International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321- 4767 P-ISSN: 2321-4759
www.ijmsi.org Volume 8 Issue 1 || Jan, 2020 || PP-07-15

Forecasting GDP Of Bangladesh Using Time Series Analysis

Arzo Ahmed[1], Md. Sifat Ar Salan[2]
1Mawlana Bhashani Science and Technology University, Department of Statistics, Santosh, Tangail-1902, Bangladesh, PH-008801717409867
2Mawlana Bhashani Science and Technology University, Department of Statistics, Santosh, Tangail-1902, Bangladesh, PH-0088 0171 035 8382
Corresponding Author: Md. Sifat Ar Salan

ABSTRACT: *Gross Domestic Product (GDP) of a country is the money value of all final goods and services produced by all the enterprises within the borders of a country in a year. It represents the aggregate statistic of all economic activity. This study concerns an analysis on a yearly data of Bangladesh's US Dollar (US$) GDP during 1968-2017 from World Bank database. The main objective is to find a suitable model for forecasting US Dollar (US$) GDP of Bangladesh. For statistical analysis, we have used graphical methods to display data distributions, Autocorrelation Function (ACF), Partial Autocorrelation Functions (PACF), Residuals and forecast, and differencing to check for stationarity. Testing for existing a data break in the data is carried out by Chow test. Evidence of a data breakpoint is found the year of 2000. Autoregressive Integrated Moving Average (ARIMA) models are constructed following the Box-Jenkins technique. With the lowest value of the Schwarz Information Criterion (SIC), ARIMA (2, 1, 1) has been obtained through an expert modeller method by considering the best fit model*

KEYWORDS: *Autocorrelation, Residuals, Forecasting, ARIMA, AIC*

Date of Submission: 04-02-2020 Date of Acceptance: 20-02-2020

I. INTRODUCTION

GDP indicates the financial health of a country as a whole-which is actually a hunting ground of researchers in the field of business in general and of economics in particular. The issues of GDP have become the most concerned amongst macroeconomy variables and data on GDP is regarded as the important index for assessing the national economic development and for judging the operating status of the macroeconomy as a whole [8].

GDP is the aggregate statistic of all economic activity and captures the broadest coverage of the economy than other macro-economic variables. It is the market value of all final goods and services produced within the borders of a nation in a year. It is often considered the best measure of how well the economy is performing. GDP can be measured in three ways. First, the Expenditure approach consists of household, business and government purchases of goods and services and net exports. Second the Production approach, it is equal to the sum of the value added at every stage of production (the intermediate stages) by all industries within the country, plus taxes and fewer subsidies on products in the period. Third 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 [1].

Since the emergence of Bangladesh in 1971 as an independent country, its economy has experiencing modest and reasonably steady annual Gross Domestic Product (GDP) growth rate per over 4.0 per cent. The rate is higher than that the pre-independence era, but somewhat low according to the standard set by contemporary South Asian countries.

GDP per capita in Bangladesh (107 US dollars) was more, than GDP per capita in Myanmar (99 US dollars), but was less, than GDP per capita in India (111 US dollars) and Pakistan (222 US dollars). Figure 1.1 shows the per capita GDP of Bangladesh, Myanmar, India, and Pakistan.

In 1970 (on the eve of Bangladesh's Independence), India's per capita income was higher by only 4.0 per cent in current dollar terms; but it increased to over 50 per cent by the year 2000. The gap with Pakistan was initially 64 per cent, but it increased to over 80 per cent by 2000. It would, therefore, appear that Bangladesh's performance in the area of economic growth has been worse than that of both India and Pakistan.

Per capita GDP of Bangladesh and China is almost the same at a time 1970-1990. But the GDP of China is suddenly chance at 20th century in 6 times (approximately). On the other hand, in 2013 GDP per capita in Bangladesh was less, than GDP per capita in United States (52392 US dollars) in 53.5 times, Germany (45091 US dollars) in 46 times, Japan (38528 US dollars) in 39.3 times and France (43339 US dollars) in 45 times. The accompanying table 1.1 compares the per capita GDP of Bangladesh, United States, Germany, France, Japan and China.

