BUILDING THE KNOWLEDGE ECONOMIES: PRINCIPAL COMPONENT ANALYSIS AND CLUSTERING OF THE EU COUNTRIES

Adela Anca FUCEC[1]

ABSTRACT
This paper shows the evolution of the European countries towards building knowledge economies, acording to the Europe 2020 Strategy. Assuming that the 8 indicators from the 5 objectives of the European Strategy have different national levels, they can accordingly show the state of development of each country towards the attainment of the main targets. The research methodology entails a Principal Component Analysis, further used as a starting point for a Cluster Analysis. The main results show three groups of European countries, acording to their stage of becoming knowledge economies. The results can also be used in order to show where each country stands and what its strenghts and weaknesses are from the knowledge economy point of view. This study is also relevant for anyone interested in a professional picture of how the main countries in Europe look like nowadays from the new economic perspective.

KEYWORDS: *Cluster Analysis, Europe 2020 Strategy, European Union, Knowledge Economy, Principal Component Analysis.*

JEL CLASSIFICATION: *C82, O11, 052.*

1. INTRODUCTION

Europe 2020 is the strategy of the European Union for the assurance of a sustainable economical future for the countries, by means of " a growth that is: smart, through more effective investments in education, research and innovation; sustainable, thanks to a decisive move towards a low-carbon economy; and inclusive, with a strong emphasis on job creation and poverty reduction" (European Commission, 2010). The objectives of the Strategy are well known and each country takes different steps and measures in order to be able to attain the objectives, through the national targets which are established for each country (European Commission, 2010). However, as far as we know, there is no way of knowing which country is close and how close to the status of being a knowledge based economy, there is no way of having a hierarchy or compare the progress that the countries are making in this dirrection. Previous studies were made in order to measure such a progress, both at national (Fucec and Marinescu, 2013) or microeconomical (Fucec, 2012; Fucec and Marinescu, 2013, Ceptureanu et. all, 2012) level, but things are evolving from year to year and we need to know where do we stand at the moment.
One other purpose of this research is to help give us a clear picture of where each country individually stands. The European Strategy has 5 objectives and 8 target indicators (European Commission, 2012). An analysis on 8 levels would be very comprehensive and complex, this is why, by means of such a research like the one we are presenting, we can see the position and the evolving state of each country in a much more clearer way, by using maximum 2 indicators.
Before briefing the research methodology we approached, we present a short conceptual frame for the knowledge based economy. The concept has been evolving since the last century and nowadays

[1]Bucharest University of Economic Studies, Romania, fucec.adela@yahoo.com

the researchers acknowledge that we are facing the third generation of knowledge management in private companies (O'Dell and Hubert, 2011). Though it is common that the private sector evolves at a faster pace than the public sector, since knowledge management is already a must in the private companies (Bate and Robert, 2002), many steps have already been made at national at international levels in order to seen and show the importance of knowledge as the key resource of this century (Geisler and Wickramasinghe, 2009).

For this research, the basic data that we used was collected from the European Commission's website for 27 EU member states, for the year 2012. The data refers to the values which each of the countries submitted to the analysis registered at the 8 indicators of the Europe 2020 strategy (Eurostat, 2012):

- Employment Rate (EmplR): expresses the employment rate in each country (%);
- Gross Expenditure on Research and Development (GERD): represents the percentage of GDP spent on R&D (%);
- Greenhouse Gas Emissions (GrGE): expressed correlated to the value from the year 1990, considered to have the value 100;
- Renewable Energy (RenEn): gives the share of renewable energy in the gross final energy cnsumption (%);
- Primary oil consumption (TOE): a measure for the real energy comsumption, expressed in " tones of oil equivalent";
- Early Leavers from Education (ELvEd): percentage of population aged 18-24 leaving school early (%);
- Tertiary Education Attainment (TrEdA): percentage of population aged 30-34 with tertiary education (%);
- People at Risk of Poversty or Social Exclusion (PrP/SE): expressed as a percentage from the total population of the country (%).

