Assessment of Region Economic Development on the Basis of Neural Network Model

S K Antipov, A A Bocharov, A Kobicheva and E E Krasnozhenova

1Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

skantipov@gmail.com

Abstract. The paper considers the possibility to use neural network modeling for assessing the economic development of regions exemplified by the Arctic region of the Russian Federation – Murmansk region. The paper presents assessing and reasoning of this usage, describes its opportunities and threats. The author analyzes four indicators as main economic characteristics: gross regional product, an amount of investments into the fixed capital, retail turnover, foreign trade turnover. The study shows which factors have the most significant effect on these characteristics and comments on the obtained results. The author describes methodology of building the model and checks it empirically. In order to assess the model more accurately, it includes autoregression elements, which allows estimating not only direct interaction, but monitoring possible temporary inertia. Assessment results based on neural network modeling are compared with the results obtained on the basis of ADL equations (autoregressive distributed lag model). Accuracy of the final calculations is analyzed with using the mean absolute percentage error (MAPE) in further preference to the model of neural networks.

1. Introduction

The world is changing. All spheres of science and technology undergo changes. More and more statistical data arise to describe in detail some processes and phenomena. More advanced methods of data aggregation, analysis and processing are expected to replace recently progressive methods and techniques. From year to year the amount of information shows a dramatic rise. In 2016 the total amount of the information retrieved and replicated by the humanity accounted for about 16 zettabytes, whereas by 2025 the information is anticipated to have reached 163 zettabytes, which is actually 10 times more. What is astonishing is that it will have happened within a decade. The shift to the Era of Big Data in the economy has occurred, and though it has almost been invisible for common people, it has affected significantly the methods of economic analysis. An increase in the amount of accessible information ensures econometric model building with higher accuracy, which leads to higher accuracy of forecasts and opportunities for point effects on elements of the economic sectors. However, despite all the advantages of this progress, there are disadvantages. Processing a great amount of information requires a substantial increase in the computing power, improved algorithms and maximum automation. The most cutting-edge technology in this area can truly be considered neural network modeling, as this approach can allow imitating the human activity with high tech devices. Neural network models based on the systematic study of big data arrays enable identifying econometric links at any levels in the most logical and effective way, considering these levels separately or as the combined models. The given paper exemplifies building and learning the neural network model of interaction between various economic sectors on the basis of links identified with the ADL-model, and its further independent development and comparison of calculational results in the learnt neural network and the classical ADL-model [1].
2. Indicators for building the general neural network model of region development

Economic development of any region can be characterized by indicators which specify factors according to various industries and sectors. In order to build the primary assessment model, which will identify main regularities and correlations, it is advisable to consider a maximum wide range of indicators as the analytical base. After the assessment most of these indicators will be excluded, due to their insignificance in relation to resulting variables. Initially, resulting from the logic of most economic significance of the region under consideration, a number of exogenous variables were selected.

Gross regional product as the main aggregate indicator of the region’s economic performance characterizes production of goods and services.

An amount of investments into the fixed capital is the most important driving force for the increase in production capacity and its economic growth. Undoubtedly, investments into the fixed capital contribute to more flexible and accurate regulation of product pricing, revenue increase, production structuring and innovating, which has a positive impact on the economic development.

Retail turnover. The cost of consumer goods sold to people fully characterizes the purchasing power of the population and shows a welfare rate of the population.

Foreign trade turnover is also one of the essential indicators which shows the potential of the region in terms of demand and supply for goods and services.

The model considered in the given paper as the analytical base is exemplified by the Arctic region of the Russian Federation – Murmansk region, with the array of input endogenous variables including 140 indicators for description of the region’s socio-economic development. After determining correlations and ratios for each of the exogenous indicators, it was revealed that most of them had an insignificant effect and could be excluded from the final form of the model. Therefore, the model includes the following factors: consumer price index, unemployment rate, natural growth rate, income per head, quantity of economically active population, retail turnover, industrial index, production index, agricultural index.

