CROSS-COUNTRY COMPARISONS AND DIFFERENTIATION OF COUNTRIES BY KEY INDICATORS OF THE SOCIO-DEMOGRAPHIC SITUATION

Natalia Sadovnikova, Plekhanov Russian University of Economics
Olga Lebedinskaya, Plekhanov Russian University of Economics
Alexander Bezrukov, Plekhanov Russian University of Economics
Leysan Davletshina, Plekhanov Russian University of Economics

ABSTRACT

One of the most important goals of harmonious and sustainable development of the world within the framework of the UN program "Transforming our World: the 2030 Agenda for Sustainable Development" is to reduce cross-country income inequality, promote social and economic integration (Goal 10). The diversity of concepts for measuring poverty and economic inequality has made it difficult to compare methodologically rigorous countries that use absolute (Russia) and relative (European countries) monetary concepts of poverty. Given that economic inequality is a multidimensional concept and includes, in addition to unequal income, unequal access to health care, education, housing, the labor market, and even inequality before the law, the authors suggested that a comprehensive assessment of the differentiation of countries can be obtained using cluster analysis. The results of clustering of 36 member countries of Organisation for Economic Co-operation and Development and Russia on 16 key indicators combined into four blocks (socio-demographic situation, demographic characteristics of the population, labor force, and standard of living), presented by the authors, allowed identifying enlarged groups of countries that characterize cross-country inequality. To ensure the comparability of the data, the k-medoid algorithm was used to minimize the total degree of differences between objects belonging to a particular cluster. It was found in the course of the study that there is a close relationship between the parameters of the socio-demographic situation and the territorial neighborhood, the general history of development.

Keywords: International Comparisons, Sustainable Development Goals, Inequality.

INTRODUCTION

The lack of a common methodology for measuring cross-country inequality and, as a result, the possibility of borrowing the experience of its assessment by non-EU countries is a *"challenge"* for statistical and analytical services of all countries. We believe that the use of the k-medoid algorithm will solve the problem of data incompatibility that arises due to changes in measurement and counting units. The use of the k-medoid technique reduces the effect of cross-country asymmetry in the socio-economic indicators of individual countries, which gives more robust clusters, and, consequently, more stable characteristics for their modeling and the formation of a strategy to overcome the differentiation of countries.

The article presents a structured picture of the cross-country inequality of the OECD member countries+Russia – clustering of countries by four groups of indicators: socio-

demographic situation, demographic situation, labor force, and standard of living, which allows assessing the level of objective well-being of the population in the OECD countries and Russia in 2018.

LITERATURE REVIEW

The assessment of the degree of differentiation of countries by key indicators of the socio-demographic situation makes it possible to make changes to the UN recommendations on the achievement of the SDGs by individual countries.

Many international organizations publish poverty statistics, including the World Bank, OECD, UNDP, and Eurostat. The failures of numerous attempts are mainly due to two reasons:

1) National data of countries are primarily focused on domestic needs and do not always meet accepted international standards;
2) The quality of country data does not meet international standards.

The main recommendations for applying different approaches to measuring poverty at the national level and improving the international comparability of poverty statistics are presented in the Guide to Poverty Measurement (UNECE, 2017). Many scholars talk about the need for cross-country comparisons. Alkire & Apablaza (2016) propose using a multidimensional poverty index based on the Alkire Foster (AF) methodology. Grigoryev & Pavlyushina (2018) propose to use data based on the World Bank's estimates of GDP by PPP in 2011 prices to analyze the uneven development of countries; there are other methods as well (Jolliffe, 2014; Ferreira et al., 2015; Atkinson, 2007).