2. RESEARCH METHODOLOGY

The research methodology which we approched is based on two sequences. First, we performed a Principal Component Analysis on the data collected (Ruxanda, 2001; Smith, 2002), in order to liberate our data from redundancy. We retain a number of maximum 2 indicators from this PCA, indicators which are informationally clean, with no redundant data among it, and then we proceed to the second sequence. The second sequence is a Cluster Analysis (Ruxanda, 2001), performed on the two principal components we found in the first phase of the analysis. The result of the Cluster Analysis will be groups of countries which are either similar or different from the point of view of this analysis, which is seeing which country stands where as for as building the knowledge economies is concerned.

3. MAIN RESULTS AND DISSCUSION

3.1. Principal Component Analysis

The first series of results comes from running a Principal Component Analysis (PCA) on the cases we defined (the countries) and the variables that characterize them (the Europe 2020 Strategy indicators). Two basic and relevant result come from here and they are (1) the eigenvalues of the initial variables and (2) the factor matrix and (3) the principal scores matrix. The result are presented and explained below, together with references to the principal scores matrix, as well.

3.1.1. The eigenvalues of the variables

The eigenvalues of the initial information, and also the variance (Ruxanda, 2001), show us how much information is still left in that indicator, after having ran the PCA. Practically, what we see in the first row of the tabel, for instance, is that, by using only one indicator (eigenvalue 1) we retrieve 88.70% of the information initially expressed by all the 8 indicators of the European Strategy. We consider this to be a great achivement, because we have 27 countries and 8 indicators to analyse, and now we have narrowed it to 27 countries and only 1 indicator, with an insignificant information loss of less than 12%. From this point on, a new research dirrection appears, as we can use this one indicator, nominate it and use it for a hierarchisation of the countries, visible in a space of only one dimension and not 8 dimensions, one for each variable, as it was before the PCA. Still, in order to be more rigurous with the present research, we retain one more principal component for our further analysis. As shown in Table 1, the first couple of eigenvalues have a cumulative procentage of variance of 97.44%, which entails a 2.56% information loss for a bidimensional space for the analysis, whish is much easier to comprehend and to express graphically.

Table 1. Eigenvalues of the original variables

Factor	Eigenvalue	% Total Variance	Cumulative Eigenvalue	Cumulative %
1.	21431.49	88.70220	21431.49	88.7022
2.	2110.92	8.73684	23542.41	**97.4390**
3.	434.84	1.79975	23977.25	99.2388
4.	84.28	0.34882	24061.53	99.5876
5.	68.52	0.28359	24130.05	99.8712
6.	19.70	0.08154	24149.75	99.9527
7.	11.11	0.04598	24160.86	99.9987
8.	0.31	0.00128	24161.17	100.0000

Source: the author, using the software tool Statistica 8

3.1.2. The Factor Matrix

The Factor Matrix, the second significant result for our analysis, serves the main purpose of giving hints as how to scientifically name the factors we decided above to retain for the study, in order to better grasp their meaning.

Table 2. The Factor Matrix

Variable	Factor 1 - " Relevance Factor "	Factor 2 - " Unemployment Rate"
EmplR	0.094451	*-0,907264*
GERD	0.433996	-0.721658
GrGE	*0.952710*	-0.011292
RenEn	-0.135286	-0.303082
TOE	*0.905832*	-0.140976
ELvEd	0.229838	0.667163
TrEdA	-0.224949	-0.311844
PrP/SE	0.882777	0.309088

Source: the author, using the software tool Statistica 8

Basically, this matrix shows the correlations between the variables and the principal components. The matrix shows that the firts principal component is strongly pozitively correlated with the GrGE indicator, but also with the TOE indicator. This says that the name for our factor should be generally related with greenhouse gas emissions and the increase of energy consumption. Research

in this domain (Pîrlogea and Cicea, 2012) shows that a high consumption of energy implies an economically strong, well developed country. Therefore, since we desire a high value for our indicator, a proper name to sugest it's meaning could be "Degree of development" or "Relevance Factor", showing us how relevant is a certain country from the point of view of its greenhouse gas emissions and its energy comsumption rates. The second principal component is strongly negatively correlated with the EmplR indicator, so it is obvious that the factor is indeed the "Unemployment Rate".

