Modelling GDP for Sudan using ARIMA

Moahmed Hassan, Hisham and Haleeb, Amin

2020

Modelling GDP for Sudan using ARIMA

Hisham Mohamed Hassan Ali[1]   Amin Mohamed Ali Haleeb[2]

[1] Department of Econometrics, University of Khartoum, Sudan, email: hishamdr@yahoo.com
[2] Department of Statistics, University of Tabuk, KSA

Abstract

This paper aims to obtain an appropriate ARIMA model for the Sudan GDP using the Box- Jenkins methodology during the period 1960-2018 the various ARIMA models with different order of autoregressive and moving-average terms were compared. The appropriate model for Sudan is an ARIMA (1,1,1), the results of an in-sample forecast showed that the relative and predicted values were within the range of 5%, and the forecasting effectiveness of this model, its relatively adequate and efficient in modeling the annual GDP of the Sudan.

 Keywords: ARIMA Modelling, Box-Jenkins methodology, forecasting, GDP, Sudan.
 JEL Classification: C53, E27

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). There are three ways in which the GDP of a country can be measured.

GDP is the sum of gross value added of the various institutional sectors or the various industries plus taxes and less subsidies on products (which are not allocated to sectors and industries) - production approach, GDP is the sum of final uses of goods and services by resident institutional units (actual final consumption and gross capital formation), plus exports and minus imports of goods and services - expenditure approach, GDP is the sum of uses in the total economy generation of income account (compensation of employees, taxes on production and imports less subsidies, gross operating surplus and mixed income of the total economy) - income approach..

Forecasting future economic outcomes is a vital component of the decision-making process in central banks for all countries. Monetary policy decisions affect the economy with a delay, so, monetary policy authorities must be forward looking, i.e. must know what is likely to happen in the future. Gross domestic product (GDP) is one of the most important indicators of national

1

economic activities for countries. Scientific prediction of the indicator has important theoretical and practical significance on the development of economic development goals.

For the forecasting of time series, we use models that are based on a methodology that was first developed in Box and Jenkins (1976), known as ARIMA (Auto-Regressive-Integrated-Moving-Average) methodology. This approach was based on the World representation theorem, which states that every stationary time series has an infinite moving average (MA) representation, which actually means that its evolution can be expressed as a function of its past developments (Jovanovic and Petrovska 2010).

The rest of the paper is organized as follows: Section 2 Literature Review Section 3 describes theoretical back ground and statistical review while in Section 4 the empirical results are presented, and finally, conclusions are provided in Section 5.

2. Literature Review

Box and Jenkins (1976) methodology has been used extensively by many researchers in order to highlight the future rates of GDP. Wei and al. (2010) use data from Shaanxi GDP for 1952-2007 to forecast country's GDP for the following 6 years. Applying the ARIMA (1,2,1) model they find that GDP of Shaanxi present an impressive increasing trend. Maity and Chatterjee (2012) examine the forecasting of GDP growth rate for India using ARIMA (1,2,2) model and a time period of 60 years. The results of their study showed that predicted values follow an increasing trend for the following years. Zhang Haonan (2013) using three models ARIMA, VAR, AR (1) examines the forecasting of per capita GDP for five regions of Sweden for the years 1993 – 2009. The results of the study showed all three models can be used for forecasting in the short run.

However, the autoregressive first order model is the best for forecasting the per capita GDP of five regions of Sweden. Shahini and Haderi (2013) test GDP forecasting for Albania using quarterly data from the first quarter of 2003 until the second quarter of 2013. For the forecasting they used two model groups ARIMA and VAR. Their results showed that the group of VAR model gives better results on GDP's forecasting rather than ARIMA model. Zakai (2014) investigates forecasting of Gross Domestic Product (GDP) for Pakistan using quarterly data from 1953 until 2012. Choosing a ARIMA (1,1,0) model he finds out the size of the increase for Pakistan's GDP for the years 2013- 2025.

3. Theoretical 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.

3.1. Autoregressive (AR) model

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

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

where ε_t is the error term in the equation; where εt 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.

3.2. 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.

3.3. 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.

3.4. 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 Xt is non-stationary series, we will take a first-difference of Xt so that ΔXt 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).

4.5. 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.

