Forecasting Egyptian GDP using ARIMA models

Article · January 2019

DOI: 10.12988/ref.2019.81023

2 authors:

Mohamed R. Abonazel
Cairo University

Ahmed Ibrahim
Egyptian Russian University

Reports on Economics and Finance, Vol. 5, 2019, no. 1, 35 - 47
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ref.2019.81023

Forecasting Egyptian GDP Using ARIMA Models

Mohamed Reda Abonazel* and Ahmed Ibrahim Abd-Elftah

Department of Applied Statistics and Econometrics
Institute of Statistical Studies and Research
Cairo University, Giza, Egypt
*Corresponding author

Abstract

The Gross Domestic Product (GDP) is that the value of all product and services made at intervals the borders of a nation in an exceedingly year. In this paper, the Box-Jenkins approach has been used to build the appropriate Autoregressive-Integrated Moving-Average (ARIMA) model for the Egyptian GDP data. Egypt's annual GDP data obtained from the World-Bank for the years 1965 to 2016. We find that the appropriate statistical model for Egyptian GDP is ARIMA (1, 2, 1). Finally, we used the fitted ARIMA model to forecast the GDP of Egypt for the next ten years.

Keywords: Box-Jenkins approach, Egypt, Forecasting, Goodness-of-fit measures, Gross domestic product, Residuals analysis

1. Introduction

GDP represents the market value of all goods and services produced by the economy during the period measured, including personal consumption, government purchases, private inventories, paid-in construction costs and the foreign trade balance (exports minus imports). GDP can be measured in three ways: (i) the expenditure approach, (ii) the production approach, and (iii) the income approach.

The issue of GDP has becomes the most concerned amongst macro economy variables and data on GDP is regarded as the important index for assessing the national economic development and for judging the operating status of macro economy as a whole (Ning et al, 2010). It is often considered the best measure of how well the economy is performing. Also, it is a vital basis for government to set up economic developmental strategies and policies.

Therefore, an accurate prediction of GDP is necessary to get an insightful idea of future trend of an economy.

In this paper, we will use the statistical techniques in time series analysis to estimate and predict the Egyptian GDP. One of the most common of these techniques is ARIMA models. There are many studies used these models for studying the GDP in different countries, such as Bhuiyan et al (2008), Ning et al (2010), Maity and Chatteriee (2012), Dritsaki (2015), Wabomba et al (2016), and Uwimana et al (2018).

The paper is organized as follows. Section 2 presents the statistical background for univariate time series analysis. In section 3, our proposed Egyptian GDP model has been presented based on Box-Jenkins approach. Finally, section 4 offers the summary and the concluding remarks for our study.

2. Statistical Background

The time series analysis can provide short-run forecast for sufficiently large amount of data on the concerned variables very precisely, see Granger and Newbold (1986). In univariate time series analysis, the ARIMA models are flexible and widely used. The ARIMA model is the combination of three processes: (i) Autoregressive (AR) process, (ii) Differencing process, and (iii) Moving-Average (MA) process. These processes are known in statistical literature as main univariate time series models, and are commonly used in many applications.

2.1. Autoregressive (AR) model

An autoregressive model of order p, AR (p), can be expressed as:

$$X_t = c + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \varepsilon_t; \;\; t = 1,2, \ldots T, \tag{1}$$

where ε_t is the error term in the equation; where ε_t a white noise process, a sequence of independently and identically distributed (iid) random variables with $E(\varepsilon_t) = 0$ and $var(\varepsilon_t) = \sigma^2$; i.e. $\varepsilon_t \sim iid\ N(0, \sigma^2)$. In this model, all previous values can have additive effects on this level X_t and so on; so it's a long-term memory model.

2.2. Moving-average (MA) model

A time series $\{X_t\}$ is said to be a moving-average process of order q, MA (q), if:

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}. \tag{2}$$

This model is expressed in terms of past errors as explanatory variables. Therefore only q errors will effect on X_t, however higher order errors don't effect on X_t; this means that it's a short memory model.

2.3. Autoregressive moving-average (ARMA) model

A time series $\{X_t\}$ is said to follow an autoregressive moving-average process of order p and q, ARMA (p, q), process if:

$$X_t = c + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}. \tag{3}$$

This model can be a mixture of both AR and MA models above.

2.4. ARIMA Models

The ARMA models can further be extended to non-stationary series by allowing the differencing of the data series resulting to ARIMA models. The general non-seasonal model is known as ARIMA (p, d, q): where with three parameters; p is the order of autoregressive, d is the degree of differencing, and q is the order of moving-average. For example, if X_t is non-stationary series, we will take a first-difference of X_t so that ΔX_t becomes stationary, then the ARIMA (p, 1, q) model is:

$$\Delta X_t = c + \alpha_1 \Delta X_{t-1} + \cdots + \alpha_p \Delta X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}, \tag{4}$$

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

2.5. Box-Jenkins Approach

In time series analysis, the Box-Jenkins (1970) approach, named after the statisticians George Box and Gwilym Jenkins, applies ARIMA models to find the best fit of a time series model to past values of a time series. For more details about Box–Jenkins time series analysis, see for example Young (1977), Frain (1992), Kirchgässner et al (2013), and Chatfield (2016). Figure 1 shows the four iterative stages of modeling according this approach.