3. Literature review

Assessment of regions’ economic development is not a new issue in the modern economics; most researchers have suggested a variety of techniques for this assessment. A number of publications are devoted to development assessment on the basis of the quality approach and analysis of empirical statistical data without mathematical modeling. [2-7] If considering more global approaches, they are usually based on building regressive models or autoregressive distributed lag model (ADL-models). [8-16] These approaches can be visual from the standpoint of structure coherence of the built model and simplicity of calculations. However, in the area of big data the apparent simplicity turns into processing large arrays of data, which requires considerable temporal and calculational resources.

Principles of building models based on neural networks are considered by a number of scientists, especially in the area of IT and mathematical research. [17-18] However, problems of using the neural networking model for forecasting the economic development of regions are not touched upon in the literature, which makes the problem raised in this paper up-to-date.

4. Analytical technique of model building

The research methodology is based on the conventional understanding of the mathematical model of the neural network. The input data in the model are economic indicators, divided into endogenous and exogenous. The Dickey-Fuller test checks whether the time series are stationary. The test is aimed at finding a unit root in an autoregressive equation

\[ Y_t = \beta_0 + \beta_1 \cdot Y_{t-1} + e_t, \quad (1) \]

Where \( Y_t \) is a time series of the indicator under analysis, \( e_t \) is an error, and \( \beta \) are coefficients of the autoregressive equation.
If \(|\beta_1| < 1\), then the time series is stationary; if \(|\beta_1| = 1\), then the time series is not stationary, but the integrated series of the first order. In this case it is necessary to check whether the time series is cointegrated with Engle-Granger cointegration tests. This is related to requirements of the model to exclude the absence of interactions in the series in a long-term run. The method consists of two stages: 1) analysis of the DF-test (Dickey-Fuller test) to check stochastic properties of random deviations in the model:

\[ Y_t = \beta_0 + \beta_1 \cdot x_t + \cdots + e_t \]  

(2)  

2) analysis of the DW-test (Durbin-Watson test) to assess cointegration in time series.

\[ DW = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \]  

(3)

where \( t \) is a value of the time series, \( T \) is a time lag, \( e_t \) shows random errors.

According to the criterion, it is assumed that autocorrelation is detected if the obtained statistical data are within the value between \( 4 - d_l < DW < 4 \) and \( 0 < DW < d_l \).

In addition, the research is based on the assumption about the linear form of potential interactions between variables. All possible non-linear interactions are excluded and regarded as insignificant for model building. The quality view of ADL-model represents a few related equations, each of which expresses one of the exogenous variables.

\[
\begin{align*}
    y_1^t &= f(y_{t-1}^k x_{t-1}^l) \\
    y_2^t &= f(y_{t-1}^k x_{t-1}^l) \\
    y_3^t &= f(y_{t-1}^k x_{t-1}^l) \\
    y_4^t &= f(y_{t-1}^k x_{t-1}^l)
\end{align*}
\]  

(4)

where \( y \) indicates exogenous variables, \( x \) indicates endogenous variables, \( k \) and \( l \) are ordinal numbers of variables, \( i \) is a lag value.

Firstly, on the basis of the collected statistics Pearson’s linear correlation values are defined.

\[
r_{xy} = \frac{n \sum_{t=1}^{T} x_t y_t - \sum_{t=1}^{T} x_t \sum_{t=1}^{T} y_t}{\sqrt{(n \sum_{t=1}^{T} x_t^2 - (\sum_{t=1}^{T} x_t)^2)(n \sum_{t=1}^{T} y_t^2 - (\sum_{t=1}^{T} y_t)^2)}}
\]  

(5)

where \( x \) and \( y \) are variables, the interaction of which has to be assessed.

Afterwards, Student’s t-test is used to assess significance of the obtained values. This allows determining a range of variables included in the equation, hereby excluding those endogenous variables which do not affect much exogenous variables, or their effect is non-linear. Furthermore, the extent of autocorrelation of the exogenous variables and cross-correlation are detected in order to specify formation of the autoregressive distributed lag model. Ultimately, the quantity view of the ADL-model is built. The quality structure of the ADL-model is the foundation for building the neural network model, with endogenous value settings as input values [19-20] and exogenous value settings as output values (fig. 1).
In order to train the neural network, it is necessary to create a final set of significant endogenous variables, similar to those which were components of the quantity ADL-model. In the process of iterations, values of the weights of the neural network will aim for the values of the ADL-model coefficients until they eventually coincide during some iteration. Hence the training of the neural network will be completed, it can be used as an independent calculation too. Furthermore, the forecast values of the trained neural network and the ADL-model will be checked and compared. The quality of the models is estimated using MAPE approach, i.e. by calculation and comparison of absolute percentage errors in forecasting.

