Forecasting GDP of Bangladesh Using ARIMA Model

K. M. Salah Uddin¹ & Nishat Tanzim²

¹Professor, Department of Management Information Systems, University of Dhaka, Dhaka, Bangladesh
²Department of Management Information Systems, University of Dhaka, Dhaka, Bangladesh

Correspondence: K. M. Salah Uddin, Professor, Department of Management Information Systems, University of Dhaka, Dhaka, Bangladesh. E-mail: salahuddin@du.ac.bd

Received: March 11, 2021              Accepted: April 20, 2021         Online Published: May 5, 2021
doi:10.5539/ijbm.v16n6p56             URL: https://doi.org/10.5539/ijbm.v16n6p56

Abstract

Gross Domestic Product (GDP) is an important indicator of a country’s economic growth. To assist in decision making process, this study intends to forecast the GDP for seven years and identify the factors affecting the GDP in context of Bangladesh using the data from World Bank. The Autoregressive Integrated Moving Average (ARIMA) model is adopted to predict the GDP from 2019 to 2025 and multiple linear regression has been used to explore the factors affecting GDP. Different types of ARIMA (P, I, Q) tested and applied the ARIMA (1, 2, 1) model are found as best appropriate for forecasting. Q-Q plot, residuals plot, PACF and ACF graphs of the residuals are drawn for checking model adequacy. From this study, it is observed that GDP trend is steadily improving over years in Bangladesh that will remain expanding in the forthcoming.

Keywords: GDP, forecast, ARIMA model, multiple linear regression, Gross national expenditure

1. Introduction

GDP is the total value of all products and services produced within a country’s geographic boundaries during a certain period of time. For instance, the production of Japan owned company in the Bangladesh will be counted as portion of Bangladesh’s GDP.

The GDP of Bangladesh is undertaking a quicker rate of development in the recent years. GDP is one of the most significant indicators of national monetary activities for countries. Methodical forecast of the indicator has vital theoretical and practical significance on the development of economic growth goals. It also becomes core to understand the nature & direction of relationship between the economic development of a country and its mechanisms. It might be very helpful in formulating new policies and programs regarding economic policy and budget making.

Therefore, an effort is made in this report to forecast the GDP and identify factors affecting GDP in Bangladesh. In Bangladesh, the amount of research done so far related to GDP is not sufficient to know a country’s economic condition. Only one study is done which is also much backdated. So, it is essential to forecast GDP for the prediction of the future economy as well as the advancement of a country based on the current data. On the other hand, a significant change in GDP, whether up or down, usually has an impact on the market economy. Therefore, it is necessary to realize the associated factors that affect GDP.

Keeping attention the significance of studying the future trend of GDP and the factors affecting it, a lot of works have been done so far. There are many approaches that can be applied for macroeconomic estimating like linear regression, AR model, MA model, ARIMA model, VAR model, etc. The specific objectives to be investigated throughout this study are forecasting GDP of Bangladesh and finding the significant factors affecting GDP in Bangladesh.

2. Literature Review

Using data from Shaanxi GDP for 1952-2007, Ning, W. et al. (2010) predicted the country’s GDP for the succeeding 6 years. Employing the ARIMA (1, 2, 1) model their study revealed that the GDP of Shaanxi presents an moving increasing trend (Ning, Kuan-jiang, & Zhi-fa, 010). Based on GDP data records of Greece from 1980-2013, Dritsaki C. (2015) established ARIMA (1,1,1) to predict the real GDP rate for the years 2015, 2016 and 2017. Numerical results presented that Greece’s real GDP rate is gradually increasing (Dritsaki, 2015).

Maity B. and Chatterjee B. (2012) examined the projecting of GDP growth rate for India applying ARIMA (1, 2,
2) model and period of time for 60 years. Their study exhibited that forecasted values for the following years is an increasing trend (Maity & Chatterjee, 2012).

