Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.7, No.11, 2017 www.iiste.org

Time Series Modeling and Forecasting GDP in the Ghanaian Economy

Theophilus Quachie Asenso [1*], Erasmus Tetteh-Bator [2] and Mbele Bidima Martin Le Doux [3]

1. African Institute for Mathematical Sciences (AIMS) Cameroon.
2. College of Science, Hohai University, No. 1 Xikang Road, 211100, Nanjing, China.
3. University of Yaounde I, Cameroon.

* E-mail of the corresponding author: theophilus.asenso@aims-cameroon.org

Abstract

In this research, empirical modeling of the Ghanaian GDP was done by using the Box-Jenkins model which is also known as the Autoregressive Integrated Moving Average model (ARIMA). We followed the 4 steps involved in using this model which include model identification, estimation of parameters, diagnostic checking and finally, model use (or forecasting). The analysis was carried out by using the GDP data of Ghana from 1970-2014. We found an original result by discovering an ARIMA (0; 1; 0) process modeling this GDP data. Next, we made forecast of the GDP of Ghana for the period 2015-2020 and compared the forecast values with that of Ghana Statistical service and other international forecasting organizations. The result from the forecast revealed that the GDP of Ghana is mostly influenced by external factors and may experience an increase for the period 2015-2020.

Keywords: Gross Domestic Product (GDP), stationarity, invertibility, Box-Jenkins, auto-regressive.

1. Introduction

The term GDP refers to the market value of the total goods and services produced by a country measured at a given period of time (Bade; 2016). The GDP plays a major role in every economy because it aids the government in budget preparation and also allows the government agencies to make forecasts which helps in studying the growth of an economy or the economic performance (Bade; 2016). GDP also forms a major part when calculating the human development index of a particular country which is used as yardstick for measuring the development of a country (Michael; 2012).

The GDP of Ghana can be grouped into three main categories, namely service, industry and agriculture (Ghana Statistical Service; 2015). The sector with the highest portion of the 2014 annual GDP is the service sector (50%), followed by the industrial sector with about (26:6%) and lastly the agricultural sector (21:5%) (State of the nation's address; 2014). The agricultural sector has experience a fall in its contribution to the GDP over the past three years. However, the industrial sector's contribution experienced an increase from 2011 to 2013. This may be due to the high contribution made by activities of crude oil production which is now one of the major exports of the country (Ghana Statistical Service; 2015, I.S.S.E.R GHANA;2014). This shows that Ghana is gradually shifting attention to the production of crude oil as it is now one of the major exports of the country, hence reducing the contribution of the agricultural sector. One of the implications of this is that, any little change in the world oil prices can affect the GDP of the country.

1.1. Ghana after the Financial Crisis

After the financial crisis, all economies are working to attain their pre-crisis growth level and Ghana is not an exception. The GDP growth rate of the country increased between 2009 and 2011 but however experienced a fall in 2013 (Ghana Statistical Service; 2015). This can be attributed to the fall in the global oil prices because oil has become one of the main products exported, contributing 40% of the total export in 2012 as compared to 2010 (I.S.S.E.R GHANA;2014). Although the country seems to be recovering from the financial crisis, according to a speech delivered by the President during the state of the nation's address (2014), it is still facing major economic challenges which include currency depreciation, energy crisis, macroeconomic issues, and other fiscal issues.

2. Methodology

In this study, the type of data used for the analysis is the secondary data obtained on the internet (http://www.kushnirs.org/macroeconomics_/en/ghana__gdp.html). Results and definitions used were taken from (Cryer; 2008, Bierens; 2004, Hsu; 1997, Dobre; 2008, Brockwell; 2002, Nau; 2014 and Shumway; 2011).

2.1 The Box-Jenkins Model Building Procedure

The steps involved in fitting the ARIMA model in this paper were taken from (Shumway, 2011, p. 144). The nature of the observed values in the GDP data is that the variables are statistically dependent, making the ARIMA (p,d,q) model the best for modeling such data. The Box-Jenkins model is based on two assumptions of the data: stationarity of the autoregressive and invertibility of the moving average.

