Advances in Computer Science Research (ACSR), volume 85
Third Workshop on Computer Modelling in Decision Making (CMDM 2018)

Economic Consequences of the Development of Digital Technologies in Russia

Marina Arkhipova* [0000−0002−9022−7385], Julia Serbova† [0000−0002−4294−8751]
National Research University Higher School of Economics, Moscow, Russia
*archipova@yandex.ru, †serbova_julia@mail.ru

***Abstract*—The article examines the literature, which determinates the factors affecting Gross Regional Product as well as broadens the analysis to different regions of the Russian Federation. The regressions modeling and cluster analysis is used for the issue. Two linear regression models are constructed based on the indicators of the Federal Statistics Survey databases for the year 2015 as well as the index of readiness of regions of Russia for the information society (ICT) is used. After the estimation the two clusters based on the values of sub-indexes were found, presenting the typical and atypical behavior towards values of sub-indexes.**

***Index Terms*—economic growth, gross regional product, effect of digital technologies, linear regression, cluster analysis**

I. Introduction

The term "digital society" is significant for both developed and developing countries on the path to rapid economic growth. Nowadays digital technologies have become firmly established in the industry as well as in everyday lives of all of us. Therefore, it is necessary to determine what indicators of ICT influence the economy. Therefore, the statistical indicators which characterize the level of development of digital technologies and the economy in each region are used to determine the economic condition of the Russian regions. The purpose is to estimate the impact of the main factors, which are responsible for the digital equipment, on the economic condition of the region.

II. The Evaluation of the Methodology of Factors Influencing Economic Growth of the Regions

Since the topic of the widespread of the digital technologies is widely discussed in mass media, the researchers find it critical to analyses the main drivers of modern economy, which tackle the digital technologies in the first place. As the new data and methods become available they are used in the more precise estimations of the current situation in the economy. One of the first authors to point the relevance of studying ICT indicators was Latysheva M.A. The author has studied the degree of economic development of Russia and its regions specifically focusing on the factors of digital economy. The choice was made to study Russia as it is on the path of becoming an influential agent of the world economy [1]. The concept of panel data regression estimation was used for the analysis, that allows us to take into account factors difficult to distinguish by simple regression, because they are not observed or are not presented in a quantitative form. The study led to the following conclusions: the value of GRP is influenced by the value of the basic funds, retail turnover per capita and the average monthly salary. In each region, the influence of individual effects was also noted, during 2001–2006, the growth of GDP was observed due to factors that are not included in the model.

A more complicated approach was presented by Postnikov E.A. The author proposes an econometric model based on panel data, but in comparison with Latysheva, there was a system of four independent equations analyzed. The dependent variables were GRP, the actual consumption of households, investment in fixed capital, and average nominal wage [2]. Firstly, all equations were evaluated by the pooled sample, then fixed and random effects model was used to estimate the coefficient of the equations. The author tests the assumption that estimates are well-established in the model with random effects. As a result, the ranking of subjects based on the value of the fixed effect and the selection of regions with the highest positive and lowest negative values of effects is presented. The advantage of this work is the estimation of three models (Pooled, FE, RE), and the choice of the type of effects. Also, the author proposes to calculate the sample correlation squared, instead of using the determination coefficient, since the latter one loses its interpretation when estimating the GLS.

Another research, which considers the economic condition of the region as the volume of the economy in value terms is made by Miroliubova and Voronchikhina. They use the set of regressors, which are determined in accordance with the macroeconomic identity. In this case, the amount of GRP is determined by the sum of the costs of four economic sectors: households, enterprises, the public sector and the outside world [3]. The advantages are that the models are tested for multicollinearity, homoscedasticity and the autocorrelation. The choice of regressors is also interesting, because the concept of the basic macroeconomic identity was tested. The authors came to the conclusion that the economic growth of the region was to the greatest extent ensured by investments in fixed assets and wages. The authors have a good point in the research; therefore, it makes sense to expand the study not only for one region, but also for all the others, to make an analysis for the whole country.