In international statistical practice, the most widely used estimates of monetary poverty are based on indicators of income or consumption expenditure (consumer spending), the conceptual definitions of which are given in the Canberra Group Manual on Household Statistics. There are other opinions: Alvaredo et al. (2016) believes that the poverty assessment can be obtained based on a comparison of the SNA data, and analysts of the European Central Bank (2020) specify that it is necessary to use distribution accounts. Bricker et al. (2016) suggests analyzing administrative tax reports and household surveys by the Federal Reserve System for Consumer Finance (SCF). The problem of data heterogeneity is also pointed out by Fesseau et al. (2013). Leulescu & Agafitei see the solution to the problem in the application of EU-SILC wage statistics as a benchmark.

The difficulty is also that most of our knowledge about the distribution of wealth comes from household surveys, for which there is no single methodology. The vast majority of research is carried out within either one country or a group of countries with similar development levels (Alvaredo & Piketty, 2014; Alvaredo & Londoño Vélez, 2013; Ziliak, 2010).

This situation makes it necessary to further search for ways to overcome the disparity of data when analyzing the differentiation of social inequality in countries to develop a strategy to overcome it in the context of globalization.

METHODS

When choosing research methods, we proceeded from the need to obtain answers to solve the main problem, which involves understanding the depth and scale of differentiation of countries by key indicators of socio-demographic conditions and identifying characteristic, homogeneous groups of territories. The method of cluster analysis was used to determine the

cross-country inequality in the key parameters of the socio-demographic development of the OECD countries and Russia.

The cluster analysis procedure belongs to the *"unsupervised"* machine learning methods. Within the framework of its application as a grouping method, it allows intelligently searching for groups of countries in such a way as to maximize the differences between country clusters. At the same time, differences between countries within clusters are minimized. The result of using clustering is obtaining homogeneous groups of countries, in which the most similar countries are combined into clusters, and separate countries are identified, for which the parameters used in clustering differ from the rest.

The application of cluster analysis, in this case, is based on the following prerequisites:

- Among the OECD countries, there are groups of countries with a similar socio-demographic situation while this division has an interpreted economic explanation. The number of these clusters is not known;
- Homogeneous groups of countries have different numbers and composition, the use of cluster analysis allows one to identify groups of similar countries while maximizing the difference between the clusters themselves;
- The use of conventional grouping methods does not allow considering the interaction between individual parameters of socio-demographic development, which determines certain differences between groups of countries. In addition, grouping methods require setting the target number of groups, while hierarchical clustering allows one to obtain the reference value of this number;
- A stable cluster solution obtained based on the v-fold cross-validation procedure allows for conclusions to be drawn about the actual number of cluster groups, since some socio-demographic parameters can be linearly inseparable. This makes it difficult to find the actual number of cluster groups, as shown below;
- The obtained cluster model allows obtaining additional information on changes in the disposition of clusters and their main parameters, classifying countries by clusters, and modeling and predicting changes in the structure and composition of countries according to the characteristics of the socio-demographic situation.

To solve the problem of determining the scale of clustering, the procedure for normalizing the initial variables was performed using the formula:

$$y'_i = \frac{\left| y_i - \bar{y} \right|}{\sigma_y},$$

Where y_i is the empirical values of the indicator;
$\bar{y}$ – the average value of the indicator;
σ_y – standard deviation of the indicator.

The assessment of inequality between the OECD countries and Russia on key parameters of socio-demographic development is based on the analysis of official statistical data presented in the open resources of the UN Statistical Commission (UNSD – https://unstats.un.org/home), the Statistical Office of the European Union (Eurostat – https://ec.europa.eu/eurostat) (Eurostat Statistics Explained, 2018), the official state statistics services of the countries and the Federal State Statistics Service (Rosstat – https://rosstat.gov.ru/) for 2018 (for several indicators, data for 2019 were used, due to the lack of data for 2018).

The clustering procedure includes 35 OECD member countries and the Russian Federation. (Note: Iceland and Colombia are excluded from the clustering procedure due to the lack of a large number of indicators or the presence of methodologically disparate indicators.)