2.2 Stationarity of the AR Models

A process $\{Y_t\}$ is said to be an Autoregressive model of order p, denoted by AR(p) if it is given in the form

$$Y_t = \sum_{i=1}^{p} \emptyset_i Y_{t-1} + e_t \qquad {\scriptstyle i=1,2\ldots p\,,} \tag{1}$$

where Y_t is stationary (will be explained later), all the $\emptyset_i$ with $(\emptyset_i \neq 0)$ are the constants and e_t is i.i.d white noise process with mean zero and variance δ^2.

Using the back shift operator, equation (1) can be written as $\left(1-\emptyset_1 B-\emptyset_2 B^2-\cdots-\emptyset_p B^p\right)Y_t = e_t$ or $\emptyset(B)Y_t = e_t$, where ϕ(B) is called the autoregressive operator or polynomial.

From equation (1), assuming p = 1, we have AR(1) which is

$Y_t = \emptyset Y_{t-1} + e_t = \emptyset(\emptyset Y_{t-2} + e_{t-1}) + e_t$

$= \emptyset^2 Y_{t-2} + \emptyset e_{t-2} + e_t$

$= \emptyset^k Y_{t-k} + \sum_{j=0}^{k-1} \emptyset^j e_{t-j} = e_t + \emptyset e_{t-1} + \emptyset^2 e_{t-2} + \cdots = \sum_{j=0}^{\infty} \emptyset^j e_{t-j}$

$E(Y_t) = 0 \ and \ Var(Y_t) = \sum_{j=1}^{\infty} \emptyset^{2j} \, Var\left(e_{t-j}\right) = \delta_e^2 + \emptyset^2 \delta_e^2 + \cdots \quad = \delta\text{^}2/(1-\emptyset\text{^}2\,)$ by taking sum to infinity.

For an AR model $Y_t = \sum_{i=1}^{p} \emptyset_i Y_{t-1} + e_{t,}$ to converge, $\emptyset^j e_{t-j}$ must keep reducing as j turns to infinity. This can only be possible if $\left|\emptyset\right| < 1$. This leads us to the stationarity of the AR(1) model. Let us write Y_t as $(1-\emptyset B)Y_t = e_t$. From this, we can say that the root of the polynomial $\emptyset(B)$ is $1\,/\,\emptyset$. Since $\emptyset(z) = 0$, where $z = {}^1\!/_{\emptyset}$. From the root of the autoregressive polynomial, the stationarity condition of the AR(1) process can be restated as: AR(1) is stationary if and only if the root of the autoregressive polynomial is greater than 1 (in absolute value terms). Let us then state the stationarity condition for AR(p) model, $\left(1-\emptyset_1 B-\emptyset_2 B^2-\cdots-\emptyset_p B^p\right)Y_t = e_t$.The autoregressive polynomial can be factored as $\emptyset(z) = (1-\lambda_1^{-1}z)(1-\lambda_2^{-1}z)\cdots(1-\lambda_p^{-1}z)$, where $\lambda_1,\cdots,\lambda_p$ are are the roots of the polynomial. Comparing this to the stationarirty condition of the AR(1), we say that for AR(p) to be stationary, all the roots must be greater than 1 for all j. If any of the roots is almost exactly equal to 1 or if the sum of all the roots is almost equal to 1, we say that the AR has a unit root.

This implies that an AR(p) model is stationary if and only if $\left|\lambda_j\right| > 1$, that is all the roots of the AR polynomial must lie outside the unit circle.

2.3 The MA Model

A time series $\{Y_t\}$ is said to follow a moving average process of order q which is denoted by MA(q) if it can be expressed as

$$Y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}, \tag{2}$$

where q is the lag in the moving average, $(\theta_1, \theta_{2,\ldots,} \theta_q)$; $\theta \neq 0$ are the constants and e_t is assumed to be white noise with zero mean and variance δ^2. The moving average process is always stationary for any value of $(\theta_1, \theta_{2,\ldots,} \theta_q)$. Using the back shift operator, equation (2) can be written as $Y_t = \theta(B)e_t$. where $\theta(B) = 1+ \theta_1 B+ \theta_2 B^2+\ldots+ \theta_q B^q$ is the moving average operator or polynomial.

We derive the invertibily condition for the MA(p) model by going through the same steps as that of the AR(p). This gives us the invertibility condition as: MA(q) model is said to be invertible if and only if $\left|\theta_j\right| < 1$ for all j.