3.1.3. The Principal Scores Matrix

The Principal Scores Matrix is another result of the PCA which can lead to another approach of this research, approach that is briefly suggested here and pursuited in another paper, not yet published. Such an approch has been used before in several researches (Fucec, 2012) and involves creating an aggreggate indicator, based on the contributions of each of the two new factors. This leads to one aggreggate indicator, which entails the 97.44% of the information from the 8 initial variables. From this point on, we can show a hierarchy of the cases we analysed (the countries), we can easily draw new maps and graphically see how Europe looks like from the point of view of developing its knowledge economies.

3.2. Cluster Analysis

After having decided to retain the two new factors for analysis, the "Relevance Factor" and the "Unemployment Rate", we continue with the Cluster Analysis of these two principal components.

3.2.1. The Distance Matrix and the Amalgamation Schedule

The Distance Matrix is one of the first results of the Cluster Analysis and is the basis of the second result which we will be explaining, the Amalgamation Schedule. The Distance Matrix shows all the Manhattan distances between the countries submitted to the analysis and the Amalgamation schedule actually puts the distances in an ascending scale, giving us a clear picture of how similar is each country to the other ones, from the point of view of developing the knowledge economies. The Matrix is as large as the number of the countries involved, it has 27 lines and 27 columns, therefore it is hard to give it here in full, but in order to be able to proceed with the explanations we will present a part of the matrix in the following figure (Figure 1).
For instance, the Manhattan distance between Lithuania and Bulgaria is 11, and between Denmark and Germany it is 390. It is obvious, therefore, that the similarities between Lithuania and Bulgaria are much more numerous than the ones between Denmark and Germany. It is highly likely that Lithuania and Bulgaria are part of the same cluster, but the chances that Germany and Denmark are in the same group are rather small. Taking a glance at this matrix allows us to assume that there are countries which are a lot alike, but also countries with very different levels of constructed knowledge economy. In the matrix, the distance 0 only appears only where we have the same country on the same line or colomn. The next distance, the smallest one, is the one between Romania and Great Britain (distance of 0.868). This means that these two countries are very similar from the point of view of this cluster analysis, which is builing the knowledge economies, and the two countries will join together in the programme and form the first cluster, which also marks the beginning of the amalgamation schedule. Further on, the amalgamation schedule shows us how each of the countries came together with the country or grup which is the closest to it (Figure 2). The graphical result of this is the dendogram, which is explained further in the paper.

Figure 1. The Distance Matrix

Case No.	City-block (Manhattan) distances (Spreadsheet5 in Workbook1-ac)												
	Belgium	Bulgaria	Czech Republic	Denmark	Germany	Estonia	Ireland	Greece	Spain	France	Italy	Cyprus	Latvia
Belgium	0	64	27	49	341	80	14	72	41	69	491	101	72
Bulgaria	64	0	41	27	405	17	52	36	34	46	524	79	41
Czech Republic	27	41	0	31	364	58	13	49	14	46	518	75	49
Denmark	49	27	31	0	390	42	37	63	45	73	509	106	68
Germany	341	405	364	390	0	421	353	413	371	410	585	435	413
Estonia	80	17	58	42	421	0	68	22	51	32	540	64	27
Ireland	14	52	13	37	353	68	0	60	27	57	505	87	60
Greece	72	36	49	63	413	22	60	0	42	10	550	42	5
Spain	41	34	14	45	371	51	27	42	0	39	532	64	42
France	69	46	46	73	410	32	57	10	39	0	560	33	5
Italy	491	524	518	509	585	540	505	550	532	560	0	592	555
Cyprus	101	79	75	106	435	64	87	42	64	33	592	0	38
Latvia	72	41	49	68	413	27	60	5	42	5	555	38	0
Lithuania	53	11	30	32	394	28	41	31	23	41	519	74	36
Luxembourg	193	216	220	189	287	230	207	252	234	262	333	295	257
Hungary	357	420	383	406	450	437	371	429	397	426	134	458	429
Malta	309	332	336	305	403	347	323	368	350	378	262	411	373
Netherlands	105	82	78	109	441	68	91	46	70	36	596	6	41
Austria	97	74	74	101	437	60	84	38	67	28	588	5	33
Poland	91	63	69	91	433	49	79	27	62	22	577	15	22
Portugal	99	76	72	104	436	62	85	40	65	30	590	2	35
Romania	62	22	39	49	403	19	50	15	32	24	535	57	20

