

31

Working Paper



Silesian University in Opava
School of Business Administration in Karvina
Institute of Interdisciplinary



**Working Papers in
Interdisciplinary Economics
and Business Research**

Cluster Analysis in Human Development of Visegrad
Group Plus Countries

Ingrid Majerová, Jan Nevima

July 2016

Working Papers in Interdisciplinary Economics and Business Research

Silesian University in Opava
School of Business Administration in Karviná
Institute of Interdisciplinary Research
Univerzitní nám. 1934/3
733 40 Karviná
Czech Republic
<http://www.iivopf.cz/>
email: iiv@opf.slu.cz
+420 596 398 237

Citation

MAJEROVÁ, I. and J. NEVIMA, 2016. Cluster Analysis in Human Development of Visegrad Group Plus Countries. *Working Paper in Interdisciplinary Economics and Business Research no. 31*. Silesian University in Opava, School of Business Administration in Karviná.

Abstract

Ingrid Majerová, Jan Nevima: **Cluster Analysis in Human Development of Visegrad Group Plus Countries**

Human Development Index measures the level of human development not only by GDP per capita, but also through the indicators of education and healthy life. The differences in the level of human development are observed not only at the national level, they are also measured at the regional level of countries. The aim of this article is, with the quantification of regional human development, to describe the potential for human development using cluster analysis at the regional level. The 46 regions of the Visegrad Group Plus countries (Czech Republic, Poland, Hungary, Slovakia, Austria and Slovenia) at NUTS II level were selected for this purpose. The research was made in the period from 2004 to 2013. In the selection of indicators, the same approaches as by the HDI were adopted, however the components of each dimension were modified. Life expectancy at birth (dimension of health), tertiary educated people and lifelong learning (dimension of education) and GDP per capita in PPS (dimension of living standards) have been chosen as indicators of human development. These components were then used in a hierarchy cluster analysis in the Ward method. Three clusters were created with different levels of development potential. Initially, a research hypothesis that there was a dynamization of human development processes in most regions has been set. This hypothesis was not confirmed and it was found that a vast majority of the regions have not changed their positions in the cluster in the monitored period.

Key words

cluster analysis, GDP per capita, human development, life expectancy at birth, lifelong education, NUTS II regions, tertiary education, Visegrad Group Plus.

JEL: O15, I31

Contacts

Ingrid Majerová, Department of Economics and Public Administration, School of Business Administration, Silesian University, Univerzitní nám. 1934/3, 733 40 Karviná, Czechia, e-mail: majerova@opf.slu.cz.

Jan Nevima, Department of Economics and Public Administration, School of Business Administration, Silesian University, Univerzitní nám. 1934/3, 733 40 Karviná, Czechia, e-mail: nevima@opf.slu.cz.

Acknowledgement

Publication of this paper was supported by the students' grant project "Influence of Selected Macroeconomic and Microeconomic Determinants on the Competitiveness of Regions and Firms in Countries of the Visegrad Group Plus". Project registration number is SGS/13/2015. The support is gratefully acknowledged.

Introduction

The most widely used indicator for measuring the state of economy is the GDP (Stiglitz, Sen and Fitoussi, 2009; Van den Bergh, 2009), although it does not include social, political, cultural and environmental aspects of development. Many alternatives can be applied for measurement of socio-economic development, the best-known and most often used is an index called the Human Development Index (HDI) as Todaro and Smith (2011) demonstrated. This index that has been used by the United Nations since 1990, brings a different perspective on development issues. It should be better able to emphasize the effect of other than just monetary (economic) factors of economy of a country.

