RESEARCH ARTICLE

MODELING AND FORECASTING GDP AT CURRENT MARKET PRICE IN BANGLADESH: AN APPLICATION OF ARIMA MODEL.

*Biplab Biswas and Md. Wahiduzzaman.
Department of Statistics, Bangabandhu Sheikh MujiburRahman Science and Technology University, Gopalganj.

Abstract

An autoregressive integrated moving average (ARIMA) model is useful to analyze time series data either for better understanding or for forecasting future points in the series. The ARIMA model is useful even when the data show evidence of non-stationarity, where an initial differencing step is applied to remove the non-stationarity. In this study, it is tried to find out the trend in GDP at current market price of Bangladesh using Box-Jenkins approach for ARIMA model based on the GDP at current market price data of Bangladesh from 1989-90 to 2014-15. Initially, evidence of non-stationarity was found in the data using the Dickey-Fuller test and the non-stationarity was removed by taking first differences. Then using the model selection criteria, AIC, \( ARIMA (0, 2, 1) \) model has been selected and it has been used to forecast GDP at current market price of Bangladesh up to 2026-27. It is hoped that the forecasts would be useful for researcher and policy makers.

Introduction:

Most of the expert and economists favor high and efficient government planning because they think that it ensures and contributes to enlarging GDP. Planning is one of the main components of public budget. Investment is one of the main components of aggregate demand. Its plays an important role of economic growth (Ahmed, et.al. 2010). Public investment is fully conducted by the government. By development plan, the government can improve economic situation of the country. Currently, we observed that government budget and private investment simultaneously plays great role to rapid economic growth. Both the public and private investments are required to boost up real GDP at current market price (main part of public investment) has a big share compared to private investment (Mahi, et.al. 2015). Bangladesh is small country but over populated. Its economy is rapidly improving based on market price. Most of the indicators of development show their positive reaction since 1971. According to Wikipedia, Bangladesh has made significant strides in its economic sector performance since independence in 1971 (Wikipedia). The economy has improved vastly after 1990s. Most of the characteristics of public investment are present in GDP at current market price. So we can consider the maintenance current market price as the public investment. Though public investment is organized in different ways but the aim and objects are same for all countries. Finally, we can infer that GDP at current market price is the public investment in our country [Wikipedia]. Gross domestic product (GDP) is a monetary measure of the market value of all final goods and services produced in a period (quarterly or yearly). Nominal GDP estimates are commonly used to determine the economic performance of a whole country or region, and to make international comparisons. Nominal GDP per capita does not, however, reflect differences in the cost of living and the inflation rates of the countries. Therefore, using a GDP...
PPP per capita basis is arguably more useful when comparing differences in living standards between nations (http://en.wikipedia). In 1993 India adopted an exchange rate system in which the exchange rate is determined by the supply and demand in the inter-bank foreign exchange market. The adoption of this floating flexible regime had not freed the Reserve Bank of India, the central bank of India, from intervening in the foreign exchange market (Branson, 2002). As the thinness of the Indian foreign exchange market as well as large transactions can cause excessive volatility, Reserve Bank of India pursues an explicit policy of intervention in the spot market and also undertakes both forward and swaps transactions in support of its exchange rate objectives (World Bank, 2001). Sri Lanka has with the adoption of the floating system the volatility increased. The State Bank of Pakistan was forced to intervene in foreign exchange market, which takes the form of outright sales of foreign exchange, swap transactions and supply of foreign exchange to banks to cover certain bulky imports (World Bank, 2001) (Krugman, 2000). In 2001 Sri Lanka adopted the free floating exchange rate regime. Immediately after the float, the exchange rate volatility arose sharply and caused massive depreciation of the Lanka rupee in spite of putting precautionary foreign exchange regulations in combination with the introduction of the float. Sri Lanka has with the free float a set of guidelines for dealing in the foreign exchange market and intervention by the central bank (World Bank, 2001) (Bukhari, et.al. 2007). In the backdrop of the discussion so far, it is intended to look at the performance of the exchange rate policy of Bangladesh and its impact on inflation and growth rate of GDP. In Bangladesh perspective, importance of GDP at current market price is relatively high compared to other developing country. Due to low infrastructure, the returns of GDP are not satisfactory and still not clear (Mahi, 2015). It is very important to know whether GDP and economic growth are related to each other especially in Bangladesh perspective or not. For any economy like Bangladesh, GDP is the vital factor of the development. From the above discussion, it is clear to us many many study was conducted to forecast some economy variables by using time series model in respect of Bangladesh but now it is needed to study about GDP. The vital objective of the study is to select the better model and forecasting GDP on overall economy of Bangladesh.