Source: the author, using the software tool Statistica 8

As seen in the picture and briefly explained above, Romania and Great Britain are the closest from the point of view of the analysis, because 0.686 is the smallest distance in the Distance Matrix. This means that they stand on similar positions from the point of view of accomplishing the objectives of the Europe 2020 Strategy, considering the two indicators we found above. The second smallest distance in the matrix is 2.238, the distance between Cyprus and Portugal. This means that the next step in the amalgamation schedule is the joining of these two countries into a new cluster. The next step is more interesting: the next smallest distance brings Austria into the cluster of Portugal and Cyprus. This means that they are the three countries with the most similarities from the point of view of evolving towards the knowledge economy.

Figure 2. The Amalgamation Schedule

linkage distance	Amalgamation Schedule (Spreadsheet5 in Workbook1-ac) Single Linkage City-block (Manhattan) distances												
	Obj. No. 1	Obj. No. 2	Obj. No. 3	Obj. No. 4	Obj. No. 5	Obj. No. 6	Obj. No. 7	Obj. No. 8	Obj. No. 9	Obj. No. 10	Obj. No. 11	Obj. No. 12	Obj. No. 13
,8680000	Romania	United Kingdom											
2,238800	Cyprus	Portugal											
2,485400	Cyprus	Portugal	Austria										
3,967900	Netherlands	Slovenia											
4,812900	Greece	Latvia											
4,944900	Greece	Latvia	France										
5,685800	Cyprus	Portugal	Austria	Netherlands	Slovenia								
10,54770	Cyprus	Portugal	Austria	Netherlands	Slovenia	Poland							
11,02070	Bulgaria	Lithuania											
12,82480	Czech Republic	Ireland											
13,83700	Greece	Latvia	France	Romania	United Kingdom								
13,94960	Belgium	Czech Republic	Ireland										
14,04360	Spain	Finland											
14,06420	Belgium	Czech Republic	Ireland	Spain	Finland								
16,41210	Bulgaria	Lithuania	Greece	Latvia	France	Romania	United Kingdom						
16,75730	Bulgaria	Lithuania	Greece	Latvia	France	Romania	United Kingdom	Estonia					
17,81880	Belgium	Czech Republic	Ireland	Spain	Finland	Bulgaria	Lithuania	Greece	Latvia	France	Romania	United Kingdom	Estonia
22,36880	Belgium	Czech Republic	Ireland	Spain	Finland	Bulgaria	Lithuania	Greece	Latvia	France	Romania	United Kingdom	Estonia
27,15000	Belgium	Czech Republic	Ireland	Spain	Finland	Bulgaria	Lithuania	Greece	Latvia	France	Romania	United Kingdom	Estonia
28,44350	Belgium	Czech Republic	Ireland	Spain	Finland	Bulgaria	Lithuania	Greece	Latvia	France	Romania	United Kingdom	Estonia
46,03440	Luxembourg	Sweden											
116,1386	Luxembourg	Sweden	Malta										

Source: the author, using the software tool Statistica 8

Further on, the Netherlands and Slovenia form a new cluster (because the distance among them is 3.967) and another new cluster is formed between Greece and Latvia (distanced by 4.812). France

