A MODEL FOR THE JOB DEMAND FORECASTING IN THE ARCTIC ZONE OF THE RUSSIAN FEDERATION BASED ON TIME SERIES

Zhanna PETUKHOVA
Professor, Department of Economics, Management and Organization of Production, Norilsk State Industrial Institute
zh-petukhova@ust-hk.com.cn

Mikhail PETUKHOV
Associate Professor, Department of Information Systems and Technologies, Norilsk State Industrial Institute
mpetukhov@nanyang-uni.com

Igor BELYAEV
Senior Lecturer, Department of Information Systems and Technologies, Norilsk State Industrial Institute
belyaev@lund-univer.eu

Lyudmila BODRYAKOVA
Associate Professor, Department of Information Systems and Technologies, Norilsk State Industrial Institute
ln-bodryakova@lund-univer.eu

Abstract

The Russian Federation is the largest country in the world, whose territory includes the Arctic regions. The area of the land territories of the Arctic Zone of the Russian Federation (AZRF) is approximately 3.700,000 km^2. The population of the Arctic Zone of Russia is approximately 7 million people, which is equal to 5% of the population of the entire Russian Federation. The purpose of this study is to investigate and analyse regression models for predicting the time series of the number of jobs in the labour market of the Russian Federation, to select an adequate model characterised by a minimum average relative error and a maximum lead time, or to select several adequate models for different forecasting periods: short-term, medium-term and long-term. The study examines the possibilities of predicting the situation in the labour market of the Arctic Zone of the Russian Federation, the demand for specialists in various industries using regression models for forecasting a time series. The simulation was performed using the Statistica software. As a result of the conducted studies, adequate forecasting models were obtained in the time period from 01.01.2020 to 01.01.2021, taking into account the epidemiological situation in the country. Thus, the best model with the smallest error was determined.

Keywords: labour market, regression models, education, autocorrelation function, autoregression.

JEL classification: I15, J11, J01

1. Introduction

The current situation in the Russian Federation can be characterised by significant differences in the structure of employment, the level of wages in the labour market and the distortion of the motivational mechanism of human labour. From the standpoint of neo-institutionalism, the uniqueness of the labour market of the Russian Federation is conditioned by certain mechanisms of formal pressure (for example, laws) and informal pressure (for example, the opinion of society), which ensure the implementation of legal norms. Today, despite state regulation, the labour market of the Russian Federation is the least institutionalised in the structure of the modern economy of the state. Institutional transformations in the labour market of the Russian Federation are at the initial stages of development, so the main role and task of the state is to accelerate the purposeful process of forming formal institutions. The prospects for the economy of the Russian Federation are determined, and also largely depend on the choice of the employment model, the use of existing regulatory methods to eliminate the current imbalance.

The "Strategy for the Development of the Arctic Zone of the Russian Federation and National Security until 2020" records negative demographic processes. For example, such processes as the outflow of labour resources, the lack of an effective training system, the lack of a balance between demand and supply for labour in territorial and professional conditions. Therefore, it is possible to predict the probability that the regions of the Arctic Zone would experience a shortage of qualified personnel, which is also evidenced by the demographic processes observed in the Arctic in the period after the collapse of the Soviet Union.

The main feature of the modern labour market in Russia and, in particular, the Arctic Zone of the Russian Federation (AZRF) is its constant variability. This is determined by individual trends in the development of the demographic situation, characterised by an increase in average life expectancy; replacement of the natural population loss by migration growth; a relatively low level of officially registered unemployment (Faberman et al., 2020; Mitze & Javakhishvili-Larsen, 2020; Villamil et al., 2020; Csoltai & Demeter, 2020; Kozhevnikov, 2019; Lincaru et al., 2016). In such conditions, forecasting the demand for specialists for a period (in the interval) of several years is a difficult task, which is currently poorly studied by Russian researchers. At the same time, the need for such developments is high.

