A statistical study of analysis of India Gross Domestic Product

Aishwarya G. Desai*

Student, Department of Statistics Sadguru Gadage Maharaj College Karad

Email:desaiaishwarya241@gmail.com

Abstract:

In today's vision of 2023 it is seen that many of the businesses, companies, manufacturing industries and agricultural welfares, etc. are the main factors which leading to the growth of economy. The Gross Domestic Product (GDP) is the monitory value of all finished goods and services made within a country during a specific period and it tells us how's the performance of a country's current aggregate production of goods and services and based on that, GDP is often considered the best measure of how well the economy is performing. This paper is work of model implementation of the time series data which is GDP of India at constant price and in forecasted models it gives snapshot of an Indian economy with its sub sequential development. This study deals with estimation of various time series models like ARIMA, ARCH and GARCH.

Keywords: GDP, Volatility, Prediction, ARCH, GARCH, ARIMA model.

1. Introduction:

GDP is long term process which involves trend phenomena also GDP count influences with various economical conditions as well as factors. Basically for long period of time, various changes will occur in volatility with a hike in probability. These factors affects the growth of GDP and in result it shows a bad impact on countries economy. At current time, the market and various industries shows volatileness with huge trends. Volatility is nothing but a condition during which the conditional variance undergoes various changes between rabidly high and low values. Basically for long period of time, various changes will occur in volatility with a hike in probability.

Although India is the countries which survive from many of the financial crisis hence, it should be more alert of the economic changes. Since economic change have the biggest contributor to GDP. In the current study, the volatility of the past GDP Properties and its

nature have been studied. For the study purpose yearly data have been used from 1950-51 to July 2021-22

For studying the data, here two time series models are preferred. From these models one can work on the non-constant volatility and that model is Autoregressive Integrated Moving Average (ARIMA) and another one is Generalized Autoregressive Conditional Heteroscedasticity (GARCH) which basically used for predicting volatility. These models will be used for fitting the data where the best model will be used to forecast the future of forecasting GDP count for next 10 years.

Researcher frequently undergoes some form of a generalized autoregressive conditional heteroscedasticity (GARCH) modelling strategy for examining the volatility of real GDP growth. And later it employ with a generalized autoregressive conditional heteroscedasticity (GARCH) modelling strategy for examining the volatility effect of GDP growth, assumes a consistent GARCH or exponential GARCH (EGARCH) process.

We have the Autoregressive integrated moving average (ARIMA) model which is the most important forecasting tool used in time-series analysis and sometimes ARIMA is also called as Box-Jenkins models. The perfect fitted ARIMA model of all GDP parameters depend on the minimum count of Akaike information criterion (AIC) as well as minimum count of Bayesian information criterion (BIC) for measuring the performance actions of the creased model. Box and Jenkins (1976) proposed methodology which has four stages as: i) Model recognition; ii) Assessment of model factors iii) Investigative examination for recognized model and iv) Forecasting. The main disadvantage of the ARIMA modelling is that it produces the chronological sequences over study from linear process and hence, it is unable to capture nonlinear patterns.

Here we get an evidence in output growth volatility that it has several structural changes. And it is combined with the high conditional volatility which is why it motivates us to revisit the issue of conditional volatility in GDP growth rates for India.

The objectives of study are as follows:

• To study trend of GDP.

•. To study stationarity behaviour of GDP data

• To develop ARIMA model on GDP data.

• To forecast GDP values for next 10 years.

2. Review of Literature:

Forecasting is very sensitive area. Many of the researchers, scientist have been worked over it by using different models and methods. I limit the review of literature as follows:

Varun Agrawal et al. [1] studied the model of time series of real GDP of Indian economy and developed the forecast model to shed light on underlying data generation process.

Kharisya Ayu Effendi [2] determined the best ARCH / GARCH model in JKSE and stock index in developed countries (FTSE, Nasdaq and STI) and then compare the JKSE with the stock index in developed countries (FTSE, Nasdaq and STI). The results obtained in this study is to determine the best model of ARCH / GARCH.

Nor Hamizah Miswan et al. [3] used two time series models which are Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heterocedasticity (GARCH) models in modelling and forecasting Malaysia property market. The capabilities of ARIMA and GARCH models in modelling and forecasting Malaysia property market will be evaluated by using Akaike's Information Criterion (AIC), Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). They concluded that Box-Jenkins ARIMA model perform better compared than GARCH model in modelling and forecasting Malaysia market properties and shares.

Palvi Jearth et al. [5] investigated the impact of Inflation and Population on Gross Domestic Product (GDP) of India. The analysis has been carried out with the help of correlation, regression analysis, t-test and ANOVA model.

Dr. Konda Hari Prasad Reddy [6] analyzed and comparing the data with regard to the contribution of the major sectors of Indian economy.Moreover, the statistical tools and tests like correlation analysis, analysis of variance and F-test have been used.

Shivani Gupta [8] This paper tries to compare Indian economy in pre and post reform period and analyses how the economic reforms of 1991 in India have affected the GDP growth rate and how it has impacted the contribution of several economic and non-economic factors towards GDP growth rate determination.