Basu and Basu (2005) consider HDIs as primarily nation level indicators, estimated for the country as a whole. The constructions of the HDI do not express the differences in regions of countries. However, the regional disparities exist and influence the regional development. Based on it, we decided to analyze the issue of human development for a group of countries of the Visegrad Group Plus (hereafter V4+) at the NUTS II level. This group includes the Visegrad Group countries (Czech Republic, Hungary, Poland and Slovakia), Slovenia and Austria, which were included to this group on the ground of the Regional Partnership Agreement from 2001. There are 46 regions on the NUTS II level – eight in the Czech Republic, seven in Hungary, sixteen in Poland, nine in Austria, four in Slovakia and two in Slovenia. The research was made in the period from 2004 (the beginning of membership of most countries V4+) to 2013 (last available data). In the selection of indicators, the same approach as by the HDI was adopted, however the components of each dimension were modified. Life expectancy at birth (dimension of health), tertiary educated people and lifelong learning (dimension of education) and GDP per capita in PPS (dimension of living standards) have been chosen as indicators of human development at regional level. These components were than used in a hierarchy cluster analysis in the Ward method.

Three created clusters include a plurality of regions based on their inner similarities that would not otherwise be apparent at first glance. These clusters include the regions with different levels of development potential – first cluster with above-average potential for human development, second cluster with average potential for human development and third cluster with under-average potential for human development. Initially, a research hypothesis that there was a dynamization of human development processes in most regions has been set. Namely, it was assumed that more than half of monitored regions in the lower group of potential to human development will shift to the higher group. This hypothesis was not confirmed and it was found that the vast majority of the regions have not changed their positions in the cluster in the monitored period.

1. Modified Human Development Index

The beginning of the Human Development Index dates back to 1990 when the UN Development Programme (UNDP) published the first report on human development (Human Development Report) which established the need of human development measurement. Human development has two forms, which should be in balance, the formation of human capabilities in terms of improving health, increasing knowledge and skills to meet human need and their own skills and competences, free time, job security, cultural, social and political events. Basically, human development is clearly and directly dependent on income. It is

therefore necessary to examine other variables that point out the potential of a country much better as well as the options currently appear in human development (Majerova, 2012).

1.1. Data of Human Development Index

The Human Development Index (HDI) is a summary measure of achievements in key dimensions of human development: a long and healthy life, an access to knowledge and a decent standard of living as has been shown in UNDP (2015).

These three dimensions have four parts - health and standard of living has one part each and education has two parts:

- health dimension – life expectancy at birth (interval 20-85 years)
- education dimension – expected year of schooling (0-18 years) and mean years of schooling (0-15 years)
- standard of living dimension – GNI per capita in USD/PPP (100-75.000).

Because of the need to improve their explanatory power, the calculation method of two of three dimensions (health indicator index is the only one which has remained unchanged) has changed over time, last change was made in 2010. The literacy rate of population has been replaced by an indicator of expected years of schooling, the combined gross enrolment by the mean number of years of education (knowledge dimension). The dimensions of living standards are now measured by GNI per capita in purchasing power parity to the USD. The main change was to switch from the original additive aggregation function (the arithmetic mean of the three components) to a multiplicative function (their geometric mean) as Ravallion (2012) claimed and is shown in Equation (1).

$$HDI = \sqrt[3]{I_{LE}^n \cdot I_E^n \cdot I_{GDP}^n} \quad (1)$$

HDI index calculation required the values in the range from 0 (the lowest level of human development) to 1 (the highest human development), and therefore they were determined for each dimension of the minimum and maximum values (more in Anand and Sen, 1994) based on historical evidence.

2. NUTS Human Development Index

For the purpose of the paper, we adopted the same principle of HDI creating for the national level – the health dimension, knowledge dimension and dimension of a living standard. Components of each dimension, however, had to be modified because of the lack of data at the regional level (NUTS II level). Data were used from a regional database of Eurostat and converted to the number of inhabitants representing the given group.

The construction of the HDI of V4 regions (NHDI) was as follows:

- Health with the value of life expectancy at birth that represents, according to Eurostat, the mean number of years that a newborn child can expect to live if subjected throughout his life, to the current mortality conditions (age specific probabilities of dying).

