AMERICAN JOURNAL OF ECONOMICS AND BUSINESS MANAGEMENT

ISSN: 2576-5973
Vol. 3, No.4, September-October 2020

Using a GMDH-type neural network and ARIMA model to forecasting GDP in Algeria during the period of 1990-2019

Abdelkader Sahed[1], Hacen Kahoui[2], Mohammed Mekidiche[3]

[1,2,3]Faculty of Economics, University Centre of Maghnia, Maghnia, Algeria

Correspondent author: sahed14@yahoo.fr DOI 10.31150/ajebm.Vol3.Iss4.200

Abstract: Forecasting is a method to predict the future using data and the last information as a tool assists in planning to be effective. GMDH-Type (Group Method of Data Handling) artificial neural network (ANN) and Box-Jenkins method are among the know methods for time series forecasting of mathematical modeling. in the present study GMDH-type neural network and ARIMA method has been used to forecasted GDP in Algeria during the period 1990 to2019 (Time series of quarterly observations on Gross Domestic Product (GDP) is used). Root mean square error (RMSE) was used as performance indices to test the accuracy of the forecast. The empirical results for both models showed that the GMDH model is a powerful tool in forecasting GDP and it provides a promising technique in time series forecasting methods.

Keywords: Gross Domestic Product, GMDH-type ANN, ARIMA Model.

Introduction

The Gross Domestic Product (GDP) is one of the most important economic indicators for any country, and is the main measuring instrument used to measure a country's economic activity of a country, and for this reason the GDP index is one of the most used and most advanced standards. When measuring the level of economic progress in most countries (Napitupulu, 2010, p. 89), Because it represents the market value of all goods and services produced by the economy during a year (Wabomba, 2016, p. 64). In other words, It is the market value of all final goods and services produced within the borders of a nation in a year, i.e. it represents the total statistics of all economic activities, and for this the economic performance can be measured with the help of a country's GDP. there are three methods of measuring GDP, namely: the Production approach which It includes the total the value added of the various institutional sectors or different industries in addition plus taxes and fewer subsidies on products in the period , and we also have the Expenditure approach , which includes the final uses of goods and services by resident institutional units, i.e. the actual final consumption and gross capital formation, in addition to exports and the reduction of imports of goods and services. Finally, is Income approach, it is equal to the sum of all factor income generated by production in the country (the sum of remuneration of employees, capital income, and gross operating surplus of enterprises i.e. profit, taxes on production and imports less subsidies) in a period (Ahmed, 2020, p. 7).

Thus, the Gross domestic product has a fundamental and vital importance for governments in setting their strategy and economic development policy. and For this, it is necessary to accurately forecast the gross domestic product in order to have a firm idea of the future trend of the country's economy.

Given the importance of this topic, it prompted researchers to develop methods used to improve prediction accuracy. The GMDH model is considered one of the techniques or sub-models of artificial neural networks, and This model has been successfully used to deal with uncertainty, linear or nonlinearity of systems in a wide range of disciplines such as engineering, science, economy, medical diagnostics, signal processing and control systems (Shabri, 2014, p. 3052). where this type of model is a powerful tool for forecasrting (Xie, 2017, p. 1333) .

The importance of this study is complemented by the use of the GMDH-Type model (Group Method of Data Handling), a group method of data handling based on the Rosenblatt's perceptron method, which the original mathematical algorithm was developed by the Russian mathematician Ivakhnenko (1968) and the implementation for R software was developed by Dag & Yozgaligil (2016) (Carvalho, 2019, p. 325) to forecast Algeria's GDP.

Based on the above, and given the great importance of the subject of prediction, the study problem revolves around the following main question:

How effective is the GMDH-Type model in forecasting Algeria's gross domestic product (GDP)?

Objectives of the study: Learn about GMDH (one the sub model of Artificial Neural Networks).Testing the capacity of the GMDH-type model in time-series analysis to forecasting of GDP, compared to the Autoregressive Model and Integrated Moving Average (ARIMA).using the GMDH type model of Forecasting Algeria's GDP.

The importance of this study lies in forecasting the GDP of Algeria, as the forecasting process provides researchers in this scope and those interested and decision-makers with the necessary information in order to develop economic plans to face any problems or crises.

Literature review

Study for (Yang, 2016),This study aimed to Application of ARIMA Model in the Prediction of Chinese GDP. The result shows that this model is effective to forecast the GDP in a short term.