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \tilde{Y}_i}{Y_i} \right|
\]

where \( Y_i \) is an actual value of the indicator, \( \tilde{Y}_i \) is an anticipated value.

### 5. Data and initial data analysis

Data preparation for further study includes two main stages:

1. upon defining the required variables, the data base has been formed in the form of tables;
2. numerical data were scaled in order to obtain more accurate results and to minimize calculation errors;
3. data were divided into endogenous and exogenous variables;
4. endogenous variables were divided into blocks to build homogeneous data sets;
5. variables underwent initial processing in compliance with requirements of the model.

Statistical data were collected within the 20-year period, from 1998 to 2017. The main resources of information were the following e-statistics resources: Federal State Statistics Service of the Russian Federation, [http://gsk.ru](http://gsk.ru); Euromonitor passport base, [http://www.euromonitor.com/](http://www.euromonitor.com/); WorldBankOpenData, [http://data.worldbank.org/](http://data.worldbank.org/).

### 6. Empirical test of the model

The obtained statistical data were analyzed to determine whether they were stationary or not. Most coefficients of autoregressive equations lay within the required range, being estimated from 0.12308 to 0.9976. The series which proved to be non-stationary were tested for cointegration. Endogenous and exogenous variables were tested to reveal whether they showed multicollinearity. As a result, several non-stationary non-cointegrated time series and several endogenous variables, due to their significant multicollinearity, were rejected. The final quality view of the model represents the following:

\[
\begin{align*}
  y_t^1 &= f(y_{t-1}^1 x_{t-5}^1 y_{t-1}^2 x_{t-1}^2 y_{t-1}^3 x_{t-1}^3 y_{t-1}^4 x_{t-1}^5) \\
  y_t^2 &= f(y_{t-1}^3 y_{t-1}^4 x_{t-1}^6 y_{t-1}^1) \\
  y_t^3 &= f(y_{t-1}^3 x_{t-1}^4 x_{t-1}^5) \\
  y_t^4 &= f(x_{t-1}^3 x_{t-1}^8 x_{t-1}^5 x_{t-1}^7 x_{t-1}^6 y_{t-1}^3 x_{t-1}^9)
\end{align*}
\]

Solving a set of equations, it is possible to determine coefficients.
\[
\begin{aligned}
y^1_t &= 0.75y^1_{t-1} + 1.58x^1_{t-5} + 0.69y^2_{t-1} - 2.62x^2_{t-1} - 0.12y^3_{t-1} + 1.77x^4_{t-1} - 0.47y^4_{t-1} + 0.02x^5_t \\
y^2_t &= 6.31y^3_{t-1} + 2.67y^4_{t-1} + 0.58x^6_t - 3.16y^1_{t-1} \\
y^3_t &= 3.13y^2_{t-1} - 4.01x^4_{t-1} - 0.11x^7_t \\
y^4_t &= 0.06x^3_{t-1} + 5.17x^8_t + 1.35x^5_t + 3.36x^7_t - 4.38x^6_t - 0.12y^3_{t-1} + 3.55x^9_{t-5}
\end{aligned}
\]

The obtained results were interpreted as the foundation of neural network training. It was built using the Python programming language. After 893 iterations calculation results of the two models coincided, the training was stopped. Further forecast results were received on the basis of the autoregressive distributed lag model (ADL) and the neural network model (NW) (table 1).

|       | Y1    | Y2       | Y3       | Y4     |
|-------|-------|----------|----------|--------|
| **ADL** | 2018  | 347190.00| 258.86   | 162689 | 219849.90 |
|       | 2019  | 394946.4 | 269.77   | 179699 | 242836.60 |
| **NW** | 2018  | 347269.85| 258.92   | 162726 | 219900.47 |
|       | 2019  | 395037.24| 269.832  | 179740 | 242892.45 |

Afterwards, using the error calculation method, the author estimated deviations.

\[
\text{MAPE}_{\text{ADL}} = 0.62 \\
\text{MAPE}_{\text{NW}} = 0.54
\]

The errors were almost similar, but the least significant was in the neural network model.

