International Journal on Emerging Technologies **10**(2a): 201-205(2019)

ISSN No. (Print) : 0975-8364
ISSN No. (Online) : 2249-3255

Development of a Regression Model of Dependence of Territorial Development of the Republic of Tatarstan on Implementation of Infrastructure Projects

Julia Stanislavovna Korchagina
Department of Geography and Cartography, Kazan Federal University, Russia.

(Corresponding author: Julia Stanislavovna Korchagina)
(Received 05 May 2019, Revised 16 July 2019 Accepted 25 July 2019)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: The article is devoted to development of a model of regression dependence of territorial development of Republic of Tatarstan on implementation of infrastructure projects. The relevance of the studied issues is due to a number of factors. The main ones are: satisfaction of infrastructure needs of the population in the form of a range of services and, as a result, improving the quality of life of the population. In addition, infrastructure projects in modern realities are tools for the implementation of strategic plans. As elements of federal and regional development programs, they are included in the group of priority national projects approved by the President of the Russian Federation. According to experts, infrastructure projects can become drivers of the socio-economic development of territories of various levels, bringing them to a new level of existence. Thus, the built-up regression model of dependence is aimed at identifying the role of infrastructure projects in the territorial development of the Republic of Tatarstan. On the way to achieving the goal, the stages of the development of the methodology are shown with a detailed description of the methods used, the rationale for their use in relation to the specifics of the Tatarstan region is presented, the difficulties encountered and the chosen ways to overcome them are described, and the conclusions are presented. The result of the study was a regression model with socio-economic indicators of the dependence of the gross territorial product of the municipal regions of the republic on the financing of infrastructure projects in them. Verification of the model showed its compliance with all Gauss-Markov conditions, which proves its applicability for predicting the process in the region.

Keywords: Infrastructure projects, territorial development, regression dependence, project management, strategic planning, econometric modeling, regional economy.

I. INTRODUCTION

Modern realities call for considering infrastructure projects not only as objects of providing services to the population and satisfying needs, but as integrated drivers of territorial development that can bring the region to a new level of socio-economic development [10]. In connection with this understanding of infrastructure projects, it is advisable, using econometric methods, to find out their role for the purposes of territorial development. In addition, the relevance of the topic is due to the fact that infrastructure projects are the basic tools for the implementation of strategic planning goals, which are currently being actively developed in Russia and the Republic of Tatarstan in particular. The basic concepts and paradigms of social and economic geography used in the work are: industry paradigm (infrastructure geography), regional paradigm, system-structural approach [6]. An analysis of the problems associated with the nature and practice of project risk management was carried out by Srinivas [8]. Issues of the methodology and conceptual framework of social design during the implementation of regional development projects have been thoroughly investigated by famous scientists - Smyth and Vanclay [9].

The study was conducted on the example of Republic of Tatarstan- one of the leading regions in the socio-economic development of Russian Federation. The object was infrastructure projects as a factor in the territorial development of the region's economy. The main goal is, on the basis of theoretical and methodological approaches to the content of infrastructure projects, to find out their roles in ensuring the ongoing socio-economic development of the territory (for example, Republic of Tatarstan).

II. METHODS

Considering the infrastructure project primarily as a driver of territorial development, it is advisable to use indicators reflecting the development of the territory (region, municipal district) to assess its implementation. In this regard, the following factors were chosen as the explanatory factors for building an econometric model: "number of infrastructure projects" and their "financing by municipal districts of the Republic of Tatarstan", and the gross territorial product (GTP) of the municipalities of the Republic of Tatarstan was adopted as the explained indicator.

Thus, the main method was the econometric model of the regression dependence of the selected indicators. The selection consisted of 27 municipal districts of the Republic of Tatarstan (including 2 urban districts: Kazan and Naberezhnye Chelny, which act as independent subjects). The choice of this number of analyzed units is due to the presence in them of implemented infrastructure projects by the reporting period of 2017. The statistics are taken from the website of the Ministry of Economy of the Republic of Tatarstan, the source is the Plan for the creation of infrastructure facilities in the Republic of Tatarstan [1].

The econometric model of paired regression was tested according to Gauss-Markov conditions, during which conclusions were made about the presence and strength of the correlation relationship between factors, the statistical significance of factors was evaluated, the overall quality level of the model was checked, and the conditions of the absence of heteroscedasticity and multicollinearity were verified. As a result, conclusions were made about the possibilities of practical use of the model for the purposes of forecasting the development of processes, taking into account the strategic goals.