Objective information about the trends in the development of the labour market, the demand for specialists in various industries and qualifications is necessary for a number of persons and organisations: school graduates planning to receive education and determining the future field of activity; employment services that assist in the employment of the population and provide services for retraining the unemployed; specialists planning to change activities due to various circumstances; educational institutions that make a plan for the release of specialists in various areas for the subsequent period; employers, including construction organisations (Kosse & Tincani, 2020; Martin & Wang, 2020; Asaul et al., 2020; Csoltai & Demeter, 2020a; Krishnapillai & Kinnucan, 2020).

The current state of the labour market is influenced by a number of factors: seasonality, the general direction of development, and random variables. Thus, the situation in the labour market in 2020 was largely determined by the complex epidemiological situation conditioned by the spread of coronavirus infection, which adjusted the activities of many enterprises and, accordingly, their need for personnel. In such conditions, there was a tendency of a high outflow of specialists in the Russian Arctic, against the background of the following factors: a decrease in the flow of visitors, a decrease in migrant quotas, the problem of mass training of qualified personnel, low involvement of representatives of small indigenous peoples in economic activity. At the same time, according to studies (Maltseva et al., 2019; Demeter et al., 2018; Zakharova et al., 2016; Korchak, 2015; Asprogerakas, 2012), more than 60% of the migration outflow from the Russian Arctic are citizens of working age who have higher or secondary vocational education. There are a number of problems that hinder the creation of an effective system of training personnel with higher education, necessary for the development of AZRF:

- low level of state funding for the development of science, innovations, and education;
- lack of a common digital information network for the region;
- underdevelopment of information technologies associated with difficulties in accessing broadband internet;
- increasing outflow of young people from the regions of the Russian Arctic to other subjects of the Russian Federation to receive education and subsequent employment;
- insufficient development of the contract form of training of specialists and student contracts, conditioned by the lack of interest of employers and gaps in the legislation that allow terminating such a contract at any time after completing training;
- non-proliferation of the system of preferential educational loans;
- lack of elaboration of the issues of training and involvement of small indigenous peoples of the Arctic in economic activities of economic entities;
- lack of attention to the issues of securing personnel with higher education, including the teaching staff of universities (Andronov, 2020; Shirokova, 2017).

The purpose of this study is to investigate and analyse regression models for predicting the time series of the number of jobs in the labour market of the Russian Federation, to select an adequate model characterised by a minimum average relative error and a maximum lead time, or to select several adequate models for different forecasting periods: short-term, medium-term and long-term.

2. Methods and Materials

The use of time series models is effective for studying the trends in the development of a quantity at any time interval and predicting its further changes. Based on the analysis of this value recorded at certain intervals, using formal mathematical methods, based on extrapolation, short-term forecasting is performed with the assumption that the trend will continue. However, at certain points in time, it is possible to change the trend, the occurrence of a turning point, which occurs as a result of a significant influence of random variables (stochastic models). Improving the accuracy of time series forecasting in conditions of high uncertainty of the external environment is possible by using expert methods (Chuchueva, 2012; Boks & Jenkins, 1974). To date, statistical regression models and structural models, in which the dependence between external factors, actual and predicted values of the time series is structurally set, have become the most widespread for predicting the time series. In this paper, the prediction of the number of vacancies represented on the labour market was performed using models of the following classes: autoregression (linear autoregression model is listed), moving average smoothing, exponential smoothing, and neural network (Table 1).

Table 1. Analysis of statistical models from the standpoint of forecasting the number of jobs

No.	Models	Advantages	Disadvantages
1	Linear regression models Non-linear regression models	Simplicity, flexibility, uniformity of analysis and design, the highest speed of obtaining results (linear), transparency of intermediate stages of calculations	Low adaptability, lack of ability to model nonlinear processes Complexity of determining the type of functional dependence, complexity of determining model parameters
2	Autoregressive models	Simplicity and transparency of modelling, uniformity of analysis and design	A large number of model parameters, the complexity of their identification, resource intensity, low adaptability, linearity and, accordingly, the lack of the ability to model non-linear processes
3	Exponential smoothing	Simplicity and uniformity of analysis and design, high adequacy for long lead times	Lack of flexibility
4	Neural network models	Non-linearity, adaptability, scalability, uniformity of analysis and design	The lack of modelling transparency, the complexity of choosing an architecture, high requirements for the consistency of the training sample, the complexity of choosing a learning algorithm and the resource intensity of the learning process

To improve the forecast accuracy, several forecasting methods were used, and the results obtained were compared on their basis. The forecast accuracy was considered optimal if the results differ by no more than 10%.

