The Ukraine and EU trade balance: prediction via various models of time series

Valerii Matskul[1*], *Diana* Okara[2], *Nataliia* Podvalna[1]

[1]Odessa National Economic University, 65082 Odessa, Ukraine
[2]Odessa State Academy of Civil Engineering and Architecture, 65029 Odessa, Ukraine

Abstract. This article is the first to study, simulate and forecast the monthly dynamics of the trade balance between Ukraine and the European Union for the period from 2005 to 2019. In the presented work, three types of models were used for modeling and forecasting: Automated Neural Networks, additive models ARIMA *ARIMAS (Autoregressive integrated moving average with season component) and Holts model with a damped trend. When modeling using the Automated Neural Networks module, various ensembles of networks and neuron activation functions in hidden layers were used. It turned out that Automated Neural Networks have poor prognostic ability (as in the case considered by us, when modeling insufficiently long series of dynamics). Therefore, when modeling and forecasting the dynamics of the Ukraine-EU trade balance, classical (so-called Box-Jenkins) time series models were used. In this case, the time series is divided into several components (in our case, terms): the main trend is the trend, the seasonal component and the random component (the so-called white noise). By smoothing the initial series, a trend was found, and an analysis of the autocorrelation functions revealed a one-year seasonality. Eliminating the trend and the seasonal component, we obtained a random component, which has a Gaussian distribution. This made it possible to apply first the ARIMA* ARIMAS additive model, and then the Holt model of exponential smoothing with a damped trend. Adequate models of Ukraine-EU trade balance dynamics have been obtained, according to which the forecast has been made. A comparative analysis of the models used. The best model was chosen for forecasting, which allowed to get a good forecast (in comparison with actual data).
Key words: trade balance, export-import, neural networks, Box-Jenkins models

1 Introduction

In recent years, international trade has begun to attract the close attention of many scientists. Among recent works, we note publication [1], which explores the dynamics of the trade balance between the EU and China. At the same time, Ukrainian scientists mainly investigate the international trade of Ukraine as a whole, not highlighting its export-import operations

---

* Corresponding author: valerii.matskul@gmail.com

with the EU. Note article [2], in which an attempt is made on the basis of annual statistical data for the period 2010-2017 years to conduct a correlation and regression analysis of the export of goods and services of Ukraine. A factor was included as one of the predictors in the model - time, and it turned out to be the most significant. Economists around the world have always paid close attention to the study, modeling and forecasting of the dynamics of economic processes. A particularly rapid development of modeling and forecasting using time series models began after the release of the classic Box-Jenkins book [3]. Note that these methods have not lost relevance at present. So in the works [4, 5], the Box-Jenkins models were used to model and predict the dynamics of unemployment. And in the work [6], Box-Jenkins models were compared with Automated Neural Networks when modeling the dynamics of Czech GDP. Examples of the application of various models of time series are described in the book [7]. The subject of this work is the study of the dynamics of trade relations between Ukraine and the EU countries. The relevance of this study is indisputable, since the EU has become the largest trading partner of Ukraine in recent years.

2 Materials and methods

Data for analysis is available on the website of the State Statistics Service of Ukraine [8]. The time period for which the data will be available is the monthly balance (as the difference between total export and imports between the Ukraine and the EU) starting from January 2005 and ending in March 2019. This is about 171 input data. The unit is the \$ million. Preliminary data processing was carried out in MS Excel spreadsheets. When modeling and computing was used DELL STATISTICA software, version 12.

2.1 Neural networks

When searching for the best model of the neural network, the "Automated Neural Networks" program module of the DELL STATISTICA software is used.

2.2 Box-Jenkins models

We use the following Box-Jenkins additive models of time series:

$$y_t = f(t) + S_t + e_t \tag{1}$$

where y_t is a level of time series at time $t = 1, 2, \ldots$; $f(t)$ - trend is traced the long-term and evolution is deterministic in time; S_t - seasonal fluctuations; e_t - random fluctuations (white noise).