Study for (Loermann, 2019),This study aimed Now casting US GDP with artificial neural networks we compare to forecasts of state of the art dynamic factor models, The result shows that this the neural network outperforms the dynamic factor model in terms of now and forecasting.

Study for (Urrutia, 2019),This study aimed Forecasting The Gross Domestic Product of The Philippines Using Bayesian Artificial Neural Network and Autoregressive Integrated Moving Average, The result shows that this model Bayesian ANN outperforms of ARIMA . and Paired T-test concludes that there is no significant difference between actual and predicted value.

Study for (Dritsaki, 2015) ,This study aimed at modeling and forecasting real GDP rate in Greece, using the Box- Jenkins methodology during the period 1980-2013.The result shows that this model is effective to forecast the Greece’s real GDP.

Study for (Giovanis, 2010) ,This study aimed Application of Feed-Forward Neural Networks Autoregressive Models with Genetic Algorithm in Gross Domestic Product Prediction and compared with of the ordinary autoregressive model ,The result shows that this the proposed regression’s forecasting outperform significant those of autoregressive model.

Study for (NAPITUPULU, 2012),This study aimed ANN is used as a tool for forecasting GDP growth in Indonesia, using some variables, such as GDP growth in the two previous periods, population growth rate, inflation, exchange rate and political stability and security conditions in Indonesia. The result shows that ANN forecasts GDP relatively better than the one issued by the government.

Methodology

ARIMA Model:

The use of the ARIMA model has been extensively known and has been applied in forecast studies, due to their attractive theoretical properties and the various empirical supporting evidence (Da Veiga, 2014, p. 610). The ARIMA model is a combination of three processes: autoregressive processes (p),differencing processes (d) , And the moving average processes (q). It takes the following formula: ARIMA(p, d, q) (Guha, 2016, p. 118).

Autoregressive Model:

The Model is a representation of a type of random process. It is defined in this model as the current value of the time series Y_t expressed in terms of the weighted sum of the values $(Y_{t-1}, Y_{t-2}, \dots)$In addition to the current random error value (ε_t) ,and in the An autoregressive (AR) model, the value of the current variable depends on the preceding values (Bakar, 2017, p. 132) .

Where the notation AR (p) refers to the autoregressive model, and The AR (p) model is defined as:

$$Y_t = \theta_0 + \sum_{i=1}^{p} \emptyset_i Y_{t-i} + \varepsilon_t \qquad (1)$$

Where $\emptyset_1, \emptyset_2, \dots . \emptyset_p$ is the parameters of the Model, and θ_0 is the constant , As for ε_t it is white noise.

Moving average Model:

In this model, the current value of the time series $(Y_t)$is expressed as a function of the current value and the previous values of the random error term $(\varepsilon_t, \varepsilon_{t-1}, \dots)$, which means that the model is based on the stochastic error of the time series.

The notation MA(q)refers to the autoregressive model, and the MA(q)model is defined as follows:

$$Y_t = \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \qquad (2)$$

Where $\theta_1, \theta_2, \dots \theta_q$ is the parametres of the model, and $\varepsilon_t, \varepsilon_1, \dots \varepsilon_q$ is are white noise error terms.

Autoregressive Moving Average:

autoregressive and Moving Average Model: The ARMA Model is a combination between autoregressive model and moving average model,in general , it takes the following form ARMA(p, q), i.e.is a combination of an the AR(p),equation (1), with MA(q), equation(2) ,It can be written as follows (Zakria, 2009, p. 2015)::

$$Y_t = \theta_0 + \sum_{i=1}^{p} \emptyset_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \qquad (3)$$

Autoregressive integrated moving average Model:

The ARIMA models can be defined as an extension of ARMA models where a difference element has been added expressing the process in which non-stationary series are converted into stationary series by allowing the differencing of the data series resulting to ARIMA (p, d, q): as it has three parameters; p is the order of autoregressive, d is the degree of differencing, and q is the order of moving-average. For example, suppose that, if $Y_t$is non-stationary series, we will take a first-difference of $Y_t$so that ΔY_tbecomes stationary, then the ARIMA (p, 1, q) model is (Abonazel, 2019, p. 37):

$$\Delta Y_t = \theta_0 + \emptyset_1 \Delta Y_{t-1} + .. + \emptyset_p \Delta Y_{t-p} + \varepsilon_t - \theta_1 \Delta \varepsilon_{t-1} - .. - \theta_q \Delta \varepsilon_{t-q} \qquad (4)$$

Where $\Delta Y_t = Y_t - Y_{t-1}$ But, in the case if $p = q = 0$ in equation (4), then the model becomes in this case a random walk model which classified as ARIMA $(0, 1, 0)$.

