

CZ GDP Prediction via neural networks and Box-Jenkins Method

Lenka Dvořáková^{1,*}

¹The Institute of Technology and Business in České Budějovice, School of Expertness and Valuation, Okružní 10, 370 01 České Budějovice, Czech Republic

Abstract. Economic indicators are nowadays ones of the most observed, their development does not only serve for comparing individual countries among each other but also show how the given country is prospering. That is why economists are trying to predict also the future development of these indicators via different statistical instruments. Neural networks or Box-Jenkins Method, able to predict future development based on data from the past, are one of the many instruments. The aim of this contribution is to find CZ GDP prediction per individual quarters using neural networks and Box-Jenkins Method, compare them mutually, and evaluate which of them is better.

Key words: GDP, prediction, neural networks, Box-Jenkins Method

1 Introduction

The development of economy is a constantly observed factor, its past proves how the specific country was prospering in the past, and some institutions are trying to predict individual economic indicators, such as Ministry of Finance in the CZ, Czech National Bank, both are trying to predict the future annual indicators such as GDP, interest rate, inflation and unemployment. The Czech Statistical Office (CSO), Czech Banking Association (CBA) or Commercial Bank's Prognosis with the next year's view deal with prognoses. The GDP indicator is one of the most observed economic indicators, and it rightly deserves this amount of attention because its amount influences the economic level of a given country.

The effort to predict a possible development of the GDP indicator using mathematic models is obvious in articles by such economists as Kriz [1]; Markovic, Petkovic, Nikolic and Milovancevic [2]; Abdelaal and Mohamed [3].

Another possibility are predictions of future GDP via Holt-Winters static tool smoothing based on the division of individual provinces, and not from the perspective of a national scale, as stated by Zhao, Liu and Cao [4]. Ways to predict economic indicators are also demonstrated by Jerabkova's article [5], who has, using the Box-Jenkins Method, predicted the unemployment development in the Czech Republic. Klimek [6] focuses on the seasonality of economic and financial indicators with the option of using Box-Jenkins Method as a useful tool for predicting this data. Vochozka [7], which his research was to compare the prediction of the GDP development in common prices within the US until

* Corresponding author: dvorakova@mail.vstecb.cz

2024 performed through neural networks and standard time series, when statistical regression has proven significantly better results of the future USA GDP development in common prices than neural networks have. Sokolov-Mladenovic, Milovancevic, Mladenovic and Alizamir [8] focus in their study was to predict the economic growth with help artificial neuron network with extreme learning machine of based on trade in services, exports of goods and services, imports of goods and services, trade and merchandise trade on the economic growth. Other authors Stevanovic, Stevanovic, Vujcic and Gajic [9] estimate gross domestic product (GDP) based on electricity estimation by artificial neural network (ANN) when their results it can be concluded that the ANN model with extreme learning method could produce the acceptable prediction of the GDP based on the electricity utilization.

With regard to the option of using both neural networks and Box-Jenkins Method for prediction, certain authors are interested in comparing these models and their prediction accuracy of the future values. Ho, Xie a Goh [10] compare these methods from both short-term and long-term views in the area of predicting repairable system failure. Yip, Fan a Chiang [11] also compare both models to predict maintenance of construction equipment costs. Gabor and Dorgo [12] have decided to compare statistical models for the prognosis of the return of an enterprise dealing with production and export in the wood industry. Or a prediction according to Lam and Oshodi [13], who state that fluctuations in the building industry influence the building industry and economy negatively, thus it is necessary to develop reliable and exact prediction models to carry out efficient strategies to mitigate the impact of these fluctuations.

Neural networks and Box-Jenkins Method are statistical programmes via which the future development of economic values may be predicted. The aim of this contribution is to compare neural networks to Box-Jenkins Method and, through the results state which of the methods is more suitable for predicting quarterly CZ GDP data for the future.

2 Materials and methods

The basic data set of the Gross Domestic Product development (GDP) for the quarterly period from 1995 until the first quarter of 2017 comes from the webpages of the Czech Statistical Office. To prepare the data set, MS Excel will be used. To process the data, the DELL Statistica software, version 12 will be used to predict neural networks as well as Box-Jenkins Method.