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 AutoCorrelation Function (ACF) and Partial Auto-Correlation Function (PACF) of the series to identification which autoregressive or movingaverage 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:

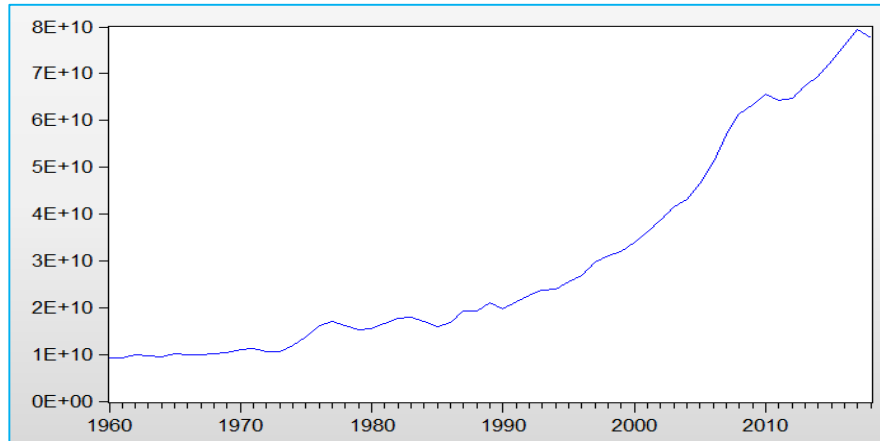$$AIC = 2m - 2\ln(\hat{L}), \ BIC = ln(n)m - 2\,ln(\hat{L}), \tag{5}$$

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.

4. Empirical Results

The variable used in the analysis is the GDP (CONSTANT 2010 US$) cover the period from 1960 to 2018 as shown in fig 1. The source of data is the World Bank (WDI) database. The ARIMA approach is an iterative four-stage process of stationary, identification, estimation and testing.
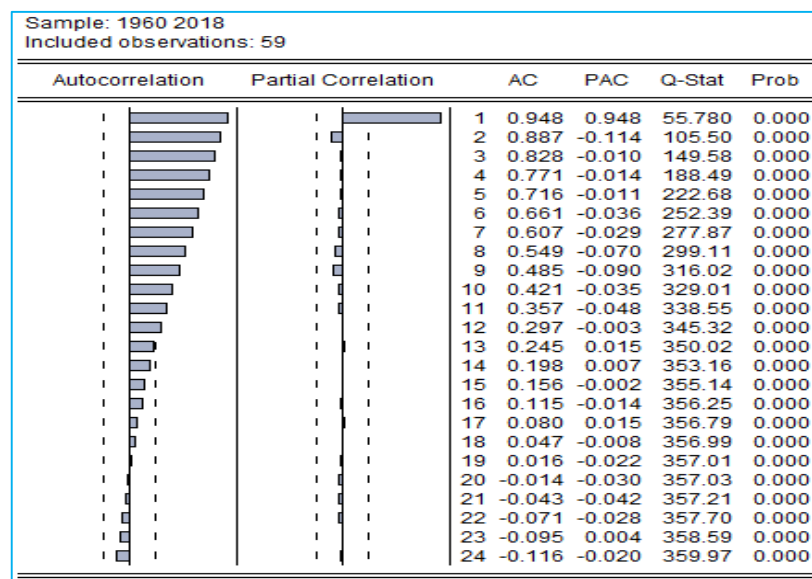
Fig 1: Sudan GDP constant 2010 US$ (1960-2018)
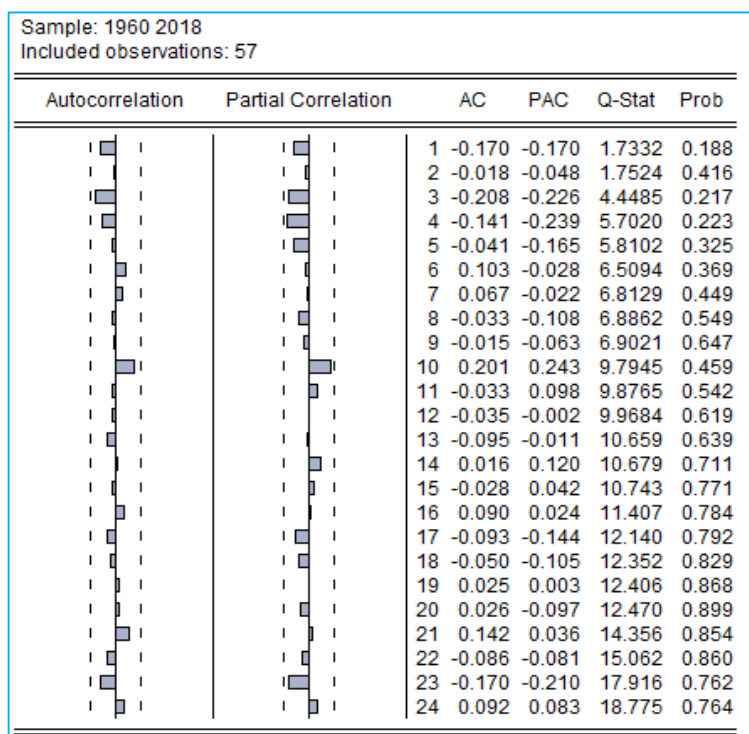