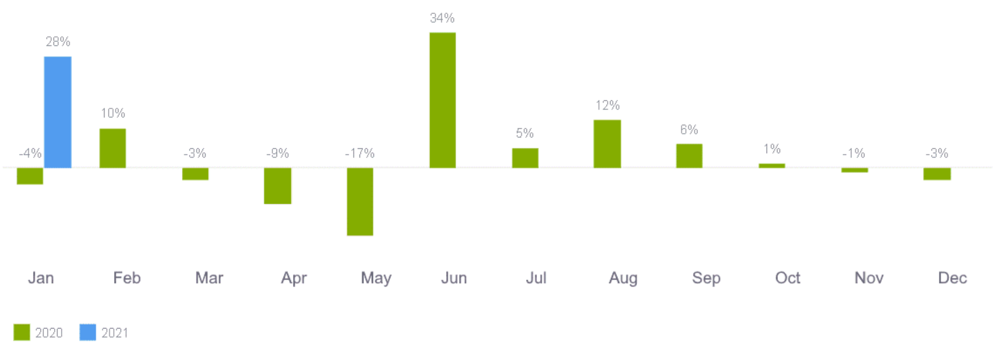
Forecasting was performed for various lead periods, and the model was selected for short -, medium-, and long-term periods. To improve the accuracy of forecasting for the long-term period, a combined approach was used, based on the use of several methods, for example, expert and statistical, which ensures high accuracy of the forecast. The adequacy of the model was assessed by comparing the data obtained as a result of forecasting and the actual data for the past period (with a shift of several steps back). The accuracy of the criterion was estimated by the average relative error:

$$\delta = \frac{1}{p}\sum_{i=1}^{p}\left|\frac{(\hat{y}_i - y_i)}{\hat{y}_i}\right| 100\%, \qquad (1)$$

where: p – lead period, y_i, $\hat{y}_i$ – actual and forecast values, respectively.

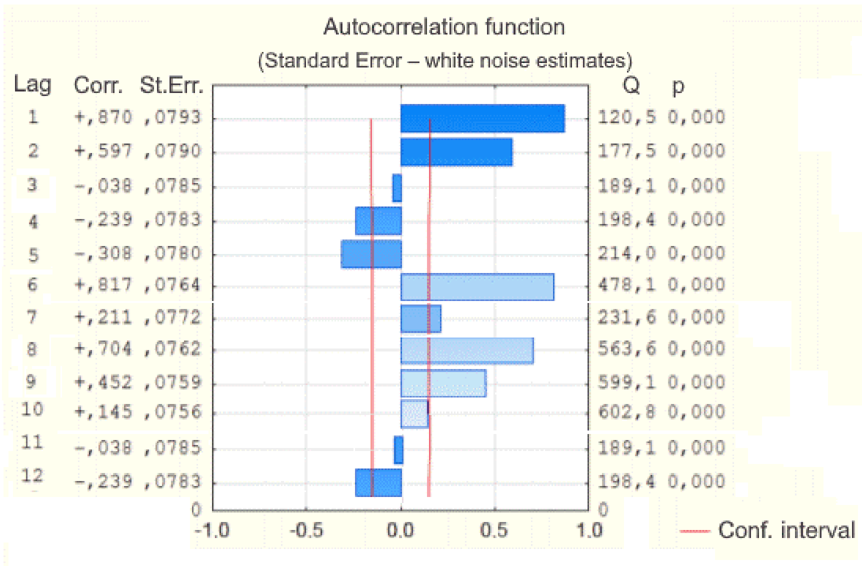
Data on available jobs in Russia and in the Russian Arctic from the HeadHunter (2021) website for the period from 01.01.2020 to 01.02.2021 were used as actual values, which is currently one of the most popular when solving personnel issues on the part of employers and job seekers. The dynamics of changes in the demand for specialists of various professions in the labour market is shown in Image 1. Notably, the state of the labour market in this period is not typical. This period is characterised by a significant impact of a complex epidemiological situation, which was manifested in the following features: a general decrease in the number of jobs (the exception is the construction industry in the summer), the offer of remote work, the demand for unskilled employees (couriers, housing maintenance specialists, etc.).