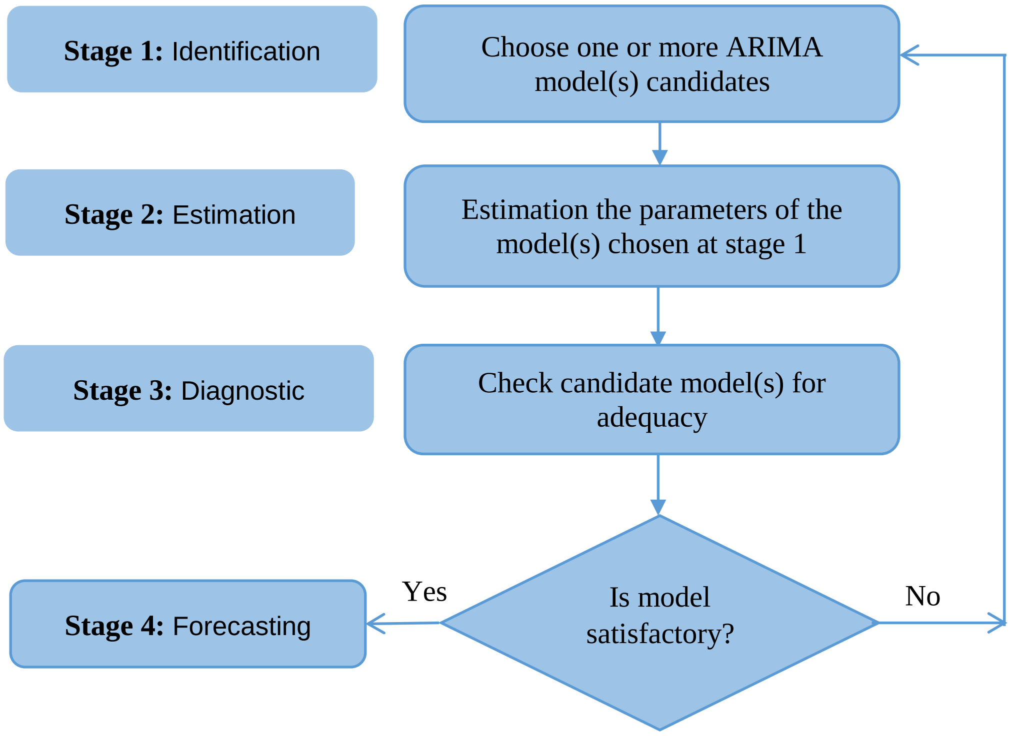


Figure 1: Stages in the Box-Jenkins iterative approach

The four stages modeling in the Box-Jenkins iterative approach:

- ***Model identification***: making sure that the variables are stationary, identifying seasonality in the series, and using the plots of the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) of the series to identification which autoregressive or moving-average component should be used in the model.
- ***Model estimation***: using computation algorithms to arrive at coefficients that best fit the selected ARIMA model. The most common methods use Maximum Likelihood Estimation (MLE) or non-linear least-squares estimation.
- ***Model checking***: by testing whether the estimated model conforms to the specifications of a stationary univariate process. In particular, the residuals should be independent of each other and constant in mean and variance over time; plotting the ACF and PACF of the residuals are helpful to identify misspecification. If the estimation is inadequate, we have to return to step one and attempt to build a better model. Moreover, the estimated model should be compared with other ARIMA models to choose the best model for the data. The two common criteria used in model selection: Akaike's Information Criterion (AIC) and Bayesian Information Criteria (BIC) which are defined by:

$$AIC = 2m - 2\ln(\hat{L}), \;\; BIC = ln(n)m - 2\,ln(\hat{L}), \qquad (5)$$

where $\hat{L}$ denotes the maximum value of the likelihood function for the model, m is the number of parameters estimated by the model, and n is the number of observations (sample size). Practically, AIC and BIC are used with the classical criterion: the Mean Squared Error (MSE).

- ***Forecasting***: when the selected ARIMA model conforms to the specifications of a stationary univariate process, then we can use this model for forecasting.

3. A Proposed Egyptian GDP Model

3.1. The Data

In this study, the annual GDP of Egypt was obtained from World-Bank from 1965 to 2016. This means that we have 52 observations of GDP that satisfies the rule of having over 50 observations in Box-Jenkins approach of your time series forecasting (Chatfield, 2016). Based on this data, we will propose the appropriate ARIMA model and then use it to forecast the Egyptian GDP for the next ten years (from 2017 to 2026).

