For reference:
Roslyakova N. The Methodology of Estimation of Indirect Economic Relations Transport Infrastructure and Economic Systems / Proceedings of the 2018 Eleventh International Conference "Management of large-scale system development" (MLSD), Edited by: A. Tsvirkun. - Moscow, 2018, pp. 1-5.
doi: 10.1109/MLSD.2018.8551822
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8551822&isnumber=8551758

The Methodology of Estimation of Indirect Economic Relations Transport Infrastructure and Economic Systems

Roslyakova Natalia
Lab. of economic dynamics and innovation management
V. A. Trapeznikov Institute of Control Sciences RAS
Moscow, Russia
roslyakovaNA@gmail.com

***Abstract*—The article shows a methodology which allows to include some data in econometric study. Based on methods mathematical and nonparametric statistics the author shows how to estimate variables and how to formulate meaningful conclusions. (*Abstract*)**

***Keywords*—regression modeling; mathematical and nonparametric statistics; assessment of the impact of transport infrastructure on the economy (*key words*)**

I. Introduction

In the economic literature devoted to the study of the relationships between transport infrastructure development and economic growth has repeatedly argued the ambiguity and inconsistency of such relations. In fact, the impact of economies and infrastructure is mutual. In reality, there is competition between infrastructure and other production factors for the sources of growth (investment resources). On the one hand, infrastructure contributes to the creation of better conditions for people's lives and economic activities. From this point of view, it can initiate the creation of enterprises, increase budget revenues, the level and quality of life of the population. On the other hand, infrastructure is largely created through investment and, therefore, its development depends on the economy situation as a whole. In the unfavorable economic conditions, significant changes can occur in the directions of investment and amounts of investment funds. From this point of view, the economy needs and opportunities can be constraints for the development of infrastructure.

This interrelation is due to the fact that the impact of transport infrastructure on regional economic growth and development is indirectly. Often in the studies using general specifications of regression modeling like a linear or Cobb-Douglas function, factors of the infrastructure are insignificant in comparison with the classical production factors. This is difficult to planning of regional development and generally leads to the estimation according to "the residual principle" [1]. In this regard, an actual scientific problem is to creating methods and models for a more complete and adequate account of the impact of indirect relations, expressing the interaction of transport infrastructure and economic systems. The key novelty is the development these methods through the using mathematical and nonparametric statistics.

II. Overview of Researches and Methods

In the practice of economic or econometric research, there are difficulties with infrastructure research. Examples can be found in many articles. When assessing the efficiency of infrastructure capital using Cobb-Douglas models, the authors note that factors of transport infrastructure availability, such as density of roads and railways, are statistical insignificant [2; 3; 4; 5]. In [6] considers different classes of models for forecasting the impacts of infrastructure projects on economic development, and notes the common problem. This is a difficulty of the impacts of transport infrastructure quantification due to the presence of indirect relationships. The classical production factors (the value of fixed assets in absolute or per capita value, the population or the number of amusing in the economy, investment levels as a whole or in individual sectors, also in absolute or per capita value) often are related with economic variables by a strong and direct relation.

It has already been mentioned that the lack of transport infrastructure effects concentration (both in space and time) leads to the fact that the factors are reflected in the economic growth and development models as insignificant.

For this reason, researchers often deviate from the original specifications of the models. This allows to catch some effects without quantity comparison their effect with the effects of production factors (capital and labor). Another way out of the situation is the estimation the full specification of the model, but important to note the different levels of the factors significance. Often main production factors reflected in the models as a significantly, while infrastructure factors reflected as insignificantly. Such results may be relevant only in some intermediate stage of research but on this basis it is impossible to make conclusions about the level of impact of the endogenous variables, the relationship of factors and the like.

III. Theoretical Principals and Models

The alternative path is the on and off method when we can find the models where transport infrastructure factors are significant into the regression models along with the main production factors. Obviously, that transport infrastructure effects will not be on a level of the effects from the fixed assets or the labor force. But significance of all included factors allows us to make some conclusions (see [7] for details).

The next step is the using natural units of measurement instead of analysis of macroeconomic indicators such as GRP measured in monetary amounts. This, in turn, removed the problem of inflation monetary values in time. For example, we used the volume of timber removal for the i-th month in the j-th areas as endogenous variable. Using the climatic data about the time of winter roads functioning (in days) it was possible to construct regression models of the following type:

$$Y_{ij} = A + B*CD_{ij} \tag{1}$$

where Y_{ij} – the volume of timber removal in the i-th month in the j-th area; A – constant; B – the coefficient of timber removal in one-day functioning of winter road; CD_{ij} – the number of days in the i-th month in the j-th area which are suitable for functioning of winter road (more in [8]).