Clustering was carried out according to 16 key indicators of the socio-demographic situation of the territories, divided into the following enlarged groups:

1. Demographic characteristics of the population, which include such indicators as central birth rate (%), the total mortality rate (%), net migration rate (%), total fertility rate, and life expectancy (years).
2. The labor force, which is characterized by the following indicators: the share of the labor force in the total working-age population (%), the unemployment rate (%), the share of unemployed women (%), the average time to find a job (months), the actual length of the working week (hours).
3. The standard of living of the population is estimated by such indicators as the consumer price index in 2019 to 2010 (%), the share of spending on food and non-alcoholic beverages in the structure of actual final consumption of households (%), the Gini coefficient (%), the number of students enrolled in tertiary (higher) education programs (per 1,000 people), the number of doctors (per 100,000 people), the number of premeditated murders (per 100,000 people).

The use of the algorithm is associated with a high variability of countries, including within the group of countries with developed economies due to the different speed and intensity of development and thus represent *"outliers"*, the peculiarities of the state structure and the economic system of a particular country.

The k-medoid method allows minimizing the total degree of differences between objects belonging to a particular cluster, while one of the cluster objects is recognized as its *"representative"*, which is the medoids. The validity of our position is also confirmed by the studies of Velmurugan & Santhanam (2008), where it is argued that the k-medoids method is more robust concerning the k-means method, and demonstrates a higher quality of classification both on homogeneous data and for data, in which there are objects sharply differing in one parameter or another, or a high intraclass variation of features (Figure 1).
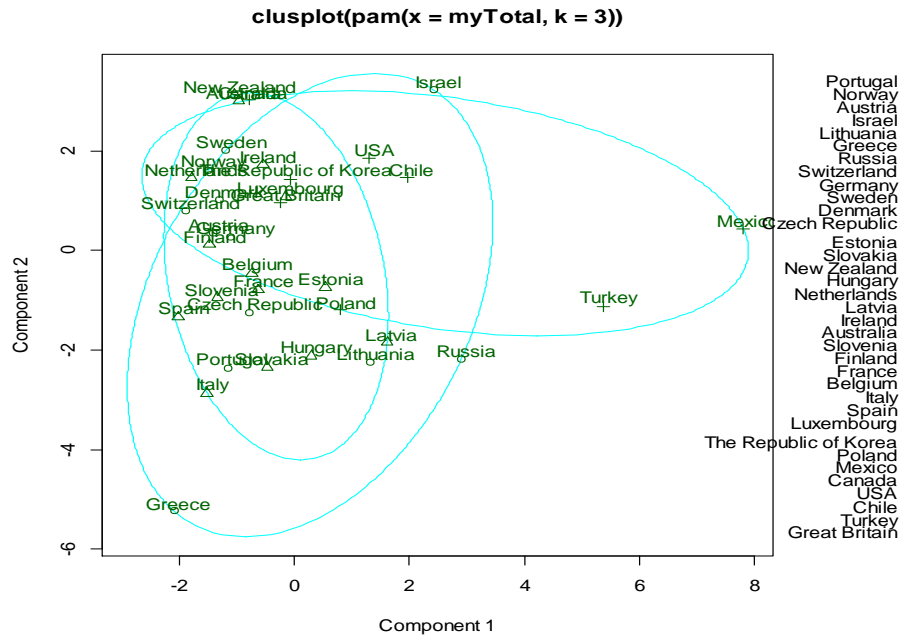
CALCULATIONS AND INTERPRETATION OF THE RESULTS

The use of the methodology described above allowed forming three clusters of countries based on indicators of the socio-demographic situation (Figure1).

The first cluster (Table 1), on the one hand, is characterized by a lower fertility rate than in other clusters against the background of a high mortality rate; in these countries, population aging occurs simultaneously with a drop in the birth rate. The low share of the labor force in the working-age population is accompanied by an increased level of unemployment (the median value is 6.1%), increased duration of finding a new job (8.6 months), which makes these countries unattractive for migration (the net migration rate is 1.6) and practically deprives some of the opportunities to equalize the standard of living. High urbanization and uneven development of industrial, social, and housing infrastructure can further differentiate the population.

