Classification of Subjects of the Russian Federation by the Level of Socio-Economic Development

Vladimir NOSOV

K.G. Razumovsky Moscow State University of technologies and management, Moscow, Russia,
Moscow Academy of the Investigative Committee of the Russian Federation, Moscow, Russia,
novlavla875@gmail.com

Alexander TCYPIN

Moscow State University of Food Production, Moscow, Russia,
tcypin@list.ru

Israil ABDULRAGIMOV

K.G. Razumovsky Moscow State University of technologies and management, Moscow, Russia,
abdulragimov@list.ru

Tatiana MAHANOVA

K.G. Razumovsky Moscow State University of technologies and management, Moscow, Russia,
mahanova@list.ru

Vyacheslav ZHENZHEBIR

K.G. Razumovsky Moscow State University of technologies and management, Moscow, Russia,
zhenzhebir@ list.ru

Abstract

The existing differentiation of subjects of the Russian Federation based on the level of socio-economic development leads to some subjects of the Russian Federation being subsidized at the expense of other subjects, which results in repression of economic development of not only subjects-donors, but of the whole country as well. Therefore, the government policy on subjects of the Russian Federation should be focused not on an automatic reduction of the inequalities in the fiscal capacities, but on the maximum use of their potential. In view of this, it seems important to identify the factors affecting the development of spatial economics with account taken of specific traits of various groups of subjects of the Russian Federation. The cluster analysis allowed for three groups of subjects of the Russian Federation: leading, main and laggard. Examination of correlation coefficients in every group has shown considerable variation of basket of factors, i.e. there are various mechanisms of formation of gross regional product (GRP) in transition from one cluster to another. The regression modeling performed with consideration of subject's belonging to one of the three clusters formed shows that the best results in socio-economic development can be achieved if values are approximate to those registered in regions described in the first cluster.

Keywords: classification, cluster analysis, regression, dummy variables.

Introduction

In their paper Dibrova et al. (2018) highlight that transition of economy from planned to market has led to changes in spatial development of the country territory and to widening of the differential gap in socio-economic development between subjects of the Russian Federation. The current situation results

in some subjects of the Russian Federation being subsidized at the expense of its other subjects. Such state regional policy approach of the government gives rise to a welfare mentality and abuse on the part of subjects authorities. Therefore, the government policy on subjects of the Russian Federation should be focused not on an automatic reduction of the inequalities in the fiscal capacities, but on the maximum use of their potential. In view of this, it seems important to identify the factors affecting the development of spatial economics with account taken of specific traits of various groups of subjects of the Russian Federation. Over the past decade there has been a strong interest in various classifications of subjects of the Russian Federation characterizing different aspects of their socio-economic development. Classification is a process of distributing and dividing objects, concepts and terms into classes, groups and categories, according to their similar features.

Different aspects of territories classification in accordance with various socio-economic parameters were presented in scientific papers by Thorbecke (2013), Tekueva et al. (2016), Shubat and Shmarova (2017), Kurushina and Petrov (2018), Ladykova and Bersenyov (2018), Ohotina et al. (2018), Shakleina and Midov (2019) and others. In their papers the authors highlight a number of factors affecting the development of regions, and classify them characterizing specific developmental features of every group.

Publications of Russian and foreign scientists on this issue confirm the relevance of the research and the need for further study of classification of Russian federal subjects. The key point of the research is development of an econometric model that describes the impact of socio-economic indicators on gross regional product in the context of division of Russian federal subjects into multidimensional groups.

Methods of Research

In most studies groups of subjects of the Russian Federation are usually singled out according to one of the studied attributes. However, subjects of the Russian Federation can be characterized by not only one attribute, but by a set of them.

GRP per capita (*Y*) in 2017 was chosen as an indicator characterizing the development of subjects of the Russian Federation.

Factors affecting the outcome variable were divided into two groups:

1. Socio-economic indicators, which describe the potential of the subject of the Russian Federation: the Industrial Production Index (*X1*), %; percentage of manufacturing enterprises in the total number of enterprises of the region (*X2*), %; amount of fee-based services per capita (*X3*), rubles/person; percentage of unprofitable enterprises (*X4*), %; fixed assets value per one enterprise (*X5*), thous.rubles/unit; degree of fixed assets depreciation (*X6*), %; investments in fixed capital per capita (*X7*), rubles/person; labor force participation rate (*X8*), %.