2.1 Neural Networks

First, linear regression will be carried out. Subsequently, automated neural networks will be used for regression purposes. Then, regression will be carried out using neural structures. We will generate multi-layer perceptron networks, and basic radial function neural networks. Time will be the independent variable. GDP per individual quarters will be determined as the dependent variable. Time series will be divided into training sets, testing, and validation sets, in the following ratio: 70%, 15%, and 15%. A delay in time series will be equal to 1 and 1000 neural networks will be generated, out of which 5, proving the best results, will be preserved. In case of a multi-layer perceptron network (MLP) at least two neurons will be set, 8 hidden neurons maximum. In case of a radial basic function (RBF) the hidden layer will consist of at least 14 neurons, 19 neurons maximum. For a multilayer perceptron network's purposes we will recognize these distributional functions in both the hidden layer and in the output layer:

- Linear;
- Logistical;
- Atanh;

- Exponential;
- Sinus.

Other settings will be left as default (according to the AN – Automated Networking tool).

2.2 Box-Jenkins method

At first, time series analysis will be carried out. Subsequently, the ARIMA model and autocorrelation function will be used. Time will be the independent variable. The GDP development per individual quarters will become the dependent variable. The ARIMA model parameters will be the following:

- p – Autoregressional will be set at 4,
- q – the Moving Average will be set at 4,
- the Seasonal Shift will be also set at 4.

The estimate method will be approximate with the option of a maximal number of 50 iterations, required accuracy of 0.00010 and the maximal amount of iterations for back-casting of 10. The amount of backups per a variable will be 20. Autocorrelation will contain standard errors of white noise, and in partial auto-correlation the p-value will be highlighted for 0.050, amount of moves equal to 15.

In the conclusion, we will compare the results of neural networks and the ARIMA model. The comparison will be carried out via the residue analysis regarding the evaluator's perspective and experience.

3 Results and Discussion

3.1 Neural Networks

In Graphics No. 1, there is the output data of training, testing and validation performances representing assymetrical results, however this division is given by the seasonality of the CZ GDP data observed, per individual quarters.