4.1 Testing for Stationarity

Figures 2 and 3 represent the correlogram of the real GDP rate series with a pattern of up to the 24 lags in level and for first differences.

Figure 2: Correlogram of Real GDP Rate Series (Level)

Sample: 1960 2018
Included observations: 59

		AC	PAC	Q-Stat	Prob
1		0.948	0.948	55.780	0.000
2		0.887	-0.114	105.50	0.000
3		0.828	-0.010	149.58	0.000
4		0.771	-0.014	188.49	0.000
5		0.716	-0.011	222.68	0.000
6		0.661	-0.036	252.39	0.000
7		0.607	-0.029	277.87	0.000
8		0.549	-0.070	299.11	0.000
9		0.485	-0.090	316.02	0.000
10		0.421	-0.035	329.01	0.000
11		0.357	-0.048	338.55	0.000
12		0.297	-0.003	345.32	0.000
13		0.245	0.015	350.02	0.000
14		0.198	0.007	353.16	0.000
15		0.156	-0.002	355.14	0.000
16		0.115	-0.014	356.25	0.000
17		0.080	0.015	356.79	0.000
18		0.047	-0.008	356.99	0.000
19		0.016	-0.022	357.01	0.000
20		-0.014	-0.030	357.03	0.000
21		-0.043	-0.042	357.21	0.000
22		-0.071	-0.028	357.70	0.000
23		-0.095	0.004	358.59	0.000
24		-0.116	-0.020	359.97	0.000

5

From the above figure we can conclude that the coefficients of autocorrelation (ACF) starts with a high value and declines slowly, indicating that the series is non-stationary. Also the Q-statistic of Ljung-Box (1978) at the 24th lag has a probability value of 0.000 which is smaller than 0.05, so we cannot reject the null hypothesis that the real GDP series is non-stationary. Thus, the series must be configured in first or second differences.

Figure 3: Correlogram of Sudan GDP Series (First Differences)

From the figure 3 we can conclude that the Q-statistic of Ljung-Box at the 24th lag has a probability value larger than 0.05, so we cannot reject the null hypothesis that the real GDP rate series is stationary. The results of Augmented Dickey-Fuller (ADF) test and Phillips-Perron (PP) test on real GDP rate series are representing on Table 1.

Table 1: ADF and Phillip-Perron's Test

	Level		First Differences	
	t-Statistic	**5% level**	**t-Statistic**	**5% level**
ADF	1.083055	-2.913549*	-2.913843	-1.946654**
PP	6.275004	-1.946549	-2.836997	-1.946654**

Note:
*MacKinnon (1996) one-sided p-values.
*** denote statistically significant at 5% significance levels.
The results in table 1 indicate that real GDP rate is stationary in first differences. Therefore for our model ARIMA (p,d,q) we will have the value d=1.

4.2 Identification of the Model

We can use the correlogram of figure 1 to determine the model ARMA (p,q), i.e. the values of parameters p and q. As already mentioned above, an AR(p) model has a PACF that truncates at lag p and an MA(q)) has an ACF

that truncates at lag q. In practice $\pm \frac{2}{\sqrt{n}}$ are the non-significance limits for both functions. We shall explore the range of models ARMA(α,b), $0 \leq a \leq p$, $0 \leq b \leq q$ for an optimum one. To do this we shall use the automatic model determination criteria AIC and SIC. The limits for both functions (ACF, PACF) are $\pm \frac{2}{\sqrt{57}} = \pm 0.264906$. From figure2, the ACF cuts off at lag 1 (q=1) and the PACF at lag 1 (p=1). Exploring the range of models {ARMA(p,q):$0 \leq p \leq 1$ 0, $0 \leq q \leq 2$}for the optimal on the basis of AIC and SIC. Thereafter we create Table 2 with the values of p and q as follows:

