The Fifth Workshop on Computer Modelling in Decision Making (CMDM 2020)  
Journal of Physics: Conference Series 1784 (2021) 012003  
doi:10.1088/1742-6596/1784/1/012003

Econometric modeling of economic factors’s impact on the labor force per capita

E Ju Liskina and E I Davydova  
Department of Mathematics and Mathematics Teaching Methodology, Ryazan State University named for S.A. Esenin, Svoboda Street, Ryazan 39000, Russia  
E-mail: katelis@yandex.ru

Abstract. In this article, we study the impact of a number of economic factors on the number of workers per capita. We use the panel data analysis method. We found that the best model is the one with fixed effects. We have identified the factors that have the most significant impact on the explanatory variable. The number of vacancies per capita, the capacity of outpatient clinics and the number of university students per capita have a positive statistically significant effect on the labor force per capita. The capital-labor ratio, the population per doctor and the total living space per inhabitant on average have a negative statistical effect on the labor force per capita. The values of individual effects of the regions were positive.

1. Introduction
The size of the labour force is an important factor of economic growth of the country and one of the special features of labour potential. As it is noted in articles [1, 2] for a variety of reasons in many Russian regions there is a decrease of this factor, and this results in a number of economic problems such as increased demographic burden on the able-bodied population and labour shortage. In order to develop effective regional policy in the sphere of labour force it is necessary to investigate the influence of economic factors on the size of labour force, the dynamics of this influence as well as identify regional varieties.

2. Literature Review
The problem of distribution of labour force within a country and its quality is a topical issue all over the world. For this reason a significant number of research works that use various methods of statistical analysis deal with the study of different aspects of regional labour market attractiveness.

The regional studies of the distribution of various labour market indicators with the use of dimensional regression analysis are regularly conducted in different countries [3–10]. So, Abella [3] investigates the spatial distribution of population and employment in frontier areas of northern Spain. Elhorst [4] studies the impact of different factors on variations in age-specific male and female labour force participation rates in the regions of ten European Union member states. Rusche [5] examines the spatial distribution of quality of life indicators that determine the labor attractiveness of German regions. Special attention is paid to the definition of the aggregation level of administrative units. In the studies of Morrisey [6] and Trejo-Nieto [7] the quality of labour force participation of various administrative units of Ireland and Mexico.
is evaluated. E.S. Vakulenko in her works [8, 9] analyzes different aspects of transregional migration whereas E.V. Semerikova and O.A. Demidova [10], for example) study employment and unemployment.

In a number of works panel analysis is used to estimate migration attractiveness of regions [11–13]. The effect of various economic factors on internal migration in Russia and from neighbouring countries is investigated by Vakulenko, Shabashev, [11, 13] and Lifshits [12] in their articles.

Kargi [15], Shelomentseva [16] and Popova [17] in their research works suggested developing regression equations of the dependence of the labour force on certain economic factors for a fixed year. Investigations devoted to the developing of simultaneous equations that contain labour market figures of particular areas of Russia are represented in the articles of Sarycheva [18], Khatskevich [19] and Bakhtin [20] in his work presents the system of econometric equations that reflect the dependence of the labour force participation on health indicators for older citizens.

In addition to the listed articles, there are works devoted to the analysis of the labor potential of regions by the index method. We do not analyze these articles in this review.

It should be noted that the number of Russian articles devoted to econometric modeling of the dependence of the labor force on economic factors is small [16, 17, 20]. At the same time, the works [16, 17] present regression equations for one year and a small set of factors (1–5 regressors). In Bakutin’s study [20], a model was developed for the labor force level of a particular age group and specific factors. Thereby the issue of dynamics and distribution of the labour force across regions remains understudied from the viewpoint of the impact of economic factors (panel analysis) as well as the interaction of regions (spatial regression). Although other characteristics of the labor market and economic development of the regions of the Russian Federation were studied by these methods [8–11,13,19]. Consequently, the problem of studying the influence of economic factors on the differentiation of Russian regions by the labor force participation is relevant.

3. Statement of problem
The given paper continues our studies [2, 14]. The endogenous variable $Y$ is a labour force per capita (or a labour force participation rate – ratio of employed population to mid-year population in the given region). This variable is the indicator of regional labour market attractiveness for indigenous people, and implicitly for internal migrants. The purpose of the given research is to identify regional varieties within the framework of the influence of economic factors on labour force participation rate using the method of panel analysis.

4. Data description
The objects of the given research are Russian regions except for Crimea and Sevastopol as data on these federal subjects have been included since 2015. Data on the Arkhangelsk region are obtained together with those on the Nenets Autonomous District. Data on the Tyumen region are obtained including autonomous regions (Khanty-Mansiysk Autonomous Okrug – Yugra and Yamalo-Nenets Autonomous Okrug). The number of regions $n = 79$.