3 Results and Discussion

The descriptive characteristics of the data are given in Table 1.

Table 1. Characteristics of the data file

Variable	Average	Minimum	Maximum	Standard deviation
Export	1269.819	648.6	2222.1	302.87
Import	1816.365	561.1	3302.7	526.95
Trade balance (TB)	-546.546	-1356.3	370.8	334.38

Source: Authors.

Analyzing the main characteristics of the dynamic range of the trade balance, we conclude that there are no abnormal levels (the so-called "emissions"). Figure 1 therefore shows the dynamics export-import and trade balance.

Fig. 1. The dynamics of the export-import between the Ukraine and EU

Source: Authors.

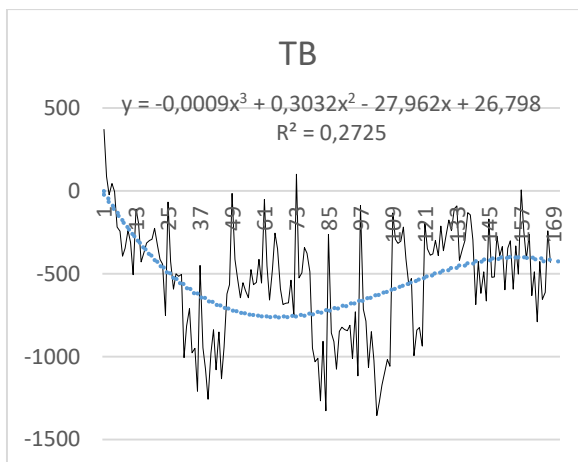Figure 2 shows the dynamics of the trade balance between Ukraine and EU

Fig. 2. The dynamics (with trend) of the trade balance (TB) between Ukraine and EU

Source: Authors.

Everywhere in the future, a dynamic series of the trade balance (variable TB) is used for modeling and forecasting. First, identify the trend. Note that some researchers are trying (see, for example, [1, 6]) to define a trend as a regression of the levels of the time series by time,

requiring a high level of determination of R2> 0,95. Numerous examples of modeling time series (see, for example, [7]) suggest that in the case of an unobvious main trend, the trend should be selected by smoothing the initial dynamic series. Figure 3 below shows the dynamic range of the trade balance after 25-fold exponential smoothing.
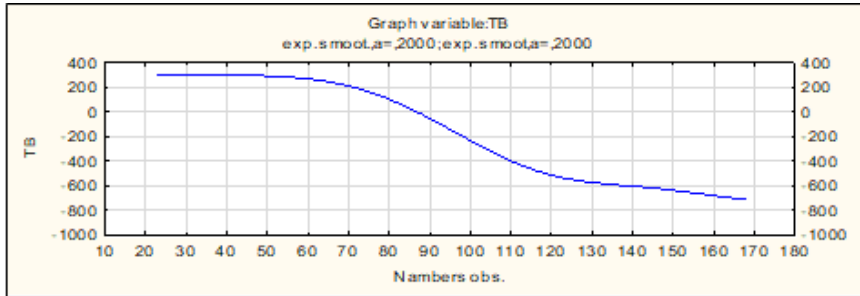

Fig. 3. Dynamic range of the trade balance after 25-fold exponential smoothing

Source: Authors.

Analyzing the smoothed series, we conclude that the trend has the form of a third-order parabola (see Figure 2).