II. THE OBJECTIVE OF THIS PAPER

The aim of this study is to apply the ARIMA model to analyze the series and use forecast the US Dollar (US$) GDP of Bangladesh. So the objectives are as follows:

To assess the trends of US Dollar (US$) GDP of Bangladesh.

Estimate the breakpoint.

To find a suitable model for predicting and estimating the parameters US Dollar (US$) GDP of Bangladesh.

To forecast (2017-2021) the US Dollar (US$) GDP of Bangladesh in the near future.

III. SOURCE OF DATA

This study uses secondary data from World Bank database, an internationally representative sample survey. The per capita GDP data from 1968 to 2017 is used for this analysis. The data used in this study is reliable and publicly available. Other published reports from the World Bank are also used in order to support the analysis. World Development Indicators (WDI) publication (Published by World Bank) is a collection of time-series data for 214 economies, with many indicators going back more than 50 years. WDI provides cross-country comparable statistics about development and people's lives around the globe. It is divided into six sections = World View, People, Environment, Economy, States and Markets, and Global Links. Data source: http://data.worldbank.org/indicator/NY.GDP.PCAP.CD

IV. DATA PROCESSING

In regard to the data processing, we have extensively uses R programming language for making graphs, assessing normality and randomness of data set. Several packages of R programming language such as "ggplot2", "tseries", "timeSeries", "forecast", "gridExtra", "TTR" and "reshape2" are used for modelling and forecasting of time series data. SPSS-20.0 and SAS -9.1.3 also used for additional analysis of data.

V. LITERATURE REVIEW

Gross domestic product (GDP) refers to the value of all final goods and services produced within a country or an area in a period of time (a quarter or a year) and is often considered the best standard of measuring national economic conditions [5].

In 2011, Wang and Wang [10] forecast the GDP of China based on the time series methods developed by Box and Jenkins [2]. They set up an ARIMA model of the GDP of China from 1978 to 2006. They then choose the best ARIMA model based on statistical tests and forecast the GDP from 2007 to 2011. The result shows that the error between the actual value and the predicted value is small which indicates that the ARIMA model is a high precision and effective method to forecast the GDP time series.

In 2006, Sheng [9] analyze and forecast the GDP per capita development in the Zhejiang province in China based on the backpropagation (BP) neural network and an ARIMA model. The result indicates that, from 2006 to 2010, the average GDP per capita of Zhejiang province during these five years will be 40624.53 Yuan, and the average growth rate of GDP per capita is 10.01% per year.

In 2010, Wei, Bian and Yuan [11] forecast the GDP of the Shanxi province in China based on the ARIMA model. Using GDP data from 1952 to 2007, they set up an ARIMA (1, 2, 1) time series model, and compare the actual and predicted values from 2002 to 2007. The result indicates that the error between the real GDP value and the predicted value is within 5%.

In 2011, Mei, Liu and Jing [7] constructed a multi-factor dynamic system VAR forecast model of GDP by selecting six important economic indicators, which include the social retail goods, fiscal revenue, investment in fixed assets, secondary industry output, tertiary industry output, and employment rate, based on data from the Shanghai region in China. The analysis shows that the significance of the model is high 5 and the results show that the relative forecast error is quite small, leading the authors to conclude that the VAR model has a considerable practical value.

VI. MATERIALS AND METHODS

Graphical Analysis

The single most important thing to do when first exploring the data is to visualize the data through graphs. The basic features of the data including patterns and unusual observations are most easily seen through graphs. Sometimes graphs also suggest possible explanations for some of the variation in the data. The first step in the analysis of any time series is to plot the data [6].

Partial auto-correlation function

The estimated partial autocorrelation function (PACF) is broadly similar to an estimated ACF. An estimated PACF is also a graphical representation of the statistical relationship between sets of order pairs drawn from a single time series. A partial autocorrelation is the amount of correlation between a variable and a lag of itself that is not explained by correlations at all lower-order-lags. The autocorrelation of a time series Y at a lag 1 is the coefficient of correlation between Y and Y, which is presumably also the correlation between Y and Y. But if Y is correlated with Y and Y is equally correlated with Y, then it is necessary to find the correlation between Y and Y. thus the correlation at lag 1 “propagates” to lag 2 and presumably to higher-order lags.