Source of data and Methods:-
The data is collected from different editions of Statistical Yearbook published by Bangladesh Bureau of Statistics (BBS). The editions are: (i) Statistical Yearbook of Bangladesh-1975(ii) Statistical Yearbook of Bangladesh-1990(iii) Statistical Yearbook of Bangladesh-2000 (iv) Statistical Yearbook of Bangladesh-2010. There are a lot of sectors that impact our GDP; Current market price is one of them. In order to, find out the probable relationship between economic growth and Current market price by using auto-regressive and moving average models and other statistical method. This study analyzes the Modeling and forecasting GDP at current market price of Bangladesh. We know that, the general purpose of ARIMA method is to know more about the proper relationship between some explanatory variable. We use annual data for the period from the fiscal year 1989/90 to 2014/2015. We will use secondary data that are collected from different sources. All data are time series data. All variables are measured in millions of us dollar in core in taka. Bangladesh is also not different from like other developing countries. Some data are clearly vague such as same data but from different sources are inconsistent. That’s why, careful attention was implied during self-compiled period. Here, data are to be used in this analysis will be standard and reliable because of all sources of data are well known, recognized, widely used and accepted by government and others. So, data, which will be used in this study, must be reliable.

Autoregressive Integrated Moving Average (ARIMA) model:-
To develop the Box and Jenkins (1972) method for analyzing stationary univariate time series data. The Box-Jenkins methodology for ARIMA models differences the series to stationary and then combines the moving average with autoregressive parameters to yield a comprehensive model useful for forecasting. The underlying goal is to find an appropriate formula so that the residuals are as small as possible and exhibit no pattern. Box-Jenkins forecasting models are based on statistical concepts and principles and are able to model a wide spectrum of time series behavior.

Augmented Dickey-Fuller Test (ADF Test):-
The Augmented Dickey-Fuller Test (ADF Test) is used to determine the stationarity of the data. The ADF test is used to test for unit root [9]. The null-hypothesis for an ADF test: $H_0: \gamma = 0 \text{ vs } H_1: \gamma < 0$. Where $H_0$: is the null hypothesis (has unit root) and $H_1$: Does not have unit root, $\gamma$ is the estimated coefficient. The testing procedure for
the ADF test is the same as for the Dickey–Fuller test but it is applied to the model. A random walk with drift and trend is represented as:
\[ \Delta y_t = \alpha + \beta_t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_{p-1} \Delta y_{t-p+1} + \epsilon_t \] (1)

Where, \( \alpha \) is a constant, \( \beta \) the coefficient on a time trend and \( p \) — the lag order of the autoregressive process. Imposing the constraints \( \alpha = 0 \) and \( \beta = 0 \) corresponds to modeling a random walk and using the constraint \( \beta = 0 \) corresponds to modeling a random walk with a drift. The test statistics, \( \tau = \frac{\bar{y}}{\sigma_y} \), Where: \( \bar{y} \) is the estimated coefficient and \( \sigma_y \) is the standard error in the coefficient estimate. The test statistics value \( \tau \) is compared to the relevant critical value for the Dickey Fuller Test. If the test statistic is less than the critical value, we reject the null hypothesis and conclude that no unit-root is present. The ADF Test does not directly test for stationarity, but indirectly through the existence (or absence) of a unit-root. Decision rule: If \( t^* > ADF \) critical value, \( =\rightarrow \) reject null hypothesis, i.e., unit root exists. If \( t^* < ADF \) critical value, \( =\rightarrow \) reject null hypothesis, i.e., unit root does not exist. Using the usual 5\% threshold, differencing is required if the \( p \) value is greater than 0.05.