A process Y_t is said to be an auto regressive integrated moving average (ARIMA (p,d,q)) process if

$$\Delta^d Y_t = (1-B)^d Y_t \tag{3}$$

is ARMA (p,q) process. d is a non-negative integer representing the differencing order.

The ARIMA model can be written using the back shift operator as $\phi(B)(1-B)^d Y_t = \theta(B)\, e_t$.

We went through the four major steps which include model identification, estimation of parameters, diagnostic checking and finally, model use (or forecasting).

2.4 Model Identification

This is the first step one needs to take when building the ARIMA(p,d,q) model. This stage involves testing for stationarity and the seasonality of the data and also identifying the order of the p,d,q components of the model. In testing for stationarity, we used the ACF and the PACF plots as well as the unit root test known as the Augmented Dickey-fuller (ADF) test. The ADF test uses the unit root test to test the hypothesis if differencing is required or not. The fitted model is given as

$$\nabla Y_t = \alpha + \widehat{\emptyset} Y_{t-1} + e_{t,} \quad e_r \sim WN(0, \delta^2), \widehat{\emptyset} = \emptyset - 1.$$

The test hypothesis is that H_0 : the model is not stationary and H_1 : the model is stationary.

If the data is not stationary, it needs to be differenced to introduce the integrated (I(d)) component of the Box-Jenkins model. The model contains some seasonal component if there exist some spikes (at regular interval) in the time plot. To check for the seasonal component of the data, we need to study the ACF plot. If there exist some spikes at regular intervals or lags like 1*l*, 2*l*; 3*l*,…, where *l* represents the lag, then the data contains a seasonal component. This can be corrected by taking a l moving average on the raw data.

The next step in the model identification is to identify the order of the autoregressive and the moving average components of the ARIMA (p.d.q) model. This can be done by studying the sample autocorrelation (SACF) and the sample partial autocorrelation (SPACF) correlogram. Whenever p = 0 and q > 0, the SACF cuts off after lag q and the PACF tails off, which indicates the order of the MA component. Therefore if the SACF cuts off after lag q or the SACF at lag 1 is negative, then the series might be over differenced. It can be corrected by introducing a MA component and the lag after which the SACF cuts off is the order of the MA component.

Also, whenever p > 0 and q = 0, the SPACF cuts off after lag p while the SACF tails off. Therefore if the SPACF cuts off after lag p or the SACF at lag 1 is positive, then the series might be under differenced. It can be corrected by introducing an AR component and the lag after which the SPACF cuts off becomes the order of the AR. Although this method gives a fair idea on the possible order for the ARIMA (p,d,q) model, the best order is selected by using the AIC, AICc or the BIC to check for the model with the minimum value of AIC, AICc or the BIC.

2.5 Parameters Estimation

This stage involves the estimation of the parameters identified in the ARIMA (p,d,q) model. This is done by using the maximum likelihood estimation. We calculate for the likelihood of the joint probability distribution function of the model and then estimate the parameters by taking partial derivative with respect to each parameter.

2.6 Model diagnostics

The Box-Jenkins model diagnostics follows the assumption that a good model should have an error term which is stationary and which follows a white noise process. When these conditions are met, then the model could be used to make predictions. If not, then we fit another model. The test for stationarity is done by studying the residual time plot. The test for the independence and identically distribution is done by studying the behavior of the SACF and SPACF of the residuals and the Ljung-Box Test. The Ljung-Box test is given as: $Q^*=T(T+2)\sum_{k=1}^{h}(T-k)^{-1}\rho_k^2$, where h is the maximum lag under consideration, T is the number of observations and k is the number of parameters in the model. The test is to check if the first h autocorrelation coefficients differ from that which is expected from the error terms. A large Q* value indicates that the correlation was not obtained from a white noise series. The assumption here is that, if the correlation was obtained from a white noise process then Q* should follow the Chi-square distribution. That is $Q^* \sim \chi^2$ with (h-k) degrees of freedom. The hypothesis is given as follows: H_0: The error terms are i.d.d random variables, and H_1 : The error terms are not i.d.d random variables. If we realize that the model does not follow the white noise process, then a new model is required instead of making forecast with the old one. The test for the white noise component and the normality assumption is done by studying the normal Q-Q plot.