The partial autocorrelation at lag 2 and the expected correlation due to the propagation of correlation at lag 1. Exponential declining nature of PACF plots also helps in taking the decision about the degree of moving average [4].

Auto-regressive

If an observation of a time series is dependent on the previous terms in the series, then it is called Auto-regressive. For example,

$$X_t = \phi X_{t-1} + \phi X_{t-2} + W_t$$

The equation represents a regression of the current value t x on the immediate two successive past values and hence the term auto-regression is suggested for this model.

Moving Average

We might replace the white noise series t w by a moving average that smoothes the series. For example, consider replacing t w in white noise by an average of its current value and its immediate neighbours in the past and future, that is,

$$V_t = \frac{1}{3}(W_{t-1} + W_t + W_{t+1})$$

Inspecting the series shows a smoother version than white noise, reflecting the fact that the slower oscillations are more apparent and some of the faster oscillations are taken out.

Autoregressive Moving Average (ARMA) Process

The time series $\{X_t\}$ is a $ARMA(p,q)$ process if $\{X_t\}$ is stationary and if for every t,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_p Z_{t-q},$$

Where $\{Z_t\} \sim WN(0,\sigma^2)$ and the polynomials $(1-\phi_1 z - \cdots - \phi_p z^p)$ and $(1+\theta_1 z + \cdots + \theta_q z^q)$ have no common factors. The process is said to be an $ARMA(p,q)$ process with mean μ if $|X_t - \mu|$ is an $ARMA(p,q)$ process [3].

In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. These models are fitted to time series data either to better understanding of the data or to predict future points in the series (forecasting). They are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to reduce the non-stationarity.

Model Selection Criteria

Several criteria are used for this purpose. Such as

Akaike Information Criterion (AIC)

The idea of imposing a penalty for adding regressors to the model has been carried further in the AIC criterion, which is defined as:

$$AIC = e^{2k/n} \frac{\sum \hat{u}_i^2}{n} = e^{2k/n} \frac{RSS}{n}$$

Where K is the number of regressors and n is the number of observations. For mathematical convenience, (1) is written as

$$\ln AIC = \frac{2k}{n} + \ln \frac{RSS}{n}$$

Where $\ln AIC$ is the natural log of AIC and $\frac{2k}{n}$ is the penalty factor.

In comparing two or more models, the model with the lowest value of AIC is preferred. One advantage of AIC is that it is useful for not only in-sample but also out of sample forecasting performance of a regression model. Also, it is useful for both nested and non-nested models. It has been used to determine the lag length in an $AR(p)$ model.

Schwarz Information Criterion (SIC)
Similar in spirit to the AIC, the SIC criterion is defined as:

$$SIC = n^{k/n} \frac{\sum \hat{u}_i^2}{n} = n^{k/n} \frac{RSS}{n}$$

Or in log-form:

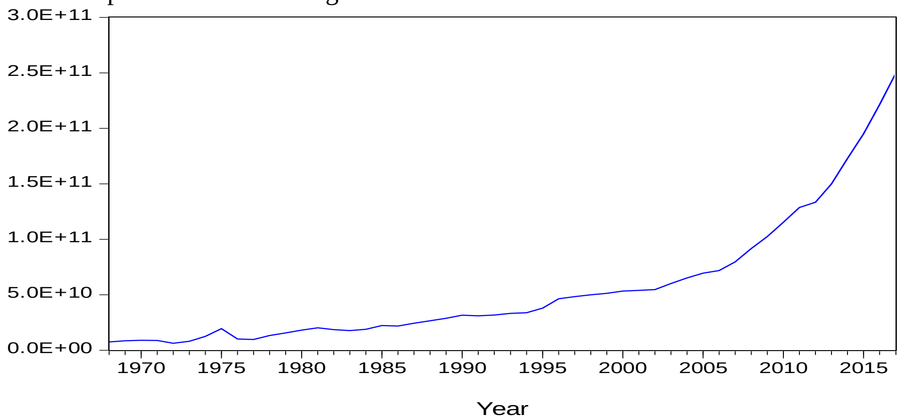
$$\ln SIC = \frac{k}{n} \ln n + \ln \frac{RSS}{n}$$

where $\frac{k}{n} \ln n$ is the penalty factor. SIC imposes a harsher penalty than AIC. Like AIC, the lower the value of SIC, the better the model. Again, like AIC, SIC can be used to compare in the sample or out of sample forecasting performance of a model.