GMDH Neural Networks:

The GMDH algorithm developed by Dag & Yogaligil in 2016, as mentioned previously assumes a time series data set that containswith t time units and p inputs, are constructing models for the data with lags or time intervals, where the numbers of observations in columns is represented by $t - p$, where p It represents the number of entries of the lagged time series, and the variable denoted z is placed in the models as a response variable, with the values that best estimate the y values and the other variables are taken into the model with lags of the time series x_i, with $i = 1, 2, \dots, p$ After this, the algorithm selects the best model which explains the relationship between the response or reaction and lagged time series (Carvalho, 2019, p. 326).

This is done by the via transfer functions that are mainly used. which Explains the relationship between response and lagged time series in the algorithms of the GMDH neural network. The via transfer functions used are as follows (Yahya, 2019, p. 322), (Dağ, 2015, p. 16):

-GMDH model using the Sigmoid function (GMDH-Sigmoid) :

$$y_k = \frac{1}{1+e^{-zk}} \tag{5}$$

- GMDH model using Radial Basis function (GMDH-RBF):

$$y_k = e^{-z_k^2} \tag{6}$$

-GMDH model using Tangent function(GMDH-Taangent) :

$$y_k = \tan(y) \tag{7}$$

- GMDH Model Using Polynomial function (GMDH-Polynomial):

$$y_k = z_k \tag{8}$$

GMDH neural network algorithms are modeling techniques ,Where you learn relationships between variables, in a time series perspective, this algorithm learns the relationship betwee namong the lags. After learning the relationships, the method to be following in the algorithm is determined (Dag, 2016, p. 381). First, we use the equation describing as the Ivakhnenko polynomial (1968) to construct a high order polynomial; and are given as follows (Basheer, 2016, p. 10792):

$$y = a + \sum_{i=1}^{M} b_i x_i + \sum_{i=1}^{M} \sum_{j=1}^{M} c_{ij} x_i x_j + \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{M} d_{ijk} x_i x_j x_k + \dots \tag{9}$$

Where:

M Represents the number of variables. a, b, c, d Represent the parameters of the variables in polynomials (called weights). y Denotes the variable of response or reaction. and x_i, x_j represents thelagged time series to be regressed (retard). In general, the period is used in the calculation up to the square terms, as follows:

$$y = a + \sum_{i=1}^{M} b_i x_i + \sum_{i=1}^{M} \sum_{j=1}^{M} c_{ij} x_i x_j \qquad (10)$$

The GMDH algorithm takes into account all the pairwise combinations of p for lagged time series, and for this each combination of each neuron enters, and by using these two inputs a model is created to estimate the desired output, and to explain more, two variables are entered into the neuron meaning m = $2(x_i \text{and } x_j)$, which gives one result Graduated as an output, and the structure of the model is determined by the Ivakhnenko polynomial in Equation 7, where this specification requires estimation of six coefficients and this in each model and is estimated by least mean square error (MSE):

$$\hat{y} = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \qquad (11)$$

The coefficients of Equation (11) are to be estimated in each neuron and by using the estimated coefficients and input variables in each neuron, the desired output is forecasted. And By means of the standard mean square error (MSE) (used to forecasted), the p-neurons are selected and the h-neurons are removed from the network.

Thus the output from the specified neuron be the input of the last layer, and this method continues to it reaches the last layer, in this layer (the last layer), only one neuron is specified, and finally the output from the last layer is the forecasting value of the time series.

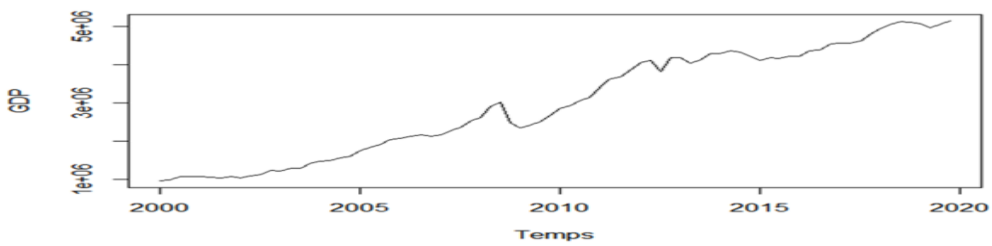
Results and Discussion

The dataset used in this paper is obtained from by the ONS .Where they represent The time series of GDP Quarterly in Algeria from 1990 to 2019.This implies that the study dealt with GDP time series of Algeria with 120 observations.