3. Implementation of the Box-Jenkins Method on the Data

3.1 Model Identification

The time plot in figure 1 shows a clear sign of non-stationary time series data. It follows an upward trend. The ACF correlogram in figure 2 also follows a slow decay and the p-value for the ADF test is 0.9123 confirming the non-stationarity of the data. From the first differenced plot (Figure 4), we can realize that the GDP data shows some stationarity after the first difference although there are three outliers at 2000, 2010 and 2014 which we can ignore. The result obtained from the ADF unit root test gave a p-value of 0.0422 which is less than 0.05, we reject the null hypothesis that states that the data is not stationary and accept the alternative hypothesis that the data is stationary. This implies that the ARIMA(p,d,q) model will be ARIMA(p,1,q), with 1 representing the first order difference of the data.

3.2 Order Selection for p,q

From the SACF correlogram in figure 5, it can be seen that right from lag 1, all the coefficients do not exceed the significant limit. That is they are assumed to be equal to zero. Looking at the SACF plot suggest a zero MA term. We proceed by looking at the SPACF correlogram which is shown in figure 6. The SPACF correlogram, also like that of the SACF, has all the coefficients not exceeding the significant limit. From the SPACF plot we can also say that the first non-seasonal differenced data also has a zero order for the AR component. The results show that we have the AR(0) and the MA(0) term of the ARIMA(p,1,q) model. However, we need to check for the order which minimizes the AICc. This is done by taking different possible combinations of ARMA(p,q) model and then check for their corresponding AICc.

3.3. Model Finding and GDP Forecasting

The table shown in Table 1 gives a summary of the possible ARMA(p,q) models with their AICc values. The part containing AIC and BIC are to give a general view of how the various criterion behave. From the table (A1), it can clearly be seen that the model with the minimum AICc is the ARIMA(0,1,0) model.

3.4 Diagnostics Testing of the Found Model

The figures for the residual plot (figure 8), the histogram (figure 8) and the normal Q-Q plot (figure 9) of the ARIMA(0,1,0) model are shown in the appendix. It can be seen from the residual time plots in figure 7 and the residual is also stationary like the differenced data. The plot shows no clear pattern except that there are few outliers in the year 2000 and 2010.

The normality assumption is seen in the histogram and the normal Q-Q plot (figure 8) and (figure 9). The plots show that the error term follows the white noise process with zero mean and a constant variance. However, there exist some slight changes (outliers) in the variance at the ends of the lag on the Q-Q plot but since this is a small sample size (20), we use the central limit theorem of Hsu (1997)[p. 127] to show that as the sample size increases, the distribution of the error term will follow the normal distribution. From the histogram, we can see that the mean of the error term is near zero.

We proceed to check for the autocorrelation coefficients of the residuals. This is shown using the ACF and the PACF correlogram of the residual presented in the figures 10 and 11 respectively. .

The ACF correlogram shows that all the coefficients of the first 15 samples of the residuals are within the significant limit ($\pm 2/\sqrt{45} = \pm 0.298$), implying that they are all near zero. Therefore the assumption that the ACF of the error term should have zero coefficients is also satisfied.

The PACF correlogram of the residual also confirms the result of the ACF, that all the coefficients between lag 1 to 15 are all within the significance region, hence assumed to be uncorrelated. This implies that the error term of the first 15 lags are independent and identically distributed. We need to perform the Ljung-Box text on the entire sample to confirm this since the ACF and the PACF considered the correlation between only two lags. Table 2 shows the result obtained after performing the Box-Ljung test. The result form the tests gave a p-value greater than 0:05 which allows us to fail to reject the null hypothesis that the error terms are independent and identically distributed. From this, we can conclude that none of the coefficients of the autocorrelation of the first 20 samples of the error term is non-zero.