At the same time, the high scale of the unemployment rate for the countries participating in this cluster indicates that the indicator *"unemployment rate"* is not typical for the analyzed cluster, which unites a group of countries according to the state of the labor market and characterizes the labor force. Such significant differences became possible since one cluster included countries with a very low unemployment rate – the Czech Republic (2.2%) and Germany (3.4%) and countries with a significant unemployment rate – Greece (19.3%). The

values of the unemployment rate for the Czech Republic and Greece characterize not only this cluster but also the total set of countries.

Note: The results of clustering explain 49.5% of the spread of feature values, are highly variable in at least one attribute (above 33%), which indicates intra-group heterogeneity.

FIGURE 1
CLUSTERS OF OECD COUNTRIES BY INDICATORS OF SOCIO-DEMOGRAPHIC CONDITIONS IN 2018

Table 1 COMPOSITION OF CLUSTERS OF OECD COUNTRIES BY KEY INDICATORS OF THE SOCIO-DEMOGRAPHIC SITUATION IN 2018							
Cluster No.	Number of countries	Name of countries	Central birth rate, ‰	Total mortality rate, ‰	Net migration rate, ‰	Total fertility rate	Life expectancy, years
1	12	Russia, Lithuania, Germany, Czech Republic, Israel, Sweden, Denmark, Norway, Switzerland, Portugal, Greece, Australia	11.29	11.32	0.89	1.66	76.33
2	15	Austria, Finland, Spain, Italy, Latvia, Slovakia, Estonia, Ireland, Luxembourg, New Zealand, Netherlands, Hungary, Belgium, Slovenia, France	12.98	6.6	9.3	1.86	81.72
3	9	Great Britain, Mexico, Poland, USA, Chile, Republic of Korea, Canada, Turkey, Japan	9.35	9.66	3.29	1.52	82.15

The countries included in the second cluster are characterized by both natural and migration population growth, this is possible in the case of a high standard of living of the population and the attractiveness of the country as a source of family income. Luxembourg and New Zealand, with migration growth rates of 16.3% and 15.6%, are the centers of attraction for

migrants, including from Lithuania (-2.5%). Negative values of the excess in demographic factors (birth rate, mortality, demographic growth) indicate a left-sided asymmetry of the distribution.

The main problem of the countries included in the third cluster is related to migration. This is the most variable indicator selected for clustering (110.3%, with an average value of 3.5% and a standard value of 3.86%). The reasons for this spread are identical to the reasons for the second cluster – an even greater differentiation in living standards provokes migration sentiments primarily to the nearest countries, forming pairs (Mexico (-0.42%) – the United States (2.7%) and the United States (2.7%) – Canada (11.3%) and exacerbating the problem of social inequality.

Thus, the clustering of 36 countries by key indicators of the socio-demographic situation makes it clear that neither territorial differences nor the degree of economic development is a significant prerequisite for socio-demographic inequality (similarity).

The analysis of the clustering results for individual blocks gave the following results.

The Distribution of OECD Countries into Clusters by Indicators that Characterize the Demographic Situation

