. The results coincide in large measure: countries with a negative value of the index are part of the "Developing" cluster and vice versa. This comparison is possible because the optimal number of clusters is two. Otherwise, the index value would not be used to classify countries, but just to rank them. Also, the composite index is a good measure of the evolution of countries and it is also used as a sign to warn a potential decrease of the economic development level of a country.

## Conclusion

The main purpose of this paper is to provide an analysis of the economic development of countries in an objective way. The proposed methodology is based on three steps: creating a composite index (by applying the principal component analysis), establishing the optimal number of clusters for the data grouping (based on the number of principal components and on the hierarchical clustering) and ranking countries according to their level of economic development (with the help of k-means clustering). Therefore, this approach solves the difficulty of classifying the countries, difficulty that is mentioned in the specialized literature. An important attention is given to the selection of indicators as economic development must be analyzed based on variables that influence the level of economic growth and the quality of life of the population. As the size of the economic development is very big, it cannot be measured by a one-dimensional indicator. The studies that analyze GDP per capita or income per capita, analyze in fact the economic growth, that influence economic development.

Therefore, the analysis of the main components, k-means and hierarchical clustering were applied on a basis of ten macroeconomic indicators with latest values for 2015, existent for 60 countries. The level of economic development is measured by the composite index and the division into groups is done in three ways that led to approximately identical results. Countries with a negative value of the composite index have turned to the group of developing countries and vice versa. The other ways of grouping involve the application of k-means clustering on the composite index, respectively on the whole database. The number of groups is according to the number of principal components and also confirmed by the results of hierarchical clustering.

In future research, we propose to use the same methodology to a database formed by distinct variables that also influence the economic development and to compare the results

with the ones from this paper. The scope is to find if different indicators change the results, even if they are covering the same group (income, inequality, health, education and living conditions). As this study proposed a methodology of analyzing the economic development level, the following paper needs to be oriented on the economical results and on the needed actions to correct the economic development decrease of a country.

## References

- Davidescu, A., Paul, A., Gogonea, & R. M., Zaharia, M. (2015). Evaluating Romanian Eco-Innovation Performances in European Context. *Sustainability*, 7(9), 12723-12757. Retrieved from <http://www.mdpi.com/2071-1050/7/9/12723/html#B21-sustainability-07-12723>.
- Davidescu, A. (2017). Measuring the social market economy. A composite index approach for EU countries. *Sciendo* 11(1). Retrieved from <https://content.sciendo.com/view/journals/picbe/11/1/article-p227.xml>.
- Ionescu, Ș.-A. (2015). Clusterizarea ierarhică cu aplicații în analiza financiară. *Revista Română de Statistică*, 2015 (8). Retrieved from [http://www.revistadestatistica.ro/supliment/wp-content/uploads/2015/09/RRSS\\_08\\_2015\\_A01\\_ro.pdf](http://www.revistadestatistica.ro/supliment/wp-content/uploads/2015/09/RRSS_08_2015_A01_ro.pdf).
- Nielsen, L. (2011). Classifications of Countries Based on Their Level of Development: How it is Done and How it Could be Done. IMF Working Paper, 11(31). Retrieved from <https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Classifications-of-Countries-Based-on-their-Level-of-Development-How-it-is-Done-and-How-it-24628>.
- Nielsen, L. (2013). How to Classify Countries Based on Their Level of Development. *Social Indicators Research*, 114 (3), 1087-1107. Retrieved from <https://www.jstor.org/stable/24720297>.
- OECD (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. *OECD Statistics Working Papers from OECD Publishing*. Retrieved from <https://www.oecd.org/sdd/42495745.pdf>
- OECD (2018), Beyond GDP: Measuring What Counts for Economic and Social Performance, OECD Publishing, Paris. Retrieved from <https://doi.org/10.1787/9789264307292-8-en>.
- Scutaru, L. (2013). Economic development versus sustainable development, *Ecoforum*, 2(1), 35-40. Retrieved from [http://scholar.google.ro/scholar\\_url?url=http://ecoforumjournal.ro/index.php/ecoforum/article/download/19/17&hl=ro&sa=X&scisig=AAGBfm0bQrc8KyFnFSYmPrS-9bFwxkFYNW&nossl=1&oi=scholar](http://scholar.google.ro/scholar_url?url=http://ecoforumjournal.ro/index.php/ecoforum/article/download/19/17&hl=ro&sa=X&scisig=AAGBfm0bQrc8KyFnFSYmPrS-9bFwxkFYNW&nossl=1&oi=scholar).
- Vaggi, G. (2017). The rich and the poor: A note on countries' classification. *PSL Quarterly Review*, 70 (279), 59-82. Retrieved from [https://doi.org/10.13133/2037-3643\\_70.280\\_4](https://doi.org/10.13133/2037-3643_70.280_4).
- Vazquez, S.T., & Sumner, A. (2012). Beyond low and middle income countries: What if there were five clusters of developing countries? *Poverty and inequality research cluster*, 2012 (404). Retrieved from <https://www.ids.ac.uk/files/dmfile/Wp404.pdf>.
- Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: how to use principal components analysis. *Health Policy and Planning*, 21 (6), 459-68. Retrieved from <https://academic.oup.com/heapol/article/21/6/459/612115>.