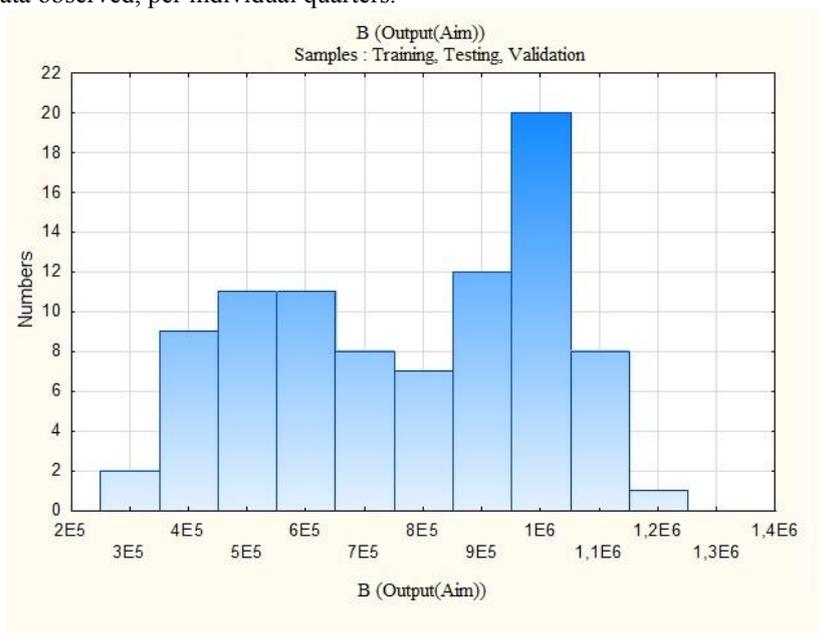


Fig. 1. GDP CZ Output Aim

Table No. 1 called 'The Overview of GDP Neural Networks' shows five of the best neural networks for prediction. The results of training, testing, and validation data have a very high-quality corresponding value of performances. Nevertheless, data performances have been very similar, that why it was necessary to add the results of residues and absolute residues up to evaluate which one of them is closest to reality and that there is a prerequisite of the most exact prediction to it. All five of the best networks may be included among the three-layer network with a forward signal spreading, and learning with a lecturer, known as the Radial Basic Function (RBF). The first RBF 1-19-1 network has got 1 input of 19 hidden layers and 1 output layer, the add-up of the residues is -681262, and the add-up of absolute residues is 4209248.1. The second RBF 1-15-1 network has 1 input of 15 hidden layers and 1 output layer, the add-up of residues is 494839, and the add-up of absolute residues is 4622729.8. The third RBF 1-17-1 network has 1 input of 17 hidden layers and 1 output layer, the add-up of residues is 448907, and the add-up of absolute residues is 4617032.8. The fourth RBF 1-15-1 network has 1 input of 15 hidden layers and 1 output layer, the add-up of residues is -135730, and the add-up of absolute residues is 3636381.1. The fifth RBF 1-16-1 network has 1 input of 16 hidden layers and 1 output layer, the add-up of residues is 208485, and the add-up of absolute residues is 3697098.9. The residue results thus prove that the RBF 1-15-1 network, which is the most suitable one for predicting time series, has the smallest residue reaching 0 and the lowest value of absolute residues.

Table 1. An Overview of GDP Neural Networks

An Overview of Active Networks (data HDP)											
Index	Name of Network	Train. Perf.	Test. Perf.	Valid. Perf.	Train. Error	Testing Error	Valid. Error	Training Algorithm	Error Funct.	Hidden Layer Activation.	Input Activation Funct.
1	RBF 1-19-1	0.971287	0.9391	0.98778	1.40E+09	2.70E+09	2.17E+09	RBFT	Quarter Sum	Gauss	Identity
2	RBF 1-15-1	0.957353	0.94728	0.98543	2.17E+09	2.22E+09	1.64E+09	RBFT	Quarter Sum	Gauss	Identity
3	RBF 1-17-1	0.957808	0.94642	0.98764	2.15E+09	2.07E+09	1.07E+09	RBFT	Quarter Sum	Gauss	Identity
4	RBF 1-15-1	0.982321	0.95041	0.98623	8.64E+08	2.25E+09	2.28E+09	RBFT	Quarter Sum	Gauss	Identity
5	RBF 1-16-1	0.975471	0.96121	0.98593	1.20E+09	1.38E+09	1.40E+09	RBFT	Quarter Sum	Gauss	Identity

Time series prediction for CZ GDP is visible in Graphics No. 2, where a curve reaching the blue-marked letter B (CZ GDP development) is being looked for. Five of the best neural networks for future prediction have been identified from the previous Table No. 1, and from the residue and absolute residue calculations we already know that the best option is offered by the fourth RBF 1-15-1 network, displayed in black colour in the Graphics.