M. S. Wabomba et al. (2016) developed a class of ARIMA models and ARIMA (2, 2, 2) model was identified as the best for forecasting the Kenyan GDP based on the recognition rules and stationary test of time series under the AIC criterion. Their investigation results indicated that the predicting effect of this model was relatively adequate and practiced in modeling the annual Kenyan GDP. Finally, they applied the fitted ARIMA model to predict the GDP of Kenya for the following five years (Wabomba, Mutwiri, & Mungai, 2016).

D. Jain et al. (2015) examined the influence of numerous macroeconomic issues on GDP components of India. In this study, the dependent variable GDP components which was expressed in terms of several macroeconomic measures of growth. Multiple regression analyses were applied to develop the relationship. In this analysis, their showed a major relation of FDI, Net FII equity and Import on GDP component (Jain, Nair, & Jain, 2015). A. R. Kira (2013) intended to investigate the issues that affect the GDP of Tanzania. The findings of this study are, Tanzanian GDP is affected by Exports and consumption (Government Expenditure and Household Expenditure) (Kira, 2013).

In the study of S. Anam & M. Hossain (2012), they used ARIMA and GARCH model and compare the performance of GDP forecasting. Their investigation presented that the result from ARIMA model is better than the GARCH model provides. From their forecasted result, it was observed that the GDP of Bangladesh had an increasing trend. They considered the data of 1971. At that time, liberation war was taking place and there was a shocking movement in GDP. That’s why the results of forecasted value can’t be properly fitted as the ARIMA model was inappropriate in this situation (Anam, & Hossain, 2012).

In Bangladesh, there has been minimal empirical work performed that looks into factors that affect GDP. S. Mehmood (2012) studied the impact of thirteen selected issues (independent variables) on Gross Domestic Product (GDP) in the economy of Pakistan and Bangladesh. Their study revealed that issues such as external debts stock total, gross national expenditures, imports, and exports have a positive impact on the GDP of Bangladesh but the issues as final consumption expenditure shows a negative impact on the Bangladesh’s GDP (Mehmood, 2012).

Nine years of quarterly real-time data which was seasonally adjusted GDP of Sweden has been used for the purpose of evaluating forecast models by Pär Österholm (2014). Their found that the survey data is useful to improve the forecasts (Österholm, 2014). R. In the study of Bhattacharyaa et al. (2019) used the data during a period of time from 1980-1981 to 2016-2017 to represent the forecasted value of GDP of India for the year 2017-2018. They used a principal component augmented Time-Varying Parameter Regression (TVPR) approach and projected the model using a mix of fiscal, economic, trade, and creation side-specific variables (Bhattacharya, Chakravartti, & Mundle, 2019).

T. Liu et al. (2018) forecasted Chinese GDP using the government data are derived from the National Bureau of Statistics of the People’s Republic of China during the time period 2005-2014. Their study focused macroeconomic predicting model and proved the impact of the two-step technique. A simple single linear equation method is used here. To predict, an intuitive approach is to place all explanatory variables into the model equally, employ a dimension reduction method, and then select explanatory variables with the best forecast ability to determine the forecast model (Liu, Xu, & Fan, 2018).

A. F. Schiff et al. (2000) used recent time series technique to forecast the GDP of New Zealand. In their study, the prediction of New Zealand real GDP produced by the AR Model and VAR model is explored to be competitive with predictions from fixed-format models and predictions created by the NZIER (Schiff, & Phillips, 2000). Miah et al. (2019) forecasted the GDP of Bangladesh and used the data for the period of time from 1960 to 2017. The time series plot of GDP demonstrations a non-stationary arrangement which have been converted into stationary by taking the data difference twice. Using model selection criterion and checking model adequacy and based on the ACF and PACF a time series model ARIMA (1, 2, 1) was chosen. Using the model, it is observed that the values of GDP forecast in Bangladesh are gradually increasing over the next thirty years (Miah, Tabassum, Rana, 2019).