The following economic factors are used as original explicative variables:

- $X_1$ – fixed capital investment per capita (rubles per capita);
- $X_2$ – gross regional product per capita (rubles per capita);
- $X_3$ – capital-labour ratio (rubles per capita);
- $X_4$ – fixed capital value per capita (rubles per capita);
- $X_5$ – number of vacancies per capita (units per capita);
- $X_6$ – number of registered enterprises per capita (units per capita);
- $X_7$ – passenger traffic of public buses (million people by km);
- $X_8$ – density of hard-surfaced public roads (km per 10,000 km$^2$ of area);
$X_9$ – density of public railway lines (km per 10,000 km$^2$ of area);
$X_{10}$ – capacity of outpatient clinics (thousand visits in shift per thousand people);
$X_{11}$ – number of doctors of all medical specialities per capita (thousand people per thousand people);
$X_{12}$ – number of inhabitants per doctor (persons);
$X_{13}$ – university admission per capita (thousand people per thousand people);
$X_{14}$ – number of university students per capita (thousand people per thousand people);
$X_{15}$ – number of educational organizations for higher education and scientific institutions per capita (units per capita);
$X_{16}$ – number of full-time graduates of state and municipal general educational institutions per capita (thousand people per thousand people);
$X_{17}$ – total living space per person on average (square m);
$X_{18}$ – average monthly nominal wage of corporate employees (rubles per month);
$X_{19}$ – average population income (rubles per month).

The factors are grouped according to the main spheres of the regional economy characterizing their labor attractiveness: macroeconomic ($X_1$–$X_5$), labor market ($X_5$, $X_7$), transport security ($X_7$–$X_9$), medical care ($X_{10}$–$X_{12}$), education ($X_{13}$–$X_{16}$), material well-being ($X_{17}$–$X_{19}$). All variables are relative in order to neutralize the effects associated with the uneven distribution of the population, enterprises and infrastructure facilities throughout the country.

Compared to the work of [2,14] explicative variables were taken with the time lag equal to 1 in order to prevent endogeneity and the observation period has been lengthened from the year 2006 to 2017 ($T = 12$). All together, $nT = 948$ values are considered for each variable. Federal state statistics service [21] is a source of statistical data. The spreadsheet processor MS Excel and statistical package Eviews were used for calculations. The significance level is 0.05.

Some special features of descriptive statistics for dependent variable $Y$ and explanatory factors $X_j$, $j = 1,19$ are illustrated in table 1.

| Variable | Average | Standard error | Median | Asymmetry | Minimum | Maximum |
|----------|---------|----------------|--------|-----------|---------|---------|
| $Y$      | 0.52    | 1.3 $\times$ 10$^{-3}$ | 0.52   | -0.58     | 0.23    | 0.66    |
| $X_1$    | 73825.8 | 2272.5         | 56163  | 3.58      | 6525    | 62324   |
| $X_2$    | 278252.1 | 7709.2        | 216122 | 3.12      | 21922.4 | 1900323 |
| $X_3$    | 1356320.9 | 41417.6    | 1022455.1 | 4.29     | 164991.9 | 12237059 |
| $X_4$    | 718210.6 | 23410.1 | 536066.2 | 4.38     | 23785.2 | 6983848 |
| $X_5$    | 7.8 $\times$ 10$^{-3}$ | 1.8 $\times$ 10$^{-3}$ | 6.8 $\times$ 10$^{-3}$ | 3.02     | 8.6 $\times$ 10$^{-5}$ | 5.1 $\times$ 10$^{-2}$ |
| $X_6$    | 2.7 $\times$ 10$^{-2}$ | 4.2 $\times$ 10$^{-4}$ | 2.4 $\times$ 10$^{-2}$ | 3.01 | 7.0 $\times$ 10$^{-3}$ | 0.11 |
| $X_7$    | 1691.8   | 52.2         | 1083.5 | 1.54      | 5.0     | 9064    |
| $X_8$    | 208.2   | 10.1         | 154.5  | 5.66      | 0.80    | 3572    |
| $X_9$    | 173.8   | 8.0          | 139.0  | 7.70      | 0.0     | 3082    |
| $X_{10}$ | 7.9 $\times$ 10$^{-2}$ | 4.0 $\times$ 10$^{-3}$ | 3.6 $\times$ 10$^{-2}$ | 5.65 | 9.0 $\times$ 10$^{-3}$ | 1.4 |
| $X_{11}$ | 4.8 $\times$ 10$^{-3}$ | 3.4 $\times$ 10$^{-5}$ | 4.7 $\times$ 10$^{-3}$ | 0.78 | 1.3 $\times$ 10$^{-3}$ | 8.6 $\times$ 10$^{-3}$ |
| $X_{12}$ | 219.4   | 1.4          | 112.4  | 0.29      | 114.7   | 352.7   |
| $X_{13}$ | 8.6 $\times$ 10$^{-3}$ | 1.5 $\times$ 10$^{-3}$ | 8.1 $\times$ 10$^{-3}$ | 4.38 | 0.0     | 5.8 $\times$ 10$^{-2}$ |
| $X_{14}$ | 3.9 $\times$ 10$^{-2}$ | 6.9 $\times$ 10$^{-4}$ | 3.7 $\times$ 10$^{-2}$ | 5.11 | 0.0     | 0.28 |
| $X_{15}$ | 5.6 $\times$ 10$^{-6}$ | 1.0 $\times$ 10$^{-7}$ | 4.9 $\times$ 10$^{-6}$ | 2.41 | 0.0     | 2.6 $\times$ 10$^{-5}$ |
| $X_{16}$ | 8.4 $\times$ 10$^{-3}$ | 5.6 $\times$ 10$^{-5}$ | 8.2 $\times$ 10$^{-3}$ | 0.72 | 3.6 $\times$ 10$^{-3}$ | 1.7 $\times$ 10$^{-2}$ |
| $X_{17}$ | 23.5     | 0.11         | 23.7   | -0.47     | 10.7    | 33.7    |
| $X_{18}$ | 22340.2 | 393.6         | 20656  | 1.78      | 4530    | 91995   |
| $X_{19}$ | 19254.1 | 313.2        | 17841  | 1.34      | 3002    | 70904   |
5. Calculations