**Formulation of ARIMA Model:**

**ARIMA\((p,d,q)\)** Completely ignores independent variables and assumes that past values of the series plus previous error terms contain information for the purposes of forecasting. The integers refer to the Auto regressive\((AR)\), Integrated \((I)\) and Moving Average \((MA)\) parts of the data set respectively. The ARIMA models are applicable only to a stationary data series, where the mean, the variance, and the autocorrelation function remain constant through time [8]. Autoregressive process \( AR \) presses a dependent variable as a function of past values of the dependent variable. An \( p \) th order process is of the form:
\[ y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t \] (2)

Where; \( y_t \) is the stationary dependent variable being forecasted at time \( t \). \( y_{t-1}, y_{t-2}, \ldots y_{t-p} \) is the response variable at time lags \( t-1, t-2, \ldots t-p \) respectively, \( \alpha = \mu(1- \phi_1 - \ldots - \phi_p) \). \( \phi_1, \phi_2, \ldots \phi_p \) are the coefficients to be estimated. \( \epsilon_t \) is the error term at time \( t \) with mean zero and a constant variance. Using the backshift operator, we can write the \( AR(p) \) model as:
\[ (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)y_t = \varphi(B)y_t = \epsilon_t \] (3)

The moving average model of order \( MA(q) \) is defined as:
\[ y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} \] (4)

Where; \( q \) is the number of lags in the moving average and \( \theta_1, \theta_2, \ldots \theta_q \) are parameters to be estimated the moving average operator is given by; \( \theta(B) = \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q \). To create an ARMA model, we begin with an econometric equation with no independent variables \( Y_t = \beta_0 + \epsilon_t \) and add to it both the \( AR \) process and the \( MA \) process.
\[ y_t = \beta_0 + \varphi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} \] (5)

\( \beta_0, \varphi_1, \phi_2, \ldots \phi_p \) are the \( AR \) parameters and \( \theta_1, \ldots \theta_q \) are the \( MA \) parameters, \( \beta_0 \) is the intercept of the AR process. Where, \( \varphi_1, \phi_2, \ldots \phi_p \) are the coefficients of the autoregressive and moving average processes respectively. The integrated ARMA or ARIMA model is a broadening class of ARMA model which includes a differencing term. A process is said to be \( ARIMA(p,d,q) \) if
\[ \nabla^d y_t = (1-B)^d y_t \] (6)

Is an \( ARIMA(p,q) \). This is generally written as; \( \varphi(B)(1-B)^d y_t = \theta(1-B) \epsilon_t \). A first- differenced GDP at current market price series is of the form:
\[ GDP_t = (N GDP_t) = GDP_t - GDP_{t-1} = \Delta GDP_t - \Delta GDP_{t-1}. \] Thus \( ARIMA(p,d,q) \) model may be specified as:
\[ GDP_t = \beta_0 + \varphi_1 GDP_{t-1} + \phi_2 GDP_{t-2} + \ldots + \phi_p GDP_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} \] (7)

Where, \( GDP_t \) is the differenced GDP at current market price series of order \( d \), and \( \varphi_1, \phi_2, \ldots \phi_p \) are the parameters to be estimated. Shumway & Stoffer (2011). The equation must assume stationarity before applying to a time series. In case of non-stationarity successive differences are taken until the series is stationary. In practice the differences are rarely more than two. The aim of this methodology is to find the most appropriate \( ARIMA(p,d,q) \) model and to use it for forecasting. It uses an iterative six-stage scheme: (1) A priori identification of the differentiation order \( d \) (or choice of another transformation) (2) A priori identification of the orders \( p \) and \( q \) (3) Estimation of the parameters, \( (\varphi, \beta, \theta, \delta^2) = \text{Var}(\epsilon_t) \) (4) Validation (5) Choice of a model (6) Prediction.
Selection of best model:-
When a model is identified, the work is to estimate the AR and MA parameters to fit the time series that is being modeled. For ARIMA model, the method of least squares and the method of maximum likelihood can be used. The method of maximum likelihood finds the values of the parameters, which maximizes the likelihood. To find a tentative model, the pattern in the ACF and PACF are considered which are summarized by Markridakis et al (1998). Spikes at lag 1 top, and cuts off to zero in PACF indicates $AR(p)$ model. Again, Spikes at lag 1 toq, and cuts off to zero in ACF indicates $MA(q)$ model. To select the best model among the plausible model, the Akaike information criterion ($AIC$) can be used, which is given by