3.1 Neural networks

The module "Automated neural networks" DELL STATISTICA software was used. We will generate multi-layer perceptron networks, and basic radial function neural networks. SALDO will be determined as the dependent variable. Time series will be divided into training sets, testing and control sets, in the following ratio: 70%, 15% and 15%. A delay in time series will be equal to 1 and 1,000 neural networks will be generated, out of which 5, proving the best results, will be preserved. As the activation functions of neurons of the hidden layer, a set of all functions available in the module was used. Since the number of observations is small, we used the multiple subsample method. As a result, the best neural network with 30 neurons in the hidden layer and radial basis activation functions was selected. Figure 4 presents the results of modeling and forecasting.
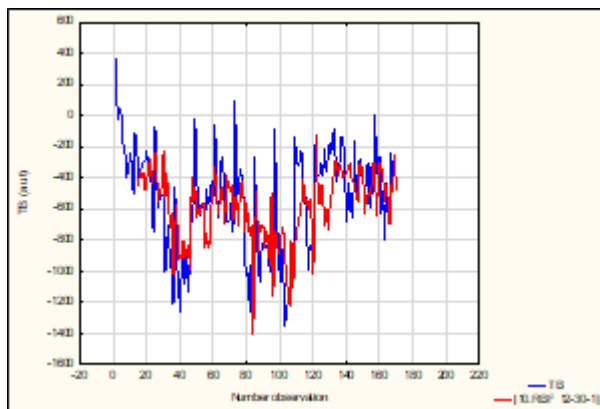

Fig. 4. Neural network modeling and prediction results

Source: Authors.

Note that the neural network showed poor prediction results compared to the actual data for 2019. Therefore, we do not provide numerous intermediate results obtained by modeling neural networks.

3.2 Box-Jenkins models

First, we will identify the time series. Analyzing the graph of the autocorrelation (Figure 5) and partial autocorrelation functions presented below, we conclude that there is a second-order autocorrelation at lag 2 and annual seasonality with a lag of 12.
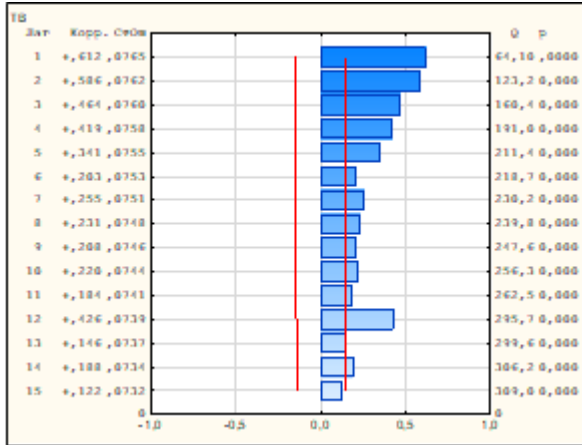

Fig. 5. The graphs of the autocorrelation function

Source: Authors.

Partial autocorrelation function is presented below (Figure 6).
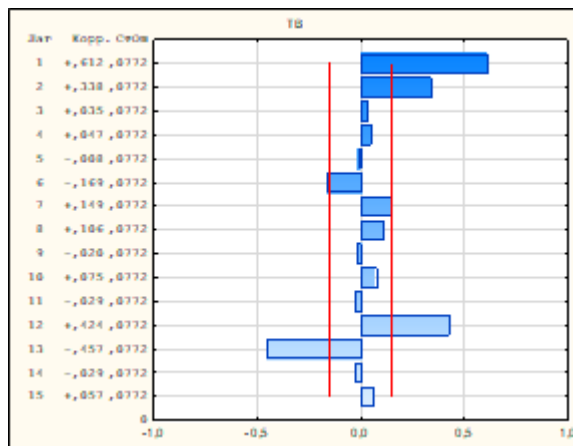

Fig. 6. The graphs of the partial autocorrelation function

Source: Authors.