5.1. Model specification
At the first stage of this study, we analyzed feature scatter diagrams \(Y_{t(t+1)}, X_{j(t)}\), \(j = 1; 19\), and selected the best paired regression model for each pair \(Y_{t(t+1)}, X_{j(t)}\), \(j = 1; 19\). The logarithmic function \(Y_{t(t+1)} = a_0 + a_1 \ln(X_{j(t)}), j = 1; 6\) and \(j = 17\) is the best for the factors \(X_1\)–\(X_6\) and \(X_{17}\). The linear function \(Y_{t(t+1)} = a_0 + a_j(X_{j(t)}), j = 7; 6\) and \(j = 18, j = 19\) is the best for the factors \(X_7\)–\(X_{16}\) and \(X_{18}, X_{19}\).

Then we estimated general (pooled) regression equation

\[
Y_{i(t+1)} = a_0 + \sum_{j=1}^{19} a_{i,j} X_{i(t),j} + \varepsilon_{i(t)}, \quad t = 2006; 2017, \tag{1}
\]

where \(X_{i(t),j}\) which is the score of factor \(X_j\) for \(i\)-region at a time \(t\), \(\varepsilon_{i(t+1)}\) is a random component of \(i\)-region. We also estimated the general (pooled) regression equation with logarithms of some factors

\[
Y_{i(t+1)} = a_0 + \sum_{j \in \{1; 6; 17\}} a_{i,j} X_{i(t),j} + \sum_{j \in \{7; 18; 19\}} a_{i,j} \ln(X_{i(t),j}) + \varepsilon_{i(t)}, \quad t = 2006; 2017, \tag{2}
\]

Basic statistical characteristics of the constructed equations (1), (2) are presented in table 2.

| Characteristics                      | equation (1) | equation (2) |
|--------------------------------------|--------------|--------------|
| \(R^2\)                              | 0.5229       | 0.4829       |
| Residual sum of squares (RSS)        | 0.7611       | 0.8248       |
| \(F\)-statistic                     | 85.3941      | 45.6285      |
| \(F\)-statistic probability         | 4.3 \cdot 10^{-141} | 1.0 \cdot 10^{-118} |
| Durbin – Watson statistic           | 0.5549       | 0.4743       |
| Jarque – Bera statistic             | 6028.6       | 5525.6       |
| Akaike info criterion               | -4.2620      | -4.1669      |
| Schwarz criterion                   | -4.1954      | -4.0645      |
| Hannan – Quinn criterion            | -4.2366      | -4.1279      |
| Number of insignificant factors      | 8            | 8            |

Since the statistical characteristics of equation (1) are better than the statistical characteristics of equation (2), we chose equation (1) for further research.

5.2. Elimination of collinear factors
At the second stage, we calculated the elements of the correlation matrix of the full set of factors. The following factors are collinear with each other:

\(X_1\) and \(X_2\) \((r_{X_1X_2} = 0.8877)\), \(X_1\) and \(X_3\) \((r_{X_1X_3} = 0.8555)\) \(X_1\) and \(X_4\) \((r_{X_1X_4} = 0.8720)\), \(X_1\) and \(X_{18}\) \((r_{X_1X_{18}} = 0.7243)\), \(X_2\) and \(X_3\) \((r_{X_2X_3} = 0.8968)\), \(X_2\) and \(X_4\) \((r_{X_2X_4} = 0.9162)\), \(X_2\) and \(X_{18}\) \((r_{X_2X_{18}} = 0.8662)\), \(X_2\) and \(X_{19}\) \((r_{X_2X_{19}} = 0.8197)\), \(X_3\) and \(X_4\) \((r_{X_3X_4} = 0.9888)\), \(X_3\) and \(X_{18}\) \((r_{X_3X_{18}} = 0.7530)\), \(X_4\) and \(X_{18}\) \((r_{X_4X_{18}} = 0.7648)\), \(X_8\) and \(X_9\) \((r_{X_8X_9} = 0.8044)\), \(X_{11}\) and \(X_{12}\) \((r_{X_{11}X_{11}} = -0.9469)\), \(X_{13}\) and \(X_{14}\) \((r_{X_{13}X_{14}} = 0.9663)\), \(X_{18}\) and \(X_{19}\) \((r_{X_{18}X_{19}} = 0.9348)\).

We then applied the method of eliminating statistically insignificant factors from equation (1). The values of the indices of determination, levels of significance for the variables’s coefficients and Akaike information criteria (AIC) for the obtained equations are shown in table 3.
Table 3. Some statistical characteristics of the general regression equations obtained using the elimination method.