N.M.F Rahman (2010) conduct research to estimate growth pattern and also observe the best ARIMA model that forecast rice (Aus, Aman and Boro) production in Bangladesh efficiently. Their investigation indicated that the short-term forecast from ARIMA model were well-predicted compared to the deterministic models. The uncertainty of rice production could be reduced if production were forecasted well and necessary actions were taken against damages (Rahman, 2010).
J.C. Paul et al (2013) tried to determine the suitable ARIMA model and used the average market price of data series for Square Pharmaceuticals Limited for forecasting. To check the stationary condition of the data series were monitored by ACF and PACF plots, then it is also observed by using other statistical methods. It is identified that that the average market price indices of the data series are non-stationary. Even after log-transformation, the average market price indices for the data series are found non-stationary. Finally their investigation showed that ARIMA (2, 1, 2) is the appropriate model for their forecasting (Paul, Hoque, & Rahman, 2013).

P. C. Padhan, (2012) had employed ARIMA model to predict the annual output of particular agricultural product. For his study, a set of 34 various products has been measured, depending upon accessibility of required data. The rationality of the model is confirmed with several model selection principles such as Adj R2, lowest of AIC and minimum MAPE values. Therefore, estimating output of agricultural crops is not only tedious but also indispensable, as large chunk of people be contingent on agriculture for their livelihood. Numerous uni-variate and multi-variate time series methods can be used for predicting such variables (Padhan, 2012).

N.M.F Rahman et al (2016) tried to forecast the production area of black gram pulse in Bangladesh data was collected from Bangladesh Bureau of Statistics over the period of 47 years. To determine the model adequacy Ljung-Box Q is used as a test statistic while for checking the normality of model it is used Jarque-Bera test. Based on AIC and BIC criterion the ARIMA (0, 1, 0) modes was found as best for forecasting the black gram pulse production (Rahman & Baten, 2016). S.H. Bari et al. (2015) predicted the long term precipitation in Sylhet with a seasonal ARIMA model that was developed using Box and Jenkins method. For this objective precipitation data was collected from Sylhet station over a period of years from 1980-2010. The developed model was verified using the data from 2007-2010 (Bari, Rahman, Hussain, & Ray, 2015).

Hossain, M. and Abdulla, F., (2015) carried out the research to recognize an appropriate ARIMA model that could be applied to predict the output of tea in Bangladesh. In this investigation considered the data collected over the period 1972 to 2013 of yearly tea production. Based on the minimum values of AIC, AICC and BIC, it is selected the best suitable model to predict the tea output in Bangladesh is ARIMA (0,2,1). The model adequacy is determined by using Run test and Jarque and Bera test standards (Hossain, & Abdulla, 2015).

Khan and Rahman (2013) investigated the influence of industry to the gross domestic product of Bangladesh (GDP) which is a vital concern in economics. It plays major role in the government and industrial sector of Country. A time series model can provide a reasonable benchmark to evaluate the value added of economic theory relative to the pure explanatory power of the past behavior of the variable; recent developments in time series analysis suggest that more sophisticated time series models could provide more accurate benchmarks for economic models. In this article the data was collected over the fiscal year from 1979 to 2012 on the industrial impact to GDP to be examined (Khan & Rahman, 2013). V. L.C. Rahman et. al (2019) used ARIMA model and exponential smoothing technique to investigate the stationary character of the data and to forecast the GDP growth of Bangladesh. The ARIMA (1,1,1) model was selected by the criterion of the lowest values of AIC and BIC (Voumik, Rahman, Hossain, & Rahman, 2019).

3. Research Methodology

In this study, the secondary data was used and collected over the period 2001-2018 from World Bank Data Bank. For predicting GDP of Bangladesh over the period of 2019 to 2025, ARIMA model is used for forecasting. For identifying the factors which affects GDP, regression analysis method is used. GDP was considered as dependent variable which was expressed as a function of several macroeconomic measures of growth. Such variables could be employment, gross national expenditure, final consumption expenditure, gross domestic savings, net exports of goods and services and foreign direct investment. Multiple regression analysis was used to develop the correlation.

4. Analysis and Results

4.1 Model Development

To stabilize the variability, the GDP is plotted for main datasets, log transformation and square root transformation of the data.