- Knowledge, which includes two components:
 - Tertiary educated people in the age of 25-64, where the indicator is defined as a percentage of population aged 25-64 who have successfully completed tertiary studies (e.g. university, higher technical institution, etc.). This educational attainment refers to ISCED (International Standard Classification of Education) 1997 level 5-6, that includes the first stage of tertiary education (bachelor and master or equivalent) and second stage of tertiary education (doctoral or equivalent).
 - Lifelong learning in the form of participation rate in education and training covers participation in formal and non-formal education and training. The reference period for the participation in education and training is at least four weeks. Participation rates in education and training for the age group of 25-64 are presented. The data are calculated as annual averages of quarterly EU Labour Force Survey data (EU-LFS).
- Standard of living, measured through GDP per capita in PPS – Purchasing Power Standards (PPS), is a common currency that eliminates the differences in price levels between countries and regions allowing meaningful volume comparisons of GDP between them.

These indicators were chosen for their greatest explanatory power in relation to human development. The life expectancy at birth reflects the level of health and quality of life and measures the qualitative aspects of living a healthy life. It correlates positively with human development – the higher the healthy life expectancy of region, the more developed it is.

The share of tertiary educated people in productive age on the population in this age group is connected with the ability of people to reflect the needs of knowledge of economy and to contribute to it and human development. Lifelong learning, in the form of participation in education and training, encompasses all learning activities undertaken throughout life (after the end of initial education) with the aim of improving knowledge, skills and competences, within personal, civic, social or employment-related perspectives as Eurostat (2015) demonstrates. Due to lifelong learning people extend their possibilities for increasing their incomes. As a dimension of health, both indicators of education are positively correlated with human development.

The last but not least dimension is the GDP per capita. The implementation of this indicator was influenced by the opinion of Sen (1999) who considered the income (product) as a primarily mean to achieve human development. The GDP per capita reflects the economic level better than its absolute value. The indicator is measured by an artificial European currency unit, the purchasing power standard (PPS).

3. Methodology of Cluster Analysis

Cluster analysis is primarily focused on the search for similarities or differences between examined objects. Cluster analysis provides one, empirically based, means for explicitly classifying objects (Punj and Stewart, 1983, p. 134). If the research object is the region, as in our case, it is clear that only by applications of cluster analysis we can confirm our assumption about the most or the least developed regions in the area of human development and its modifications.

Blashfield and Aldenderfer (1988) consider that cluster analysis method has a long history – the earliest known procedures were suggested by anthropologists, later these ideas were picked up in psychology. Clustering analysis became one of the qualifying methods in the 20th century, the usefulness of which immediately had an impact on practically all fields of science. The first comprehensive work dealing with cluster analysis was created in 1939 by Tryon (1939). The main motivation for the use of clustering is uncovering of hidden similarities or differences. For this reason, a cluster analysis is now widely used by all scientific disciplines (for us is most interesting use in the field of economy, see e.g. Halásková and Halásková (2015)).

The basis of cluster analysis is sorting (for details see Meloun, 1994), of which we appoint two basic approaches. The first is called hierarchy cluster method and is most widely used by software. It is based on the use of once formed clusters. Thus formed clusters are then used to create other clusters from the rest of the data file. This manner is then preceded until all elements of the data file are a part of the cluster. This procedure has been chosen for our analysis of the regional level of human development in the V4+ countries.

The second method is a non-hierarchy cluster approach, which is based on cluster search on the principle of the smallest difference from the average. The procedure, however, is advantageous only if the number of clusters we want to achieve, is determined beforehand. However, this may become a significant limitation in a further research, because only such a number of clusters are finally formed, which we determined beforehand and for example, some extreme values may merge with average ones (K-means).

The selection of cluster methods is necessary after determining the clustering process. There are seven methods (Caliński and Harabasz, 1974). The first two methods are based on the Between Groups Linkage or Within Groups Linkage. Their use depends on good knowledge of the data file and information about the number of clusters that we want to achieve. In the case of ignorance of the total number of clusters we want to achieve, both methods are limitations in further research. The third method, Nearest Neighbour is based on the shortest distance between clusters. The fourth method, Furthest Neighbour method, searches the values in the data file that are furthest apart. The fifth method, Centroid Clustering method, may look at first glance like the most ideal. It is based on the Euclidean distance between the centroids of clusters. The closest are those clusters which have the smallest distance between the centroids. But it does not solve the differences that may occur due to different weights for equally large clusters. The sixth method - Median clustering - solves the problem of weights variance that the previous method gives to differently large clusters.