VII.RESULT AND DISCUSSION

The first step in the analysis of any time series is to plot the data. Such a plot gives an initial clue about the likely nature of the time series or shows an upward or downward trend, seasonal or cyclical variations etc.

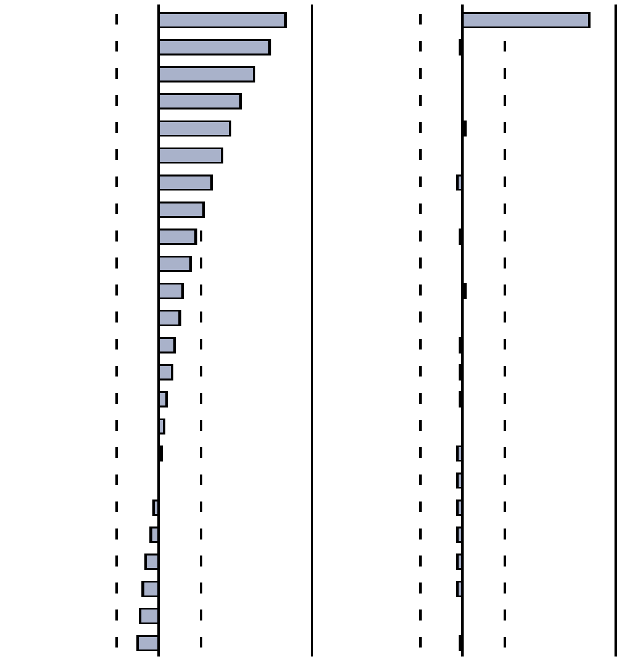
Figure-1.1: Time series plot of GDP for Bangladesh



A time series is stationary if it's mean, variance, and autocovariance (at various lags) remain the same no matter at what point of time we measure them. That is, they are time-invariant. Here from the above Figure-1.1, we can see that the mean and variance are approximately same at various levels of time. So we may say that the time series data is stationary.

ACF and PACF

Figure-1.2: ACF and PACF of GDP for Bangladesh.



	AC	PAC	Q-Stat	Prob
1	0.868	0.868	40.009	0.000
2	0.752	-0.010	70.608	0.000
3	0.651	0.001	94.046	0.000
4	0.563	-0.004	111.95	0.000
5	0.490	0.013	125.81	0.000
6	0.427	0.004	136.59	0.000
7	0.361	-0.046	144.48	0.000
8	0.304	-0.003	150.21	0.000
9	0.252	-0.015	154.25	0.000
10	0.207	-0.007	157.04	0.000
11	0.172	0.010	159.01	0.000
12	0.143	0.000	160.40	0.000
13	0.113	-0.019	161.30	0.000
14	0.086	-0.012	161.84	0.000
15	0.061	-0.011	162.11	0.000
16	0.039	-0.007	162.23	0.000
17	0.016	-0.025	162.25	0.000
18	-0.008	-0.026	162.25	0.000
19	-0.032	-0.027	162.34	0.000
20	-0.058	-0.030	162.63	0.000
21	-0.083	-0.031	163.25	0.000
22	-0.110	-0.036	164.36	0.000
23	-0.127	0.001	165.93	0.000
24	-0.142	-0.015	167.95	0.000

From the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) in Figure-1.2 we observe that the Autocorrelation Function (ACF) has significant spikes at many lags. Also, we observe that the Partial Autocorrelation Function (PACF) has significant spikes occurs at different lags. Autocorrelation coefficient starts at a very high at lag 1 (0.868) and declines slowly. Hence the autocorrelation coefficient are increasing or decreasing at different lags. So we conclude that the data is stationary.

Unit Root Test

A test of stationarity that has become widely popular over the past several years is the unit root test. Let us consider the following time series model

$$y_t = \rho\, y_{t-1} + \mu_t .$$

where μ_t is a white noise error term.

