The application of models of panel data analysis for the estimation of the volume of innovative goods, works and services in the Russian Federation

M. M. Tsvil[1], V. E. Shumilina[2]

[1]Russian customs academy (Rostov affiliate), Rostov-on-Don;
[2] Don State Technical University, Rostov-on-Don.

Abstract: This article deals with the carrying out of econometric analysis by means of panel data models (pooled model, fixed effects model) of the volume of innovative goods, works and services in the Russian Federation (RF) from 2010 up to 2014 years. The dependence has been revealed between the volume of innovative goods, works, services and such explanatory factors as the number of personnel involved in scientific research and development; internal expenses for the scientific researched and development of the Russian Federation; used advanced manufacturing technologies; coefficient of inventive activity; innovation activity of organizations.
Keywords: econometric analysis, panel data, innovation, science, integrated model of panel data (pooled model), a model of panel data with fixed effects (fixed effects model).

1. Introduction

In recent years, in terms of sanctions policy from Western countries more and more attention in the Russian Federation (RF) is paid to the creation and sale of innovative goods and to the provision of innovative works and services.

As is known, innovations are the central factor of production and productivity growth. The emergence of innovations is directly connected with the development of science. Science provides the economy and society with highly qualified personnel, revolutionary technological solutions and new technical ideas for everyday practical usage.

However, according to the report of the Russian Association of the science promotion (RASP) about the state of science in the Russian Federation, presented in 2013, the current state of the Russian science according to a number of objective indicators is characterized as catastrophic. One of the main reasons named was the organizational reason, i.e. the lack of a distinct state strategy in scientific and technical sphere. The problems of a lower level are arisen from this main problem, e.g. the rapid delay of Russia in the global field of scientific researches, the problem of reproduction of qualified scientific and engineering personnel, the problem of introducing of scientific and technical innovations into the production and their incorporation into the chain of sectoral and intersectoral economic connections (bringing to the final product and to the final consumer); hence the problem of low share of small and medium enterprises in the structure of high-tech production and low share of industries with high added value in the GDP structure. Consequently, there is no large-scale demand for scientific developments by private companies and that factor exacerbates all the above-described problems and leads to the development of negative tendencies in self-reinforcing spirals.

At the St. Petersburg international economic forum on June 16-18, 2016 at the session «Big challenges – the stimulus of science development» the participants of which were the leading scientists and organizers of domestic and foreign science, the Chairman of the Board of the managing company «RUSNANO» A. Chubais mentioned that «Science itself has not been demanded by the production, science disturbs production as it requires big changes from any production». In his opinion, innovative economy should become the connecting link between science and production: «We need not expect that the production will create a request for science. It is needed to consider the chain of three elements: science - innovation economy - production. The demand for science from the side of manufacturers can be provided by innovative economy».

At that session, the member of the Expert Council under the Government of Russian Federation, Ekaterina Shapochka, pointed out the large data amount in the modern world, which is difficult to process by means of traditional ways; the Chairman of one of the world's largest scientific publishers, Elsevier B. V., Ionsuk Chi, noted that the value of data was increasing when they were structured, and even more – when analyzed.

The actuality of this topic determined the purpose of the investigation described in the given article.

2. The aim and methods of the investigation

The object of study is the volume of innovative goods, works and services in the Russian Federation (RF) from 2010 to 2014 (The data of Federal Statistics Service[1]). The objective of the investigation is to build adequate and substantial models to examine the true cause-effect relations between the volume of innovative goods, works, services in Russia (Y) in million rubles, and the explanatory variables: X_1 – number of employees involved in scientific research and development (person); X_2 – internal costs on scientific investigation and development of the Russian Federation in million rubles; X_3 – used advanced manufacturing technologies (units); X_4 – coefficient of inventive activity (number of domestic patent applications for inventions submitted in Russia per 10 thousand persons of population); X_5 – innovative activity of organizations (specific gravity of organizations in % carrying out technological, organizational and marketing innovations in the reporting year, in the total number of the surveyed organizations) in eight Federal districts.

The experience of the econometric methods is of great interest in terms of uncertainty. Econometrics has become a powerful tool of economic research, especially in recent years as a result of the development of computer systems and special applied programs. The constantly complicated social and economic processes have led to the necessity of econometric analysis usage. The study of these processes is carried out by means of econometric models. The majority of new research methods are based on the econometric models, concepts and techniques. The application of econometric methods towards the real Russian statistical data will allow to understand deeper the goals and objectives of the state economic policy (or the company) and also to learn how to evaluate the results of this policy [2 - 6]. One of the possible solutions is the usage of the panel data. The application of estimation methods of the paneldata to solve the formulated task seems to be reasonable as the temporary numbers data for Russia on the whole are insufficient to obtain good parameter estimations.