3.2. In-Sample: Estimation Results

The preliminary analysis of the data was done by using time plots of the series as shown by Figures 2 and 3 respectively. From Figure 2, a visual inspection of the time plots indicates that Egyptian GDP is not stationary series. Moreover, the non-stationary behavior of the series is also confirmed by the ACF and PACF plots of the series, as in Figure 3, since all p-values of Q-test are less than 0.05. To reach stationary we will take the differencing, as practiced, in developing ARIMA model.

A visual examination of Figure 3 confirms that the Egyptian GDP series contains a seasonal trend can often be carried out by logarithmic transformation. The result is that the exponential trend will be transformed into a linear trend. Before embarking on further analysis using the Box-Jenkins approach, the data has to be transformed to achieve stationary.

The series was transformed by taking the second differences of the natural logarithms of the values in the series so as to attain stationary in the second moment. The equation representing the transformation is given by:

$$X_t = \ln \mathrm{GDP}_t - \ln \mathrm{GDP}_{t-2}, \tag{6}$$

where $\ln \mathrm{GDP}_t = \log_e(\mathrm{GDP}_t)$.

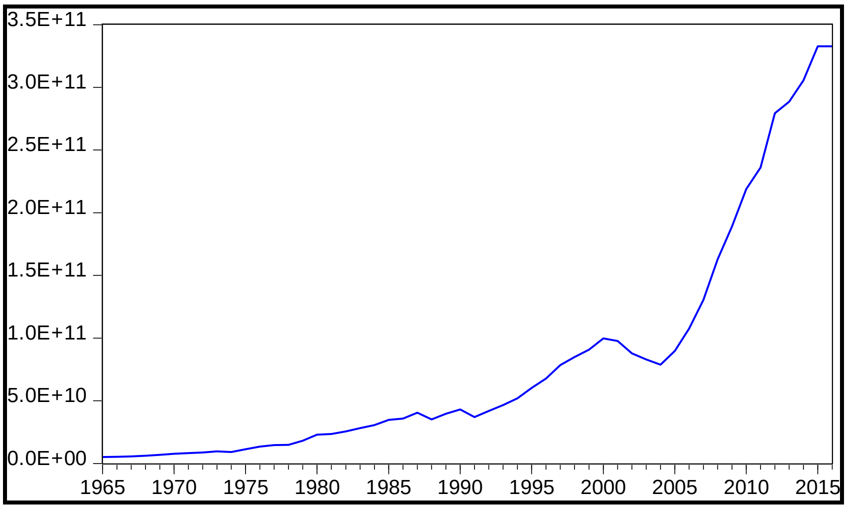
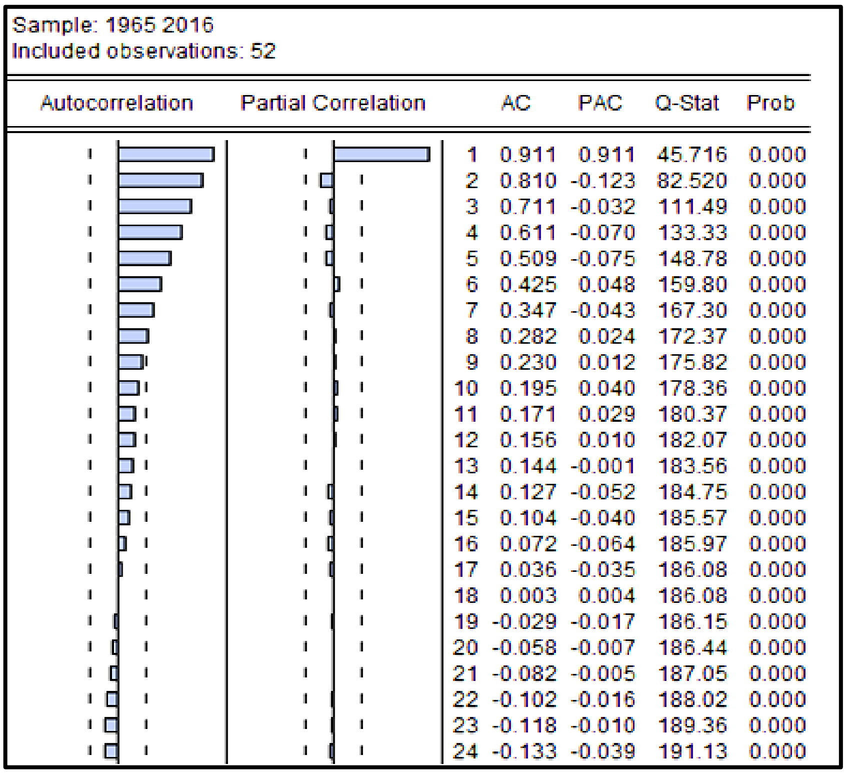