2. Social indicators, which characterize living standards in the Russian Federation: per capita income (*S1*), rubles/person; per capita meat and meat products (including category 2 by-products and raw tallow) consumption (*S2*), kg/person; floor area per capita (*S3*), sqm/person; the Gini coefficient (*S4*), coefficient.

Therefore, to construct a classification it is appears more appropriate to use statistical methods of polythetic classification whose main feature is that all available attributes are used when forming a group.

The set of subjects of the Russian Federation forms an attribute space. If the federal subject is characterized by *m* attributes, then it is considered as a point in the *m*-dimensional attribute space. The solution is to identify points that are close to each other in this attribute space. This multidimensional grouping is solved by means of cluster analysis, when the entire set of objects is divided into homogeneous groups (clusters). Subjects of the Russian Federation, belonging to the same cluster,

should be similar to each other, and the degree of similarity between them within each cluster should be greater than between subjects of the Russian Federation included in other clusters.

At present, statistics has a significant range of clustering algorithms. Most often, researchers use hierarchical algorithms, among which the Ward method is most common. According to Mooi and Sarstedt (2001), this method involves using dispersion analysis procedures to estimate distances between clusters. With this, at each step of clustering, elements leading to the smallest increase in intracluster dispersion, will be combined into one cluster. Therefore, the smallest clusters are gradually merging into larger ones.

After identifying the number of clusters, based on the logic of the research, it is proposed to write a regression equation with dummy variables.

Research Results and Discussion

Subjects of the Russian Federation are quite diverse, their sizes are different, which is why they are not comparable. It is possible to narrow disparities only by shifting from the original set to subsets consisting of homogeneous objects.

Basing on the tree structure of clustering it is proposed to create a multidimensional grouping of subjects of the Russian Federation. As a result of processing the data file a dendrogram presented on Figure 1 was obtained.