The panel data are those data, which contain statistical information about one and the same object set about for the number of consecutive time periods. Due to the structure the panel data allow to construct more flexible and substantial models and obtain answers to the questions that are unavailable only in the framework of models based on spatial data, for example. For economists, the panel data are valuable because it appears a possibility to consider and analyze the individual differences between the economic units that can't be done within the framework of standard regression models. Panel data models allow us to obtain more accurate estimated parameters [7 - 10].

3. The pooled models construction

In our case, the econometric analysis is based on panel data for the eight regions in the period from 2010 to 2014. The source of these data was the statistical data of the State Federal Service.

Table № 1.

Panel data in the Federal districts of the Russian Federation in the years 2010-2014

Federal districts	Time (year)	Y	X_1	X_2	X_3	X_4	X_5
1. Central	2010	290 757.6	381795	288960.0	68945	3.8	8.6
	2011	480 327.4	380363	331758.9	63078	3.27	10.2
	2012	938 153.2	373461	369069.5	62796	3.71	10.9
	2013	1 164 102.4	375087	398597.2	60829	3.77	10.7
	2014	1 091 170.3	381047	447161.2	65591	3.0	10.9
2. North-west	2010	120 105.5	95826	70737.3	17920	1.66	9.4

	2011	196 049.1	97221	81504.9	19308	1.87	11.2
	2012	298 020.1	97710	100002.7	18840	1.67	11.0
	2013	409 750.4	95674	108026.7	19697	1.55	10.7
	2014	354 113.0	96726	118612.3	20840	1.57	10.3
3. South	2010	86 558.4	28109	13027.3	7623	1.2	7.5
	2011	59 811.8	27738	15906.0	7394	1.07	6.5
	2012	51 801.6	23964	18618.0	7848	1.14	7.4
	2013	70 281.9	24263	19987.0	8290	1.16	7.2
	2014	102 845.3	25361	29274.3	9580	1.12	7.7
4. North-Caucasian	2010	27 682.6	6053	2639.8	3194	2.01	6.2
	2011	31 941.8	8585	4017.7	1993	1.24	5.2
	2012	27 010.1	7188	3448.1	1833	1.61	6.4
	2013	23 889.8	6330	3695.5	2113	1.74	5.9
	2014	27 961.5	6628	4197.3	2215	0.71	6.5
4. North-Caucasian	2010	27 682.6	6053	2639.8	3194	2.01	6.2
	2011	31 941.8	8585	4017.7	1993	1.24	5.2
	2012	27 010.1	7188	3448.1	1833	1.61	6.4
	2013	23 889.8	6330	3695.5	2113	1.74	5.9
	2014	27 961.5	6628	4197.3	2215	0.71	6.5
5. Privolzhskiy	2010	545 954.9	116285	74942.4	57394	1.38	12.3
	2011	781 944.9	111579	91012.1	55822	1.5	12.7
	2012	950 604.8	114204	109155.0	54976	1.55	11.9
	2013	1 128 642.7	114013	114194.6	57076	1.49	11.7
	2014	1 179 545.3	107656	126552.5	59643	1.36	11.4

6. Ural	2010	109 584.6	42672	29441.8	35596	0.96	11.5
	2011	179 708.9	43586	34408.9	30323	1.05	11.5
	2012	148 696.2	43879	40420.2	31962	1.03	10.6
	2013	189 234.1	44382	45167.0	31217	1.03	9.6
	2014	169 373.1	45037	48800.0	29617	0.91	8.9
7. Siberian	2010	46 890.0	53024	33870.0	16335	1.25	8.2
	2011	88 866.0	52794	40713.4	15079	1.25	8.8
	2012	117 118.0	52685	47011.7	15897	1.29	8.5
	2013	151 362.7	53769	47666.3	16643	1.23	9.1
	2014	186 025.2	54151	58435.9	18063	1.13	8.8
8. Far-Eastern	2010	16 178.9	12776	9758.7	5589	0.89	8.6
	2011	288 090.7	13407	11104.7	6595	0.89	11.2
	2012	341 501.1	13227	12144.6	5810	0.93	10.8
	2013	370 602.1	13227	12144.6	6801	1.04	9.5
	2014	468 731.8	13204	13714.3	6956	0.95	8.9

For the econometric models with panel data the empirical analysis begins with the choice between models with a common effect (pooled model) and models with fixed effects (fixed effect model).

Pooled model – is the usual linear regression model, which in matrix form looks like this:

$$\hat{y} = X\beta + \varepsilon, \quad (1)$$

for the coefficients estimation of which the ordinary least-squares method (LSM) may be used.