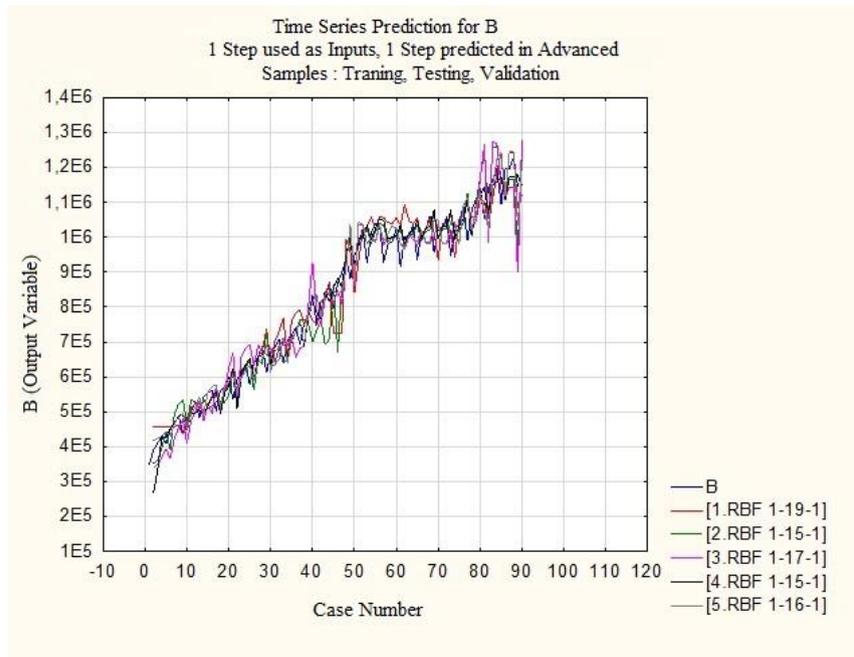


Fig. 2. Time Series Prediction

3.2 Box-Jenkins Method

Graphics No. 2 called ' CZ GDP Histogram' clearly proves that resulting values are not copying Gauss ordinary curve, but the data appears as assymetrical. It may be stated that this display is given by the seasonality of data analysed.

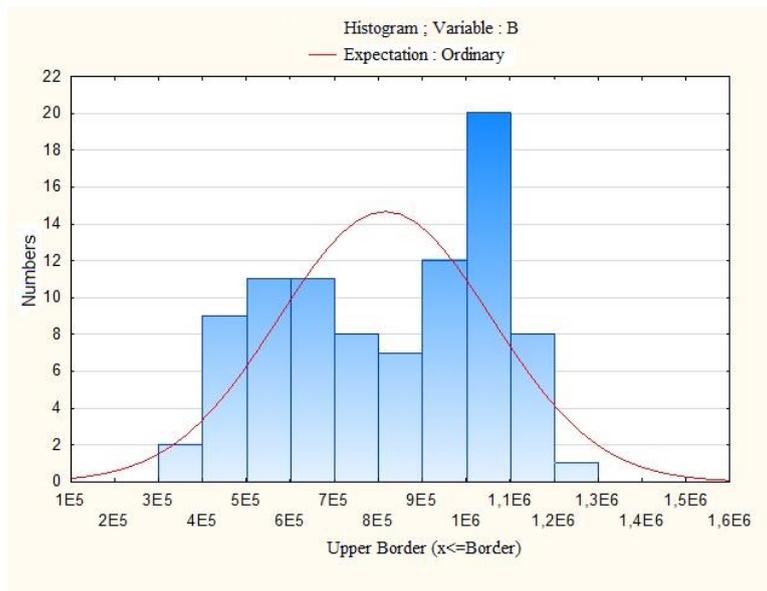


Fig. 3. GDP CZ Histogram

Table No. 2 called 'Descriptive Statistics of CZ GDP ' shows how many cases have been processed to evaluate the future CZ GDP prediction using the Box-Jenkins Method. Moreover, the average value, equal to CZK Million 815, 777.1, the standard deviation of CZK Million 241, 840.5, the minimum of CZK Million 349 683 and the maximum of CZK Million 1,229, 963 have been calculated.

Table 2. Descriptive GDP CZ Statistics

Variable	Descriptive Statistics (data_HDP)						
	Average	Stand. Dev.	Minimum	Maximum	First Case	Last Case	N
B	815777.1	241840.5	349683.0	1229963	1.000000	89.00000	89.00000

Via values presented in Graphics No. 4 we can see that the Series proves an obvious exponential trend as the series variability is not constant. Ordinary probability is relatively distanced from the expected values (the red growing curve), mostly at the beginning and at the end of the CZ GDP observed period. It may be assumed that the resulting prediction of future CZ GDP will not be responding to the real future assumption.