| Characteristic | No | $X_{19}$ | $X_{13}$ | $X_{7}$ | $X_{2}$ | $X_{18}$ | $X_{11}$ | $X_{9}$ |
|----------------|----|----------|----------|--------|--------|--------|--------|--------|
| $R^2$          | 0.5256 | 0.5255 | 0.5254 | 0.5252 | 0.5250 | 0.5248 | 0.5245 | 0.5229 |
| $P_{X_3}$      | 0.0172 | 0.0144 | 0.0139 | 0.0142 | 0.0111 | 0.0067 | 0.0064 | 0.0074 |
| $P_{X_5}$      | 0.5830 | 0.5266 | 0.4964 | 0.5673 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $P_{X_7}$      | $1 \cdot 10^{-32}$ | $8 \cdot 10^{-33}$ | $7 \cdot 10^{-34}$ | $7 \cdot 10^{-34}$ | $7 \cdot 10^{-34}$ | $7 \cdot 10^{-34}$ | $9 \cdot 10^{-34}$ | $1 \cdot 10^{-34}$ |
| $P_{X_8}$      | $3 \cdot 10^{-27}$ | $2 \cdot 10^{-27}$ | $1 \cdot 10^{-28}$ | $2 \cdot 10^{-28}$ | $8 \cdot 10^{-30}$ | $9 \cdot 10^{-30}$ | $7 \cdot 10^{-30}$ | $1 \cdot 10^{-30}$ |
| $P_{X_9}$      | $4 \cdot 10^{-5}$ | $3 \cdot 10^{-5}$ | $4 \cdot 10^{-5}$ | $3 \cdot 10^{-5}$ | $2 \cdot 10^{-5}$ | $5 \cdot 10^{-6}$ | $6 \cdot 10^{-6}$ | $6 \cdot 10^{-6}$ |
| $P_{X_{10}}$   | $9 \cdot 10^{-16}$ | $9 \cdot 10^{-16}$ | $1 \cdot 10^{-15}$ | $1 \cdot 10^{-16}$ | $7 \cdot 10^{-17}$ | $6 \cdot 10^{-17}$ | $3 \cdot 10^{-18}$ | $2 \cdot 10^{-19}$ |
| $P_{X_{11}}$   | 0.4771 | 0.5205 | 0.4974 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $P_{X_{12}}$   | 0.0012 | 0.0006 | 0.0006 | 0.0006 | 0.0005 | 0.0005 | 0.0005 | 0.0005 |
| $P_{X_{13}}$   | 0.0035 | 0.0823 | 0.0865 | 0.0750 | 0.0789 | 0.0848 | 0.0786 | 0.0000 |
| $P_{X_{14}}$   | $4 \cdot 10^{-5}$ | $4 \cdot 10^{-5}$ | $3 \cdot 10^{-5}$ | $4 \cdot 10^{-5}$ | $4 \cdot 10^{-6}$ | $4 \cdot 10^{-6}$ | $5 \cdot 10^{-6}$ | $5 \cdot 10^{-6}$ |
| $P_{X_{15}}$   | 0.4906 | 0.4974 | 0.4996 | 0.4946 | 0.4519 | 0.0000 | 0.0000 | 0.0000 |
| $P_{X_{16}}$   | 0.2306 | 0.2258 | 0.2424 | 0.2152 | 0.2183 | 0.2210 | 0.2210 | 0.2210 |
| $P_{X_{17}}$   | 0.6230 | 0.6463 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $P_{X_{18}}$   | 0.0174 | 0.0181 | $4 \cdot 10^{-9}$ | $3 \cdot 10^{-9}$ | $3 \cdot 10^{-9}$ | $4 \cdot 10^{-9}$ | $4 \cdot 10^{-9}$ | $3 \cdot 10^{-9}$ |
| $P_{X_{19}}$   | $4 \cdot 10^{-5}$ | $2 \cdot 10^{-5}$ | $2 \cdot 10^{-5}$ | $2 \cdot 10^{-5}$ | $2 \cdot 10^{-5}$ | $3 \cdot 10^{-5}$ | $2 \cdot 10^{-5}$ | $2 \cdot 10^{-5}$ |
| $P_{X_{20}}$   | $7 \cdot 10^{-5}$ | $7 \cdot 10^{-5}$ | $8 \cdot 10^{-5}$ | $5 \cdot 10^{-5}$ | $5 \cdot 10^{-5}$ | $4 \cdot 10^{-5}$ | $5 \cdot 10^{-5}$ | $2 \cdot 10^{-5}$ |
| $AIC$          | -4.2530 | -4.2528 | -4.2567 | -4.2583 | -4.2601 | -4.2617 | -4.2632 | -4.2620 |

On the basis of values of the significance levels for the variables’s coefficients we eliminated statistical insignificant factors $X_2$, $X_7$, $X_9$, $X_{11}$, $X_{13}$, $X_{18}$, $X_{19}$ from equation (1). The elimination of collinear factors $X_3$, $X_4$ together and separately led to a decrease in the value of $R^2$ and an increase in the values of information criteria. Therefore, factors $X_3$, $X_4$ were left in the final set of variables.

5.3. Panel data analysis

At the third stage of this research the following econometric model has been examined through the panel data analysis

$$Y_{i(t+1)} = a_0 + \sum_{j \in \{1;3;4;5;6;8;10;12;14;15;16;17\}} a_{i,j} X_{i(t),j} + u_i + \varepsilon_{i(t)}, \quad t = 2006; 2017,$$

where $a_0$ is an average value of hidden effect, which is the score of factor $X_j$ for $i$-region at a time $t$, $u_i$ is a normalized individual effect of $i$-region (that is $u_i = \bar{u}_i - a_0$ where $\bar{u}_i$ is an individual effect of $i$-region), $\varepsilon_{i(t+1)}$ is a random component of $i$-region. The choice of the best equation among three regression models (random effect model (RE-model), fixed effect model (FE-model), and a model without individual effect (OR-model)) was made with the help of Hausman specification test, Breusch–Pagan LM-test, and the Wald $F$-test. Basic statistical characteristics of the constructed equations are presented in table 4 and the results of testing hypotheses are shown in table 5 which indicates that fixed effect model (FE) is the preferable one for the significance value 0.05.
Table 4. Basic statistical characteristics of regression models.