Since all the assumptions of the residual term has been satisfied, we can proceed to make predictions with the ARIMA(0,1,0) model. The result obtained from table 1 and the diagnostic tests imply that ARIMA(0,1,0) (which is the same ARMA(0,0)) because the model contains no AR or MA term) is the best model for forecasting the GDP data of Ghana. Since the ACF and PACF coefficients of the first differenced are zero, it implies that the series is random after the first differenced hence follows a random walk process. This implies that changes from one period to the other are purely random. Observing the time plot of the GDP data in figure 1, we can also observe the presence of an upward trend with a slop or average growth. This average growth is called the drift and needs to be considered in our model. It is estimated as

$\mu = \frac{1}{44}\sum_{t=2}^{45}(Y_t - Y_{t-1}) \approx 0.7659.$

Consequently the ARIMA(0,1,0) model we found after the first difference, is written in compact form as $(1 - B)Y_t = \mu + e_t$; $e_t \sim WN(0, \delta^2)$, and is explicitly written as

$$Y_t = \mu + Y_{t-1} + e_{t,}, \qquad t = 2, \cdots, 45, \qquad Y_1 = 3.5$$

3.5 Making Forecast with the Selected Model

The forecast equation for our model is given as

$$\widehat{Y_t} = \mu + Y_{t-1} + e_t$$

Table 3 shows the forecast values for the Ghanaian GDP for the next 6 years. The table also includes the low and high values for 80%, 95% and 99.5% confidence intervals. The figures 12 shows the forecast for the next 6 years of the GDP of Ghana using the ARIMA(0,1,0) model. Figure 13 shows the result obtained by using the fitted model and the actual or the observed data. Figure 12 and the table 3 show that the average GDP of Ghana for the next 6 years will experience an increase. The red line represents the actual GDP from 1970-2014 and the blue line represents the forecast for the six years. The shaded region represents the confidence intervals values. The plot in figure 13 shows the fitted values and the actual GDP values. Both plots seem to be close to each other. Therefore we can say that the model is a good one for forecasting the GDP of Ghana.

4. Conclusions

The study revealed that ARIMA(0,1,0) is the best model. This implies that the GDP of Ghana follows a random walk process with a drift. The result from the essay is confirmed by the different changes (ups and downs) that occur in the GDP growth of the economy. ARIMA (p,d,q) model was used because of the nature of the time series data we used. There is a correlation between successive values of the data.

The result shows that the GDP for 2015 will be 37:365 billion USD (±3.566, ±5.438, ±7.788 for 80%; 95%; and 99:5% confidence interval respectively).

That of 2016 will be 38:086 billion USD (±5.16; ±7.892; ±11.308 for 80%; 95%; and 99:5% confidence interval respectively). This result is in line with the annual report given by the Ghana statistical service, which gave a GDP of 36.66 billion USD for the first two quarters of 2015 service (2015). For our forecast for 2016, it can be seen that if adequate measures are put in place, although the oil prices have been forecasted by Knoema (A research institution based in USA) to have a fall in 2016, the GDP can still rise. However, if measures are not put in place, then GDP can fall. Looking at the formula for GDP (C + I + G + (X−M)), if the export component is not able to generate enough income so as to cover the import, the value of the GDP will be affected. When oil price falls, it affects the X component of the GDP and this turns

to reduce the value of the GDP.

References

Bade Perkin. (Retrived:April, 2016). Economics : Canada in the global environment. http://web.uvic.ca/~aahoque/VIU/Chapter%2020.pdf.

Bierens, H. J. (2004). Introduction to the mathematical and statistical foundations of econometrics. *Cambridge University Press* (138-150).

Brockwell, P. J. and R. A. Davis (2002). Introduction to Time Series and Forecasting. *Springer Verlag New York, Inc.*

Cryer, J. D. and K. Chan (2008). Time series analysis with application. *Springer Science+Business Media LLC* (11-19).

Dobre I. and A. A. Alexandru. (2008). Modelling unemployment rate using box-jekings procedure. *Journal of applied quantitative methods* 3 (156-166).

Ghana Statistical Service (2014,2015). Gross domestic product. http://www.statsghana.gov.gh/gdp_new.html.

Hsu H. P. (1997). Theory and problems of probability random variables and random processes. *Schaum's outline series*

I.S.S.E.R GHANA. (2014). The state of the Ghanaian economy in 2014. http://isser.edu.gh/index.php/sger/50-the-state-of-the-ghanaian-economy-report-2014..

Nau, R. The mathematical structure of arima models. people.duke.edu/~rnau/411arim3.htm.