In this part, we will discuss the stages of the B-J methodology, as follows:

We will The preliminary analysis of the data was done by use of time plot for the series as shown by Figure 1:

Figure 1. Time series plot of GDP in Algeria



From figure 1 above, we note by visual inspection of the time series plot indicates that Algeria's GDP has shown Secular Trend. This implies that both the mean and the variance are not constant. Therefore we regard it as a non-stationary time series.Before dealing with the time series stability study, we divided the time series data into 20% for the purpose of testing with the expected values and the rest i.e. 80% of the data for training.

In order to validate our decision. The enhanced Dickey Fuller test (ADF) and Phillips-Perron (PP) test and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test are used to test for stability in time series. Table 2 shows the results:

Table 2. Stationarity Test of original series

Test	P-Value	Decision
ADF	0.909	non-stationary
PP	0.7348	non-stationary
KPSS	0.01	non-stationary

From the previous results of the ADF test, we note that the test is not significant, as the p-value 0.909 > 0.05 and indicating that the series is non stationary . Also, when performing the pp test, the test is not significant, and also when using the KPSS test, the results confirmed that the data series is non stationary , so we reject the null hypothesis and accept the alternative hypothesis indicating that the series is non stationary.

In order to make the series constant in variance and medium, the first difference of the data series were taken, The results are shown in Table 3:

Table 3. Stationarity Test of original series

Test	P-Value	Decision
ADF	0.01	Stationary
PP	0.01	stationary
KPSS	0. 1	stationary

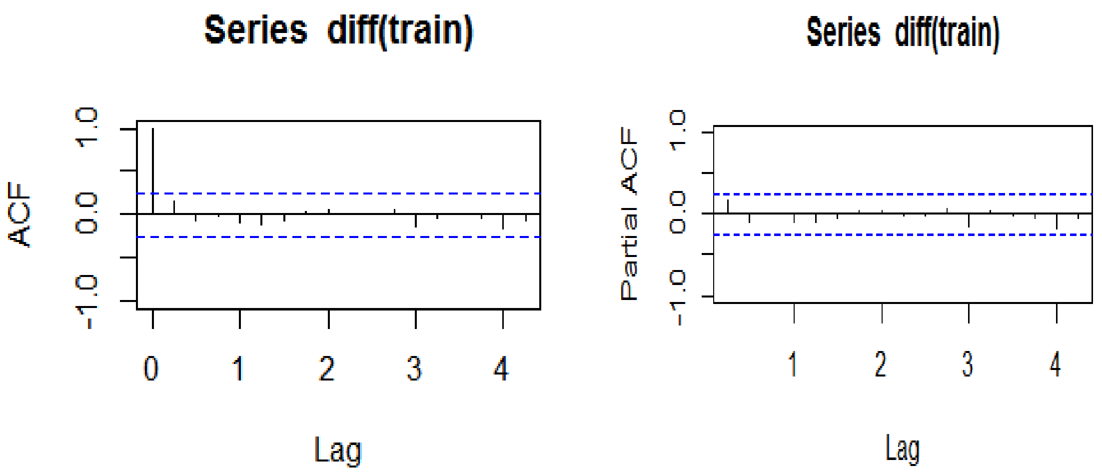
Through Table 3 we note that the ADF test is significant, as the p-value 0.101< 0.05 indicates that the series is stationary. Also when doing PP test ,as the p-value0.01< 0.05, the test is significant, and so is KPSS test, as the p-value 0.1< 0.05 , so the data series is stationary.

After time series stationary comes the identification stage and this implies the necessity to use the first-difference form data. Now, the time series are integrated of order 1, which means that we will have $d = 1$ in $ARIMA(p, d, q)$ model for time series.

Next, we need to decide determine the many autoregressive (p) and moving average (q) parameters are necessary to give an effective model of the process. The correlograms given in figure (2) enables us to estimate these parameters, In order to make a decision; it is not easy and in less common cases does not only require experience, which also requires a great deal of experimentation with alternative models in order to choose the best model.

By looking at the ACF and PACF chart for a time series, there are three models as shown in Figure.2.