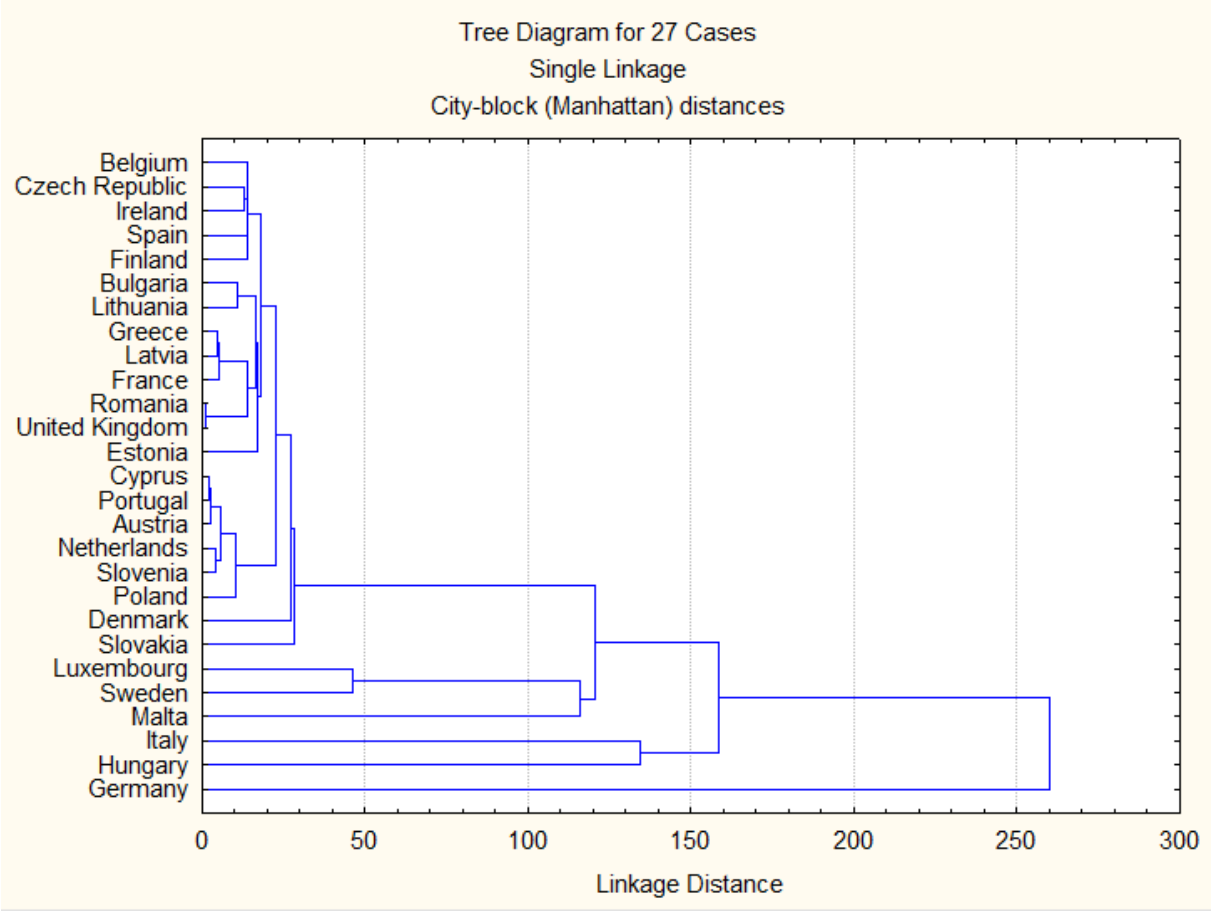
joins in the last cluster and then comes another interesting step in the amalgamation schedule: two former clusters join together. This is how the amalgamation schedule proceeds, by sideing together each small cluster to another cluster which is close (similar) to it. This way, from the initial 27 single clusters (each country was one cluster at the beginning), the countries join in clusters, based on their similarities and in the end we reach only one final cluster. With each step of the amalgamation schedule, the number of clusters decreases, because each country form or join a formerly created cluster, based on the minumum Manhattan distance. The next phase is to closely look at this amalagamation schedule and then separte the most different clusters, so as to have significant different groups of countries. This is where the dendogram comes in and is explined in the next paragraph.

3.2.2. The Horizontal hierarchical tree plot (dendogram)

By this point of the analysis, the Manhattan distances showed us how to group the countries along the amalgamation schedule. The horizonthal hierarchical plot is a picture of the amalgamation schedule and by looking at the picture above we can say which are the clusters that we are looking for. In Figure 3, we can observe 3 clusters, based on the accomplishment of the objectives of the European strategy:

- Cluster 1: Belgium, Czech Republic, Ireland, Spain, Finland, Bulgaria, Lithuania, Grecee, Latvia, France, Romania, Great Britain, Estonia;
- Cluster 2: Cyprus, Portugal, Austria, Netherlands, Slovenia, Poland, Denmark, Slovakia;
- Cluster 3: Luxembourg, Sweden, Malta, Italy, Hungary, Germany.