The plots are given in the following:
In Figure 1 above, the log transformation graph looks more linear. So, if we fit the time series with log transformation of the data, it will give better results.

**Exploring d Value:** Difference is taken to transform non-stationary data into a stationary time series. Here in this study, first, second and third differences are taken. From the Figure 2a, 2b and 2c, it can be seen that, the first and second graph looks more stationary than the third one. So, these plots suggest for d=1 or d=2.

**Exploring p and q for d=1:** Here, to explore the value of p and q for d=1, the example ACF and PACF of the first differenced data are plotted. It can be realized that the sample ACF and PACF do not cut off the blue boundary. So the plots suggest, p=0 and q=0 for d=1. So, the proposed model is ARIMA (0,1,0).

**Exploring p and q for d=2:** Here, to explore the value of p and q for d=2, the example ACF and PACF of the first differenced data are exhibited. It can be observed that the sample ACF and PACF cut off the blue boundary at lag 1. So the plots suggests p=1, q=1 for d=2. So, the proposed model is ARIMA (1,2,1). Based on the root mean square error, ARIMA (1,2,1) that has lowest RMSE. So, best model for predicting the data is ARIMA (1,2,1).

**4.2 Model Adequacy Checking**

For checking model adequacy, Q-Q plot, residuals plot, ACF and PACF graphs of the residuals are drawn.
Figure 3 illustrates the normal Q-Q plot, the points seem to fall about a straight line. So, it can be said that the model residuals is normally distributed. From the residuals plot, there is not any systematic pattern. So, it can be said that residuals are independently distributed. From ACF plot and PACF plot, it can be seen that, at any lag residuals do not have any significant spike. It indicated the lag independency of residuals. As, the fitting results of the model is convincing and residuals of the fitted model satisfies all the condition. Now, the model can be used for forecasting.

4.3 FIT ARIMA (1,2,1) for Full Data and Forecast Ahead

On the basis of the developed model, the forecasted value of GDP from 2019 to 2025 are given below:

Table1. Forecasted values of GDP

<table>
<thead>
<tr>
<th>Year</th>
<th>Forecasted Value of GDP (in US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019</td>
<td>304858245644</td>
</tr>
<tr>
<td>2020</td>
<td>339839216116</td>
</tr>
<tr>
<td>2021</td>
<td>378925613374</td>
</tr>
<tr>
<td>2022</td>
<td>422520440610</td>
</tr>
<tr>
<td>2023</td>
<td>471132614695</td>
</tr>
<tr>
<td>2024</td>
<td>525338015260</td>
</tr>
<tr>
<td>2025</td>
<td>585779966928</td>
</tr>
</tbody>
</table>

Table 1 consists of the forecast values of GDP. By using the developed models, the GDP growth rates are predicted. Projected results exhibit that Bangladesh’s GDP growth rate is a growing trend that will stay expanding in the forthcoming. This results will assist decision makers and academicians to frame economic and business policies more specifically.

The real GDP (from 2001-2018) and forecasted GDP (from 2019-2025) are plotted below:

![Figure 4. Real and forecasted value of GDP](image)

From Figure 4, it is clearly shown that the real GDP (from 2001-2018) and forecasted GDP (from 2019-2025) are plotted above. In the above Figure 4, the GDP growth sequences typically follows an increasing trend. This suggests that both the mean and the variance are changing. It follows a non-stationary shape.

4.4 Checking Linearity between Dependent and Independent Variables

First, multiple linear regression needs the association between the dependent and independent variables to be linear. The linearity hypothesis can be confirmed with scatterplots.
Figure 5 illustrates the fitting every independent variable individually to see the association between GDP and independent variables.

4.5 Fitting Each Independent Variable Separately

To see the relationship between GDP and independent variables, simple linear regression is done to fit each variable separately.