In our case the built model on the basis of the panel data the pooled model (1) has the following form:

$$\hat{y} = -451151.421 - 7.794X_1 + 6.52X_2 + 12.625X_3 + 176997.2X_4 + 361112243X_5 \qquad (2)$$

The absence in panel structure data and the possibility to receive consistent and effective assessments by means of the pooled sample with the help of LSM is formulated as a null hypothesis in F-test.

The determination coefficient of the given model (2) is $R^2 = 0.82$. The dependence of Y from X_1, X_2, X_3, X_4, X_5 is characterized as close in which 82% of the volume variations of innovative goods, works and services is determined by the variation considered in the model factors: the number of personel involved in scientific research and development; internal expenses for the scientific research and development throughout the Russian Federation; the used advanced manufacturing technologies; inventive activity coefficient; innovative activity of organizations in eight Federal districts.

The regression equation according to F–Fisher criterion is statistically significant. Let's give the main parameters of the model (2):

Table № 2.

Main parameters of the model (2)

Index	Coefficient	t-statistics	P-meaning
β_0	-451151.421	-1.961548322	0.058045124
β_1	-7.794	-4.629470975	5.16537E-05
β_2	6.520	5.352415239	5.98749E-06
β_3	12.625	4.383314752	0.000106584
β_4	176997.200	1.621723842	0.114100742
β_5	36112.243	1.735211339	0.091760779

As is seen from the table 2 the coefficients of the model (2) are mainly all statistically significant according to Student's test with a significance level $\alpha = 0.1$ and the number of degrees of freedom equal to 34 ($t_\alpha = 1.68$).

It is evident from the given model that the growth of volume of innovative goods, works and services is certainly positively influenced by a growth of the coefficient of inventive activity (X_4) and innovative activity of organizations (X_5). With the values X_4 and X_5 equal to 1 and X_1, X_2, X_3 equal to 0, the volume of innovative goods, works and services is equal to 86958.022 million rubles.

With a significance level $\alpha = 0.05$ the coefficients β_0, β_4 and β_5 are not significant according to Student's criterion. It is explained by the fact that only the Central Federal district has simultaneously significant coefficients of inventive activity and innovative activity of organizations.

Let's compare the pooled model with the fixed effect model allowing to estimate the influence of values of quantitative attributes in each region separately.

In matrix form the model looks like this:

$$У = X\beta + Z\alpha + \varepsilon, \tag{3}$$

where $Z = (Z_1, Z_2, ..., Z_8)$.

The model (3) presupposes the introduction of dummy variables $Z = (Z_1, Z_2, ..., Z_8)$ for the sample objects. The coefficients in dummy variables will give the valuations of the individual effects. Fixed effects model ensures the guaranteed receiving of unbiasedand consistent valuations. In our case, Z_1 takes the value 1 for the data of the Central Federal district, for other districts it is 0; Z_2 takes the value 1 for the data of the North-Western Federal district and for the other districtsit is 0; Z_3 takes the value 1 for the data of the Southern Federal district and for other districts it is 0 etc.

Built on the basis of the panel data the fixed effects model has the following form:

$$У = 5534088.579 - 19.437X_1 + 6.26X_2 - 1.34X_3 + 106182.495X_4 + 2749.936X_5 + 0 \cdot Z_1 -$$
$$- 4159517.582Z_2 - -5207747.819Z_3 - 5562511.087Z_4 - 3182346.686Z_5 - 4861327.279Z_6 -$$
$$- 4797898.101Z_7 - 5173066.687Z_8 \tag{4}$$

The coefficient of determination of the model (4) $R^2 = 0.96$ is statistically significant. From the constructed model it is followed that the factor Y is mainly positively influenced by the internal costs (X_2), innovative activity of organizations (X_4), the coefficient of inventive activity (X_5).

So, for the Central Federal district the model (4) has the following form:

$$У = 5534088.579 - 19.437X_1 + 6.26X_2 - 1.34X_3 + 106182.495X_4 + 2749.936X_5, \quad (5)$$

where $Z_1 = 1$, $Z_2 = Z_3 = Z_4 = Z_5 = Z_6 = Z_7 = Z_8 = 0$.

For The North-West district the model has the following form:

$$У = 1374579.997 - 19.437X_1 + 6.26X_2 - 1.34X_3 + 106182.495X_4 + 2749.936X_5 \quad (6)$$

where $Z_2 = 1$, $Z_1 = Z_3 = Z_4 = Z_5 = Z_6 = Z_7 = Z_8 = 0$.

For the South district the model has the following form:

$$У = 326340.76 - 19.437X_1 + 6.26X_2 - 1.34X_3 + 106182.495X_4 + 2749.936X_5 \quad (7)$$

where $Z_3 = 1$, $Z_1 = Z_2 = Z_4 = Z_5 = Z_6 = Z_7 = Z_8 = 0$.