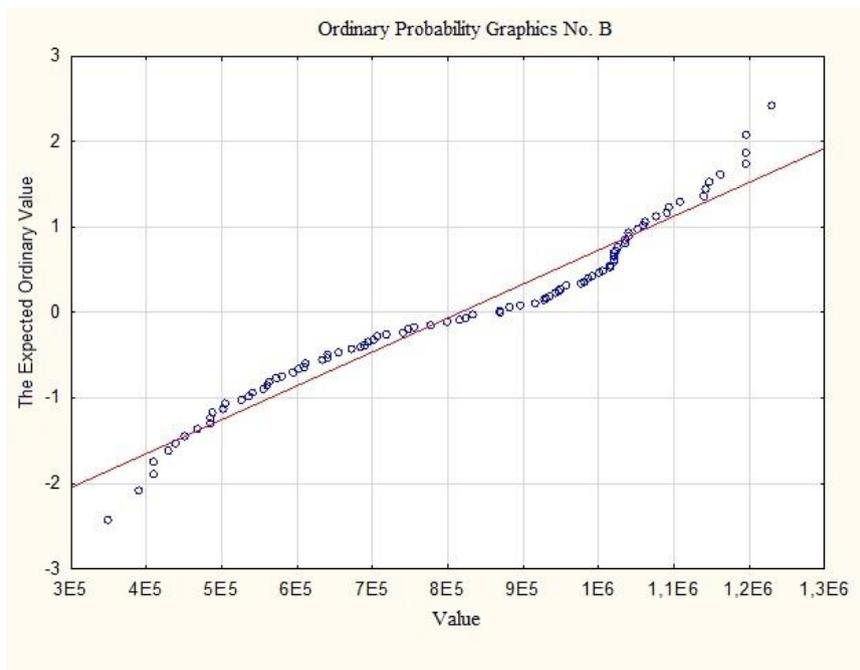


Fig. 4. Ordinary Probability

The values presented in Graphics No. 5 'Ordinary Probability without Trend' are really very distanced from the expected value (red decreasing curve), the deviation from the expected value tells us that prediction values of the CZ GDP period observed will not be responding to the real future development.

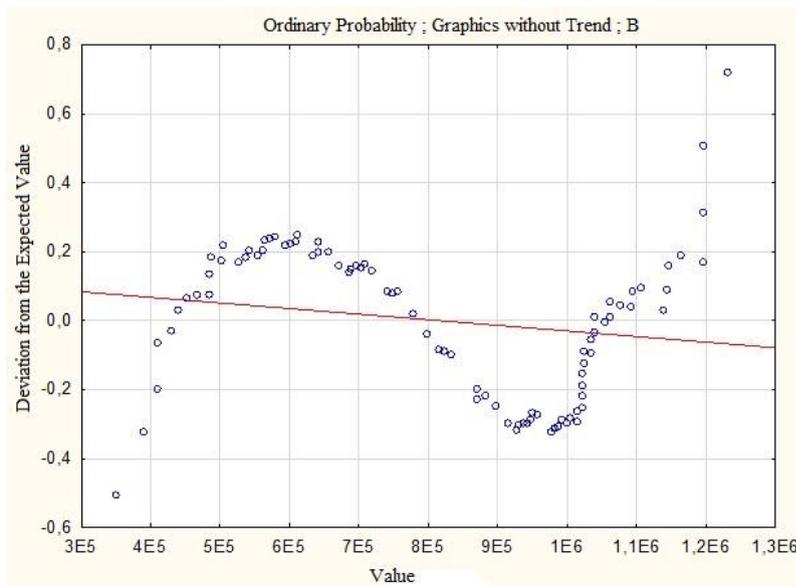


Fig. 5. Ordinary Probability without Trend

In the p1 - p4 a q1 - q4 parameter table they are highlighted because they point at the moving average ($p=0$) and at the autoregression model ($q=0$), other parameters are so-called 'positive'. The given data proves that the p2 and p3 parameters are characterised as moving averages, while q1 and q3 are classified as autoregression models. Due to comparing both methods, even here the ARIMA model residues have been added up, having the quality of 21820482,7. The residue result reaches high values, due to which it is impossible to predict the future CZ GDP values exactly.

Table 3. Parameter Results

Paramet.	Input: B (data_HDP) Transformation: None Model: (4,0,4) PC Resid. = 8081E7					
	Param.	Asympt. SmCh	Asympt. t(81)	p	Lower 95% spol	Upper 95% spol
p(1)	0.296555	0.303906	9.758109E-01	0.332064	-0.30812	0,901233
p(2)	0.967356	0.146480	6.604034E+00	0.000000	0.67591	1.258804
p(3)	0.466198	0.144058	3.236191E+00	0.001755	0.17957	0.752827
p(4)	-0.876520	0.526549	-1.664649E+00	0.099847	-1.92419	0.171148
q(1)	-0.919758	0.196606	-4.678182E+00	0.000011	-1.31094	-0.528574
q(2)	0.225150	0.154531	1.456994E+00	0.148984	-0.08232	0.532618
q(3)	1.014520	0.000000	2.471890E+13	0.000000	1.01452	1.014520
q(4)	0.247900	0.409295	6.056770E-01	0.546423	-0.56647	1.062268

4 Conclusion

The aim of this contribution has been to compare the methods and evaluating time series, and based on the results of a future prediction to compare whether neural networks or the Box-Jenkins Method are more suitable to be used to predict the development of CZ GDP per individual quarters.