Figure 2: Time series plot for Egyptian GDP data



Sample: 1965 2016
Included observations: 52

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.911	0.911	45.716	0.000
		2	0.810	-0.123	82.520	0.000
		3	0.711	-0.032	111.49	0.000
		4	0.611	-0.070	133.33	0.000
		5	0.509	-0.075	148.78	0.000
		6	0.425	0.048	159.80	0.000
		7	0.347	-0.043	167.30	0.000
		8	0.282	0.024	172.37	0.000
		9	0.230	0.012	175.82	0.000
		10	0.195	0.040	178.36	0.000
		11	0.171	0.029	180.37	0.000
		12	0.156	0.010	182.07	0.000
		13	0.144	-0.001	183.56	0.000
		14	0.127	-0.052	184.75	0.000
		15	0.104	-0.040	185.57	0.000
		16	0.072	-0.064	185.97	0.000
		17	0.036	-0.035	186.08	0.000
		18	0.003	0.004	186.08	0.000
		19	-0.029	-0.017	186.15	0.000
		20	-0.058	-0.007	186.44	0.000
		21	-0.082	-0.005	187.05	0.000
		22	-0.102	-0.016	188.02	0.000
		23	-0.118	-0.010	189.36	0.000
		24	-0.133	-0.039	191.13	0.000

Figure 3: ACF and PACF plots for Egyptian GDP data

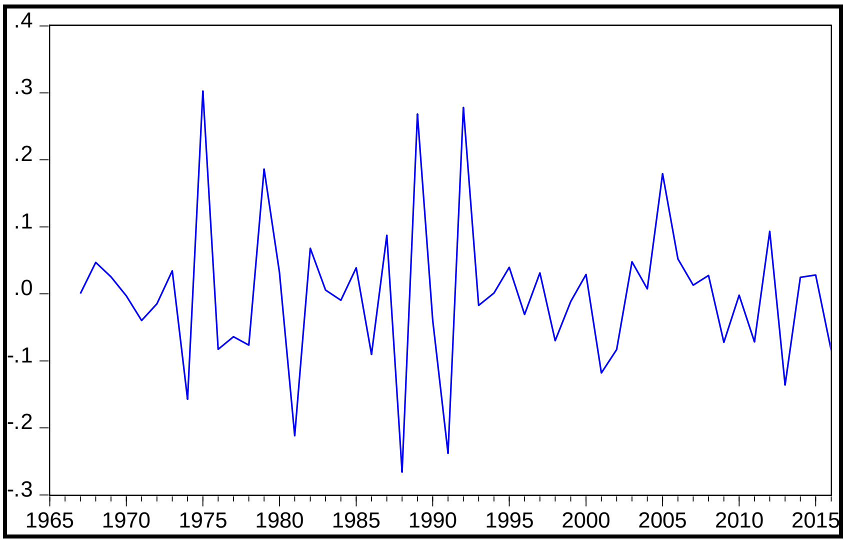
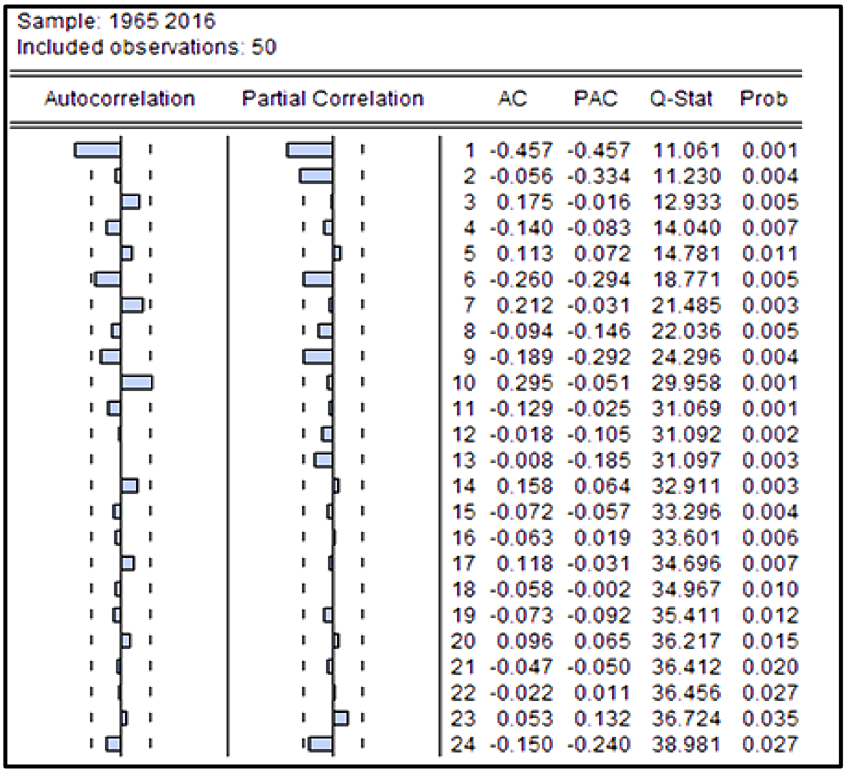