| Characteristics                              | OR-model | FE-model | RE-model |
|----------------------------------------------|----------|----------|----------|
| $R^2$                                         | 0.5229   | 0.8333   | 0.3008   |
| Residual sum of squares (RSS)                | 0.7611   | 0.2659   | 1.115    |
| $F$-statistic                                | 85.3941  | 47.5973  | 14.5355  |
| $F$-statistic probability                    | $4.3 \cdot 10^{-141}$ | $3.0 \cdot 10^{-275}$ | $3.7 \cdot 10^{-28}$ |
| Durbin–Watson statistic                      | 0.5549   | 1.0135   | 0.2617   |
| Jarque–Bera statistic                        | 6028.6   | 156064.0 | 2867.1   |
| Akaike info criterion                        | $-4.2620$ | $-5.1490$ | $-3.7152$ |
| Schwarz criterion                            | $-4.1954$ | $-4.6830$ | $-3.2492$ |
| Hannan–Quinn criterion                       | $-4.2366$ | $-4.9714$ | $-3.5376$ |

Table 5. Observed and critical test values on selecting the best model.

| Test                                           | Observed value | Critical value | Conclusion |
|------------------------------------------------|----------------|----------------|------------|
| Wald $F$-test (OR-model is more preferable than FE-model) | 20.46          | 1.29           | rejected   |
| Breusch–Pagan $LM$-test (OR-model is more preferable than RE-model) | 983.86         | 3.84           | rejected   |
| Hausman specification test (RE-model is more preferable than FE-model) | 149.40         | 5.99           | rejected   |

According to the findings of the research equation (3) is represented as follows:

$$Y_{i(t+1)} = 0.61 - 5.0 \cdot 10^{-9} X_{1(t)} - 1.1 \cdot 10^{-8} X_{3(t)} + 1.2 \cdot 10^{-8} X_{4(t)} + 0.70 X_{5(t)} + 0.14 X_{6(t)} + 6.5 \cdot 10^{-6} X_{6(t)} + 0.02 X_{10(t)} - 2.0 \cdot 10^{-4} X_{12(t)} + 0.12 X_{14(t)} - 739.79 X_{15(t)} - 0.80 X_{16(t)} - 1.9 \cdot 10^{-3} X_{17(t)} + \varepsilon_{i(t)}.$$  \hspace{1cm} (4)

The statistical characteristics of the coefficients of equation (5.3) are presented in table 6

5.4. Study of the coefficients’s stability

The study of the data graphs for each factor showed that the values of the factors for some regions differ significantly from the average values. The average values of factors $X_1$, $X_3$, $X_4$, $X_5$ are exceeded in the Sakhalin Region. The average values of factors $X_1$, $X_3$, $X_4$, are exceeded in the Tyumen Region. The average values of factors $X_6$, $X_8$, $X_{14}$, $X_{15}$ are exceeded in Moscow City, and the values of factors $X_6$, $X_8$, $X_{15}$ are exceeded in the St. Petersburg City. In addition, there are no values for factor $X_{15}$ for the Chukotka Autonomous Region.

To study the stability of the coefficients of equation (5.3), we calculated the coefficients of model without individual effect, fixed effect model and random effect model in cases when some regions were excluded from the data panel. We have calculated the coefficients of the complex of tree models for the following data panels:

(i) for a panel without the Chukotka Autonomous Region;
(ii) for a panel without the Chukotka Autonomous Region and the cities of Moscow and St. Petersburg;
(iii) for a panel without the Chukotka Autonomous Region, Tyumen and Sakhalin Regions;
(iv) for the panel without the Chukotka Autonomous Region, Moscow City and St. Petersburg City, Tyumen and Sakhalin Regions.

The fixed effects model turned out to be the best for all data panels. The coefficients for the factors and some statistical characteristics of the constructed regression equations are presented in table 7.

### Table 7. Comparison of the coefficients of equation (5.3) and equations (i)–(iv).

| Characteristic | equation(5.3) | equation(i) | equation(ii) | equation(iii) | equation(iv) |
|----------------|---------------|--------------|--------------|---------------|--------------|
| $a_0$          | 0.61          | 0.61         | 0.63         | 0.60          | 0.62         |
| $X_1$          | $-5.0 \cdot 10^{-9}$ | $-1.4 \cdot 10^{-8}$ | $-1.8 \cdot 10^{-8}$ | $4.3 \cdot 10^{-9}$ | $-1.8 \cdot 10^{-9}$ |
| $X_3$          | $-1.1 \cdot 10^{-8}$ | $-1.5 \cdot 10^{-8}$ | $-1.2 \cdot 10^{-8}$ | $-1.4 \cdot 10^{-8}$ | $-1.1 \cdot 10^{-8}$ |
| $X_4$          | $1.2 \cdot 10^{-8}$ | $2.3 \cdot 10^{-8}$ | $1.8 \cdot 10^{-8}$ | $1.6 \cdot 10^{-8}$ | $1.2 \cdot 10^{-8}$ |
| $X_5$          | 0.70           | 0.55         | 0.56         | 0.54          | 0.55         |
| $X_6$          | 0.14           | 0.16         | -0.09        | 0.15          | -0.08        |
| $X_8$          | $6.5 \cdot 10^{-6}$ | $5.4 \cdot 10^{-6}$ | $4.0 \cdot 10^{-5}$ | $6.1 \cdot 10^{-6}$ | $4.0 \cdot 10^{-5}$ |
| $X_{10}$       | 0.02           | 5.7 $\cdot 10^{-3}$ | 0.02         | 2.3 $\cdot 10^{-3}$ | 0.01         |
| $X_{12}$       | $-2.0 \cdot 10^{-4}$ | $-1.6 \cdot 10^{-4}$ | $-1.4 \cdot 10^{-4}$ | $-1.3 \cdot 10^{-4}$ | $-1.1 \cdot 10^{-4}$ |
| $X_{14}$       | 0.12           | 0.14         | 0.13         | 0.14          | 0.12         |
| $X_{15}$       | -739.79        | -771.05      | -523.37      | -965.83       | -685.23      |
| $X_{16}$       | -0.80          | -0.60        | -0.89        | -0.58         | -0.87        |
| $X_{17}$       | $-1.9 \cdot 10^{-3}$ | $-2.4 \cdot 10^{-3}$ | $-3.4 \cdot 10^{-3}$ | $-2.2 \cdot 10^{-3}$ | $-3.2 \cdot 10^{-3}$ |