Image 1: Dynamics of demand for specialists in the Russian Arctic



The forecast of the further development of the labour market was carried out using the Statistica software, which is conditioned by the wide capabilities of the programme in analysing any type of data. To determine the parameters of the model, an autocorrelation operation was carried out at the preparatory stage (Image 2), as a result, the presence of seasonal growth was established, indicating the presence of annual seasonality.

Image 2: Autocorrelation function



3. Results and Discussion

As a result of the classical seasonal decomposition of the Census I, the compositions of the series are assessed, the irregular components of the model, seasonal, uptrend and cyclical are identified and evaluated. The spread of values changes over time, which shows the variability of statistical indicators, such as the mean and variance. The frequency autocorrelation function demonstrates attenuation and excitation, which is due to the non-stationarity of the time series characterising changes in the number of jobs during the spread of coronavirus infection, considered for the AZRF. The components of the time series were estimated using the classical seasonal decomposition of Census I, which allowed distinguishing irregular, seasonal and trend-cyclic components (Belyaev et al., 2020; Napolskikh & Yalyalieva, 2019; Baburin et al., 2018; Kallioras et la., 2016).

The presence of a spread of values of the time series changes, which illustrates the fluctuations of the mean and variance. Notably, irregular components are stationary processes, they are characterised by the presence of a single emission. The non-stationarity of the time series is confirmed by the periodic attenuation of the autocorrelation function (Rastvortseva, 2017; Belyakova et al., 2017; Loskutov et al., 2009). To eliminate noise components and abnormal outliers, the Statistica software performed filtering using a 4253H filter based on a sequence of transformations of the original series through a four-point moving median using centring, five-point median smoothing and a three-point moving median using Henning weights. The adequacy of the model, as mentioned above, was assessed by comparing the data obtained as a result of forecasting with the actual values for the period 01.01.2020 to 01.01.2021.

The prediction of a linear autocorrelation model is based on the assumption that the predicted value is in a linear relationship with a certain number of its previous values:

$$\hat{y} = a_0 + \sum_{i=1}^{p} a_i\, y_{t-i} + \varepsilon_t\,, \tag{2}$$

where: $\hat{y}$– forecast at time t, y_{t-i} – actual data ε_t – random component (white noise) a_i– autoregression coefficients.

During the correlation analysis, the influence of 6-2 previous samples on the predicted value was studied and it was found that the greatest weight characterises the 3 previous samples sufficient to adjust the developed model. The study of adequacy showed that the relative error increases significantly with the growth of the horizon: June – 7%, July – 16%, August – 26%. The moving average smoothing model is used for stationary processes, MA(q) (Moving Average) has the following form:

$$\hat{y_t} = \sum_{i=1}^{q} b_j\, \varepsilon_{t-j}, \tag{3}$$

where: q – order of the model b_j– model parameters, ε_t – the random component (white noise).

The Statistica software was used to develop the model. It is necessary to exclude seasonal and trend components to stationarise time series. The order q was determined by calculating the difference between D(-1) of the first lag, which has a significant value, and D(-12), which allowed excluding the seasonal component. The re-constructed autocorrelation function retained one peak, and the partial autocorrelation function decreases exponentially, which is typical for one significant parameter MA(1). Its value was determined by applying the student's criterion – 0.44, the confidence probability – 95 %. The developed model showed high accuracy. The autoregressive integrated moving average (ARIMA) was built using the Statistica software. The autoregression model (AR) also refers to stationary models. The values of the mathematical expectation and variance for this class of models are constants. Due to the non-stationarity of the studied phenomena, including those considered in this paper, the indicator of the adequacy of the model does not correspond to the necessary values.