TABLE I
DIGITAL REGRESSORS

Variable	Mark	Unit of measure
Organizations which use broadband Internet	INTERNET_SHIR	%
Percentage of high-tech industries	HIGHTECH_OTR	%
Information activity of enterprises	INN_ACTIVITY	%
Percentage of innovative products in total number of products	INN_GOODS	%
Percentage of enterprises that use the Internet	INTERNET	%
Percentage of enterprises that use computers	COMPUTER	%
The level of digitalization of the local television network in cities	DIGITAL_TEL	%

TABLE II
ECONOMIC REGRESSORS

Variable	Mark	Unit of measure
Retail trade	TORG	Rub per capita
Average annual level of unemployment	BEZRAB	%
Percentage of investments to fixed capital in GRP	DOLYA_INVEST	%
Percentage of investments in fixed capital in GRP	R_AND_D	%
Investments to fixed capital	INVEST	Rub per capita
Average monthly wage	ZAR_PLATA	Rub
Percentage of population with incomes below the subsistence minimum	PROJ_MIN	%

III. THE INFLUENCE OF ECONOMIC AND DIGITAL INDICATORS ON GRP ESTIMATED BY LINEAR REGRESSION MODELS

After analyzing literature, two groups of regressors were made which influence on the value of GRP:

1) regressors which characterize digital technologies;
2) other regressors which affect the GRP.

The description of the first group of factors is given in Table I.

The second group of factors is given in Table II. The monetary figures were adjusted for the population in the region to offset the differences associated with the large scale of the region. As well as following the economic theory, monetary indicators follow the lognormal law of distribution, therefore, to normalize the values, all monetary indicators were standardized.

The dependent variable is Gross Regional Product of the subjects of the Russian Federation. The figures were taken from the statistical database of Russian Federal State Statistics Service for the year 2015.

Before estimating the model, the dependent variable was testes for outliers using box-and-whiskers plot. The results are shown in Fig. 1. Asterisks indicate the extreme values, which go beyond the limits 3IQR. In order not to lose the observations and not to shift the estimates of regression coefficients, dummy variables were introduced to indicate emissions in the sample. Thus, additional discrete indicators were included:

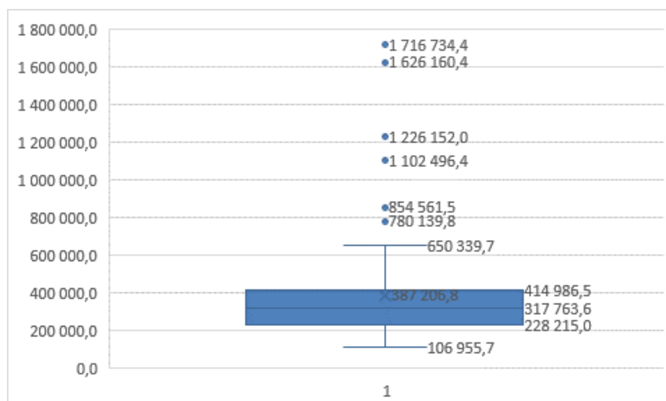

Fig. 1. Test Y (GRP of region) for anomalous observations. Source: Federal State Statistics Survey 2015

- d1 is a dummy variable which takes the value 1 if the subject is Moscow;
- d2 is a dummy variable which takes the value 1 if the subject is Chelyabinsk region;
- d3 is a dummy variable which takes the value of 1 if the subject is the Sakhalin region;
- d4 is a dummy variable which takes the value 1, if the subject is Chukotka Autonomous Okrug.

The method of step-by-step inclusion of variables was chosen as a method of constructing a regression linear model.

TABLE III
THE DEPENDENCE OF GRP FROM DIGITAL AND ECONOMIC INDICATORS

	Variable	Coefficient	t-statistics	p-value
Dependent variable – GRP (without dummy variables)	const	5.1016	10.306	0.0000
	INN_GOODSDS	0.0066	1.4817	0.1427
	BEZRAB	-0.0373	-5.2396	0.0000
	R_AND_D	0.0422	1.5741	0.1197
	INVEST	0.6985	16.4996	0.0000
	R^2	F-statistics	BIC	HQ
	0.846	101.6375	-0.0458	-0.1357
Dependent variable – GRP (with dummy variables)	const	6.3151	11.954	0.0000
	INN_GOODSDS	0.0088	2.1691	0.0335
	BEZRAB	-0.0386	-6.1556	0.0000
	R_AND_D	0.04611	1.9388	0.0566
	INVEST	0.5895	12.7998	0.0000
	d1	0.7266	3.7559	0.0004
	d2	0.3437	1.6227	0.1081
	d3	0.6729	3.2036	0.0020
	d4	0.5184	2.6206	0.0108
	R^2	F-statistics	BIC	HQ
	0.89	70.858	-0.1617	-0.3235

After iterations, the model includes the indicators BEZRAB (Average annual unemployment rate), INVEST (investment in fixed capital), R_and_D (Percentage of internal costs for R and D), INN_GOODS (percentage of innovative products in the total number of products), COMPUTER (percentage of enterprises that use computers), as well as INTERNET_SHIR (percentage of enterprises which use broadband Internet).