Figure 2. ACF and PACF plot of 1nd difference of GDP series



were selected and in order to choose the best among them The Akaike information criteria (AIC) (Akaike, 1974) was used in order to fit the best model. The model that gives the lowest AIC value would be selected as the best.

Table 4. Different ARIMA Models for GDP in Algeria

Model	AIC	Log likelihood
ARIMA(0,1,1)	-198.18	101.01
ARIMA(1,1,1)	-200.77	103.39
ARIMA(1,1,0)	-199.05	101.52

according to the AIC estimator is given in Table 4. Based on the results of this estimation ARIMA(1,1,1) is selected as the best model because it shows the lowest AIC value.

After fitting ARIMA(1,1,1) model has been fitted to the series of GDP. where an Investigating the results of this fit, resulted that all coefficients are significant and the diagnostic model suggests that this model is suitable.

GMDH Model:

As previously mentioned, the time series of GDP has been divided into two parts, training and testing data. The training (80 percent) part is used to obtain the model parameters , As for the testing(20 percent) part it was used in order to compare all models generated.

In order to determine the optimal input neuron number, and the number of layers in the network, and the transfer functions that we discussed in the theoretical side were used (GMDH-Sigmoid, GMDH-RBF, GMDH-Taangent and GMDH-Polynomial). Among the four models, GMDH which used radial basis function (RBF) performed the best in terms of RMSE, with a value of 0.042539272 in training step.

The structure of the selected model included the following, the input is 4, layers are 3 and transfer function is the Radial Basis function (RBF).

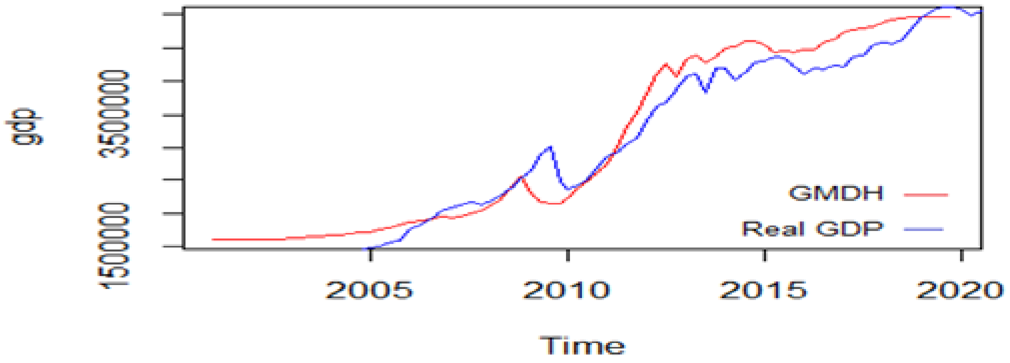
Comparison between ARIMA model and GMDH-neural network model:

In order to determine the better model and appropriate for forecasting GDP. we compare the model statistic of ARIMA and GMDH-Type (ANN) in terms Training and Testing using RMSE . Model with lower values of RMSE as compare to the other model, is better. The model statistic of GMDH-Type(ANN) and ARIMA both are presented in Table 5. The table indicates that GMDH-Type (ANN) is better model than ARIMA in the training and testing step.

Table 5. The results of models computed over the Training and Testing period

Model No.	Description	Training RMSE	Testing RMSE
1	ARIMA(1,1,1)	0.046181	0.023262
2	GMDH Model	0.042539	0.010858

Figure 3. Real GDP vs predicted GMDH



The Figure 3 gives the comparison of the forecasted and actual GDP data of the GMDH-Type (Group Method of Data Handling) artificial neural network (ANN).

From the Figure.3, it is concluded that the GMDH-Type is much closer to the actual values. This model can be more useful and appropriate in predicting and forecasting the future values.

Conclusion

The major purpose of studies on forecasting accuracy is to help the forecasters in selecting best forecasting model. This study has proposed two efficient approaches forecasting models. In the first model GMDH is used, the second model ARIMA is used, on real data for GDP in Algeria. Based on the result, can be conclude that the best forecasting model to predict the GDP in Algeria prediction is model for GMDH-Type (Group Method of Data Handling) artificial neural network (ANN) outperform the ARIMA .

In addition, GMDH-type offer consistent prediction performance compared to ARIMA and performs very well in economic and financial data, and thus it makes a great contribution as an efficient tool for of forecasting. By knowing predicted values, it helps the decision makers to take decision on GDP.