Figure 4: Time series plot for X_t



Sample: 1965 2016
Included observations: 50

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.457	-0.457	11.061	0.001
		2	-0.056	-0.334	11.230	0.004
		3	0.175	-0.016	12.933	0.005
		4	-0.140	-0.083	14.040	0.007
		5	0.113	0.072	14.781	0.011
		6	-0.260	-0.294	18.771	0.005
		7	0.212	-0.031	21.485	0.003
		8	-0.094	-0.146	22.036	0.005
		9	-0.189	-0.292	24.296	0.004
		10	0.295	-0.051	29.958	0.001
		11	-0.129	-0.025	31.069	0.001
		12	-0.018	-0.105	31.092	0.002
		13	-0.008	-0.185	31.097	0.003
		14	0.158	0.064	32.911	0.003
		15	-0.072	-0.057	33.296	0.004
		16	-0.063	0.019	33.601	0.006
		17	0.118	-0.031	34.696	0.007
		18	-0.058	-0.002	34.967	0.010
		19	-0.073	-0.092	35.411	0.012
		20	0.096	0.065	36.217	0.015
		21	-0.047	-0.050	36.412	0.020
		22	-0.022	0.011	36.456	0.027
		23	0.053	0.132	36.724	0.035
		24	-0.150	-0.240	38.981	0.027

Figure 5: ACF and PACF plots for X_t

After the transformation, Figure 4 shows that X_t is stationary and not have trend. So d = 2 in ARIMA model. Next to identify the value of the other two parameters p and q of ARIMA model, we study the appearance of the ACF and

PACF of the differenced series and comparing the two plots that given in Figure 5. We expect that the properly values of the two parameters p and q are p = 1 and q = 1. In other words, the properly model is ARIMA (1, 2, 1). Modeling results of an ARIMA (1, 2, 1) process have been estimated by MLE and are presented in the Table 1. The coefficient estimate of AR (1) is not significant, but the coefficient estimate MA (1) is statistically significant at 1% level of significance, and the model overall are statistically significant at 1% level of significance.

Table 1: Parameter estimates of ARIMA (1, 2, 1) model

Variable	Estimate	Standard Error	t-statistic	P-value
AR(1)	0.1081	0.1477	0.73	0.468
MA(1)	1.0478	0.0437	23.98	0.001
Constant	0.0005	0.0005	0.88	0.384
Model Summary				
R-squared	0.9953	Adj. R-squared		0.9952
F-statistic	5029.502	p-value of F		< 0.0001

The above model was compared with different ARIMA models to select the best model for the data using different goodness-of-fit measures (MSE, AIC, and BIC). The results are presented in Table 2.

Table 2: Evaluation of various ARIMA models

Model	Goodness-of-fit Measure		
	AIC	BIC	MSE
ARIMA (1,2,0)	-82.01	-76.28	0.0104
ARIMA (0,2,1)	-91.68	-85.94	0.0077
ARIMA (2,2,0)	-85.86	-78.21	0.0094
ARIMA (2,2,1)	-83.89	-74.33	0.0096
ARIMA (2,2,2)	-90.07	-80.51	0.0078
ARIMA (0,2,2)	-91.13	-83.49	0.0081
ARIMA (1,2,1)	-93.38	-87.64	0.0076

According to the results in Table 2, the best model is ARIMA (1, 2, 1), because have the minimum values of MSE, AIC, and BIC. Moreover, the predictive power of the model is very high as indicated by the small difference between actual and fitted values as presented in Figure 6. The estimated regression equation of ARIMA (1, 2, 1) model is:

$$X_t = 0.0005 + 0.1081\, X_{t-1} + 1.0478\, \varepsilon_{t-1} + \varepsilon_t \tag{7}$$