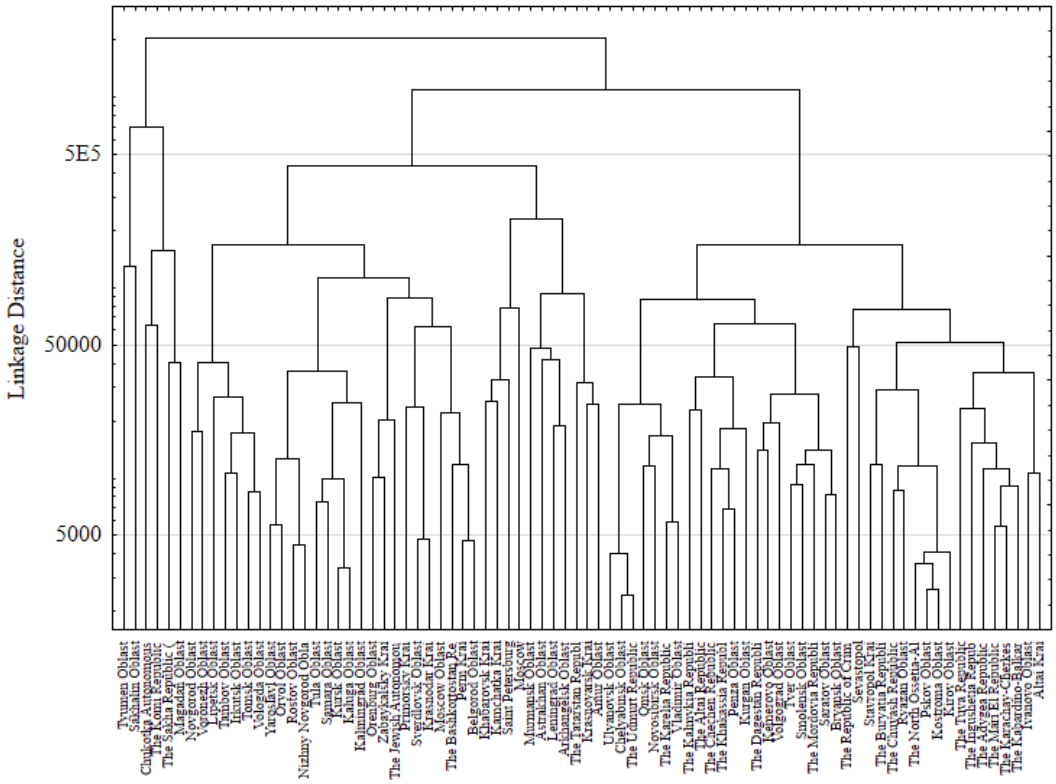

Fig. 1: Results of clustering of indicators describing the developmental level of Russian regions economy. Drafted by authors in the STATISTICA package basing on data from the "Russian Regions" Yearbook

According to the results presented in the Figure 1, the original set of federal subjects is divided into 3 groups. The first group includes six objects with the highest parameters, which are the main donor-regions: Tyumen Oblast, Chukotka Autonomous Okrug and Sakhalin Oblast. The second group consists of subjects with average values, which form the basis of the country economy: Moscow Oblast, the Tatarstan Republic and Samara Oblast, almost all of them belong to the Central Federal District and Far Eastern Federal District. The third group includes 39 subjects characterized by minimum values, which are recipient-regions: Bryansk Oblast, the Kalmykia Republic and the Dagestan Republic.

Average values for clusters are presented in Table 1. To assess the significance of mean values contrast in groups F-test values were calculated.

Table 1: Intracluster mean values of indicators describing stages of economic development of Russian regions

Variables	Cluster #1	Cluster #2	Cluster #3	F-test	*p*-significance level
Number of objects in cluster	6	37	39	X	X
Y	1179.6	436.91	239.22	98.52	0.00
S1	45969	29999	21659	49.11	0.00
S2	72.0	75.9	68.2	3.80	0.03
S3	25.5	25.9	24.1	2.54	0.09
S4	14.5	13.0	11.2	18.50	0.00
X1	100.2	103.7	105.0	0.95	0.39
X2	5.4	8.2	9.1	9.18	0.00
X3	82467	57111	37372	27.21	0.00
X4	35.1	30.5	33.4	3.42	0.04
X5	133740	41766	29974	56.22	0.00
X6	47.8	46.7	50.2	1.82	0.17
X7	347659	99771	46711	110.66	0.00
X8	74.2	69.4	67.2	15.91	0.00

Drafted by authors in the STATISTICA package using data from the "Russian Regions" Yearbook

According to the F-test values presented in Table 1, there are only three cases when mean values are the same: *S3*, *X1* and *X6*. In other cases mean values differ significantly when shifting from one cluster to another.

These differences appear to be huge: the first cluster is 4.5 times larger than the third cluster on the fixed capital cost per enterprise rate, 4.9 times larger on GRP per capita rate, and 7.4 times larger on investments to fixed capital per capita rate.

It would be logical to assume that the high variation of indicators values would have a negative impact on the regression model if it was built for the whole set. The solution to this situation would be to estimate three separate regressions or one regression with dummy variables. This research is presented in the next paragraph.

According to data in Figure 2, there is a considerable variation of the correlation coefficient values depending on the cluster, i.e. every cluster includes a certain set of variables which significantly affect GRP per capita:

cluster #1 - variables S2, S4, X5, X6 and X7 have an impact above 0.5

cluster #2 - is affected by variables S1, X3, X7 and X6

cluster #3 - is mostly affected by the X7 factor.

Moreover there is a strong connection between independent variables, i.e. one could assume the existence of multicollinearity. Thus, use of shared variables in the model (based on the first and second clusters) will result in an unreliable estimation of the regression equation parameters.

Since there is an issue of multicollinearity in this case, it is necessary to develop a model with factors independency.