## Appendix. The description of the variables

Indicator Name	Long definition
<b>Inflation, consumer prices (annual %)</b>	Inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly.
<b>Unemployment, total (% of total labor force)</b>	Unemployment refers to the share of the labor force that is without work but available for and seeking employment.
<b>GINI index (World Bank estimate)</b>	Gini index measures the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution. A Lorenz curve plots the cumulative percentages of total income received against the cumulative number of recipients, starting with the poorest individual or household. The Gini index measures the area between the Lorenz curve and a hypothetical line of absolute equality, expressed as a percentage of the maximum area under the line. Thus, a Gini index of 0 represents perfect equality, while an index of 100 implies perfect inequality.
<b>Life expectancy at birth, total (years)</b>	Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.
<b>Human capital index (HCI) (scale 0-1)</b>	The HCI calculates the contributions of health and education to worker productivity. The final index score ranges from zero to one and measures the productivity as a future worker of child born today relative to the benchmark of full health and complete education.
<b>Access to electricity (% of population)</b>	Access to electricity is the percentage of population with access to electricity. Electrification data are collected from industry, national surveys and international sources.
<b>Cause of death, by communicable diseases and maternal, prenatal and nutrition conditions (% of total)</b>	Cause of death refers to the share of all deaths for all ages by underlying causes. Communicable diseases and maternal, prenatal and nutrition conditions include infectious and parasitic diseases, respiratory infections, and nutritional deficiencies such as underweight and stunting.
<b>People using at least basic drinking water services (% of population)</b>	The percentage of people using at least basic water services. This indicator encompasses both people using basic water services as well as those using safely managed water services. Basic drinking water services is defined as drinking water from an improved source, provided collection time is not more than 30 minutes for a round trip. Improved water sources include piped water, boreholes or tube wells, protected dug wells, protected springs, and packaged or delivered water.
<b>Cost of business start-up procedures (% of GNI per capita)</b>	Cost to register a business is normalized by presenting it as a percentage of gross national income (GNI) per capita.
<b>Time required to start a business (days)</b>	Time required to start a business is the number of calendar days needed to complete the procedures to legally operate a business. If a procedure can be speeded up at additional cost, the fastest procedure, independent of cost, is chosen.
<b>GDP per capita, PPP (current international \$)</b>	GDP per capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the U.S. dollar has in the United States. GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current international dollars based on the 2011 ICP round.
<b>HCI: Learning-Adjusted Years Of School</b>	Learning-Adjusted Years of School are calculated by multiplying the estimates of Expected Years of School by the ratio of most recent Harmonized Test Score to 625, where 625 corresponds to advancement attainment on the TIMSS (Trends in International Mathematics and Science Study) test.