Since the areas considered in this study are very much differed in the level of density of years-round used roads, there was a hypothesis about the significant impact of road density factor.

However, a key limitation of the regression modeling was made by the data structure. For the length and density of roads the statistics are annual in nature and cannot be included in the regression models with monthly data. The assumption that all 12 months of one year, the figure of transport infrastructure is the same, and with the new year the level of roads is changing and further until the end of the year remains unchanged, violates the meaning of the study. The aggregation other variables from monthly into annual data can lead to insufficient number of observations and the leveling of regional differences. It is unacceptable as the main interest is in identification of regional differences. If factors with such limitations are included in the regression model, there is a risk of getting wrong estimates of coefficients due to the lack of variation in the exogenous variables or getting untenable estimates due to the lack enough number of observations in the sample.

The solution to this problem and all class of similar problems, when the data have an irrelevant structure for the regression study, in our opinion, is the using of the methods of mathematical and nonparametric statistics.

Depending on the task and completeness of the data, can be applied different tools. The analysis of variance (ANOVA) allows to answer the question about the significant difference between groups of selected based on the level of some factor.

In turn, the estimation of the Spearman coefficient allows you to understand the strength and direction of the factors impacts and to assess its significant. Additional conclusions may reflect the assumptions about the nonlinear nature of the relationship between variables. The specific of Spearman coefficient is that not the absolute values of the variables are correlated, but their ranks. Its interpretation is similar to Pearson's linear correlation coefficient. It shows the strength of the relation between variables. If the coefficient is: from ±0,7 to ±1-the relationship is considered strong; from ±0,3 to ±0,699-the relationship is considered average; from 0 to ±0,299 – the connection is considered weak. The sign indicates the direction of the relationship: direct, in the case of a positive sign, and reverse, in the case of a negative sign. The main difference from the Pearson linear correlation coefficient is that the Spearman coefficient requires an additional Student's criterion (at a given level of error probability and number of freedom degrees). Only on the basis it is possible to make a conclusion about significance or insignificance of the relationship. That is, in contrast to the Pearson coefficient, for Spearman coefficient there may be a weak but significant relationship.

IV. Estimations and Conclusions

The essence of the analysis of variance used for such problems is the allocation of groups of observations that differ in the level of variable. In particular, in the example of the Republic of Karelia research, it was performed a ranking procedure in each period (month). Depending on the roads density each area of the Republic of Karelia was assigned the rank: 1 – the area with the highest density, 5 – the area with the least density (an example of ranking for one period can be seen in table 1).

As a result of this procedure should be obtained a new set of data, where the real data of the endogenous variable (in the example it is the timber removal) correspond to the ranks of the exogenous variable (in the example it is the road density). The very procedure of ANOVA can be carried out in the setting Data Analysis of MS Excel or in a specialized program Statistica. Output of MS Excel (see table. 2) allows to see values of mean-square (MS) for intergroup and intragroup variance.

TABLE I. THE EXAMPLE OF RANKING TO ALLOCATION THE OBSERVATIONS IN GROUPS WITH DIFFERENT ROAD DENSITY

Areas	Parameters			
	Period	***Road density, km per km²***	***Range***	***Timber removal (Y_{ij})***
North	2009, oct	0.04904	5	33.6
Centre	2009, oct	0.06761	4	116.4
South-East	2009, oct	0.08724	3	98.3
South-West (Sortavala)	2009, oct	0.16766	1	33.7
South-West (Syoyarvi)	2009, oct	0.16191	2	62.6
North	2009, sep	0.04904	5	26.2
Centre	2009, sep	0.06761	4	111.1
South-East	2009, sep	0.08724	3	127.4
South-West (Sortavala)	2009, sep	0.16766	1	38.2
South-West (Syoyarvi)	2009, sep	0.16191	2	83.5

TABLE II. THE RESULTS OF ONE-WAY ANOVA FOR DENSITY OF ROADS

Parameters				
	SS	***MS***	***F***	***p***
Intergroup	3354631	3354631	3850.54	0.00
Intragroup for road density	838761	209690	240.69	0.00