Table 2: Comparison of Models within the Range of Exploration Using AIC and SIC

p	Q	AIC	SIC
0	1	50.30624	50.37666
0	2	50.39189	50.46231
1	0	45.80916	45.87959
1	**1**	45.40885	45.51449

The results from table 2 indicate that according to the criteria of Akaike (AIC), and Schwartz (SIC) the model ARMA is formulated to ARMA(1,1). As the model is stationary on first differences, i.e. (d=1) our ARIMA model will be ARIMA (1,1,1).
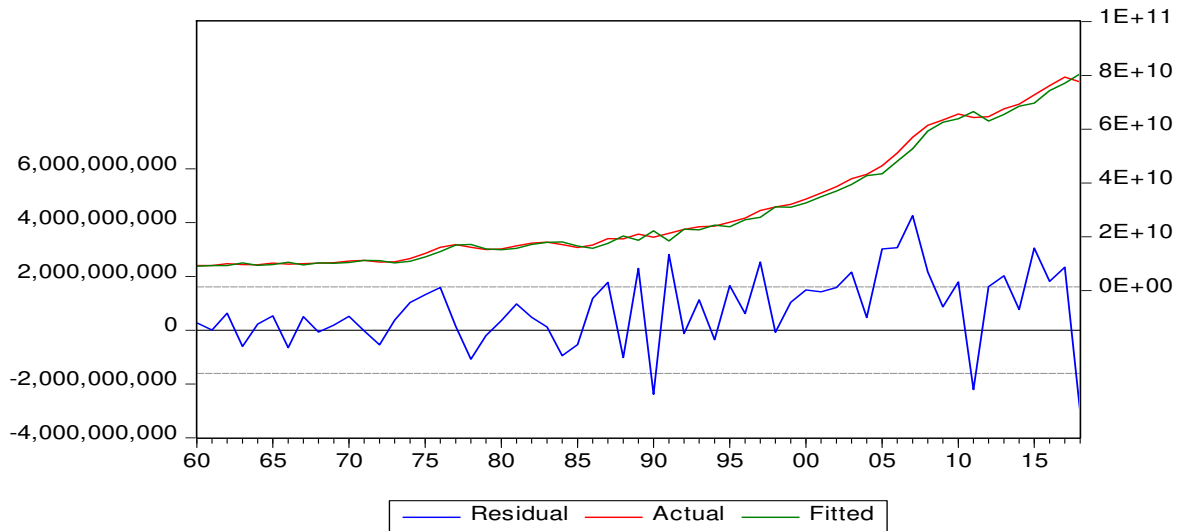
4.3 Estimation of the Model

Thereafter we can proceed to estimating the above model. The following table 3 presents the results of this model.

Table 3: Method: ARMA Maximum Likelihood (OPG - BHHH)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.998961	0.002123	470.4446	0.0000
MA(1)	0.548916	0.101568	5.404401	0.0000
SIGMASQ	2.45E+18	4.73E+17	5.183695	0.0000

R-squared	0.994834	Mean dependent var	3.01E+10
Adjusted R-squared	0.994650	S.D. dependent var	2.20E+10
S.E. of regression	1.61E+09	Akaike info criterion	45.40885
Sum squared resid	1.45E+20	Schwarz criterion	45.51449
Log likelihood	-1336.561	Hannan-Quinn criter.	45.45009
Durbin-Watson stat	1.393882		

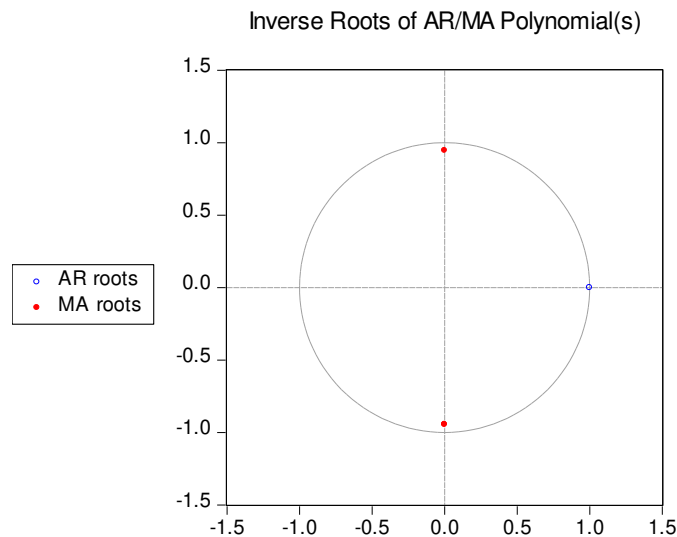|	
Inverted AR Roots	1.00
Inverted MA Roots	-.55