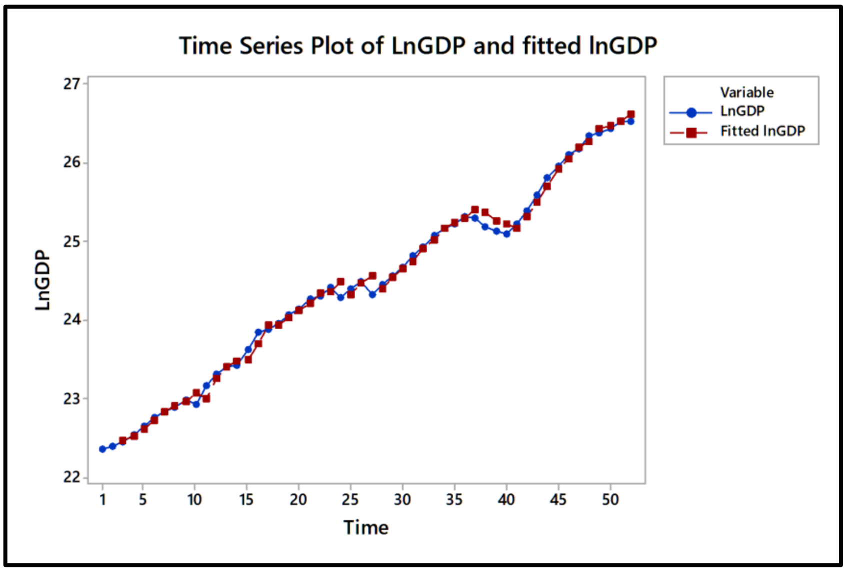


Figure 6: Time series plot for actual and fitted lnGDP values

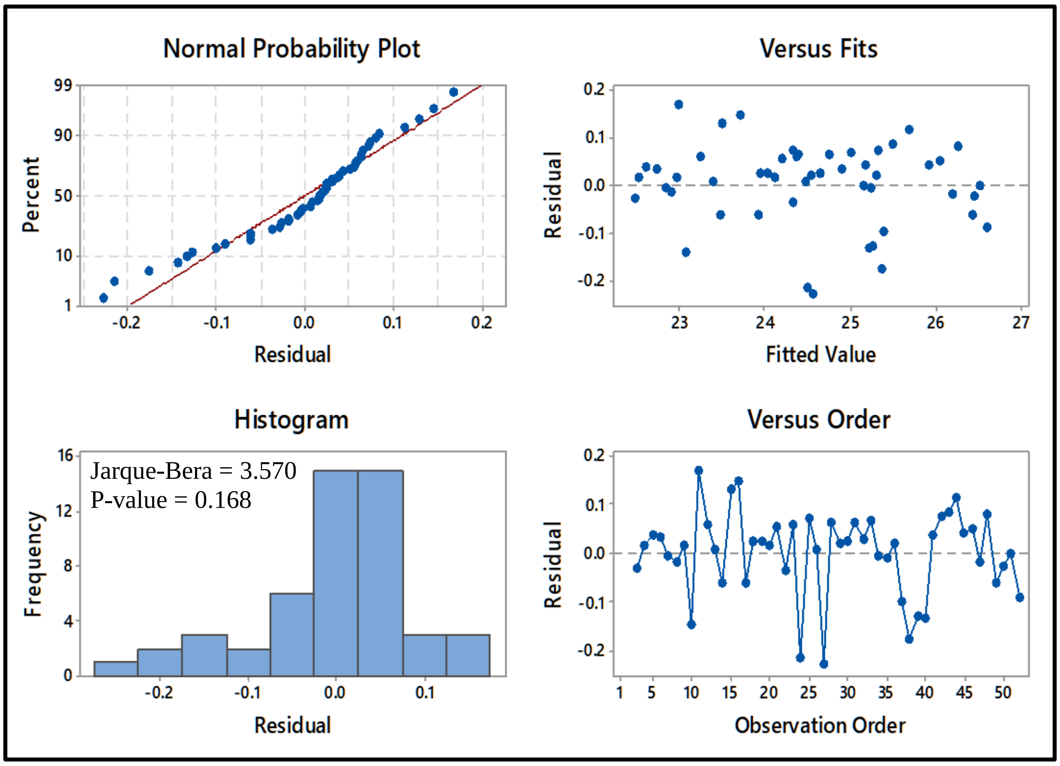
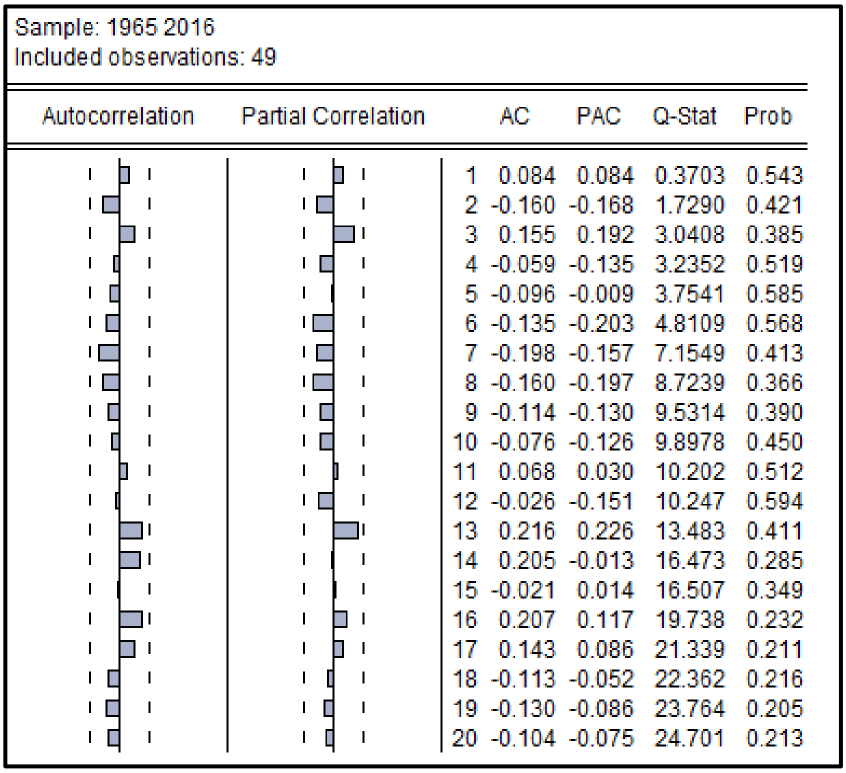