For The North Caucasus district the model has the following form:

$$У = -28422.508 - 19.437X_1 + 6.26X_2 - 1.34X_3 + 106182.495X_4 + 2749.936X_5, \quad (8)$$

where $Z_4 = 1$, $Z_1 = Z_2 = Z_3 = Z_5 = Z_6 = Z_7 = Z_8 = 0$.

For the Privoljski district the model has the following form:

$$У = 2351741.893 - 19.437X_1 + 6.26X_2 - 1.34X_3 + 106182.495X_4 + 2749.936X_5, \quad (9)$$

where $Z_5 = 1$, $Z_1 = Z_2 = Z_3 = Z_4 = Z_6 = Z_7 = Z_8 = 0$.

For the Ural district the model has the following form:

$$У = 672761.3 - 19.437X_1 + 6.26X_2 - 1.34X_3 + 106182.495X_4 + 2749.936X_5, \quad (10)$$

where $Z_6 = 1$, $Z_1 = Z_2 = Z_3 = Z_4 = Z_5 = Z_7 = Z_8 = 0$.

For The Siberian district the model has the following form:

$$У = 736190.478 - 19.437X_1 + 6.26X_2 - 1.34X_3 + 106182.495X_4 + 2749.936X_5, \quad (11)$$

where $Z_7 = 1$, $Z_1 = Z_2 = Z_3 = Z_4 = Z_5 = Z_6 = Z_8 = 0$.

And for the Far –Eastern district the model has the following form:

$$У = 361021.892 - 19.437X_1 + 6.26X_2 - 1.34X_3 + 106182.495X_4 + 2749.936X_5 \quad (12)$$

where $Z_8 = 1$, $Z_1 = Z_2 = Z_3 = Z_4 = Z_5 = Z_6 = Z_7 = 0$.

4. Conclusion

Thus, the practical significance of the models (5-12) is that it will allow to predict and calculate the volume of innovative goods, works and services taking into account the number of personnel involved in scientific research and development; internal costs for the research and development; used advanced manufacturing technologies; coefficient of inventive activity and innovation activity of organizations in each of the above mentioned regions of the Russian Federation.

Notes: In the course of writing of this article on the site of Federal State Statistics (Official Rosstat Statistics [1]) there have appeared the actual data for the year 2015, according to which, the meaning Y, for example, for the Central Federal district was 1491536.1 million rubles, and the prognostic value according to the model (5) in accordance with the data of 2015 was 1562198.7678 million rubles, that evidences about the high prognostic quality.

On the basis of the above mentioned it is possible to make a conclusion that the innovation development has the great influence on the economic development of the Russian Federation under present-day conditions. The volume increase of the innovative goods, works and services is positively influenced by the innovation activity of organizations, inventive activity and domestic costs on research and development.

References

1. Dannye Federal'noy sluzhby gosudarstvennoy statistiki. URL: gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/science_and_innovations/science/.

2. Tsvil' M.M., Kolesnikova I.V. Inženernyj vestnik Dona (Rus), 2016, №4 URL: ivdon.ru/ru/magazine/archive/n2y2016/3591.

3. Tsvil' M.M., P'yanova Yu.S. Aktual'nye problemy gumanitarnykh i estestvennykh nauk, 2016. № 6-2, pp. 96-99.

4. Tsvil' M.M., Shumilina V.E. Inženernyj vestnik Dona (Rus), 2014, №4 URL: ivdon.ru/ru/magazine/archive/ N4y2014/2555.

5. Tsvil' M.M., Shumilina V.E. Inženernyj vestnik Dona (Rus), 2015, №2, URL: ivdon.ru/uploads/article/pdf/ivd_10_Tsvil1.pdf_ebd0610868.pdf.

6. Tsvil' M.M., Shumilina V.E. Inženernyj vestnik Dona (Rus), 2014, №1 URL: ivdon.ru/ru/magazine/archive/n1y2014/2241.

7. Baltagi B.H. Econometric Analysis of Panel Data. B.H. Baltagi. 3rd Edition. Chichester: John Wiley &Sons, Ltd, 2005, 356 p.

8. Matyas L. The Econometrics of Panel Data. Fundamentals and Recent Developments in Theory and Practice. L. Matyas, P. Sevestre (eds.). 3rd Edition, 2008, Berlin, Springer, 955 p.

9. Ekonometrika: uchebnik dlya magistrov [Econometrics: the manual for masters], I.I. Eliseeva [i dr.]; pod red. I.I. Eliseevoy. M.: Izdatel'stvo Yurayt, 2012. 453 p.

10. Greene W.N. Econometric Analysis. W.H. Greene. 4th Edition. New Jersey: Prentice Hall, 2002, 272 p.