The results in table 3 and as shown in fig 4 indicate that both coefficients are statistically significant at 1% level of significance. The non-linear techniques used by Eviews 10 involved an iterative process that is converged after 21 iterations. The roots are 1 and 0.55, both inside the unit circle indicating stationarity and invertibility respectively. The chosen model as summarized in Table 3 is ARIMA (1,1,1) and figure 3 shows the fitted, actual and residual of the selected model.

Figure 4: the fitted, actual and residual of the ARIMA(1,1,1)

On the figure 5 the inverse roots of AR and MA characteristic polynomials for the stability of ARIMA model are presented. From the diagram we can see that the ARIMA model is stable since the corresponding inverse roots of the characteristic polynomials are in the unit circle.

Figure 5: Inverse Roots of AR and MA

Inverse Roots of AR/MA Polynomial(s)

4.4 Diagnostic Checking of the Model

Diagnostic checking of the model, help us to check if the estimated model is acceptable and statistical significant that means that the residuals are not auto correlated and follow normal distribution. For checking autocorrelation we use Q statistic of Ljung-Box (1978) and normality test using Jarque-Bera (JB) test (1980). The figures below represents the tests of the autocorrelation and normality of the residuals of the model ARIMA (1,1,1).

The results of figure 6 indicate that the residuals of ARIMA (1,1,1) model follow normal distribution. Moreover, the results of figure 7 indicate that the Q statistic of Ljung–Box for all the 24lags has values greater than 0.05 thus the null hypothesis cannot be rejected i.e. there is no autocorrelation for the examined residuals of the series. The while Figure 8 shows the response to One SD innovation, the result indicates the stability of the model.

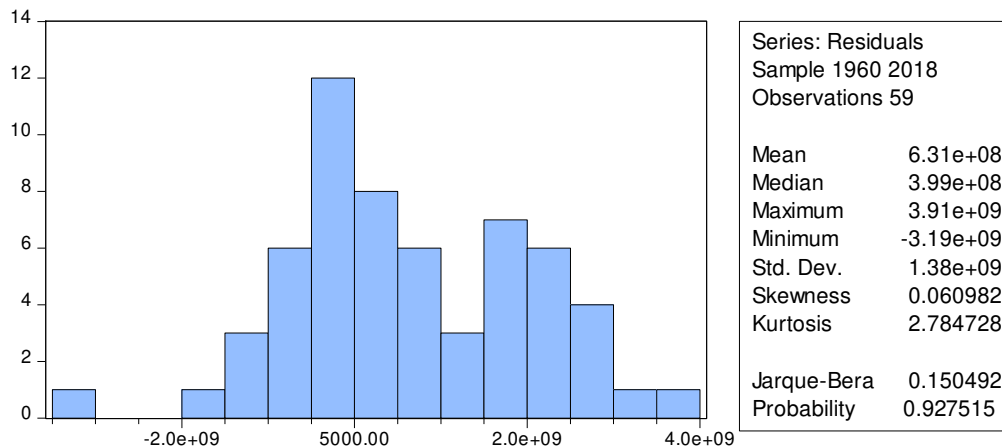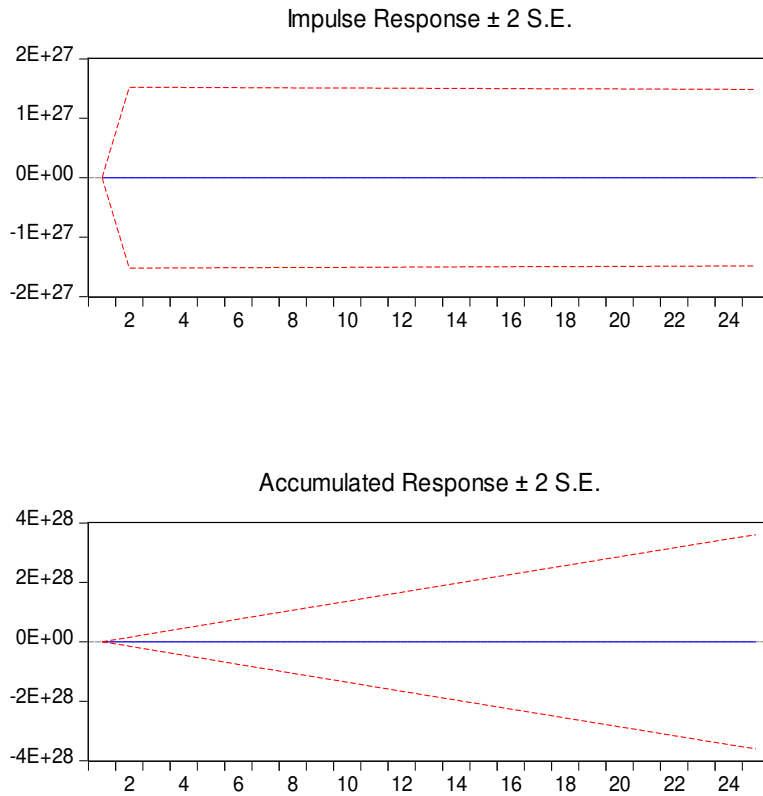**Figure 6: Histogram of the residuals of model ARIMA (1,1,1)**