Table-1.1: Unit Root Test of GDP for Bangladesh

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		5.191984	1.0000
Test critical values:	1% level	-3.574446	
	5% level	-2.923780	
	10% level	-2.599925	

*MacKinnon (1996) one-sided p-values.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDP_US$	0.113043	0.021773	5.191984	0.0000
D(GDP_US$	0.146679	0.166831	0.879209	0.3840
C	-1.64E+09	7.80E+08	-2.10159	0.0412

*D (GDP_US$) =First difference

R-squared	0.797225	Mean dependent var	5.03E+09
Adjusted R-squared	0.788213	S.D. dependent var	7.56E+09
S.E. of regression	3.48E+09	Akaike info criterion	46.83748
Sum squared resid	5.44E+20	Schwarz criterion	46.95443
Log likelihood	-1121.100	F-statistic	88.46035
Durbin-Watson stat	1.942741	Prob(F-statistic)	0.000000

From the above Table-1.1, we use the Augmented Dickey-Fuller (ADF) test and Schwarz Information Criterion (SIC) for model selection. Here the null hypothesis states that GDP has a unit root. This means that the data is stationary.
If we take an absolute value and ignore the sign, we see the value of t-statistic is greater than all other test critical values at 1% level, 5% level and 10% level of significance. So we reject the null hypothesis. Now we conclude that the data is stationary.

Model Selection
From Schwarz Information Criterion (SIC) for model selection, we get the best model ARIMA$(2, 1, 1)$ which contain two Auto-regressive (AR) parameters and one Moving Average (SA) parameter. The estimated parameters are shown below:

Table-1.2: Model Selection of GDP for Bangladesh

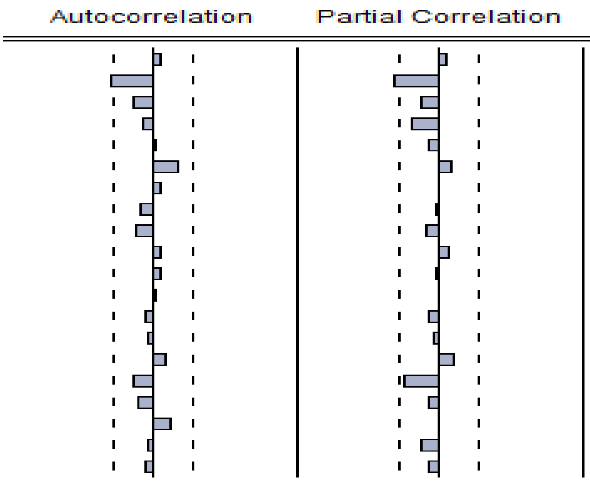
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.82E+10	1.28E+11	0.455057	0.6513
AR(1)	2.129740	0.095862	22.21675	0.0000
AR(2)	-1.125951	0.105916	-10.63059	0.0000
MA(1)	-0.999459	0.087020	-11.48535	0.0000

R-squared	0.996831	Mean dependent var	5.82E+10
Adjusted R-squared	0.996615	S.D. dependent var	5.86E+10
S.E. of regression	3.41E+09	Akaike info criterion	46.81598
Sum squared resid	5.11E+20	Schwarz criterion	46.97191
Log-likelihood	-1119.584	Hannan-Quinn criteria.	46.87491
F-statistic	4613.358	Durbin-Watson stat	1.887372
Prob(F-statistic)	0.000000		

From the above Table-1.2, we use t-Statistic to test the null hypothesis that the parameters are equal to zero. The p-value shows that all the coefficients are highly significant. The R-squared and Adjusted R-squared of ARIMA $(2, 1, 1)$ model is 0.996831 and 0.996615 respectively.
The entire coefficient of the estimated model is significant at 5% level of significance. The observed R-Squared suggested that the sample regression line fit the data as well. Also, all of the estimated coefficients are satisfied invariability and stability condition.

Residual Analysis
The autocorrelations and partial autocorrelations of the squared residuals up to any specified number of lags and computes the Ljung-Box Q-statistics for the corresponding lags. The correlograms of the squared residuals can be used to check autoregressive conditional heteroskedasticity (ARCH) in the residuals.