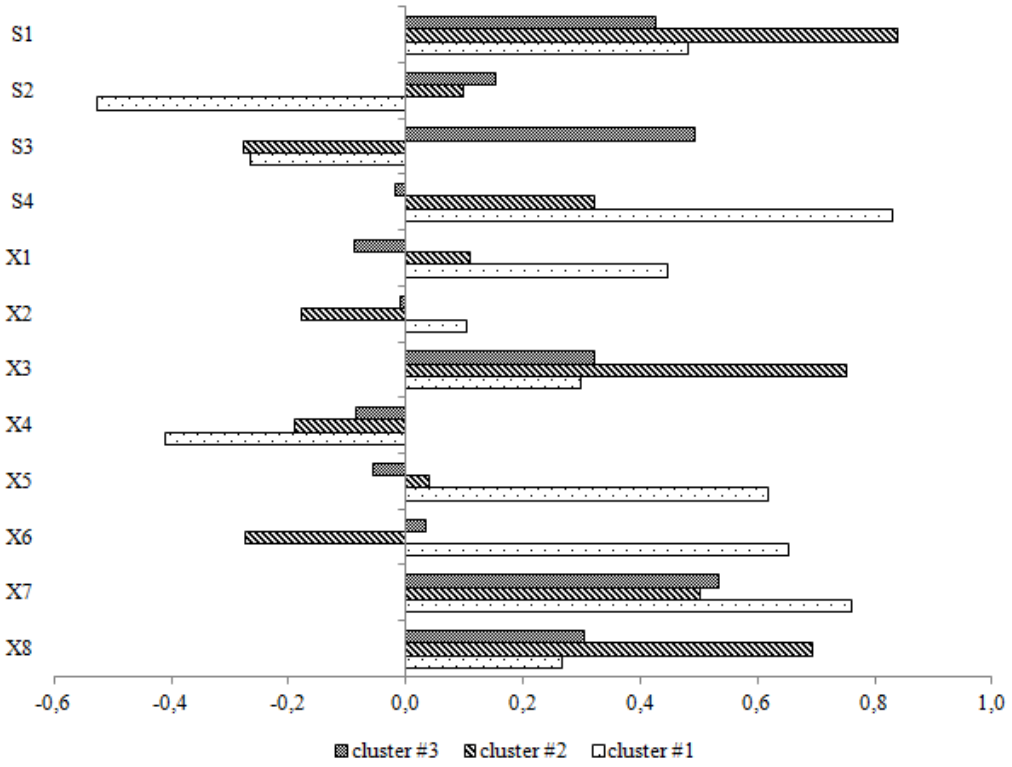

Fig. 2: The correlation coefficient values for three clusters. Drafted by authors in the STATISTICA package basing on data from the "Russian Regions" Yearbook

In all three cases the *X7* variable appears, hence it is possible to build a unified model with dummy variables reflecting belonging of objects to a certain cluster. As there are three gradations, two binary variables should be introduced:

D1 - a dummy variable taking on a value of 1, if the subject of the Russian Federation belongs to the second cluster, and a value of 0 in other cases;

D2 - a dummy variable taking on a value of 1, if the subject of the Russian Federation belongs to the third cluster, and a value of 0 in other cases.

This classification assumes that the first cluster will act as a basic one, therefore partial regressions will be calculated in relation to this group.

Results obtained from the least squares estimation of the multiple regression equation parameters in the STATISTICA package are presented in Table 2.

Table 2: Parameters of the regression equation, impact of factor on GRP per capita in Russia

Indicators	Regression equation coefficient	Standard error of the regression coefficient	Student's *t*-test	*p*-significance level
The intercept	469.682	115.421	4.069	0.000
D1	-236.505	91.769	-2.577	0.012
D2	-325.851	104.894	-3.106	0.003
X7	0.002	0.000	6.831	0.000

Drafted by authors in the STATISTICA package using data from the "Russian Regions" Yearbook;
R2 = 0,821; F(3,78)= 119,199; p<0,05

According to data presented in Table 2, there is a strong correlation relationship between the dependent variable and factors of the regression equation. At that, the 82.1% variation of GRP per capita can be explained by variation of model indicators, which attests to a high accuracy of history matching.

In econometric modeling to evaluate the statistical significance of the regression equation, the F-test is used. The F-test table value is 2.72, which is considerably lower than the actual value equal to 119.2, therefore, it can be concluded that the regression equation is statistically significant.