III. RESULTS AND DISCUSSION

In our work, we used the multiple regression model (in this case, it is given by two explanatory variables x_1 and x_2), which has the form [5]:

$$y = \hat{f}\,(x_1, x_2, \ldots x_k)$$

where y is the explained variable (GTP); x_1, x_2 - explaining variables (number of infrastructure projects, financing of infrastructure projects in municipal districts of the Republic of Tatarstan).

Baseline data are presented in Table 1 [1, 4].

The selection consists of 27 units. Accordingly, n = 45, m = 2.

First, we construct the matrix R [5].

$$R = \begin{pmatrix} 1 & & \\ Ryx_1 & 1 & \\ Ryx_2 & Rx_1Rx_2 & 1 \end{pmatrix}|$$

Table 1.

Municipalities		GTP	Number of projects	Project financing
S.No.	**Name**	**y**	**x 1**	**x 2**
1	Kazan	604300	48	3288999,687
2	NaberezhnyeChelny	163300	24	1394336.3
3	Agryzsky	6553.7	1	1814
4	Aznakaevsky	45330	1	598
5	Aktanyshsky	9000	3	3065,181
6	Almetyevsky	253,000	6	692,336.69
7	Bugulminsky	47461	1	1018.2
8	Buinsky	9792.9	2	3002.5
9	Verkhneuslonsky	4895.8	3	35
10	Vysokogorsky	14037	1	2270.6
11	Drozhzhanovsky	5420	2	2302.5
12	Yelabuzhsky	35,500	19	1666729,79
13	Zainsky	29300	2	48320
14	Zelenodolsky	33200	11	24647,8
15	Kukmorsky	8898.2	1	1815
16	Laishevsky	19608,4	15	124546,4466
17	Leninogorsk	58800	1	28866.9
18	Mamadyshsky	7370	1	eight
19	Mendeleevsky	12,400	4	45632,657
20	Menzelinsky	5081	1	6000
21	Nizhnekamsky	172100	19	482844.92
22	Pestrechinsky	6183.3	3	2452.24
23	Rybno-Slobodsky	4746	8	2719,308
24	Sabinsky	10210.75	4	1819.1
25	Sarmanovsky	22851.5	1	70
26	Tukaevsky	31700	3	1714.04
27	Chistopolsky	14329.8	2	4347.5
Σ		1635369.35	187	7832312.36
Average		60569,2352	6.9259	290085,6429

(1) Calculate multicollinearity and correlation coefficients using the formulas [5]:

$$\mathrm{Rx1x2} = \frac{\Sigma(\mathrm{xi1} - \dot{\mathrm{x}}1)(\mathrm{xi2} - \dot{\mathrm{x}}2)}{\sqrt{\Sigma(\mathrm{xi1} - \dot{\mathrm{x}}1)^2 * \Sigma(\mathrm{xi2} - \dot{\mathrm{x}}2)^2}} = 0{,}9249$$

$$\mathrm{Rx1y} = \frac{\Sigma(\mathrm{xi1} - \dot{\mathrm{x}}1)(\mathrm{yi} - \bar{\mathrm{y}})}{\sqrt{\Sigma(\mathrm{xi1} - \dot{\mathrm{x}}1)^2 * \Sigma(\mathrm{yi} - \bar{\mathrm{y}})^2}} = 0{,}8394$$

$$\mathrm{Rx2y} = \frac{\Sigma(\mathrm{xi2} - \dot{\mathrm{x}}2)(\mathrm{yi} - \bar{\mathrm{y}})}{\sqrt{\Sigma(\mathrm{xi2} - \dot{\mathrm{x}}2)^2 * \Sigma(\mathrm{yi} - \bar{\mathrm{y}})^2}} = 0{,}8723$$

It is worth noting the significant link between the explanatory variables and the factor y ($r_{xy} \geq 0.7$), however, due to the fact that the multicollinearity of the variables x_1 and x_2 was revealed ($\mathrm{Rx1\ x2} = 0{,}9249$) , then one of the factors must be excluded. The connection x_2 with y is slightly higher than with x_1, therefore the further model will be lined up in a steam regression format with variables *x 2* and *y* [5].

Paired regression is a model with two variables - *y* and *x*, i.e. view model [5]:

$$(y = \hat{f}(x)$$

where *y* is the explained variable (GTP); *x* - explanatory variable (financing of infrastructure projects).

Next, you need to find *b0* and *b1* from the formulas shown below [5].

$$b0 = \bar{y} - b_1\bar{x} = 17773{,}8647$$

$$b1 = \frac{\overline{xy} - \bar{x} * \bar{y}}{\overline{x^2} - (\bar{x})^2} = 0{,}1475$$

The next step is the verification of the resulting model to Gauss-Markov conditions.