Figure 7: Residuals analysis plots of ARIMA (1, 2, 1) model

According to Box-Jenkins approach, the diagnostic tests of the model are checking the normality and the stationary of the residuals. Figure 7 shows the values of residuals are distributed normally because the p-value of Jarque-Bera (1980) test is 0.168 and more than 0.05. And the residuals are stationary because all p-values of Q-test are more than 0.05 as in Figure 8. Therefore we can use this model for the forecasting.



Sample: 1965 2016
Included observations: 49

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.084	0.084	0.3703	0.543
		2	-0.160	-0.168	1.7290	0.421
		3	0.155	0.192	3.0408	0.385
		4	-0.059	-0.135	3.2352	0.519
		5	-0.096	-0.009	3.7541	0.585
		6	-0.135	-0.203	4.8109	0.568
		7	-0.198	-0.157	7.1549	0.413
		8	-0.160	-0.197	8.7239	0.366
		9	-0.114	-0.130	9.5314	0.390
		10	-0.076	-0.126	9.8978	0.450
		11	0.068	0.030	10.202	0.512
		12	-0.026	-0.151	10.247	0.594
		13	0.216	0.226	13.483	0.411
		14	0.205	-0.013	16.473	0.285
		15	-0.021	0.014	16.507	0.349
		16	0.207	0.117	19.738	0.232
		17	0.143	0.086	21.339	0.211
		18	-0.113	-0.052	22.362	0.216
		19	-0.130	-0.086	23.764	0.205
		20	-0.104	-0.075	24.701	0.213

Figure 8: ACF and PACF plots of ARIMA (1, 2, 1) residuals

3.3. Out-of-Sample: Forecasting Results

Since ARIMA (1, 2, 1) model is fit to the GDP data, therefore we can use equation (7) directly to forecast GDP values for the next ten years out-of sample (from 2017 to 2026). The forecasted values of GDP are given in Table 3. Figure 9 presents the trend of the actual and the forecasted lnGDP values with their 95% confidence limits.

The forecasted values indicate that the Egyptian GDP will continue to rise. Keep in mind that this result is only a predicted value, but the national economy is a complex and dynamic system. Therefore, we should pay attention to the risk of adjustment in the economic operation and maintain the stability and continuity of the microeconomic regulation and control too prevent the economy from severe fluctuations and adjust the corresponding target value according to the actual situation, see Wabomba et al (2016).

Table 3: Forecasted values of Egyptian GDP

year	Forecasted lnGDP	Forecasted GDP	95% Confidence Interval Lower	95% Confidence Interval Upper
2017	26.6171	3.6279E+11	3.0586E+11	4.3032E+11
2018	26.7133	3.9943E+11	3.1145E+11	5.1221E+11
2019	26.8109	4.4038E+11	3.2507E+11	5.9659E+11
2020	26.9092	4.8587E+11	3.4389E+11	6.8638E+11
2021	27.0079	5.3627E+11	3.6710E+11	7.8347E+11
2022	27.1073	5.9231E+11	3.9447E+11	8.8930E+11
2023	27.2071	6.5448E+11	4.2613E+11	1.0052E+12
2024	27.3074	7.2352E+11	4.6231E+11	1.1323E+12
2025	27.4083	8.0034E+11	5.0348E+11	1.2722E+12
2026	27.5097	8.8575E+11	5.5006E+11	1.4263E+12