As it is known, for correct application of the ARIMA*ARIMAS model it is necessary that the time series be stationary. Therefore, we eliminate the trend by taking the third-order

difference on lag 3. After this operation, we really get a stationary series, which is clearly shown by the following graph (Figure 7).
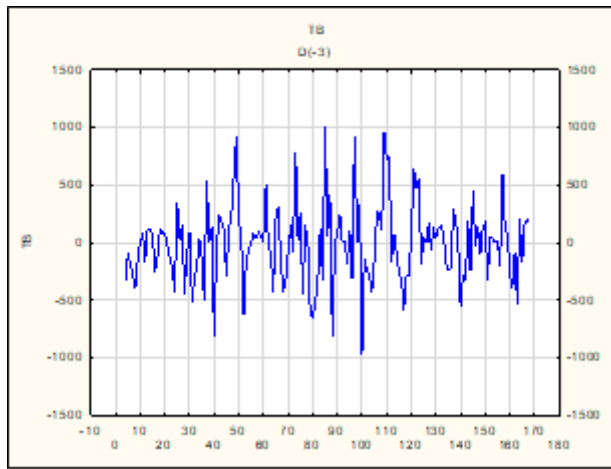

Fig. 7. The graph of the variable TB after elimination the trend

Source: Authors.

After the analysis of the dynamics, we use an additive model *ARIMA*ARIMAS* with the trend and seasonal lag 12:

$$y_t = p_{t-2}y_{t-2} + p_{t-1}y_{t-1} + f(t) - q_{t-1}e_{t-1} + S_t \qquad (2)$$

where $y_t, t = \overline{1,168}$; is the levels of time series;

$$f(t) = -0,0009x^2 + 0,3032x^2 - 27,962x + 26,798 \qquad (3)$$

is a trend component; p_{t-2} and p_{t-1} is the auto-regression coefficients of the second and of the first order; q_{t-1} is a coefficient the moving average model; e_{t-1} is an irregular component (random deviations or so-called white noise). We apply the Smoothed Moving Average of first order $S_t = c + Q_{t-1}S_{t-1}$ to find seasonal coefficients. Parameters of the model one should estimate at 95% confidence level (or 5% risk) with the condition of minimizing the MSE (mean square error):

$$MSE = \sum_{i=1}^{n} \frac{\left(y_i - \tilde{y}_i\right)^2}{n} \qquad (4)$$

where y_i is the actual value and $\tilde{y}_i$ estimates of indicator y at the period i. As a result, we obtained the following parameter estimates (Table 2).

Table 2. Parameter estimates of the model ARIMA*ARIMAS

Parameter	Estimate	Standard deviation
p_{t-2}	0.35976	0.07908
p_{t-1}	-0.6137	0.08725
q_{t-1}	-0.8768	0.05805
Q_{t-1}	-0.4498	0.06147

Source: Authors.

We are convinced of the adequacy of the obtained ARIMA*ARIMAS model by analyzing the histogram of residuals (Figure 8).
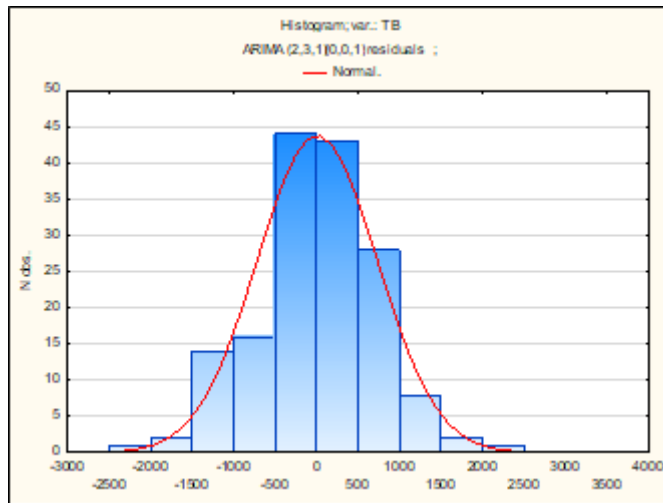

Fig. 8. The histogram of the residuals

Source: Authors.

Let us compare the results of forecasting according to the ARIMA*ARIMAS model with the actual data (for the first quarter of 2019) of the trade balance between Ukraine and the EU (Table 3).