### Table 6. Statistical characteristics of the coefficients of equation (5.3).

| Variable | Value | Standard error | t-statistic | Prob. |
|----------|-------|----------------|-------------|-------|
| $a_0$    | 0.61  | 0.02           | 0.52        | 1.7 $\cdot 10^{-4}$ |
| $X_1$    | $-5.0 \cdot 10^{-9}$ | $2.6 \cdot 10^{-8}$ | -0.2       | 0.84 |
| $X_3$    | $-1.1 \cdot 10^{-8}$ | $4.2 \cdot 10^{-9}$ | -2.6       | 0.01 |
| $X_4$    | $1.2 \cdot 10^{-8}$ | $8.0 \cdot 10^{-9}$ | 1.5        | 0.14 |
| $X_5$    | 0.70  | 0.17           | 4.0         | 6.0 $\cdot 10^{-5}$ |
| $X_6$    | 0.14  | 0.20           | 0.7         | 0.47 |
| $X_8$    | $6.5 \cdot 10^{-6}$ | $3.4 \cdot 10^{-6}$ | 1.9        | 0.05 |
| $X_{10}$ | 0.02  | $7.5 \cdot 10^{-3}$ | 3.3        | 8.9 $\cdot 10^{-4}$ |
| $X_{12}$ | $-2.0 \cdot 10^{-4}$ | $5.7 \cdot 10^{-5}$ | -3.3       | 1.1 $\cdot 10^{-39}$ |
| $X_{14}$ | 0.12  | 0.04           | 2.9        | 4.1 $\cdot 10^{-3}$ |
| $X_{15}$ | -739.79 | 582.2         | -1.2       | 0.20 |
| $X_{16}$ | -0.80 | 0.53           | -1.5       | 0.13 |
| $X_{17}$ | $-1.9 \cdot 10^{-3}$ | 0.02         | -3.5       | 4.9 $\cdot 10^{-4}$ |

6. Discussion of results

Let’s analyze the impact of each group of factors on the labour force participation rate.
Macroeconomic factors – fixed capital investment per capita ($X_1$) and the fixed capital value per capita ($X_4$) – make a statistically insignificant negative impact on $Y$. Gross regional product per capita ($X_5$) does not have a statistically significant effect on $Y$. The capital-labor ratio ($X_3$) has a significant negative impact, which can presumably be explained by the fact that the growth of the capital-labor ratio in the region reduces the number of jobs, increases unemployment and stimulates the outflow of labor to neighboring regions. Since factors $X_1$, $X_4$ are strongly correlated with $X_3$, we assume that fixed capital investments are aimed at increasing the productivity of fixed investments while maintaining or reducing the number of jobs, which leads to an outflow of labor from the region. Since the gross regional product $X_2$ is calculated by production, it does not affect the labor force participation rate.

Labor market factors – the number of vacancies per capita ($X_5$) and the number of registered enterprises per capita ($X_6$)) have a positive impact on the labor force per capita, as they characterize the attractiveness of the region in terms of employment. Factor $X_5$ has a significant impact, factor $X_6$ is insignificant.

The infrastructure factor of the density of hard-surface public roads ($X_9$) has an insignificant positive impact on the variable $Y$, since it characterizes the presence of a good road infrastructure in the region. Factor $X_9$ positively correlates with $X_8$. Factor $X_7$ has no significant effect on variable $Y$, presumably because there is no labor migration to neighboring regions by bus transportation.

The capacity of outpatient clinics ($X_{10}$) has a statistically significant positive impact on the labor force per capita, since it characterizes the presence of a good medical infrastructure in the region. Number of inhabitants per doctor ($X_{12}$) has a statistically significant negative effect on $Y$. The increase of population per doctor degrades the quality of medical infrastructure and stimulates migration to more prosperous regions. The indicator $X_{11}$, which is not included in equation (5.3), negatively correlates with $X_{12}$.

The statistically significant positive impact of factor $X_{14}$ on the labor force per capita $Y$ shows that the number of university students per capita increases the labor force per capita. This fact can be explained by the fact that full-time students, according to the Rosstat methodology, are not a labor force, and at the same time, the majority of full-time students work. The factors – number of educational organizations for higher education and scientific institutions per capita ($X_{15}$) and the graduation of students by state and number of full-time graduates of state and municipal general educational institutions per capita ($X_{16}$) – have a negative statistically insignificant impact. This is due to the fact that full-time graduates of state and municipal general educational institutions seek opportunities for further education hence they are not in the labour force.