For a start, it is worth noting that the positive sign of the coefficient *b1* shows the unidirectionality of *x2* and *y* changes, that is, with an increase in funding for infrastructure projects, the GTP in the municipality also increases, and vice versa [5].

(1) Already calculated the sample correlation coefficient (r_{x2y} = 0.8723) indicates a significant relationship of variables [5]. Thus, we conclude that on the basis of the correlation coefficient, x_2 must be present in the model.

(2) Let us turn to the calculation of the statistical significance of the analyzed model of the pair regression.

To do this, it is necessary to calculate the indicator t_{calc}, having previously calculated such indicators as: $Sbj, S^2{}_{bj}, S_0{}^2$. The formulas and calculation results are presented below [5].

$$t_{calc} = \left|\frac{b_j}{S_{bj}}\right| = 8{,}9215$$

$$Sbj = \sqrt{S^2{}_{bj}} = 0{,}165$$

$$S^2{}_{bj} = \frac{1}{\sum(x_i - \bar{x})^2} * S_0{}^2 = 0{,}0003$$

$$S_0{}^2 = \frac{\sum e_i{}^2}{n-m-1} = 3829537863{,}6944$$

t_{calc} allows to make a conclusion about the statistical significance of the explanatory variable. Since $t_{calc} \geq 3$, therefore the factor x_2 is strongly significant [2].

3) The model is checked for the level of overall quality by calculating [5]:

$$F_{calc} = \frac{R^2}{1-R^2} * \frac{n-m-1}{m} = 79{,}5766$$

$$R^2 = 1 - \frac{\sum e_i{}^2}{(\sum(y_i - \bar{y})^2)} = 0{,}7609$$

From the obtained R^2 value, we can conclude that the factor *x2* explains the behavior of *y* by about 76%.

From Table Fischer define F_{cr} (for α = 0,05, V1 = *m* = 1, V2 = n-m-1 = 25)

F_{kr} = 4,245

$F_{calc} > F_{kr}$, therefore the model has a good level of quality [2].

4) It is also necessary to make a check for the absence of autocorrelation of residuals. For this, there are the DW method and the series method, we will use both.

(a) DW method. We calculate this indicator by the formula [5]:

$$DW = \frac{\sum(e_i - e_i^1)}{\sum e_i^2} = 1{,}9741$$

According to the Durbin-Watson table (*n* = 27, *m* = 1), we find critical points, they are schematically shown in Fig. 1.

Dl = 1,316

Du = 1,469

4 - *Dl* = 2,684

4- *Du* = 2.531

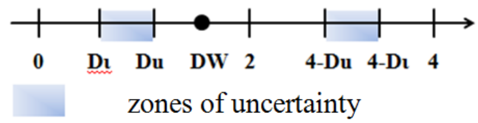

Fig. 1. Critical points and areas of uncertainty for the DW model.

As can be seen from the graph, DW does not fall into the zone of uncertainty, but enters the interval ($d_u \leq DW \leq 4 - d_u$), there is therefore no autocorrelation [2; 5].

b. Row method.

For this method, it is necessary to write out the characters in a row with the indicator *e* for the entire sample population. We get the following sequence:

+ - + - ++ ----- ++ - + --- + --- ++ -

The number of sequences: k = 14.

The number of positive signs: n1 = 10.

The number of negative characters: n2 = 17.

According to the table of critical values for the series method, we find the lower (k1 = 8) and upper (k2 = 19) boundaries.

k = 14 falls into the interval (8; 19); therefore, there is no autocorrelation of residuals [2, 5].

Both methods confirmed the absence of autocorrelation of residuals.

5) We will check for heteroscedasticity. To do this, fill in the following Table 2 [5].

Table 1.

x2	N_x2	|ei|	Nei	dì	di²
3288999,687	27	101310,9488	25	2	4
1394336,3	25	60175,6596	23	2	4
1814	7	11487,7734	10	-3	9
598	4	27467,9190	20	-16	256
3065,181	15	9226,0560	8	7	49
692336,69	24	133088,0095	26	-2	4
1018,2	5	29536,9283	21	-16	256
3002,5	14	8423,9088	6	8	64
35	2	12883,2234	15	-13	169
2270,6	10	4071,8341	1	9	81
2302,5	11	12693,5402	14	-3	9
1666729,79	26	228160,9657	27	-1	1
48320	21	4397,6510	3	18	324
24647,8	18	11789,9320	11	7	49
1815	8	9143,4209	7	1	1
124546,4466	22	16539,3833	19	3	9
28866,9	19	36767,5022	22	-3	9
8	1	10405,0402	9	-8	64
45632,657	20	12105,8942	13	7	49
6000	17	13578,0201	17	0	0
482844,92	23	83093,6339	24	-1	1
2452,24	12	11952,3308	12	0	0
2719,308	13	13429,0305	16	-3	9
1819,1	9	7831,4758	5	4	16
70	3	5067,3131	4	-1	1
1714,04	6	13673,2734	18	-12	144
4347,5	16	4085,4322	2	14	196

For further calculations, we need to calculate the sum of squares in *dì*. In our example, it is equal to: $\sum d_i^2$ = 1778 [5].