Michael P. Todaro and Smith Stephen (2012). Economic Development (11 ed.). Number Page 37-73. *Pearson Education, Inc*.501 Boylston Street, Suite 900, Boston, USA.

President of the republic of Ghana (2014). The State of the nation address. www.presidency.gov.gh/sites/default/files/photos/sona2016.pdf.

Shumway, R. H. and D. S. Stofer (2011). Time Series Analysis and Its Applications, With R Examples. *Springer-Science, New York*.

Appendix

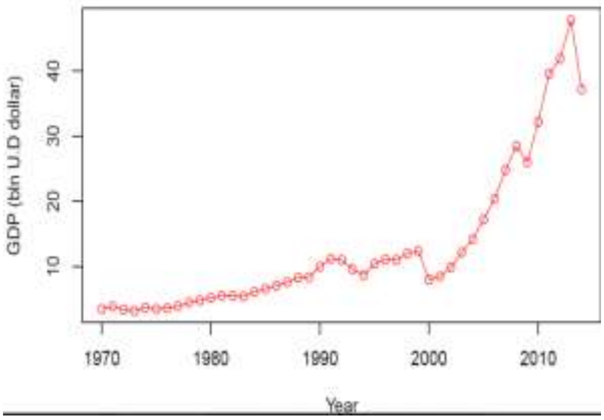

Figure 1: Time series plot of GDP data

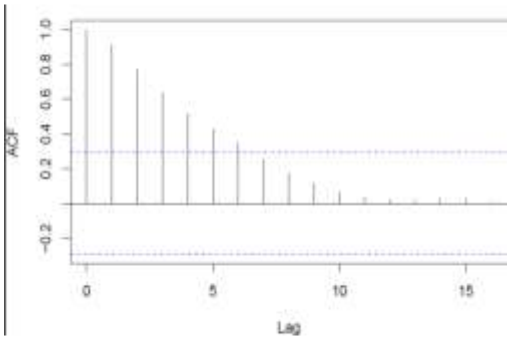

Figure 2: ACF plot of the GDP data

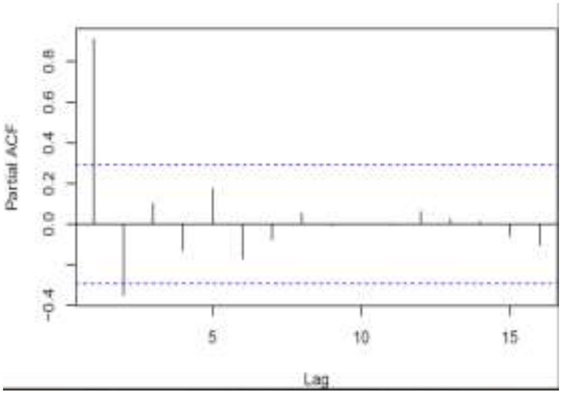

Figure 3: PACF plot of the GDP data

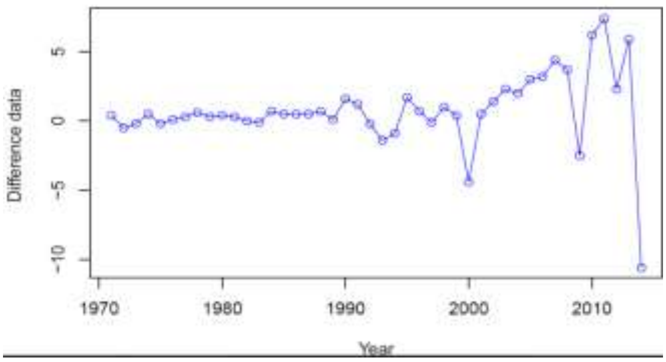

Figure 4: Time series plot of the differenced GDP data

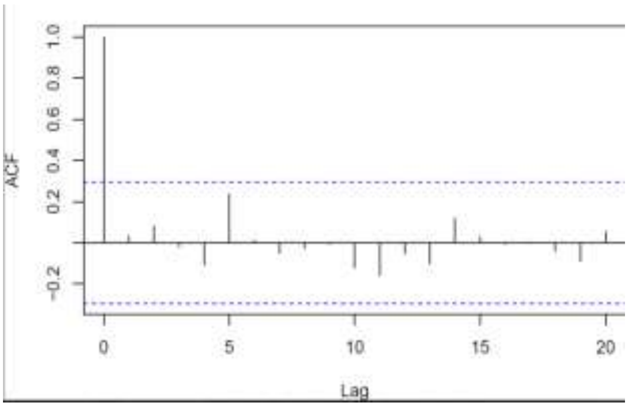

Figure 5: ACF plot of the differenced GDP data

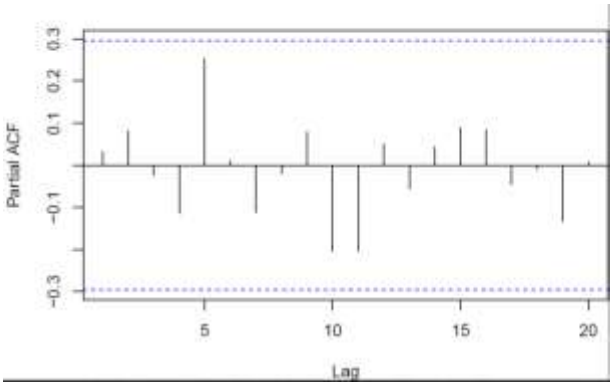

Figure 6: PACF plot of the differenced GDP data

Table 1: Possible ARMA(p,q) models with their AIC values

ARIMA model	Coefficients AR 1	AR 2	MA 1	MA 2	σ^2 EST	Log likelihood	AIC	AICc	BIC
(1,1,0)	0.0517				7.699	−106.31	218.63	219.23	223.98
(1,1,1)	0.8016		-0.6857		7.99	-107.16	220.33	220.93	225.68
(1,1,2)	-0.3387		0.3943	0.4978	7.579	-105.2	220.39	221.97	229.31
(0,1,2)			0.1506	0.3748	7.638	-105.77	219.53	220.56	226.67
(2,1,0)	0.0036	0.1444			7.799	-106.09	220.18	221.2	227.31
(0,1,0)					7.533	-106.35	216.7	216.99	220.27