Total living space per person on average ($X_{17}$) has a statistically significant negative impact on the value $Y$. This can be explained by the fact that the shortage of residential accommodation generates demand which gives a boost to construction industry and as a result increases size of the labour force in this sphere of economy. The average monthly nominal wage of corporate employees ($X_{18}$) and average population income ($X_{19}$) were not included in the final equation. We assume that their nominal growth lag behind the rate of inflation; therefore, some categories of the population that may not be a labor force (pensioners, full-time students) do not leave their jobs.

Calculations have shown that factors $X_3$, $X_5$, $X_{14}$ $X_{17}$ are statistically significant for all equations (5.3) and (i)–(iv). Factor $X_{10}$ has became statistically insignificant for equations (i)–(iv), factor $X_{12}$ has become insignificant for equation (iv). Factor $X_4$ has became significant for equations (i) and (ii), and factor $X_8$ has became significant for equations (ii) and (iv). It follows from table 7 that the coefficients for statistically significant factors $X_3$, $X_{14}$ $X_{17}$ are stable for all equations (5.3) and (i)–(iv). The coefficient at factor $X_5$ is stable in the absence of the Chukotka Autonomous Region in the panel. Coefficients for insignificant factors $X_6$, $X_8$, $X_{11}$, $X_{12}$, $X_{15}$, $X_{16}$ were not included in the final equation.
\(X_{15} \) and \(X_{16}\) are unstable. The coefficient at factor \(X_{10}\) is also unstable.

Average value of the hidden fixed effect was calculated as \(a_0 = 0.61\). According to the value of normalized individual effect all regions can be divided into two groups: regions with positive normalized effect \(u_i\) (accordingly \(u_i > a_0\)) and regions with negative statistically insignificant impact (accordingly \(u_i < a_0\)) (table 8). In addition, the greatest value of normalized individual effect is attained for the Magadan region (0.987) and the least value is indicated for the Republic of Tyva (−0.1256). On the basis of value \(a_0 = 0.61\) we obtain that all individual effects of the regions are positive.

Table 8. Allocation of Russian regions according to the value of normalized individual effect.

| Regions: Belgorod (8.0 \(-10^{-4}\)), Vladimir (2.9 \(-10^{-2}\)), Kaluga (3.1 \(-10^{-2}\)), Kostroma (1.4 \(-10^{-2}\)), Lipetsk (1.4 \(-10^{-2}\)), Moscow (4.7 \(-10^{-2}\)), Smolensk (2.8 \(-10^{-2}\)), Tver (2.1 \(-10^{-2}\)), Tula (1.95 \(-10^{-2}\)), Yaroslavl (8.7 \(-10^{-3}\)), Arkhangelsk (8.5 \(-10^{-3}\)), Vologda (2.5 \(-10^{-2}\)), Kaliningrad (2.3 \(-10^{-2}\)), Leningrad (4.8 \(-10^{-2}\)), Murmansk (7.6 \(-10^{-2}\)), Novgorod (1.6 \(-10^{-2}\)), Pskov (1.6 \(-10^{-2}\)), Kirov (9.4 \(-10^{-3}\)), Nizhny Novgorod (2.0 \(-10^{-2}\)), Samara (2.4 \(-10^{-2}\)), Ulyanovsk (1.0 \(-10^{-2}\)), Sverdlovsk (1.1 \(-10^{-2}\)), Tyumen (4.2 \(-10^{-2}\)), Chelyabinsk (1.5 \(-10^{-2}\)), Magadan (9.8 \(-10^{-2}\)), Sakhalin (6.2 \(-10^{-2}\)), Republics: Kareliya (6.2 \(-10^{-3}\)), Komi (5.0 \(-10^{-2}\)), Mariy-El (1.4 \(-10^{-2}\)), Mordovia (2.6 \(-10^{-2}\)), Tatarstan (1.25 \(-10^{-2}\)), Udmurtiya (5.5 \(-10^{-3}\)), Chuvashiya (3.6 \(-10^{-3}\)) | Regions: Bryansk (−6.8 \(-10^{-6}\)), Voronezh (−2.25 \(-10^{-2}\)), Ivanov (−2.0 \(-10^{-4}\)), Kursk (−5.7 \(-10^{-3}\)), Orlov (−7.2 \(-10^{-3}\)), Ryazan (−3.2 \(-10^{-2}\)), Tambov (−4.2 \(-10^{-3}\)), Astrakhan (−1.6 \(-10^{-2}\)), Volgograd (−1.2 \(-10^{-2}\)), Rostov (−8.1 \(-10^{-3}\)), Orenburg (−6.9 \(-10^{-3}\)), Penza (−1.9 \(-10^{-3}\)), Saratov (−1.4 \(-10^{-2}\)), Kurgan (−1.3 \(-10^{-2}\)), Irkutsk (−6.1 \(-10^{-2}\)), Kemerovo (−1.5 \(-10^{-2}\)), Novosibirsk (−8.7 \(-10^{-3}\)), Omsk (−2.5 \(-10^{-3}\)), Tomsk (−3.1 \(-10^{-2}\)), Amur (−8.8 \(-10^{-3}\)), Jewish Autonomous Region (−1.8 \(-10^{-2}\)) |
| | | Regions: Adygeya (−6.0 \(-10^{-2}\)), Kalmykiya (−1.7 \(-10^{-2}\)), Dagestan (−1.7 \(-10^{-2}\)), Ingushetia (−8.1 \(-10^{-2}\)), Kabardino – Balkariya (−5.1 \(-10^{-2}\)), Karachaevo-Cherkessiya (−4.3 \(-10^{-2}\)), North Ossetiya – Alaniya (−3.5 \(-10^{-2}\)), Bashkortostan (−2.2 \(-10^{-2}\)), Altai (−6.5 \(-10^{-2}\)), Tyva (−0.13), Buryatiya (−7.5 \(-10^{-2}\)), Khakasiya (−2.4 \(-10^{-2}\)), Sakha (Yakutia) (−2.5 \(-10^{-2}\)), Krasnoyarsk (−2.6 \(-10^{-2}\)), Stavropol (−5.0 \(-10^{-2}\)), Perm (−1.5 \(-10^{-2}\)), Zabaikalskii (−4.1 \(-10^{-2}\)) |