Figure 7: Q statistics probabilities adjustment

Figure 8: Response to One SD innovation
Response to One S.D. Innovation

Impulse Response ± 2 S.E.

Accumulated Response ± 2 S.E.

In figure 9 we represent the criteria for the evaluation of the forecasts of the model ARIMA(1,1,1)

Figure 9: **Forecast Accuracy Test on the Model ARIMA (1,1,1)**

Forecast: GDP__CONSTF	
Actual: GDP__CONSTANT_2010_US$_	
Forecast sample: 1980 2018	
Included observations: 39	
Root Mean Squared Error	3.27E+10
Mean Absolute Error	2.48E+10
Mean Abs. Percent Error	49.61799
Theil Inequality Coefficient	0.547048
Bias Proportion	0.573051
Variance Proportion	0.413419
Covariance Proportion	0.013530
Theil U2 Coefficient	9.431106
Symmetric MAPE	73.83411

10

6. Conclusion

The aim of this paper to model and forecast Sudanese GDP based on BoxJenkins approach based on the annual data from 1960 to 2018. The four stages of Box-Jenkins approach are conducted to obtain an appropriate ARIMA model for the Sudan GDP, time series 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 the appropriate ARIMA (p, d, q) process. The corresponding correlogram helped in choosing the appropriate p and q for the data series. An ARIMA (1,1,1) model.

References

Box, G. E. P. & Jenkins G. M. (1976). Time series analysis. Forecasting and control. Holden-Day, San Francisco.

C. Chatfield, (2016). The Analysis of Time Series: An Introduction. CRC Press.

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

C. W. Granger, P (1986). Newbold, Forecasting Economic Time Series, Academic Press.

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

Dickey, D. A. & Fuller W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root.Journal of the American Statistical Association, 74(366), 427–431. The European System of Accounts ESA 1995, Eurostat, 1996.

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

Jarque, C. M.,& Bera A. K.(1980).Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters6(3), 255–259.

Jovanovic, B. & Petrovska M. (2010). Forecasting Macedonian GDP: Evaluation of different models for short-term forecasting. Working Paper, National Bank of the Republic of Macedonia.

Ljung, G. M., & BoxG. E. P. (1978). On a measure of a lack of fit in time series models. Biometrika,65(2), 97–303.

Maity, B., & ChatterjeeB. (2012). Forecasting GDP growth rates of India: An empirical study. International Journal of Economics and Management Sciences, 1(9), 52-58.

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

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

J. Frain, (1992). Lecture Notes on Univariate Time Series Analysis and Box Jenkins Forecasting, Economic Analysis, Research and Publications.

Phillips, P. C. B. & Perron P. (1988). Testing for a unit root in time series regression. Biometrika,75(2), 335–346.

Shahini, L. &Haderi S. (2013). Short term Albanian GDP forecast: One quarter to one year ahead. European Scientific Journal, 9(34),198-208.

Wei N, Bian K. & Y. Zhi-fa (2010). Analysis and forecast of Shaanxi GDP based on the ARIMA model. Asian Agricultural Research, 2(1), 34-41.

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

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

Zakai, M. (2014). A time series modeling on GDP of Pakistan. Journal of Contemporary Issues in Business Research,3(4), 200-210.

Zhang, H. (2013). Modeling and forecasting regional GDP in Sweden using autoregressive models. Working Paper, Högskolan Dalarna University, Sweden.