Figure-1.3: Residuals Test of GDP for Bangladesh



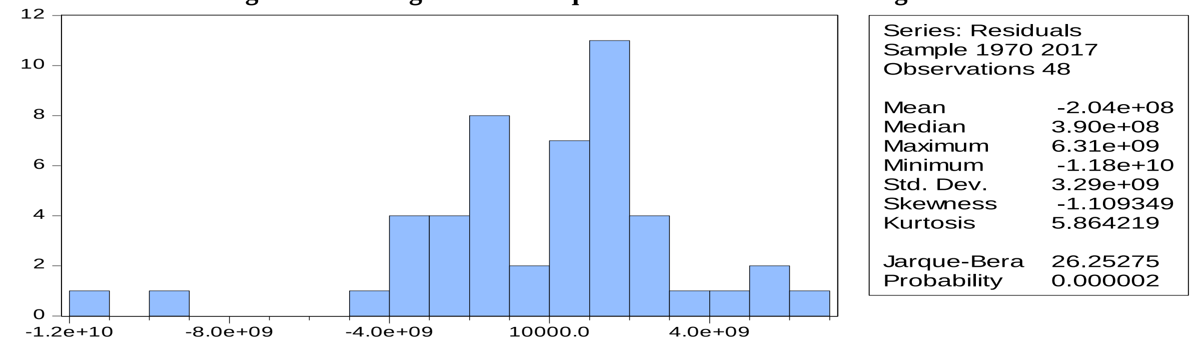
	AC	PAC	Q-Stat	Prob
1	0.050	0.050	0.1298	
2	-0.319	-0.323	5.4439	
3	-0.155	-0.131	6.7238	
4	-0.081	-0.195	7.0833	0.008
5	0.021	-0.081	7.1068	0.029
6	0.188	0.088	9.1169	0.028
7	0.054	-0.004	9.2847	0.054
8	-0.086	-0.013	9.7317	0.083
9	-0.133	-0.093	10.815	0.094
10	0.048	0.068	10.958	0.140
11	0.056	-0.015	11.164	0.193
12	0.011	-0.007	11.173	0.264
13	-0.058	-0.077	11.405	0.327
14	-0.042	-0.030	11.531	0.400
15	0.097	0.111	12.215	0.429
16	-0.154	-0.250	14.000	0.374
17	-0.100	-0.078	14.769	0.394
18	0.116	-0.008	15.846	0.392
19	-0.029	-0.122	15.918	0.459
20	-0.056	-0.080	16.189	0.511

Autocorrelation and Q-test for different lags support the hypothesis that there is no autocorrelation in the residual. Thus the model is fully specified.

Histogram and Jarque-Bera test

The histogram and descriptive statistics of the residuals, including the Jarque-Bera statistic for testing normality. If the residuals are normally distributed, the histogram should be bell-shaped and the Jarque-Bera statistic should not be significant.

Figure-1.4:Histogram and Jarque-BeraPlotof GDP for Bangladesh.



Histogram and Jarque-Bera test indicate that the residual is normally distributed. Thus the model is fully specified.

Table-1.5: Breusch-Godfrey Serial Correlation LM Testof GDP for Bangladesh

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	3.491832	Prob. F(2,42)	0.0396
Obs*R-squared	6.682509	Prob. Chi-Square(2)	0.0354

From the Table-1.5 we can see that both LM tests support the hypothesis that there is no autocorrelation in the residual. Thus the model is fully specified.

Stability test

Stability test are another test for model validation. For the stability test, the most used test is Chow's test. The Chow's test is two categories, they are Chow's Breakpoint test and Chow's Forecast test. They are given below-