$$AIC = -2\log L + 2m$$

Where, $L$ is the maximum likelihood, $m$ is the number of parameters estimated in the model. The model with the minimum AIC is to be chosen.

Analysis and results:-

Time series plot:-
At first the data are plotted, where the variable is the GDP at current market price of Bangladesh. The following figure 3.1 shows the time plot of the variable.

![Time plot of GDP at current market price](image)

**Figure: 3.1:-** Time plot of GDP at current market price of Bangladesh
From the figure 3.1, it is seen that the data are showing an upward trend that is the data is non-stationary. As a result, to achieve stationarity differencing is needed.

Checking and obtaining stationarity:-
Before taking any difference, it is convenient to examine stationarity by using auto-correlation function (ACF) and partial auto-correlation function (PACF) graphs or by using different tests like augmented Dickey-Fuller test or L-Jung-Box test. The following figure shows the ACF and PACF plots of GDP at current market price of Bangladesh.

![ACF and PACF of GDP at current market price](image)

**Figure: 3.2:-** ACF and PACF of GDP at current market price of Bangladesh.
Here we see that, the ACF plot shows a slowly decreasing and for several lags. Thus the result shown by ACF and PACF are that the data is non-stationary. Stationarity can also be checked by several statistical tests. **Augmented Dickey-Fuller (ADF) test about unit root or stationarity:** A statistical test to check stationarity of a time series dataset is ADF test. Here, the null and alternative hypothesis is given by $H_0$: Data is non-stationary versus $H_1$: Data is stationary. For lag order 3, the calculated Dickey-Fuller test statistics is -1.3251 and the p-value is 0.6489. As a result, the null hypothesis cannot be rejected at 5 percent level of significance. Thus the test concludes the series is non-stationary. **White noise test for checking independence (L-Jung-Box test):** If the data are correlated or not, then we can use the L-Jung-Box test. Here, the null and alternative hypothesis is as: $H_0$: Data points are independent of one-another Vs $H_1$: Data points are correlated. For 52 degrees of freedom, the test shows the calculated Chi-squared statistic is 18.4972 and the p-value is 1.702e-05 which is less than .01. So, at 5 percent level of significance, the conclusion is, dataset is not independent. Since the GDP at current market price of Bangladesh is non-stationary, here differencing is used to achieve stationarity. At first, the first difference has been taken and the data is again checked whether it is stationary or not. The following plot shows the time plot of after taking first difference of the GDP at current market price data of Bangladesh.

![Time plot of first difference of GDP at current market price](image1)

![ACF of first difference of GDP at current market price](image2)

![PACF of first difference of GDP at current market price](image3)

**Figure: 3.3:** Time plot, ACF and PACF of first difference of GDP at current market price

From the above time plot, it is seen that non-stationarity in mean is achieved. Where the ACF plot shows only five spikes fall outside the confidence limits. Again in the PACF plot only one spike falls outside the confidence limits, which are the indications of stationarity. But, to become assured, the statistical tests are applied.