Next, we calculate the indicators r_{xe} (Test Spearman) and t_{calc} according to the following formulas [5]:

$$r_{x|e|} = 1 - 6\frac{\sum d_i{}^2}{n(n^2-1)} = 0{,}4573$$

$t_{calc} = \left|\frac{r_{xi|e|} * \sqrt{n-2}}{\sqrt{1-r_{xi|e|}^2}}\right|$=2,5708

Rx|e| ϵ [0.3; 0.7), we can conclude that the relationship exists. According to a rough estimate t_{calc} ϵ [2; 3), which means that the factor x_2 is significant. Calculate t_{cr} on Student's table (α = 0.05, v = n - m -1 = 25). t_{cr} = 2.06 [5].

Conclusion: $t_{calc} > t_{cr}$ -heteroscedasticity is present [2].

This indicates the presence of sufficient data scatter. Attempting to eliminate heteroscedasticity by eliminating units from the model that make up less than 0.5% of the mean value in *x*2 did not produce results. Therefore, in order to level this problem and eliminate heteroscedasticity, we use the Box-Jenkins method [5].

According to him, it is considered that the magnitude of the deviations depends on the magnitude of the explanatory variables.

$$\sigma i^2 = \sigma^2 * xi^2$$

Consequently, the model is converted from the form: for y = b0 + b1* x_i to a model of the following form [5]:

$$\frac{yi}{xi} = b0 * \frac{1}{xi} + b1$$

To estimate it, it is necessary to recalculate the initial data (xi; yi) to (x_i^*; y_i^*) as follows [5]:

$$x_i^* = \frac{1}{Xi}; \qquad y_i^* = \frac{yi}{xi}$$

For the new selection, we calculate the model, which is given by the following equation: $y^* = b_0^* x_i^* + b_1^*$,

where b_0^* is the coefficient of the explaining variable

b_1^* is the free term of the equation.

The formulas and calculations for b_0^* and b_1^* are presented below [5].

$$b1^* = \frac{\overline{xy}^* - \bar{x}^* * \bar{y}^*}{\overline{x^2} - (\bar{x}^*)^2} = \frac{113,7981}{0,0155} = 7337,2897$$

$$b0^* = \bar{y} - b_1\bar{x} = 58,1935 - 7337,2897 * 0,0065 = 10,4783$$

Now that we have found the values of b_0^* and b_1^*, we will calculate this model, first of all we are interested in checking for heteroscedasticity. Similar calculations, but with the substitution of new values of the parameters, given the result of d_i^2 is equal to 2470.5 [2].

Calculate the indicator $r_{x|e|}$:

$r_{x|e|} = 1 - 6\frac{\sum d_i^2}{n(n^2-1)}$=0,2459

Check the result of student statistics.

$t_{calc} = \left|\frac{r_{xi|e|} * \sqrt{n-2}}{\sqrt{1-r_{xi|e|}^2}}\right|$=1,2683

According to a rough estimate t_{calc} ϵ [1; 2), which means that the factor x_2 is weak. Calculate t_{cr} on Student's table (α = 0.05, v = n -m -1 = 25). t_{cr} = 2,060.

Conclusion: $t_{cr} > t_{calc}$ - heteroscedasticity is absent [2].

For the initial dependence, the equation takes the following form: y = b_0^* + b_1^* x_i . According to the theory of econometrics, it is guaranteed that there is no heteroscedasticity. Thus, the applied Box-Jenkins method allowed us to eliminate heteroscedasticity in the model [5].

We will also make a prediction on the model obtained using the example of the Nizhnekamsk municipal district. For the planned period until 2020, this municipality plans to finance infrastructure projects in the amount of 3 million rubles [3]. Thus, calculating the value of *y* by the model equation, we obtain a value equal to 22022.3474 (y is also measured in million rubles), thus, it can be seen that funding for this amount will allow an increase in GTP by about 13% (calculated relative to *y* = 172100) (GTP in the prices of the reporting period of 2017 excluding inflation and the forecast of GTP dynamics for the planning period up to 2020) [7].