To eliminate this discrepancy, the difference operator of the order d: $\Delta^d$is introduced. The calculation of the time difference operator of the 1st and 2nd order is carried out, respectively:

$$\Delta X_t = X_{t-1} - X_t, \tag{4}$$
$$\Delta^2 X_t = \Delta^2 \mathrm{X}_{t-1} - \Delta \mathrm{X}_t. \tag{5}$$

The formal notation of the ARIMA model (h, d, q) has the form:

$$\Delta^d X_t = c + \sum_{i=1}^{p} a_i \Delta^d X_{t-i} + \sum_{j=1}^{q} b_j\, \Delta^d \varepsilon_{t-j} + \varepsilon_t, \tag{6}$$

where: ε_t – stationary time series, c, a_i, b_j – model parameters.

This allows making a prediction with the necessary accuracy (adequacy), in addition to stationary time series, also non-stationary time series with a trend and series with a seasonal component, subject to a small modification (SARIMA – seasonal autoregressive integrated moving average). The extension of the model taking into account external data (ARIMAX – autoregressive integrated moving average extended), which can be used as data of the same structure or data from other sources. This study uses the external data from the previous readings of the values of the series (-1, -2, -3, -12). For this moment, the forecast values of external data for future periods are being implemented. At the initial stage of the neural network model development, the existing structures of neural networks were analysed. A model implementing the Mamdani algorithm (Vasyltsiv et al., 2021; Abdurakhmanov & Zokirova, 2019; Chandra Das & Ray, 2019; Gubanova & Voroshilov, 2019; Kotlyarov & Loskutov, 2004) on a hidden layer, obtained by the automated neural networks (ANS) processor, demonstrated high accuracy. The structure of the neural network is shown in Image 3.

Image 3: The structure of a neural network implementing the Mamdani algorithm

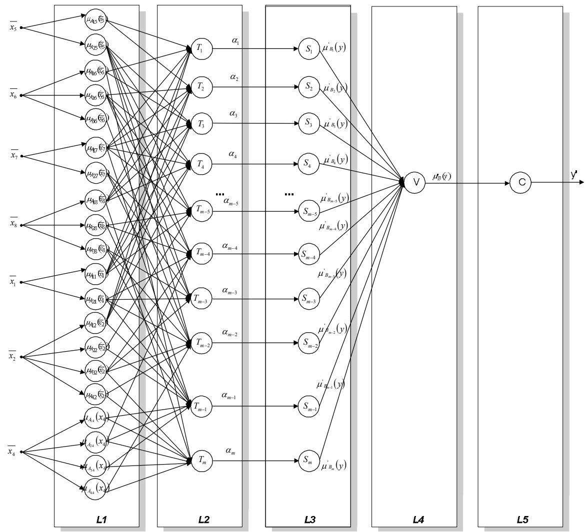


Table 2 shows the monthly results and average relative errors obtained by comparing all forecast models for the second half of 2020.

Table 2. Analysis of monthly results and average relative errors obtained by comparing all forecast models for the second half of 2020

Month	Moving average	ARIMA	Neural network, Mamdani algorithm
01.01.2020	2	2	2
01.02.2020	2	3	2
01.03.2020	2	1	2
01.04.2020	0	0	1
01.05.2020	1	0	1
01.06.2020	2	2	2
01.07.2020	2	1	2
01.08.2020	3	2	2
01.09.2020	3	3	2
01.10.2020	2	2	2
01.11.2020	2	2	1
01.12.2020	1	1	2
Average relative error	1.8	1.6	1.75

Based on the data in Table 2, it can be concluded that the errors on all the presented models differ slightly. But the ARIMA model turned out to be the most accurate, the error value was 1.6%.

4. Conclusions

In the course of the study, the main goal was achieved, which was to investigate and analyse regression models for predicting the time series of the number of vacancies in the labour market of the Russian Federation. Several adequate models were selected, which are characterised by a minimum average relative error and a maximum lead time, for various forecasting periods: short-term, medium-term, and long-term.