Figure 3. The Dendogram



Source: the author, using the software tool Statistica 8

Due to this specific working methodology, the clustering of these countries is in the eye of the researcher (Estivill-Castro, 2002), just like a business opportunity exists only in the eye of the entrepreneur (Nicolescu and Nicolescu, 2008). We could have, for example, chosen to separate Germany from any other cluster, because of the big distance between it and all the other countries, but this would not have been relevant for our research. We could also have chosen to join two of the clusters and only have two final clusters for the analysis, as a second option. Still, for the purpose of the analysis, we consider it more relevant to retain three clusters, in order to be able to illustrate the traits of these groups.

4. CONCLUSIONS

Using the research methodology which we presented here gave us the following important results: first, we found two principal components (the degree of development and the unemployment rate) which are nonredundant and which hold a percentage of 97.44% of the information from the 8 initial target indicators of the Europe 2020 Strategy. This gives us hints on which indicator is more important and shows each country, according to the values that they have for the indicators, how to emphasis their development so that they are both efficient and effective. Secondly, we used this "clean" indicators and divided the countries in 3 groups, according to how their development of knowledge economies is going. By knowing in which group you stand as a country, it becomes easy to share best practices with the other countries in your group, to compare situations and find practical and applicable solutions from the countries in more evolved groups.
Of course, further research options are open, so as to find the specificity of each cluster, by deepening the analysis of the development rate of each country and also the unemployment rate. The main flaw of this research is that it analysis the previous situation, the current one and gives recommandations for the following years, but the changes in the countries are constant and appear every year, this is why such a research must be performed annually so that the solutions can always be ajusted to the local realities of each country.

REFERENCES

Europe 2020 in a nutshell – Priorities. In European Commission (2010). Retrieved August 24, 2014, from http://ec.europa.eu/europe2020/europe-2020-in-a-nutshell/priorities/index_en.htm

Europe 2020 in a nutshell – Targets. In European Commission (2010). Retrieved August 24, 2014, from http://ec.europa.eu/europe2020/europe-2020-in-a-nutshell/targets/index_en.htm

Fucec, A. A. & Marinescu (Pîrlogea), C. (2013). Cluster Analysis of the European Countries: the Europe 2020 Point of View. *Proceedings of the 10th International Conference on Intellectual Capital, Knowledge Management & Organisational Learning (ICICKM 2013), 2*, 507.

Fucec, A. A. (2012). Is Romania a favourable environment for the development of knowledge-based organizations?. *Review of International Comparative Management*, 13 (5), 768 – 777.

Fucec, A. A. & Marinescu (Pîrlogea) C. (2013). Knowledge economies in the European Union: Romania's position**.** *Proceedings of the Emerging Markets Queries in Finance and Business International Conference*.

Ceptureanu, S. I., Ceptureanu, E. G., Tudorache, A. &Zgubea, A. (2012). Economic Survey on Knowledge Based Management in Romanian Companies. *Review of International Comparative Management, 13, 2,* 325-336.

Europe 2020 indicators – Headline Indicators. In Eurostat (2010). Retrieved August 24, 2014, from http://epp.eurostat.ec.europa.eu/portal/page/portal/europe_2020_indicators/headline_indicators

O'Dell, C. & Hubert, C. (2011). *The new Edge in Knowledge: How Knowledge Management is Changing the Way we do Business*. USA: John Wiley & Sons, Inc.

Bate, S.P. & Robert, G. (2002). Knowledge Management and communities of practice in the private sector: lessons for modernising the National Health Service in England and Wales, Public Administration, 80(4), 643-663.

Geisler, E. & Wickramasinghe, N. (2009). *Principles of Knowledge Management: Theory, Practice and Cases.* New York: M.E.Sharpe.

Ruxanda, G. (2001). *Analiza Datelor.* Bucharest: Editura ASE.

Smith, L. (2002). *A tutorial on Principal Componets Analysis.*, p. 12.

Pîrlogea, C. & Cicea, C. (2012). Econometric perspective of the energy consumption and economic growth relation in European Union. *Renewable and Sustainable Energy Reviews,* 16 (8), 5718 – 5726.

Estivill-Castro, V. (2002). Why so many clustering algorithms: a position paper. *ACM SIGKDD Explorations Newsletter,* 4 (1), 65-75.

Nicolescu, O. & Nicolescu, C. (2008). *Intreprenoriatul si managementul inteprinderilor mici si mijlocii.* Bucharest: Editura Economica.