Check the result of student statistics

IV. SUMMARY

Thus, the developed model of the regression dependence of the gross territorial product on the costs of infrastructure projects in the municipal regions of the Republic of Tatarstan has proved its worth by checking for compliance with the Gauss-Markov conditions. The study proved the statistical significance of the explanatory indicator, significant correlation of indicators. The direct dependence of the selected coefficients was revealed: financing of infrastructure and gross territorial product in the municipalities of the republic. For a specific example, a predictive calculation is made. The role of infrastructure projects for the progressive socio-economic development of the region is proved.

V. CONCLUSIONS

In conclusion, we note that with the recalculation of the regression model, she went through all the stages of checking the Gauss-Markov conditions, during which it was determined that there was no autocorrelation of residuals in the model (proved by DW and row methods) and heteroscedasticity (this item was made possible by recalculating Box-Jenkins models).

Estimates proved the unidirectionality of the changes explaining and explaining the variable (with an increase in the financing of infrastructure projects in the municipality the gross territorial product grows), a high level of correlation of variables (rx2y = 0.8723) and statistical significance of the parameter (t_{calc} = 8.9215; $t_{calc} \geq 3$).

Conclusions about a good quality level of the model were made ($F_{calc} > F_{cr}$). It was also determined that the financing of infrastructure projects by about 76% determines the value of the gross territorial product in the municipal districts of the Republic of Tatarstan.

Thus, this model can be used in practical activities, including for the calculation of projected indicators of GTP changes for the municipal districts of the Republic of Tatarstan, depending on the allocated funding for infrastructure projects.

ACKNOWLEDGEMENTS

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

We are also grateful for the help in the work of the candidate of economic sciences, associate professor of the department of economic theory and econometrics of Kazan (Volga region) Federal University – Balashova Elizaveta Yakovlevna.

REFERENCES

[1]. Plan for the creation of infrastructure facilities in the Republic of Tatarstan [Electronic resource]. – URL. http://mert.tatarstan.ru/rus/plan-sozdaniya-investitsionnih-obektov-i-obektov.htm (date of the application: 30.09.2018).

[2]. Statistical tables [Electronic resource]. URL: http://www.rsatu.ru/sites/vmat/?doc=1381995869 (date of the application: 19.10.2018).

[3]. Strategy of social and economic development of the Republic of Tatarstan until 2030 [Electronic resource]. – URL: http://mert.tatarstan.ru/strategiya-sotsialno-ekonomicheskogo-razvitiya.htm Table of critical values of Student's t-test. URL: https://statpsy.ru/t-student/t-test-tablica/ (date of the application: 10.10.2018).

[4]. Territorial agency of the Federal State Statistics Service for the Republic of Tatarstan [Electronic resource]. URL: http://tatstat.gks.ru/ (date of the application: 25.09.2018).
[5]. Econometrics. Lecture notes, E. I. Kadochnikova; Kaz Feder. un-t - Kazan, 2013, 106 p.
[6]. A Guide to the Project Management Body of Knowledge. PMI. Edition 6, (2016.
[7]. A rating system for sustainable infrastructure / Institute for sustainable infrastructure[Electronic resource]. – URL: https://www.colorado.edu/center/mortenson/sites/default/files/attached-files/envision_v6_isi.pdf (date of access: 13.03.2019).
[8]. Srinivas, K. (2019). Process of Risk Management, Perspectives on Risk, Assessment and Management Paradigms, Ali G. Hessami, IntechOpen, DOI: 10.5772/intechopen.80804. Available from: https://www.intechopen.com/books/perspectives-on-risk-assessment-and-management-paradigms/process-of-risk-management.
[9]. Eddie S. and Vanclay, F. (2017). The Social Framework for Projects: a conceptual but practical model to assist in assessing, planning and managing the social impacts of projects, Impact Assessment and Project Appraisal, *35*: 1, 65-80. DOI: 10.1080/14615517.2016.1271539
[10]. Rostow, W. W. (1962). The Stages of Economic Growth. London: Cambridge University Press, 324 p.
[11]. Kamberov, R., Granell, C., & Santos, V. (2017). Sociology Paradigms for Dynamic Integration of Devices into a Context-Aware System. *Journal of Information Systems Engineering & Management*, *2*(1), 2. https://doi.org/10.20897/jisem.201702.
[12]. Salimi, M., Khoddam, K., Morakkabatchy, D., & Pornadem, M. (2015). Optimization of carbon nanotube field-effect transistors (CNTFET) and compare them to CMOS silicon. *UCT Journal of Research in Science, Engineering and Technology*, *3*(4), 10-16.