Table 2. Results of the fitted models

<table>
<thead>
<tr>
<th>Variable</th>
<th>β Estimate</th>
<th>P value for β Estimate</th>
<th>P value for F test</th>
<th>R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment</td>
<td>2.5e+10</td>
<td>.335</td>
<td>.335</td>
<td>.058</td>
</tr>
<tr>
<td>Gross National Expenditure</td>
<td>.932</td>
<td>.001</td>
<td>.001</td>
<td>.999</td>
</tr>
<tr>
<td>Net Export</td>
<td>-11.800</td>
<td>.001</td>
<td>.001</td>
<td>.856</td>
</tr>
<tr>
<td>Final Consumption</td>
<td>1.334</td>
<td>.001</td>
<td>.001</td>
<td>.998</td>
</tr>
<tr>
<td>Gross Domestic Savings</td>
<td>3.954</td>
<td>.001</td>
<td>.001</td>
<td>.988</td>
</tr>
<tr>
<td>Net Foreign Direct Investment</td>
<td>-62.83</td>
<td>.001</td>
<td>.001</td>
<td>.809</td>
</tr>
</tbody>
</table>

Note. e+10 = $10^{10}$

4.6 Interpretation for β Estimates

Table 2 indicates the interruption for β estimates.

Employment: If employment rate is increased by 1% (in total), GDP will increase on average 2.5e+10 US$.

Gross National Expenditure: If gross national expenditure is increased by 1 US$, GDP will increase on average .932 US$.

Net Export: If net export is increased by 1 US$, GDP will decrease by on average 11.8 US$.

Final Consumption: If final consumption is increased by 1 US$, GDP will increase by on average 1.334 US$.

Gross Domestic Savings: If gross domestic savings is increased by 1 US$, GDP will increase by on average 3.954 US$.

Net Foreign Direct Investment: If net foreign direct investment is increased by 1 US$, GDP will decrease by on average -62.83 US$.

4.7 Interpretation of P Value for β Estimates

At 5% level of significance employment is not a significant factor for affecting GDP as p = .335 > .005.

For Gross National Expenditure, net export, final consumption, gross domestic savings, net foreign direct investment, p value is less than .05. They are highly significant factors affecting GDP in Bangladesh.
4.8 Interpretation of \( R^2 \)

From the table, it seems that except employment other variables captures 80 to 99.9 percentage of total variation of GDP. As, gross national expenditure captures 99.9 percentage of total variation of GDP, it provides the evidence that gross national expenditure should be included into the model.

From the value of \( R^2 \), it can be assumed that, there exists high multicollinearity between the independent variables.

4.9 Checking Multicollinearity

For checking multicollinearity, the correlation matrix is given below:

### Table 3. Correlation between the factors

<table>
<thead>
<tr>
<th>Employment</th>
<th>Gross national expenditure</th>
<th>Net export</th>
<th>Total consumption</th>
<th>Gross domestic savings</th>
<th>Foreign direct investment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross national expenditure</td>
<td>-0.097</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net export</td>
<td>0.260</td>
<td>-.928</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total consumption</td>
<td>-0.116</td>
<td>.999</td>
<td>-.935</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Gross domestic savings</td>
<td>0.013</td>
<td>.991</td>
<td>-.873</td>
<td>.987</td>
<td>1</td>
</tr>
<tr>
<td>Net foreign direct investment</td>
<td>0.377</td>
<td>-.905</td>
<td>.928</td>
<td>-.912</td>
<td>-.860</td>
</tr>
</tbody>
</table>

From the Table 3, it can be seen that, without employment, each variable is highly correlated with each other. This data produce extremely high multicollinearity. If all variables are kept in the model, it will violate the assumption of linear model that there is a little or no multicollinearity in the independent variables. As gross national expenditure should be included into the model (because of high \( R^2 \)), other variables which are highly correlated with this must be excluded into the model because of multicollinearity. Correlation between gross national expenditure and net export, gross national expenditure and total consumption, gross national expenditure and gross domestic savings, gross national expenditure and net foreign direct investment are -.928, .999, .991, -.905 respectively which are quite high. So, except employment, other variables can’t be included when gross national expenditure must be need to include into the model. Projected results exhibit that Bangladesh’s GDP growth rate is a growing trend that will stay expanding in the forthcoming.