TABLE III. THE EXAMPLE OF RANKING THE LEVEL OF TIMBER REMOVAL

№	Parameters			
	Area	***Period***	***Timber removal range (Y_{ij}_range)***	***Road density range ($Dens_{tr}$_range)***
427	South-West (Sor.)	2005, may	600	1
115	North	2009, may	599	5
103	North	2008, may	598	5
426	South-West (Sor.)	2005, apr	597	1
415	South-West (Sor.)	2004, may	596	1
379	South-West (Sor.)	2001, may	595	1
367	South-West (Sor.)	2000, may	594	1
403	South-West (Sor.)	2003, may	593	1
402	South-West (Sor.)	2003, apr	592	1
409	South-West (Sor.)	2003, nov	591	1
404	South-West (Sor.)	2003, jun	590	1
391	South-West (Sor.)	2002, may	589	1
19	North	2001, may	588	5
373	South-West (Sor.)	2000, nov	587	1
414	South-West (Sor.)	2004, apr	586	1
366	South-West (Sor.)	2000, apr	585	1
408	South-West (Sor.)	2003, oct	584	1
468	South-West (Sor.)	2008, oct	582	1
469	South-West (Sor.)	2008, nov	582	1
118	North	2009, aug	581	5

The null hypothesis this case is the insignificance of allocated groups, that is, when the factor of division into groups (or the ranking factor) does not affect the variation of the resulting variable. If the null hypothesis is confirmed, MS of the intergroup variance calculated for the all sample without taking into account the group allocation and the intragroup variance will differ slightly. In the example, we can see the reverse situation: the mean square of the intergroup variance (3354631) is significantly different from the mean square of the intragroup variance (838761). The significance of the differences can be estimated using the Fisher test and the probability level. If the calculated criterion (in table 2 - F) exceeds the critical value, the level of P is close to 0 and we can talk about the significance of the allocated groups. Hence the conclusion that allocation factor (exogenous variable) are significantly affect on the endogenous variable.

To determine the Spearman coefficient, the data is required another transformation. In the example, which is considered throughout this article, is studied the monthly data about timber removal from five areas of the Republic of Karelia from November 1999 to October 2009. That is, there are 120 monthly observations for each area, for all five areas is 600 observations. To estimate with Spearman's coefficient, it is necessary to make ranked series from all observations. Higher ranks will receive higher values of the timber removals, lower ranks – smaller. Thus, we get that the timber removal has a range of ranks from 1 to 600; the ranks for the density of roads have already been obtained above, they have a range of changes from 1 to 5 (by the number of areas, see above). In table 3 is a part of ranking as an example.

Such data conversion allows to calculate the Spearman coefficient of rank correlation. The significant Spearman coefficient is -0.3883. The calculated Student's test is compared with a table value for the error probability level 0.05 (which is equivalent to the confidence level p=0.95) and the number of degrees of freedom df=n-m-1. In the example (p=0.95 and df =600-1=598) Student's criterion is 44.6, which exceeds the table value for the corresponding parameters equal to 1.65.

Thus, it can be concluded that is a significant impact of the roads density to the timber removal. The relationship is negative by direction and average by strength. On the basis of this criterion, it can be argued that the increase in roads density can have a negative impact on the timber removal with a probability of 0.3883 (about 40 cases out of 100).

Using the Statistica provides additional results, for example, a test for the equality of means in the graph form (see the example in Fig. 1).

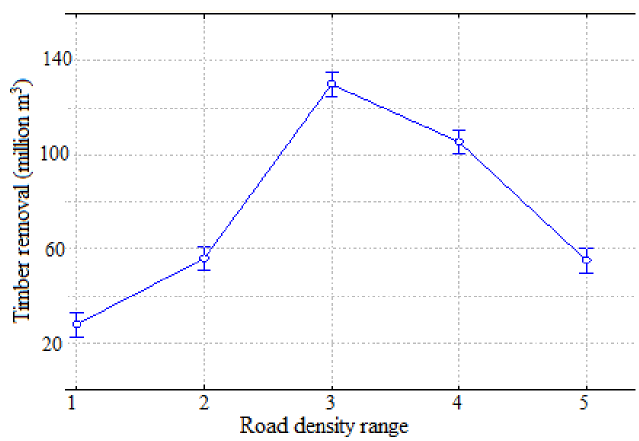

Fig. 1. The graph of the equality of means test for the exogenous variables (density of roads) and the endogenous variable (timber removal)

Fig. 1 allows make a clear visualization of the ratio between the volume of the endogenous variable (Y) and the level of the factor, the influence of which is studied. First, we see that areas which are represented as different levels of roads density are significantly differ in the level of timber removal. Moreover, the confidence interval (vertical columns in Fig. 1) shows that all observations for a particular area are close (small size of the columns in Fig. 1). To interpret the results should note that the areas with high density of roads show low level of removal. This may be due to the greater focus on other economic activities (more technological and complex), which are more than logging, allow to realize the potential of existing roads. It may also be the case that, with good road availability, all forests are quickly depleted. The forest depletion leads to a reduction in logging activities. The territory with the density level 3 (South-Eastern region of the Republic of Karelia) can act as some optimum, because not the highest road availability allows to show the greatest results in the industry. Against the background of this territory, it is obvious that the areas with even lower road density (North and Center of the Republic of Karelia) are restricted in developing of logging. In the presence of similar forest resources, these territories cannot demonstrate the same high rates of removal as in the South-East.