Table 2: Box-Ljung test statistics

Test	χ^2	Deg. of Freedom	P-value
Box-Ljung	9.4744	20	0.9767

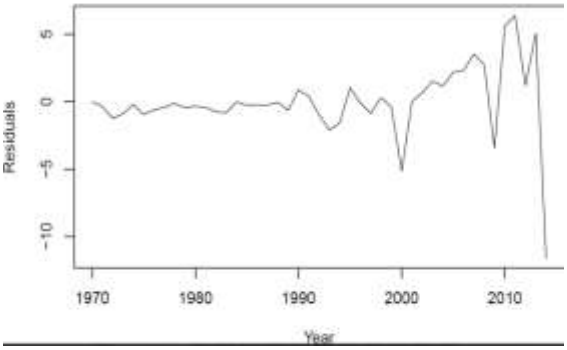

Figure 7: Residual plot of ARIMA (0,1,0) model

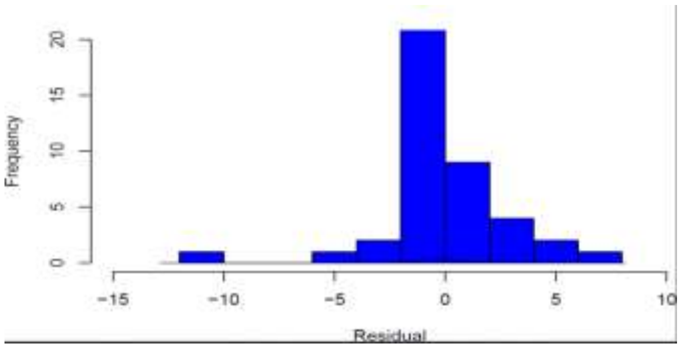

Figure 8: Histogram of the residuals

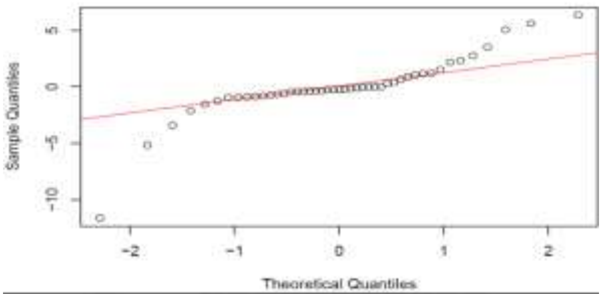

Figure 9: Normal Q-Q plot of residuals

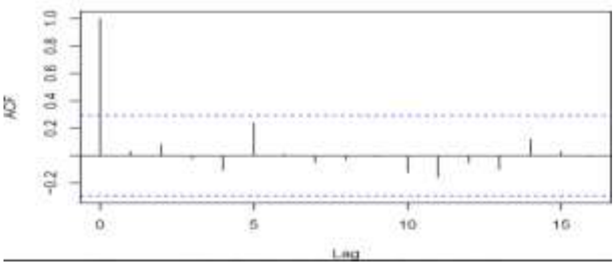

Figure 10: ACF plot of the residuals

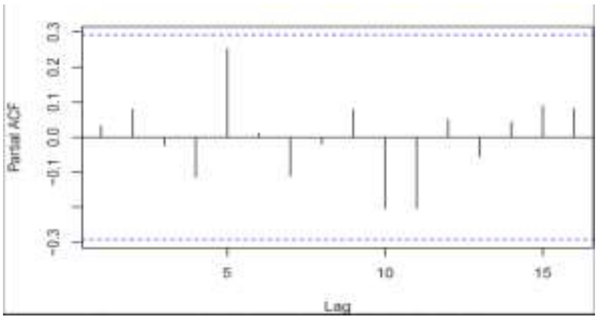

Figure 11: PACF of the residuals

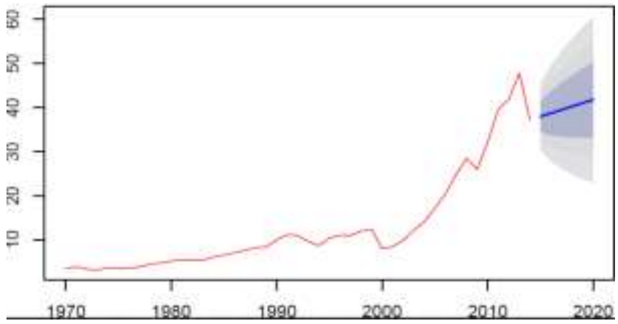