**White noise test for checking independence (L-Jung-Box test):** Hypothesis: $H_0$: Data points are independent of one-another Versus $H_1$: Data points are correlated. For 52 degrees of freedom, the test shows the calculated Chi-squared statistic is 16.6605 and the p-value is 4.47e-05. So, at 5 percent level of insignificance, the conclusion is, dataset is not independent, because the null hypothesis can be rejected. Since the first difference of GDP at current market price of Bangladesh is non-stationary, here second differencing is used to achieve stationarity. Then, the second difference has been taken and the data is again checked whether it is stationary or not. The following plot shows the time plot of after taking second difference of the GDP at current market price data of Bangladesh.
From the above time plot, it is seen that stationary in mean is achieved. The ACF plot shows that only one spike falls outside the confidence limits. Again in the PACF plot have no any spike falls outside the confidence limits, which are the indications of stationarity. But, to become assured, the statistical tests are applied. **White noise test for checking independence (L-Jung-Box test):** Hypothesis: $H_0$: Data points are independent of one another Versus $H_1$: Data points are correlated. For 52 degrees of freedom, the test shows the calculated Chi-squared statistic is 3.2491 and the p-value is 0.07146 which is less than .0.01. So, at 5 percent level of significance, the conclusion is, dataset is independent, because the null hypothesis cannot be rejected.

**Determining an appropriate ARIMA model:-**

Among the different plausible models, the best model can be selected by using the Akaike Information Criterion (AIC). Here, the model with the minimum AIC will be selected as the appropriate model. Table 3.1 shows the AIC values for different combinations of $p$ and $q$, that is, for different ARIMA $(p; d; q)$ models.

<table>
<thead>
<tr>
<th>ARIMA $(p, d, q)$</th>
<th>Values of AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA $(0, 0, 1)$</td>
<td>722.86</td>
</tr>
<tr>
<td>ARIMA $(0, 1, 1)$</td>
<td>615.79</td>
</tr>
<tr>
<td>ARIMA $(0, 2, 1)$</td>
<td>547.01</td>
</tr>
<tr>
<td>ARIMA $(0, 2, 2)$</td>
<td>548.12</td>
</tr>
<tr>
<td>ARIMA $(0, 2, 3)$</td>
<td>550.11</td>
</tr>
</tbody>
</table>

From the above table, we can find that ARIMA $(0, 2, 1)$ model is appropriate for the GDP at current market price data of Bangladesh, since it has the minimum AIC (547.01).

**Estimation and diagnostic checking:-**

We can find that, ARIMA $(0,2,1)$ is the best model for the data. Here, the ARIMA model includes no autoregressive and one moving average coefficients with two differences. Thus, we can write the ARIMA model is as $(1 - B)^2 y_t = (1 + \theta_1 B)e_t$. That is, $\phi^2 y_t = e_t + \theta_1 e_{t-1}$. Where, $\phi$’s be the autoregressive coefficient and $\theta$’s be the coefficient for moving average. Now, the parameters have to be estimated using the maximum likelihood method. Following table gives the maximum likelihood estimates of the parameters and significance of the parameters.
Table 2: The estimate and significance levels of the parameters of $ARIMA (0, 2, 1)$ model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>p-value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>-1.0000</td>
<td>0.1121</td>
<td>0.001</td>
<td>Significant</td>
</tr>
</tbody>
</table>

Here, the diagnostic checking can be started by checking whether the errors are white noise or not. The assumption of independence of residuals can be checked by L-Jung-Box test.

**L-Jung-Box test for checking independence of residuals:**

Hypothesis: $H_0$: The residuals are independent of one-another Versus $H_1$: The residuals are correlated. For 1 degree of freedom, the test shows the calculated Chi-squared statistic is 4.8824 and the p-value is 0.02713. So, at 5 percent level of significance, the null hypothesis cannot be rejected. Hence, it can be said that the residuals are independent of each other. Following plot shows the standardized residuals, ACF of residuals and p-values for L-Jung-Box statistics taking different lag values that is the diagnostic plot for the model.

**Figure 3.5:-** Diagnostic plot for $ARIMA (0,2,1)$ model

From this plot, the independence assumptions of residuals are also satisfied since none of the ACF values of lag 1 or more is significant. The plot also shows the large p-values for L-Jung-Box statistics for different lag values which is also the evidence of independent residuals.
The normal q-q plot is the following:

![Q-Q plots for the residuals of ARIMA (0,2,1) model](image)

Finally, the q-q plot shows that the residuals are almost normal.