The result of the application of the nonparametric methodology to prove the existence of significantly different groups in terms of road density and thus to verify the impact of this factor, to assess the strength and direction of relationship for the endogenous variable of the timber removal. Similarly, we can try to estimate any parameters that are not integrated into the regression model and through the proposed method to obtain meaningful conclusions about the significance of the relationship of parameters, the strength and direction of relationship with the resulting variable.

Currently, there are plans of the roads development in the South of the Republic of Karelia. The greatest contradiction is that the transport infrastructure is expanding in the southern area of Republic of Karelia where the main production and consumers are concentrated, and the need for the development of transport infrastructure for industry is existing in the North. Shifting the focus of infrastructure construction to the North seems justified, especially for the forest complex. Fig. 1 was proved that the Northern and Central regions of Karelia have restrictions of the increase of production due to the lack of roads infrastructure development.

The existing disproportion in the development of the transport infrastructure is in direct contradiction with the stated priority, because larger costs (increase in transport costs) in the delivery of raw materials increases the cost of raw materials and reduces the ability of production companies to a sustainable and profitable existence. The consequence is the displacement of the Northern areas from the suppliers of southern industries, the growth of social tension in the territories (deterioration of the state budgets of the territories), as well as an additional burden on the more accessible forests of the southern areas, which also has social consequences and the decline of these settlements after the depletion of forests.

Thus, the methodology of quantification of indirect relations make it is possible to carry out the relationship of state socio-economic and transport development programs, which will be expressed in the interconnections programs (nesting approach). Nesting approach will allow to overcome the purely formal nature of state planning (description of only controlled indicators, even if they are not relevant) the development of regional and municipal programs.

With the development of the principles of sustainable development and increasing their inclusion in governance practice and the question arises about the need for informational support of planning process at the territorial level. It is obvious that such methodology allows the Federal authorities can plan the spread of economic impulses from the Federal level projects to the regional and local levels because provide reliable and adequate information about the opportunities and needs of lower-level economic systems. Such aggregated information, is necessary to link and fine-tune the objectives of national, regional and local long-term development at the regional and state level.

[1] N. Mikheeva, "The structural constraints and opportunities for the growth of regions in new condition," Proceedings of the Granberg conference, pp. 60-67, October 2016.

[2] E. Kolomak, "Efficiency of infrastructure capital in Russia," Journal of the New Economic Association, vol. 10, pp 74-93, 2011.

[3] S. Lall, E. Schroeder and E. Schmidt, "Identifying Spatial Efficiency – Equity Trade-offs in Territorial Development Policies: Evidence from Uganda," The Journal of Development Studies, pp. 1-17, August 2014.

[4] E. Kim, G.J.D. Hewings and H. Amir, "Economic evaluation of transportation projects: An application of Financial Computable General Equilibrium model," Research in Transportation Economics, pp. 44-55, Vol. 61, March 2017.

[5] Y. Zhou, J. Yang, G.J.D. Hewings and Y. Huang, “Does High-speed Rail Increase the Urban Land Price in China: Empirical Evidence based on Micro Land Transaction Data,” (December 4, 2017). Available at SSRN: http://dx.doi.org/10.2139/ssrn.3082657.

[6] G. Weisbrod, Estimating wider economic impacts in transport project prioritisation using ex-post analysis, International Transport Forum Discussion Paper, No. 2016-03. Paris: Organisation for Economic Co-operation and Development (OECD), International Transport Forum, 2016.

[7] N.P. Goridko and N.A. Roslyakova, Factors of Russian regions development: the role of innovation and transport infrastructure, Moscow: NPAO “NIB”, 2014, pp. 440.

[8] E.A. Prokopyev, N.A. Roslyakova and P.A. Ryazantsev “Climate change impact assessment on the temporary transport infrastructure,” pp. 1894-1903, 2018 [Digests 31th IBIMA Milan, Italy, 2018].