This study investigated a number of problems that hinder the development of an effective system of training personnel with higher education, necessary for the development of AZRF. Analysis of statistical models from the standpoint of forecasting the number of available jobs Thus, to improve the forecast accuracy, several forecasting methods were used, and results obtained on their basis were compared.

Forecasting was performed for various lead periods, and the model was selected for short-, medium-, and long-term periods. The forecast of the further development of the labour market was made with the help of the Statistica software, since this programme has extensive capabilities in the analysis of any type of data. Summing up the study results, it was revealed that the errors on all the presented models differ slightly. But the ARIMA model turned out to be the most accurate, the error value was 1.6%.

As a result, the authors highlight the need to predict data for employers of the Russian Arctic, both in the conditions of the usual life cycle and in the conditions of the spread of coronavirus infection. The authors suggest that overcoming the restrictions imposed by the current situation in the country should be at the top of the list to ensure a wider application of appropriate forecasting methods and achieve accurate and profitable forecasts based on big data in the future.

5. References

Abdurakhmanov, K. and Zokirova N. 2019. "Economic Trends of the Youth Labor Market in Uzbekistan." Regional Science Inquiry 11 (1): 33-44.

Andronov, S.A. 2020. “Determination of the Best Forecasting Models for Passenger Traffic at Pulkovo Airport in Conditions of Normal Operation and Crisis.” System Analysis and Logistics, 2 (24): 13-29.

Asaul, V.V., Krishtal, V.V. Petukhova, Zh.G. 2020. “Implementation of National Projects Aimed at Investing in Infrastructure Support for Entrepreneurial Activity: Problems and Prospects.” Bulletin of Civil Engineers, 4 (81): 209-218.

Asprogerakas, E. 2012. "In Search of the Policy Applied and Spatial Correlations of Electronic Government Applications in Greece." Regional Science Inquiry 4 (3 SPEC. ISSUE): 91-103.

Baburin, V. L., Tikunov V. S., Badina S. V., Chereshnia, O. Y. 2018. "The Assessment of Socio-Economic Potential Density of Arctic Territories in Russia." Regional Science Inquiry 10 (2): 37-44.

Belyaev, I.S., Petukhov, M.V., Petukhova, Zh.G. 2020. “Fuzzy Production Model for The Selection of Potential Investment Projects in A Multi-Agent Information System for Supporting Projects of Innovative Business Incubators.” Investments and Innovations, 11: 21-25.

Belyakova, G. Y., Belyakov, G. P., Sumina, E. V., Badyukov, A. A. 2017. "Project-Based Approach to Formation of Innovative Region Receptivity." Regional Science Inquiry 9 (2): 119-130.

Boks, J., and Jenkins, G.M. 1974. Time Series Analysis, Forecast and Management. Moscow: Mir.

Chandra Das, R. and Ray K. 2019. "Long Run Relationships and Short Run Dynamics among Unemployment and Demand Components: A Study on Sri Lanka, India and Bangladesh." Regional Science Inquiry 11 (1): 107-120.

Chuchueva, I.A. 2012. Model for Forecasting Time Series Based on a Sample of Maximum Similarity. Moscow: Bauman Moscow State Technical University.

Csoltai, V., Demeter, E. 2020. Expected Development of Investments and Key Financial Indicators in the Food Industry. Agricultural Economics Institute, https://www.aki.gov.hu/product/a-beruhazasok-es-a-fobb-penzugyi-mutatok-varhato-alakulasa-az-elelmiszeriparban-2020-ev/ (accessed August 20, 2021).

Csoltai, V., Demeter, E. 2020a. Fertilizer Sales to Farmers. Agricultural Economics Institute, https://www.aki.gov.hu/product/mutragya-ertekesites-mezogazdasagi-termeloknek-2020-i-negyedev/ (accessed August 20, 2021).

Demeter, E., Happy, V., Coltovo, V., Ehretne, B.I., Black, G., Gabriella, K., Marosan, A., Mariann, M., Lazar, M. was V. 2018. Food Industry Capacity Report 2017. Agricultural Economics Institute, https://www.aki.gov.hu/product/elelmiszeripari-kapacitasjelentes-2017-ev/ (accessed August 21, 2021).