The results of evaluation of two models with and without dummy variables for anomalous observations are presented in Table III. If do not take into account outliers, the coefficients are insignificant, and the model deteriorates. The model was tested for prerequisites of the OLS model:

- the residuals are distributed normally (Jarque – Bera test);
- the standard error does not depend on the regressors (White test);
- there is no correlation between the regressors.

However, the model considers only three factors which characterize digital technologies. The coefficients of other quantitative regressors are insignificant, therefore, they are excluded from the model. Thus, it is necessary to choose a composite indicator which would combine other digital factors.

It was decided to choose the index of readiness of the subjects of Russia to the information society as an indicator which combines digital technologies [4]. This index is calculated by the Institute for the Development of the Information Society over the past seven years.

Calculating the index, the authors use 77 indicators, many of them are recommended by the international organizations. When ICT parameter was added to the previous model, the model was evaluated without using dummy variables. It means, that the regressor explains the errors of the dependent variable. The model is significant with an error's probability of 0.05, the coefficients are also significant with an error's probability of 0.1 (Table IV).

The second regression model with the index were also tested for the implementation of prerequisites of OLS. Thus, the inclusion of ICT in the model allows to indirectly explain the variation of GRP through the economic environment and human capital and other factors which reflect the degree of digitalization of the region.

The additional regression was estimated to take into account the relation between the index and the original data of the Federal State Statistics Survey, where the dependent variable is the index itself, independent ones are INTERNET_SHIR (the percentage of organizations using broadband access to the Internet), INN_ACTIVITY (Innovative activity of enterprises), DIGITAL_TEL (the level of digitalization of the local telephone network in cities and towns) and other factors considered in the construction of the first regression.

The results are presented in Table V. Using the method of elimination of variables, the above regressors were left in the model.

Coefficient estimates are significant with the probability of error $\alpha = 0.1$, the hypothesis of insignificance of the model is not rejected with the probability of error $\alpha = 0.01$. The model was also tested on the implementation of OLS prerequisites.

According to the results of the estimated regression it can be concluded that the ICT Index significantly depends on the factors INTERNET_SHIR, INN_ACTIVITY, DIGITAL_TEL, which were not previously included in the model due to their insignificance. This gives an advantage to the ICT index, because it is a composite indicator which combines the impact on the GRP value of those indicators which individually are not able to significantly affect the dependent variable

TABLE IV
THE DEPENDENCE OF GRP FROM INDEX ICT AND ECONOMIC INDICATORS

	Variable	Coefficient	t-statistics	p-value
Dependent variable – GRP (without dummy variables)	const	5.2012	10.306	0.0000
	BEZRAB	-0.03365	-5.027	0.0000
	R_and_D	0.03804	1.8694	0.0617
	INVEST	0.6800	15.8439	0.0000
	IKT	0.09004	1.9519	0.0547
	R^2	F-statistics	BIC	HQ
	0.849	104.1831	-0.0667	-0.1566

TABLE V
THE DEPENDENCE OF THE INDEX ICT FROM DIGITAL INDICATORS

	Variable	Coefficient	t-statistics	p-value
Dependent variable – ICT index, model with dummy variables	const	0.03203	0.23014	0.8186
	INTERNET_SHIR	0.00237	1.94197	0.0561
	INN_ACTIVITY	0.00678	2.99405	0.0038
	DIGITAL_TEL	0.00196	1.92912	0.0577
	d1	0.16645	2.0921	0.04
	d2	-0.06776	-0.8472	0.3998
	d3	-0.15894	-1.8759	0.0656
	d4	4.66137	59.1126	0.0000
	R^2	F-statistics	BIC	HQ
	0.98	512.6752	-1.9372	-2.081

IV. DIFFERENTIATION OF REGIONS OF RUSSIAN FEDERATION BY CLUSTER ANALYSIS

Due to the fact that the index is a composite indicator, it is necessary to analyze the differentiation of regions by its sub-index values [5]:

- ICT in health care settings
- ICT in business
- ICT in educational institutions
- ICT in cultural institutions
- The usage of ICT by households and people.