Clustering of OECD countries by indicators that characterize the demographic situation in 2018 (Figure 2) revealed four groups of countries. The allocation of the fourth cluster (Israel, Mexico, and Turkey) is since their demographic indicators differ significantly from the median ones. The clusters themselves are quite homogeneous.
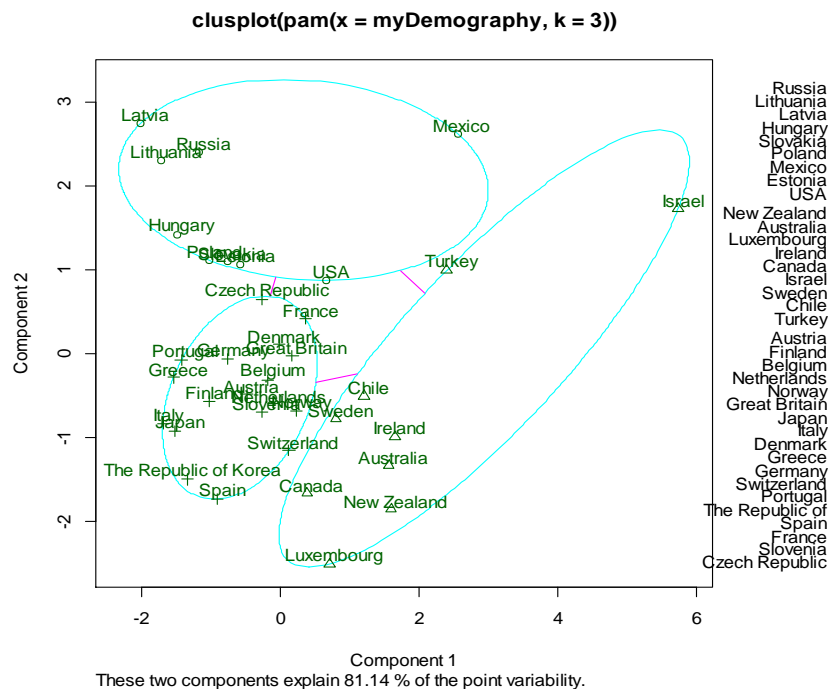

FIGURE 2
CLUSTERS OF OECD COUNTRIES BY INDICATORS CHARACTERIZING THE DEMOGRAPHIC SITUATION IN 2018

In the countries included in the first cluster, extreme values were noted for three of the five parameters: the highest median value of the mortality rate (11.9%), the minimum values of life expectancy (77.4 years), and the migration component (1.9%).

The second cluster was formed mainly by the countries of the former Soviet Union (Slovakia, Greece, Poland, Estonia, Latvia, Lithuania, Russia, Hungary) and the United States. The high birth rate (median value of 11.8%), low mortality (6.8%), advanced medicine, and social support of the population have led to the fact that the median life expectancy of the population of these countries is 82.3 years. Only the indicator of migration turnover gives a high heterogeneity (the coefficient of variation is 225.8%, the range of variation is 5.8% with an average value of the migration coefficient of 0.89%), since Latvia is the outsider of migration turnover among all 36 countries, and the maximum values of the leaders of migration turnover are only 2.96-3.3 % (Hungary – 3.3% and Estonia –2.96%). In general, the worst situation in terms of migration dynamics has developed in the Baltic countries: a total of 1.22% of the Eurozone population lives on their territory, which has declined by about 20-25% since independence.

The countries of the largest third cluster (17 countries), which includes both the countries of the Old World (France, Italy, Portugal, Greece, and Germany), as well as North Korea, which is characterized by a high degree of socialization, and Japan, are brought together by a moderate demographic policy. In addition, the four cities of Munich, Dusseldorf, and Frankfurt am Main have been recognized as the most favorable cities for life for 10 years. The presence of the Netherlands and Belgium in this group is due to the high life expectancy.

The ambivalence of migration processes is related to the dynamics of the labor force (Figure 3).