The last method, the Ward method, focuses on the allocation of profiles to groups equally. The principle of the method is not optimization, but minimization of heterogeneity. The purpose is to find the greatest similarity. In measuring the human development and its modifications, it is necessary to look for similarities among the 46 regions using this method.

3.1. Application of Cluster Analysis in NUTS II Regions of V4+

Based on the methodology described in the previous section of this paper, cluster analysis will now be practically applied on the regions at NUTS II level of the Visegrad Group Plus countries. The V4+ regions will be divided according to their development potential in terms of human development. As already mentioned, the hierarchy cluster approach by means of Ward's method was used for the classification of regions, and all performed calculations were performed by using SPSS software.

The subjects of cluster analysis are 46 NUTS II regions that have been evaluated by following metrics:

- cluster 1 is the group of regions with above-average potential for development in terms of human development and its input parameters;
- cluster 2 indicates the group of regions with average development potential in terms of human development and its input parameters;
- cluster 3 indicates the group of regions with below-average development potential in terms of human development indicators and its input.

Table 1 shows how various inputs influencing the final value of human development during the reporting period have changed. Some of the clusters remained unchanged throughout the monitored period, on the contrary some of them evolved over time. From this table, we derive whether developments in the regions when analysing the input variables are rather constant or whether the processes lead to dynamization in regions.

Tab. 1: Created and changed clusters of V4 + regions taking into account the four inputs between the years 2004-2013

NUTS 2	Cluster (period)	NUTS 2	Cluster (period)
SK01	cluster 1	PL51	cluster 3
SK02	cluster 2	PL52	cluster 2
SK03	cluster 2	PL61	cluster 2
SK04	cluster 2	PL62	cluster 2
CZ01	cluster 1	PL63	2004 – 2009 cluster 2 2010 – 2013 cluster 3
CZ02	2004-2010 cluster 2 2011-2013 cluster 1	HU10	cluster 1
CZ03	2004-2010 cluster 2 2011-2013 cluster 1	HU21	cluster 2
CZ04	cluster 2	HU22	cluster 2
CZ05	2004-2010 cluster 2 2011-2013 cluster 1	HU23	cluster 2
CZ06	2004-2007 cluster 2 2008-2013 cluster 1	HU31	cluster 2
CZ07	cluster 2	HU32	cluster 2
CZ08	cluster 2	HU33	cluster 2
PL11	cluster 3	AT11	cluster 1
PL12	cluster 1	AT12	cluster 1
PL21	cluster 3	AT13	cluster 1
PL22	cluster 3	AT21	cluster 1
PL31	cluster 3	AT22	cluster 1
PL32	cluster 2	AT31	cluster 1
PL33	cluster 2	AT32	cluster 1
PL34	cluster 2	AT33	cluster 1
PL41	cluster 1	AT34	cluster 1
PL42	cluster 2	SI01	2004-2007 cluster 2 2008-2013 cluster 1
PL43	cluster 2	SI02	cluster 1

Note: the change in the development of the region in the years 2004-2013 is indicated

Source: own creation

Conclusion

The Human Development Index is one of the indicators which can measure the socioeconomic development. This indicator has been used since 1990, it measures the above mentioned development at the national level and it is used to compare differences between economies. However, there are not only disparities between economies, but also within them. For this reason we decided to construct the modified Human Development Index (NHDI) and for this purpose countries of the Visegrad Group Plus at the NUTS II level have been selected and there are the 46 NUTS II regions. For our purpose the data had to be modified, but methodology of the NHDI was the same as for the HDI. We used three components - the health dimension (life expectancy at birth), the knowledge dimension (tertiary educated people and participation rate in education and training) and the dimension of living standard (GDP per capita).