7. Conclusions

This study made it possible to identify the impact of certain economic factors on the size of labor force per capita. As a result of applying panel data analysis we obtained that the fixed-effect model is the most appropriate. The values of individual fixed effects of the regions have appeared to be positive. Number of job vacancies per capita, capacity of outpatient clinics, and number of university students per capita have a positive statistically significant impact on size of the labour force per capita. Capital-labour ratio, number of inhabitants per doctor, and total living space per person on average have a negative statistical impact on size of the labour force per capita. Other macroeconomic factors (fixed capital investment per capita, and fixed capital value per capita) have a negative statistically insignificant impact. The number of higher education and scientific institutions per capita and number of full-time graduates of state and municipal general educational institutions per capita have the same effect.
References

[1] Sinitsa A L 2017 The number of working age population in Russia and possible responses to its decline Nation. Interests: Prior. Secur. 13:11 2005–17
[2] Liskina E Ju and Shchukina I P 2019 Study of the Dynamics of Factors Influencing Labor Attractiveness of Russian Regions (Saratov: Saratov National Research State University) pp 62–7
[3] Abella S, Duarte R and Escario J J 2003 Spatial distribution of population and employment in frontier areas: Empirical application to Northern Spain Sud-Ouest europén 2003 Tome 15. Midi-Pyrénées. Populations et territoires (ed R Marconis) 113–20
[4] Elhorst J P 2008 A spatiotemporal analysis of aggregate labour force behaviour by sex and age across the European Union J. Geograph. Syst. 10 pp 167–90
[5] Rusche K 2010 Quality of life in the regions: an exploratory spatial data analysis for West German labor markets Jahrbuch für Regionalwissenschaft 30 1–22
[6] Morrissey K and O’Donoghue C 2011 The spatial distribution of labour force participation and market earnings at the sub-national level in Ireland Rev. Econ. Analysis Digit. Initiat. at the University of Waterloo Library 3(1) pp 80–101
[7] Trejo-Nieto A 2016 A spatial analysis of urban labour markets and submarkets in the metropolitan area of Mexico City Europ. Region. Sci. Assoc. Conf. Pap. (ERSA2016)
[8] Vakulenko E 2015 Analysis of the relationship between regional labour markets in Russia using Okun’s model App. Economet. 40(4) 28–48
[9] Vakulenko E 2019 Motives for internal migration in Russia: what has changed in recent years? App. Economet. 55 113–38
[10] Semerikova E V Demidova O 2016 A Interaction of regional labour markets in Russia: spatial econometric analysis Spat. Econ. 3 57–80
[11] Vakulenko E 2012 Migrations in Russian cities: econometric analysis App. Economet. 1(25) 25–50
[12] Lifshits M L 2015 Migration between Russia and other post-soviet countries: the 2007-2015 factors and future prospects Russian Econ. Internet J. 4
[13] Shabashev V A, Shorokhov S I, Verkhozina M F and Chelombitko A N 2017 The migration attractiveness of Russian regions: economic, social and demographic factors Region. Econ. Theory Pract. 15:2 391–404
[14] Liskina E Ju and Shchukina I P 2020 Research of regional differences of the Russian Federation’s labor market by the panel data analysis Math. Nat. Sci. Theory Pract. 15 (Yaroslavl: Publishing house of Yaroslavl State Technological University) pp 168–73.
[15] Kargi B 2014 Labor force participation rate and economic growth: observations for Turkey Univers. J Manag. Soc. Sci. 4:4 46–54
[16] Shelomentseva M V 2013 Economical and statistical models of the processes of the labour resources formation: retrospective aspect Internet J. Naukovedenie 3(16)
[17] Popova M S 2017 Econometric model of dependence of labor force on the number of companies Econ. Innov. Manag. 4
[18] Sarycheva T V 2012 Econometric modeling of labour resources in the region Bulletin of the Mari State University 8 43–6
[19] Khatskevich V L, Bezrjadina G N and Kontsevaja N V 2012 Econometrical modeling of indicators of a region labor market Mod. Econ. Probl. Solut. 10(34) 192–9
[20] Bakhtin M and Aleksandrova E 2018 Health and labor force participation of elderly Russians Appl. Economet. 49 5–29
[21] Regions of Russia. Socio-economic indicators: statistical collection. 2010-2019. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_113862506156