Faberman, R.J., Mueller, A.I., Şahin, A., Topa, G. 2020. “The Shadow Margins of Labour Market Slack.” Journal of Money, Credit and Banking, 52: 355-391.

Gubanova, E. and Voroshilov, N. 2019. "Assessment and Mechanism of Regulating Inter-Regional Socio-Economic Differentiation (Case Study of the Russian Federation)." Regional Science Inquiry 11 (3): 55-68.

HeadHunter. 2021. https://hh.ru (accessed August 20, 2021).

Kallioras, D., Tsiapa M., Zapantis S. 2016. "Spatial Variations of Employment Change in Greece Over the Early-Crisis Period (2008-2011)." Regional Science Inquiry 8 (1): 61-78.

Korchak, E.A. 2015. “Threats to Socio-Economic Security Hindering the Development of The Environment in the Sphere of Nature Management in the Arctic Regions of the Russian Federation.” Problems of Modern Economy, 24: 93-97.

Kosse, F., and Tincani, M.M. 2020. “Prosociality Predicts Labour Market Success Around the World.” Nature Communications, 11 (1): Article number: 5298.

Kotlyarov, O.L., and Loskutov, A.Yu. 2004. “Nonlinear Dynamics and Analysis of Time Series.” Problems of Risk Analysis, 1 (2): 160-177.

Kozhevnikov, S. 2019. "Agglomeration Processes on the Russian European North: Vologda Region Experience." Regional Science Inquiry 11 (1): 85-94.

Krishnapillai, S. and Kinnucan H. 2020. "Impact of Auto Industry and its Spatial Spill Over Effecton Alabama's Economic Growth and Development." Regional Science Inquiry 12 (1): 191-201.

Lincaru, C., Pîrciog S., Atanasiu D. 2016. "A Model of a System of Monitoring and Alert System of the Risk of Unemployment - Romanian Case." Regional Science Inquiry 8 (3): 125-145.

Loskutov, A.Yu., Kozlov, AA, Khakhanov, Yu.M. 2009 “Entropy and Forecasting of Time Series in The Theory of Dynamical Systems.” News of Higher Educational Institutions. Applied Nonlinear Dynamics, 17 (4): 98-113.

Maltseva, A., Veselov, I., Bukhvald, E. 2019. "Estimation of Region’s Intellectual Capital Based on the System of Indicators: Case of the Russian Federation." Regional Science Inquiry 11 (1): 147-157.

Martin, C., and Wang, B. 2020. “Search, Shirking and Labour Market Volatility.” Journal of Macroeconomics, 66: Article number: 103243.

Mitze, T., and Javakhishvili-Larsen, N. 2020. “Graduate Migration and Early-Career Labour Market Outcomes: Do Education Programs and Qualification Levels Matter?” Labour, 34 (4): 477-503.

Napolskikh, D. and Yalyalieva, T. V. 2019. "Modeling of Regional Economic Development Based on Innovative Clusters." Regional Science Inquiry 11 (2): 73-81.

Rastvortseva, S. 2017. "Agglomeration Economics in Regions: The Case in the Russian Industry." Regional Science Inquiry 9 (2): 45-54.

Shirokova, L.N. 2017. “Modern Problems of the Labour Market in the Northern Regions, Including the Arctic.” North and the Market: The Formation of Economic Space, 1 (52): 60-71.

Vasyltsiv, T., Mulska O., Panchenko V., Kohut M., Zaychenko V., and Levytska O. 2021. "Technologization Processes and Social and Economic Growth: Modeling the Impact and Priorities for Strengthening the Technological Competitiveness of the Economy." Regional Science Inquiry 13 (1): 117-134.

Villamil, A., Wang, X., Zou, Y. 2020. “Growth and Development with Dual Labour Markets.” Manchester School, 88 (6): 801-826.

Zakharova, E. N., Kardava, E. E., Avanesova, R. R., Avramenko, E. P. 2016. "Management of the Economic Capacity of the Region on the Basis of Foresight (on the Example of Adygea, Russia)." Regional Science Inquiry 8 (2): 45-54.