**Forecasting:**
At the starting of the forecast, we have found the forecasted values which were included in our data by using our fitted ARIMA model. Then the initial forecasts obtained from the ARIMA model was for differencing series of the GDP at current market price data. To obtain the forecasts valid for the original series second differencing has been taken. The point forecast with 95% confidence interval of yearly total GDP of Bangladesh for the years 2015-2016 to 2026-2027 (next twelve years) using selected model is given in table 3.

**Table 3:** Forecast of total GDP for next twelve years

<table>
<thead>
<tr>
<th>Time periods</th>
<th>Point forecasts</th>
<th>95% Confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-16</td>
<td>1679054</td>
<td>(1639464, 1718644)</td>
</tr>
<tr>
<td>2016-17</td>
<td>1844508</td>
<td>(1762103, 1926912)</td>
</tr>
<tr>
<td>2017-18</td>
<td>2009961</td>
<td>(1876525, 2143398)</td>
</tr>
<tr>
<td>2018-19</td>
<td>2175415</td>
<td>(1983717, 2367113)</td>
</tr>
<tr>
<td>2019-20</td>
<td>2340869</td>
<td>(2084442, 2597296)</td>
</tr>
<tr>
<td>2020-21</td>
<td>2506323</td>
<td>(2179273, 2833373)</td>
</tr>
<tr>
<td>2021-22</td>
<td>2671777</td>
<td>(2268659, 3074895)</td>
</tr>
<tr>
<td>2022-23</td>
<td>2837231</td>
<td>(2352959, 3321502)</td>
</tr>
<tr>
<td>2023-24</td>
<td>3002684</td>
<td>(2432472, 3572897)</td>
</tr>
<tr>
<td>2024-25</td>
<td>3168138</td>
<td>(2507449, 3828827)</td>
</tr>
<tr>
<td>2025-26</td>
<td>3333592</td>
<td>(2578107, 4089077)</td>
</tr>
<tr>
<td>2026-27</td>
<td>3499046</td>
<td>(2644632, 4353460)</td>
</tr>
</tbody>
</table>

The plot of the forecast values with the 95% confidence intervals is shown in figure 3.8
Discussion and Conclusions:-
The objectives of this study were to model the GDP at current market price data of Bangladesh and to forecast it for next twelve years using that model. To achieve the objectives, Box-Jenkins approach to ARIMA model has been used. The data was the GDP at current market price of Bangladesh for the years 1989-90 to 2014-15, Taka in core was used.

Initially, non-stationarity and dependency was found in the data by augmented Dickey-Fuller and the L-Jung-Box tests. This non-stationarity was removed first by taking first differencing and then second differencing. Using the model selection criterion AIC, ARIMA (0, 2, 1) model was selected. The parameter of this model has been estimated using the maximum likelihood method and all the parameters were found to be significant assuming normal distributions of the estimators. The assumptions made on normality and independence of the residuals was checked using different plots and tests and all the results were satisfactory. The plot comparing the actual values of the data and the fitted values using the model shows very much close fit. Then the model was used for forecasting purposes.

The forecast for the original series has been obtained by taking second differencing of the initial forecasts which were obtained for the first difference data. At last the forecasts for GDP at current market price of Bangladesh with the 95% confidence interval for twelve years along with the data has been shown in a plot and it is found that the upward trend in GDP at current market price will still continue for this period and our study conclude that Bangladesh will promote his position from developing countries to high income country with appropriate time or our forecasted time.

Limitations:-
In this study, the model has been established by using ARIMA approach only. Other types of time series models such as ARCH, GARCH and SV models were not considered. The study has been done based on the secondary data. So, the limitations of secondary data remain in the study.
Acknowledgement:-
This study is conducted based on secondary data of Gross Domestic Product (GDP) of Bangladesh. We have taken data from different editions of Statistical Yearbook published by Bangladesh Bureau of Statistics (BBS). So, we have to give a special thanks to the authority of BBS for providing the data.

References:-