Table 3. Forecasts for ARIMA*ARIMAS model and actual data

Forecasts (ARIMA)	Actual
195.1039	239.4
-301.5741	-409.3
-210.9955	-125.7

Source: Authors.

As we can see, the relative error of ARIMA forecasting is big enough. Therefore, an additive Holt model of exponential smoothing is applicable for prediction. The advantage of this model is that the selection of smoothing parameters can be achieved so that the most recent

levels of the time series are taken into account when predicting. For example, the value of the smoothing parameter 0.1 means that 10% of the last values of the time series determine 90% of the forecast. To simulate again, we will use the "Time Series and Prediction" module of the STATISTICS program. Before building the Holt model, in order to find the optimal values of the smoothing parameters, we will conduct the so-called "Search on the grid". As a result, the following values were found: the smoothing parameter of the main process alpha = 0.4; smoothing seasonal component parameter delta = 0.1: trend damping parameter phi = 0.1. Table 4 shows the results of modeling and forecasting.

Table 4. Results of the modeling and forecasting (fragment) using the Holt model of exponential smoothing

Exp.smooth.: Addit.season (12) Demph.trend,ad.seas.: Alpha=,400 Delta=,100 Phi=,100 TB				
Observ.	TB	Smoothing TB	Resid.	Season comp.
1	370.80	413.82	-43.02	455.323
2	85.40	16.64	68.762	88.092
3	-23.40	-21.16	-2.241	2.954
4	44.40	-4.31	48.705	21.025
5	-3.60	99.95	-103.5	91.557
6	-218.60	2.24	-220.8	65.250
7	-244.20	-342.60	98.396	-126.23
8	-393.90	-243.84	-150.0	-94.386
9	-346.00	-362.35	16.350	-109.47
10	-238.70	-399.94	161.29	-157.6
11	-311.10	-165.05	-146.0	-34.308
12	-506.20	-433.17	-73.0	-202.19
13	-111.20	171.59	-282.7	
14	-195.70	-387.75	192.05	
15	-429.20	-344.06	-85.13	
166	-612.10	-667.12	55.022	
167	-241.50	-301.99	-60.47	
168	-432.00	-489.31	57.307	
169		233.62		
170		-407.11		
171		-131.08		

Source: Authors.

We are convinced of the adequacy of the obtained the Holt model of exponential smoothing by analyzing the histogram of residuals and Normal curve (Figure 9).
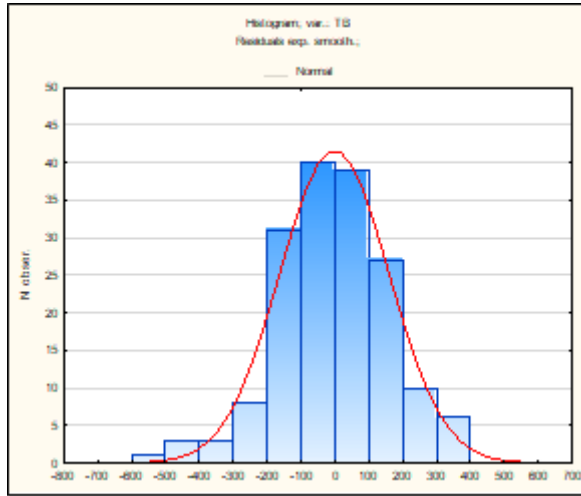
Fig. 9. The histogram of the residuals

Source: Authors.