Figure 12: Forecast of ARIMA(0,1,0) with drift

Table 3: Six years forecast for the GDP of Ghana

Year	Forecast	Low 80	High 80	Low 95	High 95	Low 99.5	High 99.5
2015	37.365	33.809	40.921	31.927	42.803	29.576	45.154
2016	38.086	32.926	43.246	30.194	45.978	26.783	49.389
2017	38.836	32.457	45.214	29.081	48.591	24.865	52.806
2018	39.587	32.188	46.986	28.272	50.903	23.381	55.793
2019	40.338	32.044	48.633	27.653	53.024	22.171	58.507
2020	41.090	31.987	50.193	27.168	55.0118	21.152	61.028

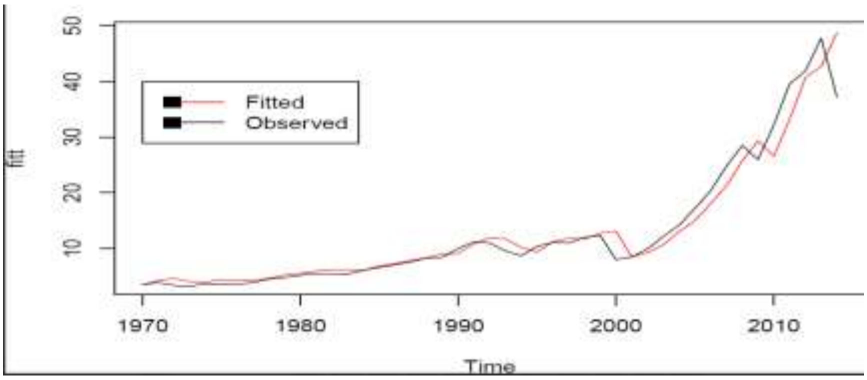

Figure 13: Fitted values verses the observed GDP values