7. Results and conclusions

Analyzing the obtained results, it is possible to draw a number of conclusions. Firstly, the assumption about the opportunity to use a neural network model for estimating the economic development of the region has been valid; actually, such a model can contribute to solving this task. Secondly, the results of the calculated forecasts of the neural network model with sufficient training and the autoregressive distributed lag model (ADL) are quite similar. This can be explained by accurate model building and by the fact that at the training stage the neural network relies on the regularities resulted from the ADL-model. Therefore, it can be noted that in the present research the neural network model represents a new step in the evolution of regression modeling due to automation as at the stage of estimating significance of endogenous variables as at the stage of calculating weights (coefficients) of these variables. Thirdly, it is worth mentioning that the neural network can be trained on the basis of other, more complicated models. In particular, it comes to situations with possible non-linear interactions, which were not considered in this paper. Having an algorithm of building a neural network, we can easily suggest any model in the capacity of data for training, depending on goals and tasks of the study. This greatly simplifies quantity estimation in the field of regional economic development and provides rapid reliable forecasting in accordance with required parameters. Thus, an opportunity to receive up-to-date information contributes to more effective management of the region.

Specifying the particular economic significance of the built model, it can be claimed that out of 140 indicators, selected for characterizing the socio-economic development of Murmansk region, only 7 indicators showed an actual contribution, they are the following: consumer price index, unemployment rate, natural growth rate, per capita income, economically active population, retail turnover, industrial production index, agricultural production index.

References

[1] Antipov S K 2018 Neural network model as a way of processing complex systems of econometric equations characterizing the interaction of the Russian Arctic Matec Web. Conf. 170 June
[2] Young S et al 1994 Multinational enterprises and regional economic development Reg. Studies 28 657–677
[3] James T et al 1996 A hybrid econometric—neural network modeling approach for sales forecasting International Journal of Production Economics 4(2-3) pp 175-19
[4] Moshtiri S and Cameron N 200 Neural network versus econometric models in forecasting inflation Journal of Forecasting 19(3) pp 201-217
[5] Gibbs D 2000 Ecological modernisation, regional economic development and regional development agencies Geoforum 31(1) pp 9-19
[6] Shenggen F A and Zhang N X 2004 Infrastructure and regional economic development in rural China, China Economic Review 15(2) pp 203-214
[7] Howells J 2005 Innovation and regional economic development: A matter of perspective? Research Policy 34(8) pp 1220-1234
[8] Didenko N I et al 2018 Innovative and technological potential of the region and its impact on the social sector development International Conference on Information Networking 2018-January pp 611-615
[9] Atroshenko S A et al 2016 Evaluation of physico-mechanical properties of high-chromium tool steels modified with harrington method Materials Physics and Mechanics 26(1) pp 26-29
[10] Didenko N I et al 2018 Innovative and technological potential of the region and its impact on the social sector development International Conference on Information Networking 2018-January pp 611-615
[11] Didenko N I et al 2018 A country competitiveness analysis. Adl-model involved International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM 18(5.3) pp 3-10
[12] Kikkas K N and Kulik S V 2018 Modelling the Effect of Human Activity on Fresh Water Extraction from the Earth's Reserves IOP Conference Series: Earth and Environmental Science 180(1) 012017
[13] Klochkov Y et al 2017 Development of the internal audit procedure ICTUS 2017 2018-January pp 738-744
[14] Skripnuk D F and Samylovskaya E A 2018 Human Activity and the Global Temperature of the Planet IOP Conference Series: Earth and Environmental Science 180(1) 012021
[15] Klochkov Y et al 2017 Development of the internal audit procedure ICTUS 2017 2018-January pp 366-369
[16] Tarkhov D A and Vasilyev N A 2014 Nonlin. Phenom Complex Syst. 17(3) pp 327–335
[17] Kainov N U et al 2018 Nonlin. Phenom Complex Syst. 17(1) pp 57–63
[18] Krogh A and Vedelsby J 1995 Neural Network Ensembles, Cross Validation, and Active Learning Advances in neural information processing
[19] Azoff E M 1994 Neural Network Time Series Forecasting of Financial Markets (John Wiley & Sons, Inc. New York)