In accordance with data presented in Table 2, as a result of the conducted regression analysis the following equation can be obtained:

$$Y' = 469.682 - 236.505 \times D1 - 325.851 \times D2 + 0.002 \times X7 + E$$

Actual values of the Student's t-test are high ($p<0.05$), which means that the assessed regression equation parameters are statistically significant.

It also should be noted that statistical significance of coefficients given dummy variables indicated the existence of structural differentiation, i.e. there is an impact of a group (cluster) on the regression equation level.

The *X7* coefficient can be interpreted in the following way: in the case of an increase in investments to the fixed capital per capita by one thous.rubles/person, GRP per capita will increase by two thous.rubles/person.

Since dummy variables introduced in the regression equation are the intercept differentiation coefficients, it seems possible to create partials for all three clusters on the basis of the existing equation:

cluster #1 (basic or simple regression):

$$Y' = 469.682 + 0.002 \times X7 + E$$

cluster #2 (this partial regression is lower than the simple one by 233.177 rubles):

$$Y' = (469.682 - 236.505) + 0.002 \times X7 + E$$

cluster #3 (this partial regression is lower than the simple one by 143.831 rubles):

$$Y' = (469.682 - 325.851) + 0.002 \times X7 + E$$

Obtained regression equations can be represented by a scatter plot in Figure 3. Partial regression lines in Figure 3 show the differences arising from shifting from one cluster to another.

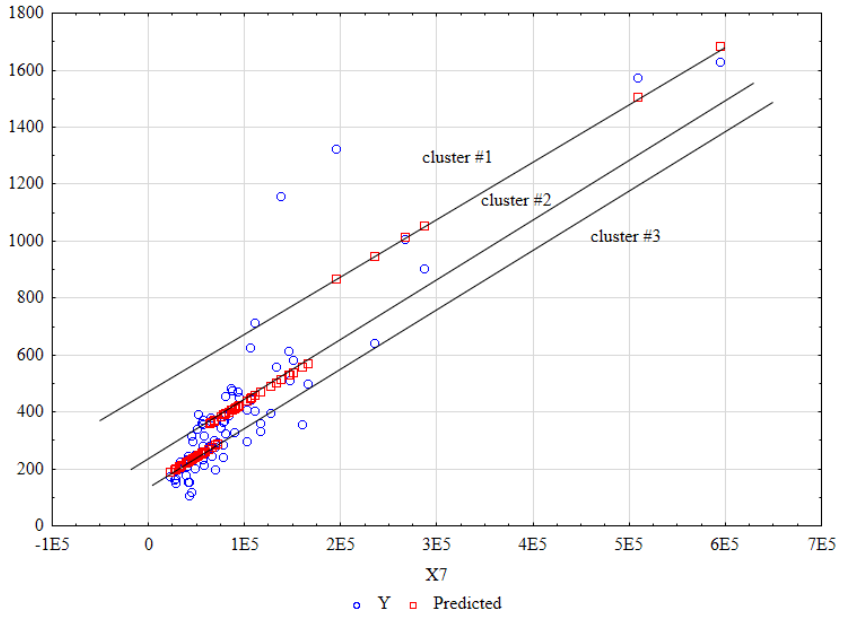

Fig. 3: Correlation plot containing results of regression equation of GRP per capita dependency on factors. Drafted by authors in the STATISTICA package basing on data from the "Russian Regions" Yearbook

At the last stage of the analysis, upon having evaluated the statistical significance of the model, it is necessary to model possible values of *Y* variable at set values of an independent factor.

It is proposed to conduct a situational forecasting using mean values of investments to the fixed capital (per capita), maximum and minimum values of *X7*. This would result in realistic, pessimistic and optimistic scenarios for economic development. The results are presented in Table 3.

The best seems the prediction with maximum values of an independent variable with basic positions, i.e. the third cluster, in which case GRP per capita increases to 1683.7 thous.rubles/person. This forecast is optimistic because it does not correspond to reality, as it is 55.8 thous.rubles higher than the maximum value of 1627.95 thous.rubles in 2017 in Tyumen Oblast.