Distribution of OECD Countries and Russia into Clusters by Labor Force Indicators
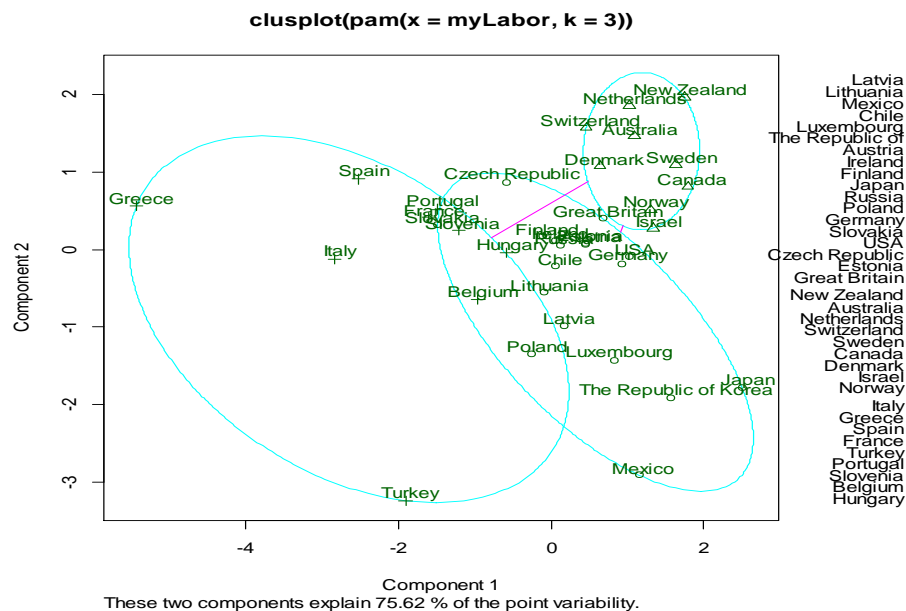

FIGURE 3
CLUSTERS OF OECD COUNTRIES BY LABOR FORCE INDICATORS IN 2018

At first glance, the results of the distribution of countries into clusters that characterize the labor force and demographic situation should be very similar. Indeed, the distribution is largely duplicated, but there are also differences. Thus, the Scandinavian countries, Israel, Switzerland, and Australia moved from the first to the second cluster, the members of which have the maximum similarity in the five indicators studied. Cluster 1 includes countries where two of the five indicators show heterogeneity: the unemployment rate (coefficient of variation of 33.8%) and the average time to find a job (34.1%). It is noteworthy that the latter indicator is heterogeneous only for this cluster and only for an enlarged group of indicators that characterize the labor force. The third cluster is characterized by a longer period of job search (9.18 months), a high level of unemployment (9.64%, but with a coefficient of variation of 52.5%, and an absolute measure of the variation of 15.6%, we can say that there is no normal distribution of countries within this cluster) with an average working week of 38.78 hours. The composition of the clusters is shown in Table 2.

Cluster number	Number of countries	Name of countries	Share of the labor force in the total working-age population	Unemployment rate, %	Rate of unemployed women, %	Average job search time, months	Actual duration of the working week, hours
		Table 2 COMPOSITION OF OECD COUNTRY CLUSTERS BY LABOR FORCE INDICATORS IN 2018					
1	18	Russia, Lithuania, Latvia, Slovakia, Estonia, Mexico, Poland, USA, Ireland, Luxembourg, Chile, Germany, Czech Republic, Austria, Finland, Japan, Republic of Korea, United Kingdom	61.38	4.82	44.69	6.39	38
2	9	Israel, Sweden, Australia, New Zealand, Canada, Denmark, Norway, Switzerland, Netherlands,	66.44	4.8	47.4	6.14	34.11
3	9	Hungary, Turkey, Portugal, Spain, Italy, Belgium, Slovenia, France, Greece	55.11	9.64	48.86	9.18	38.78

Distribution of OECD Countries into Clusters by the Enlarged Group that Characterizes the Standard of Living of the Population