Table-1.6: Stability test of GDP for Bangladesh

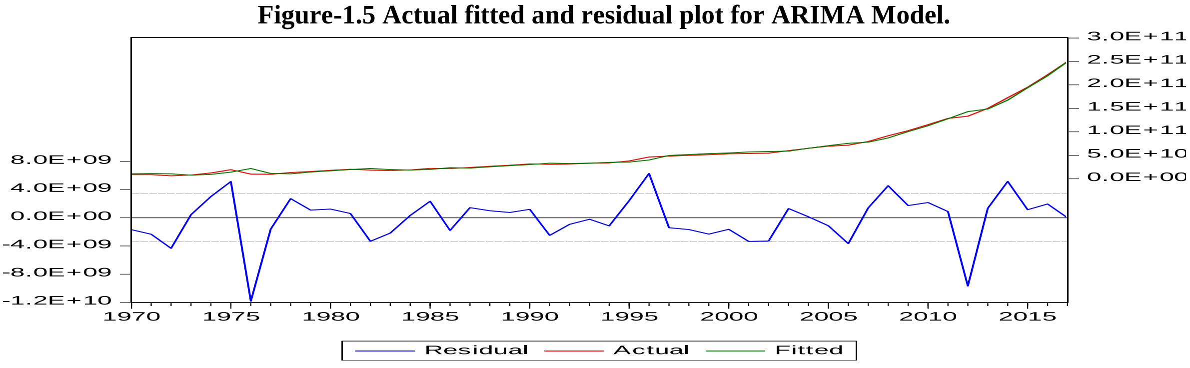
Log-likelihood ratio	14.45805	Prob. Chi-Square(4)	0.0060

Table-1.7: Chow Forecast Test of GDP for Bangladesh

	Value	df	Probability
F-statistic	1.879299	(8, 36)	0.0941
Likelihood ratio	16.75108	8	0.0328

Here for Chow's breakpoint test, we take the breakpoint sample in 2000 and for Chow's forecast test we take forecast sample from 2010 to 2017. From the Table-4.7 we can see that both Chow's tests support the hypothesis that there is no structural break in the model. Thus the model is fully specified.

Figure-1.5 Actual fitted and residual plot for ARIMA Model.



According to our above calculation or completing the Box-Jenkins process, our estimated model is ARIMA (2,1,1) .We have estimated the model using the data covered the period from 1968to 2017. Now we forecast between the observations 1968 to 2017, which is called in-sample forecast. The in-sample forecast sample GDP is from 1968 to 2017. The out-sample forecast sample of maximum temperature is in 2017 to 2021

Figure-1.6 in-sample forecast of GDP from1968 to 2017

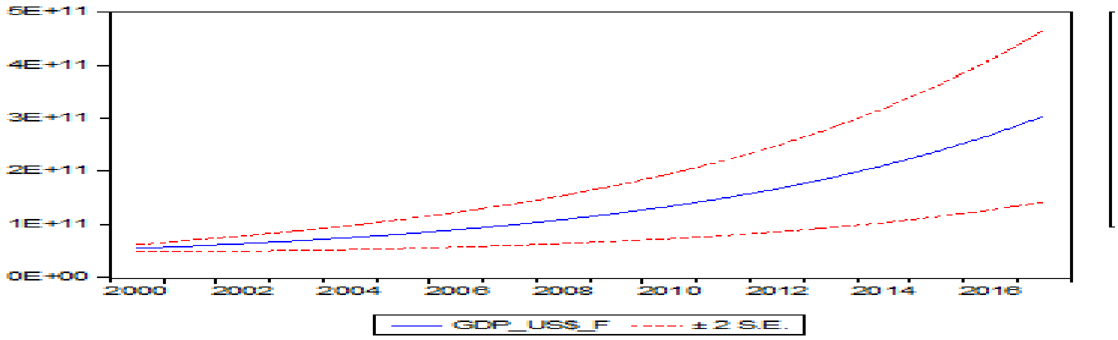
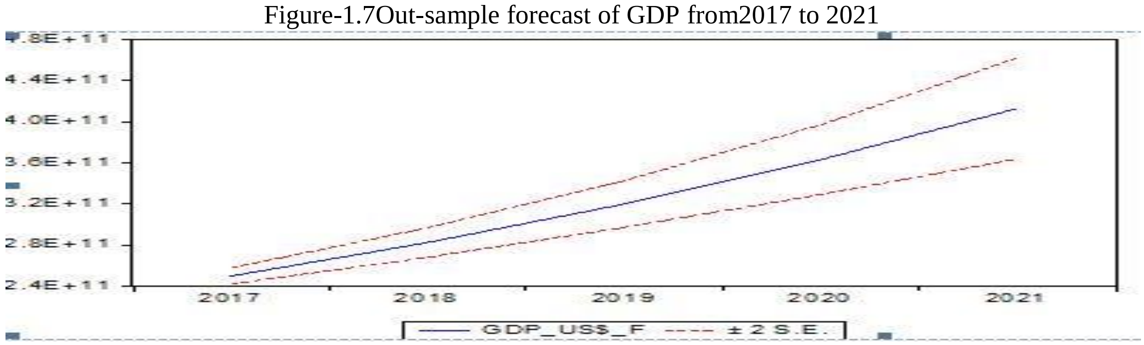