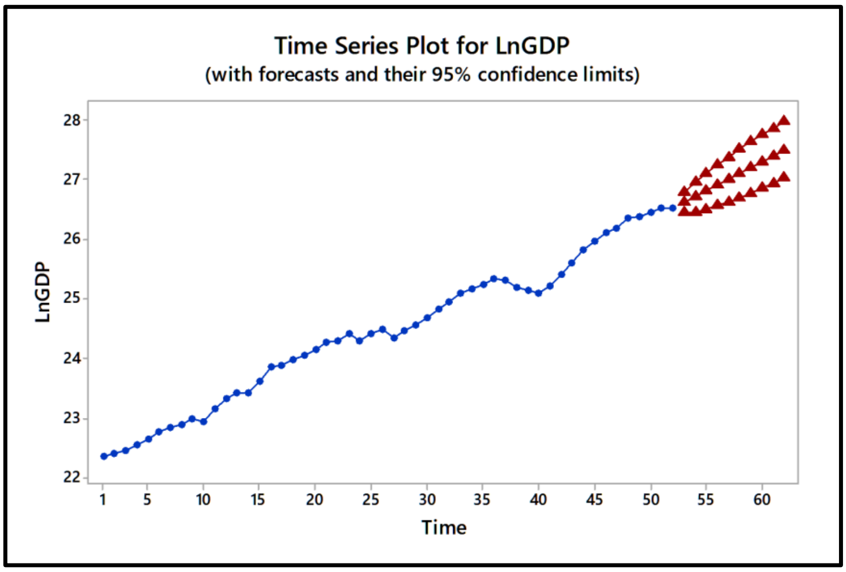


Figure 9: Time series plot for actual and forecasted lnGDP values

4. Summary and Conclusion

The aim of the study was to model and forecast Egyptian GDP based on Box-Jenkins approach based on the annual data (from 1965 to 2016). The four stages of Box-Jenkins approach are conducted to obtain an appropriate ARIMA model for the Egyptian GDP, and we used this model to forecast the Egyptian GDP for the next ten years (from 2017 to 2026). Time series plots and the correlogram plots were used for testing the stationarity of the data. Also, the MLE was used for estimating the model. Using the different goodness-of-fit measures (MSE, AIC, and BIC), the various ARIMA models with different order of autoregressive and moving-average terms were compared. We find that the best model is ARIMA (1, 2, 1), because have the minimum values of MSE, AIC, and BIC. Moreover, we expect that the Egyptian GDP will continue to raise according to the forecasted values form our model.

References

[1] M. N. A. Bhuiyan, K. S. Ahmed and R. Jahan, Study on Modeling and Forecasting of the GDP of Manufacturing Industries in Bangladesh, *Chiang Mai University Journal of Social Science and Humanities*, **2** (2008), 143-157.

[2] G. E. P. Box, G. M. Jenkins, *Times Series Analysis Forecasting and Control*, Holden-Day San Francisco, 1970.

[3] C. Chatfield, *The Analysis of Time Series: An Introduction*. CRC Press, 2016.

[4] C. Dritsaki, Forecasting real GDP rate through econometric models: An empirical study from Greece, *Journal of International Business and Economics*, **3** (2015), 13-19. https://doi.org/10.15640/jibe.v3n1a2

[5] J. Frain, *Lecture Notes on Univariate Time Series Analysis and Box Jenkins Forecasting*, Economic Analysis, Research and Publications, 1992.

[6] C. W. Granger, P. Newbold, *Forecasting Economic Time Series*, Academic Press, 1986.

[7] C. M. Jarque, A. K. Bera, Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals, *Economics Letters*, **6** (1980), 255-259. https://doi.org/10.1016/0165-1765(80)90024-5

[8] G. Kirchgässner, J. Wolters and U. Hassler, Univariate Stationary Processes, in *Introduction to Modern Time Series Analysis*, Springer, Berlin, Heidelberg, 2013, 27-93. https://doi.org/10.1007/978-3-642-33436-8_2

[9] B. Maity, B. Chatterjee, Forecasting GDP Growth Rates of India: An Empirical Study, *International Journal of Economics and Management Sciences*, **1** (2012), 52-58.

[10] W. Ning, B. Kuan-jiang and Y. Zhi-fa, Analysis and Forecast of Shaanxi GDP Based on the ARIMA Model, *Asian Agricultural Research*, **2** (2010), 34-41.

[11] A. Uwimana, B. Xiuchun and Z. Shuguang, Modeling and Forecasting Africa's GDP with Time Series Models, *International Journal of Scientific and Research Publications*, **8** (2018), 41-46. https://doi.org/10.29322/ijsrp.8.4.2018.p7608

[12] M. S. Wabomba, M. P. Mutwiri and M. Fredrick, Modeling and Forecasting Kenyan GDP Using Autoregressive Integrated Moving Average (ARIMA) Models, *Science Journal of Applied Mathematics and Statistics*, **4** (2016), 64-73. https://doi.org/10.11648/j.sjams.20160402.18

[13] W. L. Young, The Box-Jenkins Approach to Time Series Analysis and Forecasting: Principles and Applications, *RAIRO-Operations Research-Recherche Opérationnelle*, **11** (1977), 129-143. https://doi.org/10.1051/ro/1977110201291

Received: November 4, 2018; Published: March 8, 2019