The following Figure 10, which shows the original and simulated time series, clearly demonstrates the quality of the constructed model.
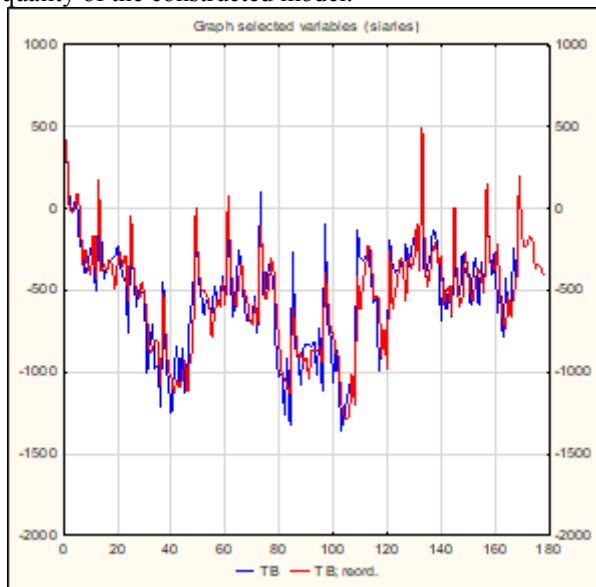

Fig. 10. Visualization of source data and results obtained using the Holt model of exponential smoothing

Source: Authors.

Let's compare the forecast results with the actual data of the trade balance between Ukraine and the EU for the first quarter of 2019. To do this, we calculate the relative prediction error by the formula:

$$\delta_{\max} = \max_{t} \left\{ \frac{\left| y_t^{predict} - y_t^{actual} \right|}{\left| y_t^{actual} \right|} \right\} = 0,0428 \tag{5}$$

So, the last model showed good prediction results, since the relative error does not exceed 0.0428 or not more than 4.28%.

4 Conclusion

The main objective of this work was the ability to predict the dynamics of the trade balance between Ukraine and the EU using various adequate mathematical models of time series. And as a result of the research, this goal has been achieved. Note that the Ukraine-EU trade balance is consistently negative. Since January 1, 2016, Ukraine and the EU have entered into an agreement on an expanded free trade zone (Deep and Comprehensive Free Trade Area). Thanks to this agreement, the EU has become the largest trading partner for Ukraine. At the same time, quotas and the requirement to comply with European standards significantly limit the export of Ukrainian goods to the EU. This is clearly seen from the studies: only in the first month of the year is the Ukraine-EU trade balance positive. That is, Ukraine very quickly "chooses" its quotas for the supply of export goods. The Ukrainian government intends to revise the agreement Deep and Comprehensive Free Trade Area. Once again, as a result of the research conducted in this work, it can be concluded that it is possible to build models of time series that are quite acceptable for predicting the trade balance. Therefore – the goals of the work are achieved.

References

1. J. Vrbka, Z. Rowland, P. Šuleř, Comparison of neural networks and regression time series in estimating the development of the EU and the PRC trade balance. *SHS Web of Conferences*: *Innovative Economic Symposium 2018 – Milestones and Trends of World Economy*, **61**, 01031 (2019)

2. G.V. Golubova, The Statistical Analysis and Forecasting of Ukraine's Export. *Statistics of the Ukraine*, **2**, 6-12 (2018)

3. G.E.P. Box, G.M. Jenkins, Time Series Analysis. *Forecasting and Control* (1976)

4. C. Dzitsaki, Forecast of SARIMA models: An application to unemployment rates of Greece. *American Journal of Applied Mathematics and Statistics*, **4**, 5, 136-148 (2016)

5. N. Dzitsakis, P. Klareglon, Forecasting Unemployment Rates in USA Using Box-Jenkins Methodology. *Internathional Journal of Economics and Financial Issues*, **8(**1), 9-20 (2018)

6. L. Dvořáková, CZ GDP Prediction via neural networks and Box-Jenkins Method. *SHS Web of Conferences: Innovative Economic Symposium 2017 – Strategic Partnership in International Trade*, **39**, 01005 (2017)

7. Y. Kozak, V. Matskul, T. Shengelia, Mathematical methods and models for master of economics. *Practical applications* (2017)

8. The State Statistics Service of Ukraine [online], Available at: http://www.ukrstat.gov.ua (2019)