Figure-1.7Out-sample forecast of GDP from2017 to 2021



The above figure (Figure-1.7) shows five years (2017 to 2021) forecasting with a 95% confidence interval of GDP. The upper red and lower red lines on the above and below the blue line shows the 95% confidence upper and lower limits respectively. According to the graph, we can forecast GDP in 2017 to 2021 and we can see that the GDP is increasing.

Forecast up-to 2017-2021.

Forecast Year	Forecast Value
2017	2.50E+11
2018	2.82E+11
2019	3.20E+11
2020	3.63E+11
2021	4.13E+11

VIII. CONCLUSION

The research was done to identify a suitable model for forecasting the money value of Bangladesh using the yearly GDP from 1968 to 2017. For this reason, a time series analysis has been done based on the characteristics of the data.

Initially, non-stationarity has been found in data by Augmented Dickey-Fuller test. This non-stationarity has been removed first by taking second differences. Using the model selection criterion AIC and BIC, ARIMA (2, 1, 1) has been selected. The parameters of this model have been estimated using the maximum likelihood method and all the parameters of this model have been found to be significant assuming normal distributions of the estimators. The assumption is made on normality and independence of the residuals has been checked using different plots and test and all the result have been satisfactory. The plot comparing actual values and fitted values using the model shows much close fit. Then the model was used for forecasting purposes and we found an increasing trend in the data set.

BIBLIOGRAPHY

[1]. Ard, H.J, den Reijer. (2010), Macroeconomic Forecasting using Business Cycle leading indicators, Stockholm: US-AB.
[2]. Box, G. E. P., Jenkins, G. W., and Reinsel, G. (1976), Time Series Analysis, Forecasting And Control, 3rd edition, Englewood Cliffs, N.J.: Prentice-Hal
[3]. Brokwell, P. J. and Davis, R. A. (1996), Introduction to Time Series and Forecasting, Springer, New York. Gujrati, Damodar N.(2003), Basic Econometrics, 4th edition, McGraw-Hill Book co; New York.
[4]. Mankiw, N. G. & Taylor, M. P. (2007). Macroeconomics. New York: Worth.
[5]. Makridakis, S., Wheelwright, S. C. and Hyndman, R. J. (1998), Forecasting Methods and Applications, 3rd edition, John Wiley, Inc., New York, NY 101858-0012.
[6]. Mei, Q., Liu, Y. & Jing, X. (2011). Forecast the GDP of Shanghai Based on the Multi-factors VAR Model. Journal of Hubei University of Technology, 26(3).
[7]. Ning, W., Kuan-jiang, B. and Zhi-fa, Y. (2010), Analysis and forecast of Shaanxi GDP based on the ARIMA Model, Asian Agricultural Research, Vol. 2 No. 1, pp. 34-41.
[8]. Sheng, Y. (2006). Forecast the GDP per Capita of Zhejiang Province Based on the BP Neural Network and ARIMA Model. Journal of Modern Shopping Malls, 23(1).Mamun, M. A. (2009).A study of GDP sectors
[9]. Wang, Z. & Wang, H. (2011). GDP Prediction of China Based on ARIMA Model. Journal of Foreign Economic and Trade, 210(12).
[10]. Wei, N., Bian, K. J. & Yuan, Z. F. (2010). Analyze and Forecast the GDP of Shaanxi Province Based on the ARIMA Model. Journal of Asian Agricultural Research, 2(1), 34-41.

Md. Sifat Ar Salan "Forecasting GDP of Bangladesh Using Time Series Analysis."International Journal of Mathematics and Statistics Invention (IJMSI), vol. 08(01), 2020, pp.07-15.