According to the UN recommendations, the standard of living of a person should be assessed by a system of indicators that indicate health, consumption, employment, education, housing, social security, and other characteristics.
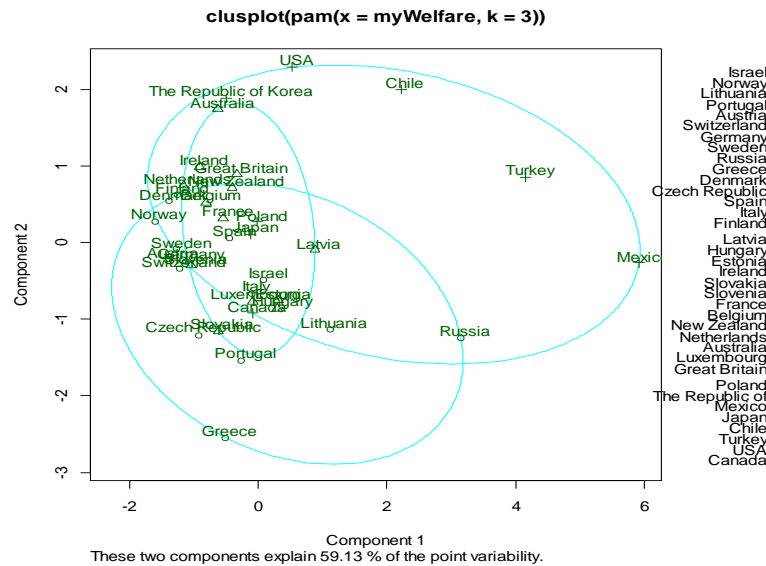
FIGURE 4
DISTRIBUTION OF OECD COUNTRIES INTO CLUSTERS BY THE ENLARGED GROUP THAT CHARACTERIZES THE STANDARD OF LIVING OF THE POPULATION IN 2018

Despite the median proximity of the countries included in each of the three clusters by 6 indicators, there are at least two indicators within each cluster that characterize these countries as heterogeneous (Figure 4). Such values may indicate significant differences in the level and quality of life of the countries under study (Table 3).

The largest first cluster. This is largely due to the policy of forming poverty criteria. For example, in Scandinavia, the minimum pension is taken as the official poverty line. The current social model, based on the redistribution of income from the rich to the poor, has reduced the poverty rate in Sweden and Denmark to 10-12%, which predetermined the placement of these countries in the first cluster. The social contract system also has a positive impact on reducing poverty, when the rich agree to pay for not seeing the poor around them. There is no doubt that a positive system will work only in the case of a small country. This is not effective in densely populated countries. The bottleneck in such a system remains insufficiently integrated migrants. If the issue of economic migrants is resolved quite effectively, then low qualifications and focus on professions that are already irrelevant for Europe almost automatically creates structural unemployment, which the migrants themselves are not interested in overcoming. Such attitudes often lead to various radical ideologies, Islamism, and an increase in crime.

The second cluster was formed mainly by countries focused on the application of an absolute approach to poverty when income is correlated with the value of the subsistence minimum, determined based on the consumer basket. The presence of France in this cluster is because, in addition to the official border (it is calculated according to the normative method, i.e. in absolute values), there is also a social border, which is lower than the first one by 75%.

The third cluster is the most variable one. A significant variation was obtained in terms of the crime rate (13%), the number of students (53.4%), and the share of food expenditures

(42.7%). The placement of the United States in this cluster, despite the high level of GDP per capita and a low share of consumer spending on food, is associated with a high crime rate of 5/100,000 of the population and a low number of doctors.

Cluster No.	Number of countries	Name of countries	Consumer price index, % in 2019 to 2010	% of expenditures on food and non-alcoholic beverages in the structure of actual final consumption	Gini Coefficient, %	Number of students enrolled in tertiary (higher) education programs per 1000 people of the population	Number of doctors per 100,000 population	Number of premeditated murders per 100,000 population
1	15	Russia, Lithuania, Germany, Czech Republic, Austria, Finland, Israel, Sweden, Denmark, Norway, Switzerland, Portugal, Spain, Italy, United Kingdom	111.85	11.47	32.41	41.38	449.54	1.22
2	13	Latvia, Slovakia, Estonia, Ireland, Luxembourg, Australia, New Zealand, Greece, Japan, Hungary, Belgium, Slovenia, France	116.18	10.97	31.19	40.73	333.45	1.36
3	8	Mexico, Poland, USA, Chile, Republic of Korea, Canada, Netherlands, Turkey	135.62	12.96	37.64	40.46	239.12	5.56