Table 3: Predicted GRP per capita values at specified value of *X7*

Alternative forecasts	Clusters	*D1*	*D2*	*X7*	GRP per capita forecast, thous.rubles	Bound	
						Lower	Upper
Average (realistic)	1	0	0	92673	658.9	477.2	840.6
	2	1	0	92673	422.4	381.9	462.9
	3	0	1	92673	333.1	285.3	380.9
Minimum (pessimistic)	1	0	0	22032	514.7	296.6	732.7
	2	1	0	22032	278.2	216.8	339.5

	3	0	1	22032	188.8	146.9	230.7
Maximum (optimistic)	1	0	0	594534	1683.7	1506.0	1861.4
	2	1	0	594534	1447.2	1150.0	1744.4
	3	0	1	594534	1357.9	1029.5	1686.2

Drafted by authors in the STATISTICA package using data from the "Russian Regions" Yearbook

The worst result is seen at a minimum value of *X7* in the third cluster and is 333.1 thous.rubles per person, which is much better than the minimum value in the Republic of Ingushetia in 2017 (106.76 thous.rubles per person).

Conclusions

Summarizing the conducted analysis the following conclusions can be made:

- based on macroeconomic indicators, subjects of the Russian Federation were divided into three clusters: leading, main and laggard;
- the highest GRP per capita values are seen in the first cluster which includes regions-donors to the budget of the Russian Federation;
- the assessment of the impact of factors on GRP per capita shows a considerable variation of basket of factors, indicating various mechanisms for formation of dependent variable values;
- the only factor appearing in all three clusters is the "investment in fixed capital (per capita)" indicator, with that, the higher the indicator, the higher the outcome variable;
- the conducted regression modeling considering subject's belonging to one of the three clusters formed shows that the best results can be achieved if values are approximate to those registered in regions described in the first cluster;
- classification of subjects of the Russian Federation and analysis of factor affecting gross regional product allows for providing a scientific base for formulation of development strategies for subjects of the Russian Federation.

References

Dibrova, Zh., Nosov, V., Ovchenkova, G., Karpenko, E., Pilyugina, A., and Erkovich, E. (2018), 'The main directions of the solution of the problem of food security in Russia,' *International Journal of Mechanical Engineering and Technology*, 9 (13), 387-394.

Kurushina, E. and Petrov, M. (2018), 'Performance criteria of spatial development projects based on interregional integration,' *Economy of Region* 14 (1), 176-189.

Ladykova, T. and Bersenyov, V. (2018), 'Typology of Macroeconomic Parameters of Population Income,' *Economy of Region* 14 (2), 380-394.

Mooi, E. and Sarstedt, M. (2011), 'A concise guide to market research: The process, data, and methods using IBM SPSS Statistics,' Springer, New York. 10.1007/978-3-642-12541-6.

Tekueva, M., Burkov, A., Nosov, V., Novoselova, S. and Nayanov, A. (2016), 'Agriculture and Agribusiness: Clustering Issues,' Resea*rch Journal of Pharmaceutical, Biological and Chemical Sciences*, 7 (6), 1634–1638.

Ohotina, A., Lavrinenko, O., Gladevich, J. and Lazdans, D. (2018), 'The investment climate in Latvia's, Lithuania's and Belarus's cross-border regions: The subjective-objective assessment,' *Entrepreneurship and Sustainability Issues*, 6 (2), 767-780.

"Russian Regions. Main Traits of Subjects of the Russian Federation: Statistical Book,' *Federal State Statistics Service.* [Online], [Retrieved August 08, 2019], http://www.gks.ru/free_doc/doc_2018/region/reg_sub18.pdf.

Shakleina, M. and Midov, A. (2019), 'Strategic classification of regions according to the level of financial self-sufficiency,' *Economic and Social Changes: Facts, Trends, Forecast* 12 (3), 39-54.

Shubat, O. and Shmarova, I. (2017), 'Cluster analysis as an analytical tool of population policy,' *Economy of Regio*n 4 (4), 1175-1183.

Thorbecke, E. (2013) 'The interrelationship linking growth, Inequality and poverty in Sub-Saharan Africa,' *Journal of African Economies*, 22 (1), 15-24.