4.10 Multiple Linear Regression Model

To see the association between GDP and independent variables, multiple linear regression is done to fit all variables.

### Table 4. Results of the multiple linear regression model

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \beta ) Estimate</th>
<th>P value for ( \beta ) Estimate</th>
<th>P value for F test</th>
<th>R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment</td>
<td>3.993e+08</td>
<td>.065</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross National Expenditure</td>
<td>1.111e-01</td>
<td>.650</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net Export</td>
<td>-1.165e-02</td>
<td>.940</td>
<td>.001</td>
<td>1</td>
</tr>
<tr>
<td>Total Consumption</td>
<td>8.660e-01</td>
<td>.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross Domestic Savings</td>
<td>9.190e-01</td>
<td>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foreign Direct Investment</td>
<td>-1.649e-01</td>
<td>.437</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. \( e+08 = \times 10^8 \)*

Here, in Table 4 exhibits the value of \( R \) square is 1. It shows that the model clarifies all the variablity of the response data about its mean. In that case, the fitted values equal the data values and, accordingly, all of the observations fall closely on the regression line. So, it can be said that the model is over fitted. Thus, overfitting a
regression model reduces its generalizability outside the original dataset. R square becomes high when there is multicollinearity. Already it is observed that, between the independent variables, there exists extremely high multicollinearity. So, the model with all variables would give misleading regression coefficients and p-values.

4.11 Final Model

For high multicollinearity, only two variables are kept into the model, employment and gross national expenditure.

Our Hypotheses for \( \beta \) estimates are:

1. \( H_01: \) Employment does not have any effect on GDP  
   \( H_a1: \) Employment has an effect on GDP
2. \( H_02: \) Gross National Expenditure does not have any effect on GDP  
   \( H_a2: \) Gross National Expenditure has an effect on GDP

Table 5. Results of the final fitted model

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \beta ) Estimate</th>
<th>P value for ( \beta ) Estimate</th>
<th>P value for F test</th>
<th>R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment</td>
<td>-2.008e+08</td>
<td>.777</td>
<td>.001</td>
<td>.999</td>
</tr>
<tr>
<td>Gross National Expenditure</td>
<td>.932</td>
<td>.001</td>
<td>.001</td>
<td>.999</td>
</tr>
</tbody>
</table>

4.12 Interpretation:

From the Table 5, it seems that gross national expenditure captures 99.9 percentage of total variation of GDP. From \( \beta \) estimate, it can be said that, if gross national expenditure increases by 1 US$, on an average GDP increases by .932 US$.

From the p value from \( \beta \) estimate, it can be seen that, for employment p>.05. So at 5% level of significance, it rejects the null hypothesis. Employment is not a significant factor affecting GDP in Bangladesh. On the other hand, for national gross expenditure, p=.001 which is less than .05. So at 5% level of significance, gross national expenditure is highly significant. It is an important factor affecting GDP.

From the above F test it can be concluded that our model is valid to capture the variation of GDP as p<.05.

4.13 Model Adequacy Checking

Model adequacy checking can be done by plotting a normal Q-Q plot. The Q-Q plot is given below:

![Normal Q-Q Plot](image)

From the Figure 6, it can be seen that the Normal Q-Q plot maximum points falls on or near the line. So, the errors are normally distributed which is one of the assumptions of linear regression.

5. Conclusion

One of the objectives of this study is to forecast GDP in Bangladesh. Findings of the study shows a higher growth trend of GDP in the next 7 years from 2019 to 2025. Another objective of this study is to recognize the issues affecting GDP of Bangladesh. From the results, it has been found that net export and foreign direct
investment are negatively correlated with GDP. In Bangladesh, imports are greater than exports and so, net exports are negative. If net exports are negative, the nation has a negative trade balance. Also, net FDI is decreasing day by day as net liabilities are greater than net assets.

On the other hand, GDP increases if employment (% of total population), gross national expenditure, total consumption and gross domestic savings increases. These factors are important for a growing economy and development of a country. Government has to investigate this issue and undertake necessary measures to increase net export and net FDI of Bangladesh.

References


**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).