Table 3
COMPOSITION OF CLUSTERS OF OECD COUNTRIES IN THE ENLARGED GROUP THAT CHARACTERIZES THE STANDARD OF LIVING OF THE POPULATION IN 2018

CONCLUSION

The study confirmed the methodological consistency of clustering of the OECD countries and Russia to compare poverty and inequality in the context of implementing the Sustainable Development Goals strategy. The results showed that in the presence of territorial similarity (natural and climatic conditions, geographical location, historical development, etc.), there are significant differences in socio-demographic development and, as a result, distribution by clusters, division of a group of countries (territory) into similar groups (clusters) according to the system of indicators does not guarantee homogeneity within the cluster for individual indicators.

ACKNOWLEDGMENTS

REFERENCES

Alkire, S., & Apablaza, M. (2016). Multidimensional poverty in Europe 2006–2012: Illustrating a methodology. *Oxford Poverty and Human Development Initiative Working Papers*.

Alvaredo, F., & Londoño Vélez, J. (2013). High incomes and personal taxation in a developing economy: Colombia 1993-2010.

Alvaredo, F., & Piketty, T. (2014). Measuring top incomes and inequality in the Middle East: Data limitations and illustration with the case of Egypt.

Alvaredo, F., Atkinson, A., Chancel, L., Piketty, T., Saez, E., & Zucman, G. (2016). *Distributional National Accounts (DINA) guidelines: Concepts and methods used in WID. world.*

Atkinson, A.B. (2007). Measuring top incomes: methodological issues. *Top incomes over the twentieth century: A contrast between continental European and English-speaking countries*, *1*, 18-42.

Bricker, J., Henriques, A., Krimmel, J., & Sabelhaus, J. (2016). Measuring income and wealth at the top using administrative and survey data. *Brookings Papers on Economic Activity*, *2016*(1), 261-331.

European Central Bank. (2020). *Understanding household wealth: linking macro and microdata to produce distributional financial accounts.* Retrieved from https://www.ecb.europa.eu/pub/pdf/scpsps/ecb.sps37~433920127f.en.pdf

Eurostat Statistics Explained. (2018). *SDG 10 – Reduced inequalities (statistical annex). Reduce inequality within and among countries (statistical annex).* Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php/SDG_10_-_Reduced_inequalities_(statistical_annex)

Ferreira, F.H., Chen, S., Dabalen, A., Dikhanov, Y., Hamadeh, N., Jolliffe, D., Narayan, A., Prydz, E.B., Revenga, A., Sangraula, P., & Yoshida, N. (2016). A global count of the extreme poor in 2012: data issues, methodology and initial results. *The Journal of Economic Inequality*, *14*(2), 141-172.

Fesseau, M., Wolff, F., & Mattonetti, M.L. (2013). *A cross-country comparison of household income, consumption and wealth between micro sources and national accounts aggregates.*

Grigoryev, L.M., & Pavlyushina, V.A. (2018). Inter-country inequality as a dynamic process and the problem of post-industrial development. *Voprosy Ekonomiki*, *7*, 5-29.

Jolliffe, D. (2014). *A measured approach to ending poverty and boosting shared prosperity: concepts, data, and the twin goals*. World Bank Publications.

United Nations Economic Commission for Europe (UNECE). (2017). *Guide on poverty measurement*. New York and Geneva: United Nations. Retrieved from https://unece.org/fileadmin/DAM/stats/publications/2018/ECECESSTAT20174.pdf

Velmurugan, T., & Santhanam, T. (2008). Performance analysis of k-means and k-medoids clustering algorithms for a randomly generated data set. In *Proceedings of the International Conference on Systemics, Cybernetics and Informatics, Jan* (Vol. 8, pp. 578-583).

Ziliak, J.P. (2010). Alternative poverty measures and the geographic distribution of poverty in the United States. *Report prepared for the Office of Assistant Secretary for Planning and Evaluation, US Department of Health and Human Services. University of Kentucky, Center for Poverty Research, Lexington*.