References

1. Napitupulu, T. A. (2010, December). Artificial neural network application in gross domestic product forecasting: an Indonesia case. In 2010 Second International Conference on Advances in Computing, Control, and Telecommunication Technologies (pp. 89-93). IEEE.p89.
2. Wabomba, M. S., Mutwiri, M. P., & Mungai, F. (2016). Modeling and forecasting Kenyan GDP using autoregressive integrated moving average (arima) models. Science Journal of Applied Mathematics and Statistics, 4(2), 64-73.p64.
3. Ahmed, A., & Salan, M. S. A. Forecasting GDP Of Bangladesh Using Time Series Analysis.p7.
4. Shabri, A., & Samsudin, R. (2014). A hybrid GMDH and box-jenkins models in time series forecasting. Applied Mathematical Sciences, 8(62), 3051-3062.p3052.
5. Xie, L., Xiao, J., Hu, Y., Zhao, H., & Xiao, Y. (2017). China's energy consumption forecasting by GMDH based auto-regressive model. Journal of Systems Science and Complexity, 30(6), 1332-1349.p1333.
6. Carvalho, R. L. D. S., & Delgado, A. R. (2019). Estimates of reference evapotranspiration in the municipality of Ariquemes (RO) using neural networks GMDH-type. Revista Brasileira de Engenharia Agrícola e Ambiental, 23(5), 324-329.p325.
7. Yang, B., Li, C., Li, M., Pan, K., & Wang, D. (2016, December). Application of ARIMA model in the prediction of the gross domestic product. In 2016 6th International Conference on Mechatronics, Computer and Education Informationization (MCEI 2016). Atlantis Press.
8. Loermann, J., & Maas, B. (2019). Nowcasting US GDP with artificial neural networks.
9. Urrutia, J. D., Longhas, P. R. A., & Mingo, F. L. T. (2019, December). Forecasting the Gross Domestic Product of the Philippines using Bayesian artificial neural network and

autoregressive integrated moving average. In AIP Conference Proceedings (Vol. 2192, No. 1, p. 090012). AIP Publishing LLC.

10. Dritsaki, C. (2015). Forecasting real GDP rate through econometric models: an empirical study from Greece. Journal of International Business and Economics, 3(1), 13-19.
11. Giovanis, E. (2010). Application of Feed-Forward Neural Networks Autoregressive Models with Genetic Algorithm in Gross Domestic Product Prediction. World Academy of Science, Engineering and Technology, 64, 638-664.
12. Da Veiga, C. P., Da Veiga, C. R. P., Catapan, A., Tortato, U., & Da Silva, W. V. (2014). Demand forecasting in food retail: A comparison between the Holt-Winters and ARIMA models. WSEAS transactions on business and economics, 11(1), 608-614.p610.
13. Guha, B., & Bandyopadhyay, G. (2016). Gold price forecasting using ARIMA model. Journal of Advanced Management Science, 4(2).p118.
14. Bakar, N. A., & Rosbi, S. (2017). Autoregressive integrated moving average (ARIMA) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction. International Journal of Advanced Engineering Research and Science, 4(11), 237311.p132.
15. Zakria, M., & Muhammad, F. (2009). Forecasting the population of Pakistan using ARIMA models. Pakistan Journal of Agricultural Sciences, 46(3), 214-223.p215.
16. Abonazel, M. R., & Abd-Elftah, A. I. (2019). Forecasting Egyptian GDP Using ARIMA Models. Reports on Economics and Finance, 5(1), 35-47.p37.
17. Yahya, N. A., Samsudin, R., Shabri, A., & Saeed, F. (2019). Combined group method of data handling models using artificial bee colony algorithm in time series forecasting. Procedia Computer Science, 163, 319-329.p322.
18. Dağ, O. (2015). GMDH-type neural network algorithms for short term forecasting (Master's thesis, MIDDLE EAST TECHNICAL UNIVERSITY).p16.
19. Dag, O., & Yozgatligil, C. (2016). GMDH: An R Package for Short Term Forecasting via GMDH-Type Neural Network Algorithms. R J., 8(1), 379.p381.
20. Basheer, H., & Khamis, A. B. (2016). A hybrid group method of data handling (GMDH) with the wavelet decomposition for time series forecasting: A review. ARPN Journal of Engineering and Applied Sciences, 11, 10792-10800.p10792.