These components were then used in hierarchy cluster analysis in the Ward's method in the period from 2004 to 2013. The three clusters were created. These clusters included a plurality of regions based on their inner similarities that would not otherwise be apparent at first glance. At the beginning of the monitored period, the situation in various regions was as follows: Regions of Austria were very homogeneous and placed in a group with above-average potential for development. Hungarian regions (except a region in which the capital is) were in the second group – with average development potential. Czech regions were placed in the first two groups and Polish regions exhibited the lowest homogeneity and were placed in all groups (most in group 2). There was initially set the research hypothesis that more than half of monitored regions in the lower group of potential to human development will shift to the higher group.

Finally, it is also necessary to note that in some regions the time offset between the individual clusters is quite obvious. As a general rule, it was a situation where regions have shifted from average to above-average potential towards of development, i.e. from cluster 2 to cluster 1. It was the case of regions in the Czech Republic – Střední Čechy (CZ02), Jihozápad (CZ03), Severovýchod (CZ05), Jihovýchod (CZ06) and Slovenia – Vzhodna Slovenia (SI01). There was, however, the reverse process, which led to some slowdown in the development potential of the region. The shift from the group of average to below-average happened only in one region, the region of Poland – Pomorskie (PL63). However, the vast majority (exactly forty of forty-six) of the examined regions did not change their position in the cluster during the reporting period. Our hypothesis about dynamization of most regions was not confirmed.

References

- [1] Anand, S. and A. K. Sen, 1994. *Human development index: methodology and measurement*. Occasional Paper 12. New York: Human Development Report Office.
- [2] Basu, S. and P. Basu, 2005. Regional Disparity in Australia: Analysis of Gender Development Index. *International Review of Business Research Paper*, vol. 1, No. 2, pp. 56-66.
- [3] Blashfield, R. K. and M. S. Aldenderfer, 1988. The methods and Problems of Cluster Analysis. In J. R. Nesselrode et al. (eds.), *Handbook of Multivariate Experimental Psychology*, pp. 447-474. New York: Plenum Press.
- [4] Caliński, T. and J. Harabasz, 1974. A dendrite method for cluster analysis. *Communications in Statistics*, vol. 3, No. 1, pp. 1-27.

- [5] Eurostat, 2015. *Regional Database*. [online]. [2016-07-15]. Available from: <http://ec.europa.eu/eurostat/web/regions/data/database>.
- [6] Halásková, M. and R. Halásková, 2015. The Structure and Financial Dimensions of Public Administration in EU Countries. *Transylvanian Review of Administrative Sciences*, vol. 45E, pp. 41-57.
- [7] Majerova, I., 2012. Least Developed Countries in Indexes of Human Development and Poverty. In Soliman, K. (ed.) *Proceedings of the 18th IBIMA Conference Innovation and Sustainable Economic Competitive Advantage: From Regional Development to World Economies*, pp. 1210-1224. Norristown: IBIMA.
- [8] Meloun, M., 1994. *Statistické zpracování experimentálních dat v chemometrii, biometrii, ekonometrii a dalších oborech přírodních, technických a společenských věd*. Praha: Plus.
- [9] Punj, G. and D. W. Stewart, 1983. Cluster analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research*, vol. XX, pp. 134-148.
- [10] Ravallion, M., 2012. Troubling tradeoffs in the Human Development Index. *Journal of Development Economics*, vol. 99, pp. 201-209.
- [11] Sen, A., 1999. *Development as Freedom*. Oxford: Oxford University Publishing.
- [12] Stiglitz, J. E., Sen, A. and J. P. Fitoussi, 2009. *Report by the commission on the measurement of economics performance and social progress*. Commission on the Measurement of Economic Performance and Social Progress.
- [13] Todaro, M. P. and S. C. Smith, 2011. *Economic Development*. Essex: Pearson Education.
- [14] Tryon, C. R., 1939. *Cluster analysis*. Ann Arbor: Edwards Brothers.
- [15] UNDP, 2015. *Human Development Report. Work for Human Development*. New York: UNDP.
- [16] Van den Bergh, J. C. J. M., 2009. The GDP paradox. *Journal of Economic Psychology*, vol. 30, pp. 117-135.