The optimal number of clusters was determined by the median method and the Ward's method. Then K-means method of clusterization was applied. The figure shows the average for each group (Fig. 2).

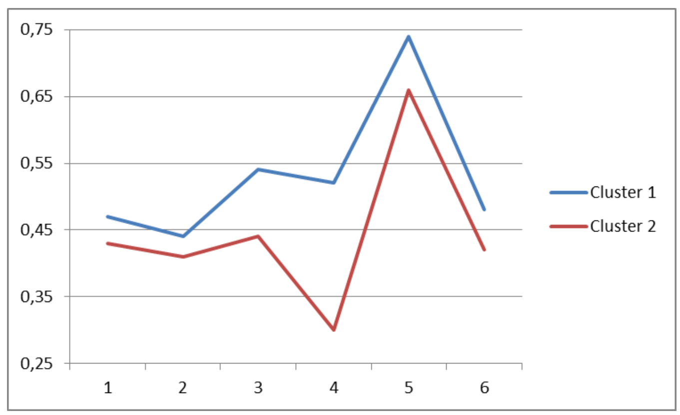

Fig. 2. The plot of averages of six sub-indexes on each cluster

The first cluster brings together the regions of the Russian Federation, which have relatively high rates of sub-indexes, so they can be called as the leading subjects, while the second one unites the regions with lower values of sub-indexes, in other words – lagging in the level of information and communication technologies (ICT).

The indicator of the division of observations between clusters is ICT in culture because there is a significant difference between the averages, while on other indicators these differences are not so much. For all indicators, the hypothesis of the equality of averages was tested, the hypothesis was rejected with the probability of error $\alpha = 0.05$.

So, the first group can be named as "Regions of the Russian Federation with a high level of ICT in culture, use of ICT by individuals and public authorities, and relatively high level of ICT in medicine, business and education". The second one is "Regions with a very low level of ICT in medicine and business and relatively low level of ICT in the use of householders and government".

Typical and atypical representatives for each cluster were identified based on the distance to the cluster center (Table VI).

It is necessary to study the dynamics of ICT values in the following years, because it is likely that atypical representatives of the first class can get to the second, and atypical representatives of the second class will be in the first cluster.

TABLE VI
TYPICAL AND ATYPICAL REPRESENTATIVES FOR EACH CLUSTER

Cluster	Typical	Atypical
The first one	Sakhalin region Omsk region	Yamalo-Nenets Autonomous Okrug Khanty–Mansi Autonomous Okrug – Yugra
The second one	Mari El Republic Saratov region	Amurskaya oblast Kabardino-Balkar Republic

V. CONCLUSION

To sum up, the indicators affecting the value of the GRP are identified, the ICT index mostly influences. The level of digitalization of the telephone network, and the percentage of use of the Internet by enterprises have become significant indicators in the model. The addition of the ICT index allows us to explain the differentiation of all regions of the Russian Federation without exception of anomalous values from the sample. Cluster division of regions by the values of sub-index which are included in the ICT has shown differences in the use of ICT in the culture.

Prospects for further study are the estimation of systems of equations which could be used to include more complex relationships between nominal regressors and composite indicators. It is also recommended to conduct a regression analysis on panel data to take into account the individual characteristics of each region.

ACKNOWLEDGEMENTS

This work was supported by a grant of Russian Foundation for Basic Research #18-010-00564 Modern Tendencies and Social and Economic Consequences of Digital Technologies Development in Russia.

REFERENCES

[1] M. Latysheva, "Econometric modeling of social and economic development of russian regions," *Upravlenie bol'shimi sistemami*, vol. 27, pp. 282–292, 2009.

[2] E. A. Postnikov, "Econometric modelling of the development of russian regions on the basis of panel data," *Upravlenie Sotsialnimi Sistemami*, vol. 2, pp. 3–12, 2016.

[3] T. V. Miroliubova and E. N. Voronchikhina, "Determination of key focus areas of regional economic policy on the basis of regional economic growth econometric modeling and forecasting," *Vestnik Permskogo Universiteta*, vol. 4, pp. 80–91, 2015.

[4] "Program for the development of the digital economy in the russian federation until 2035." [Online]. Available: http://spkurdyumov.ru/uploads/2017/05/strategy.pdf

[5] "Readiness of russian regions to the information society." [Online]. Available: http://eregion.ru/