Only RBF networks have reached the 5 of the best neural networks, these networks were really similar, and that was why it was necessary to add up the individual residues to identify the best one. The fourth RBF 1-15-1 network was, based on the residue add-up evaluated, from the perspective of performance, as the best one, it had the lowest residue value, equal to -135730 reaching 0, and the lowest absolute residue value equal to 3636381.1.

The results of the Box-Jenkins Method prove that only a half of parameters responds to moving averages or autoregression models. Specifically, they are the p2, p3, q1 a q3 parameters. The p1, p4, q2 a q4 parameters are so-called positive parameters. To compare them to neural networks it was necessary to add up the residues the value of which is 21820482.7. The resulting comparison of residue values unambiguously proves that the Box-Jenkins Method proves much higher values than neural networks. Neural network performance thus may be understood as much more exact values than using Box-Jenkins Method for future prediction of individual CZ GDP quarters.

References

1. R. Kriz, Nonlinear prediction of the GDP growth rate in the globalized world. *Proceedings of the 16th International Scientific Conference of Globalization and its Socio-Economic Consequences*, (2016)
2. D. Markovic, D. Petkovic, V. Nikolic, M. Milovancevic, Soft computing prediction of economic growth based in science and technology factor. *A: Statistical Mechanics and its Applications*, **465**, 217-220 (2017)
3. M. M. A. Abdelaal, S. F. S. Mohamed, Statistical model of Egyptian economic growth prediction. *Advances and Applications in Statistics*, **47**, 225-246 (2015)
4. N. Zhao, Y. Liu, G. Cao, E. L. Samson, J. Zhang, Forecasting China's GDP at the pixel level using nighttime lights time series and population images. *Giscience & Remote Sensing*, **54**(3), 407-425 (2017)
5. V. Jerabkova, Unemployment in the Czech Republic and its predictions based on the Box-Jenkins methodology. *Proceedings of the 12th International Scientific Conference on Applications of Mathematics and Statistics in Economy*, 189-196 (2009)
6. P. Klimek, Practical use of the BoxJenkins methodology for seasonal financial data prediction. *7th International Scientific Conference on Finance and Performance of Firms in Science, Education and Practice*, 598-611 (2015)
7. M. Vochozka, Practical Comparison of Results of Statistic Regression Analysis and Neural Network Regression Analysis. *Littera Scripta*, **9**(2), 156-168 (2016)
8. S. Sokolov-Mladenovic, M. Milovancevic, I. Mladenovic, M. Alizamir, Economic growth forecasting by artificial neural network with extreme learning machine based on trade, import and export parameters. *Computers in Human Behavior*, **65**, 43-45 (2016)

9. M. Stevanovic, S. Vujicic A. M. Gajic, Gross domestic product estimation based on electricity utilization by artificial neural network. *Physica A: Statistical Mechanics and its Applications*, **489**, 28-31 (2018)
10. S. L. Ho, M. Xie a T. N. Goh, A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. *Computers & Industrial Engineering*, **42**(2-4), 371-375 (2002)
11. H. Yip, H. Fan, Y. Chiang, Predicting the maintenance cost of construction equipment: Comparison between general regression neural network and Box-Jenkins time series models. *Automation in Construction*, **38**, 30-38 (2014)
12. M. R. Gabor, L. A. Dorgo, Neural networks versus Box-Jenkins method for turnover forecasting: a case study on the romanian organisation. *Transformation in Business & Economics*, **16**(1), 187-210 (2017)
13. K. C. Lam, O. S. Oshodi, Forecasting construction output: a comparison of artificial neural network and Box-Jenkins model. *Engineering, Construction and Architectural Management*, **23**(3), 302-322 (2016)