3. Data source:

The dataset which is used for the study has been collected from the following government website: https://www.indiabudget.gov.in/economicsurvey/doc/Statistical-Appendix

Which consists of 72 records, and each record has 9 attribute values (features), in which

1. Year
2. PFCE: Private Final Consumption Expenditure
3. GFCE: Government Final Consumption Expenditure
4. GFCF: Gross Fixed Capital Formation
5. CIS: Change in Stocks
6. Valuables
7. Export of goods and cervices
8. Import of goods and services, these are independent variables, and
9. GDP is dependent variable

4. Methodology:

ARIMA Model:

The data used in this research is secondary time series data. Hence our main focus is to build suitable ARIMA model.The model used is nothing but the technique of Box-Jenkins and this Box-Jenkins models is commonly known as Autoregressive Integrated Moving Average model (ARIMA). Box and Jenkins have popularized this method. This ARIMA method consists of main three stages for selecting a suitable model for the purpose of estimation and forecasting of univariate time series data, the model identification, parameter estimation, and forecasting. ARIMA model is developed such that it is a combination of autoregressive (AR) and moving average (MA) models. And both these models work upon the assumption that the data should be analyzed and it must moves along a constant average that is stationary. Incase, the data is not stationary, then for model building it is must to make it stationary using the log and differentiation processes. ARIMA is a combination of AR and MA models through

different processes and it have time slackness. In differencing inaction, the first time period in a process is known as autoregressive first order or autoregressive and it is abbreviated as AR (1). The symbol used for the notation of the number of inaction at the specific time of the autoregressive process is “p”. And the first time period in a process is called as moving average of first-order and it could be abbreviated as MA (1). The symbol for the number of inaction at the time when the moving average process is q. and it is possible that the values of p and q can be more than 1. Differencing process used before the ARIMA models which is aimed at obtaining data to be stationary. This process can be performed once or may be done more than once until the data is stationary. Usually, different process is not more than 2 times. Different process the data symbol is d. Writing ARIMA models for AR (p), MA (q), and different times as much d is ARIMA (p, d, q).

A nonseasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of nonseasonal differences needed for stationarity, and
- **q** is the number of lagged forecast errors in the prediction equation.

The forecasting equation is constructed as follows. First, let y denote the d^{th} difference of Y, which means:

If d=0: $y_t = Y_t$

If d=1: $y_t = Y_t - Y_{t-1}$

If d=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$

ARCH/ GARCH Model :

The GARCH model works on the property of a constant residual variance or homoskedasticity which are mainly used in the process of modeling of time series data. But in real life scenario, the residual variance is not constant for most of the times othat is heteroskedastisity will be present in many time series data, especially in the the sectors like finance it would be present. This lead to the assumption of homoskedasticity modeling cannot be used. Hence in such situation the ARCH models or Autoregressive Conditional Heteroskedasticity model can be more popularly used to analyze the time series data which

contains heteroskedasticity. Moreover, this model can be used after the proof that data have auto regression which is stated to be presence of heteroscedasticity which is nothing but an unequal variance of residual term. If an Autoregressive moving average (ARMA) model is assumed for the error variance, the model is a generalized autoregressive conditional heteroskedasticity (GARCH) model . The GARCH model is a popular extension of an ARCH model, in which an autoregressive moving average (ARMA) model, rather than autoregressive (AR), is used to model the variance of the time series (Engle, 1982; Bollerslev, 1986)

Data Pre-Processing:

Government of India publishes every year the data on economic variables in Economic Survey.The data used in this research is secondary time series data. In this study, our main focus is to study and forecast GDP. In order to achieve this we have used Time series techniques to obtain useful information from the raw data and convert raw data into a format that is suitable for analysis and modelling. The data pre-processing task can be summarized as follows:

1. Data Cleaning:
   Data cleaning is one of the important processes involved in data analysis, which is being the first step after data collection. The Dataset which is being used for the study was available in a raw format hence it is needed to clean and work upon it.
2. Missing values:
   If the dataset used for classification contains more missing values then we may end up building a biased model hence it is important to handle them. In our study, we deal with columns containing missing values in a different way depending on the condition such as, In the dataset there was a column named as Valuables which majorly had na values which was not possible to fill information, hence in this phase removing the column, this cleaning technique is used in order to ensure that the dataset is free from model.
3. Stationarity check :
   The first task is to determine suitable parameter sets for the ω, α, β which ensure that σ 2 t has finite expected value or higher moments. Another consideration which will be important when studying the asymptotic properties of ARIMA and GARCH models is whether σ 2 t converges to a stationary distribution. Unfortunately, we will

see that these conditions translate to rather severe restrictions on the choice of parameters.

Definition : A process Xt is called stationary (strictly stationary),

If for all times t1, ..., tn, h ∈ Z: FX(xt1+h, ..., xtn+h) = FX(xt1 , ...xtn)

where FX(xt1 , ..., xtn) is the joint cumulative distribution function of Xt1 , ..., Xtn

Test for checking stationarity:

1. **Augmented Dickey-Fuller (ADF) t statistic test for unit root (ADF test):** In ARIMA model the first step is to determine the number of differencing required to make the series stationary. The ADF test belongs to a category of tests called 'Unit Root Test', which is the proper method for testing the stationarity of a time series. Unit root is a characteristic of a time series that makes it non-stationary. While, technically speaking, a unit root is said to be exist in the time series if the value of alpha = 1. The null and alternative hypothesis for ADF test are:

   H0: Time series is non stationary / Time series displays unit root i.e. (b1=0)

   H1: Time series is stationary / There is no unit root in the time series i.e. (b1<0)

2. **KPSS test: Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test:** KPSS tests are also used for testing a null hypothesis that an observable time series is stationary around a deterministic trend. Additionally, in the KPSS test, the absence of a unit root is not a proof of stationarity. But by design, of trend-stationarity. This is an important distinction since it is possible for a time series to be non-stationary, have no unit root yet be trend stationary. In both unit root and trend-stationary processes, the mean can be growing or decreasing over time; however, in the presence of a shock, trend-stationary processes are mean-reverting (i.e. transitory, the time series will converge again towards the growing mean, which was not affected by the shock) while unit root processes have a permanent impact on the mean. The null and alternative hypothesis for ADF test are:

   H0: Time series is trend stationary.

H1: Time series is not trend stationary.

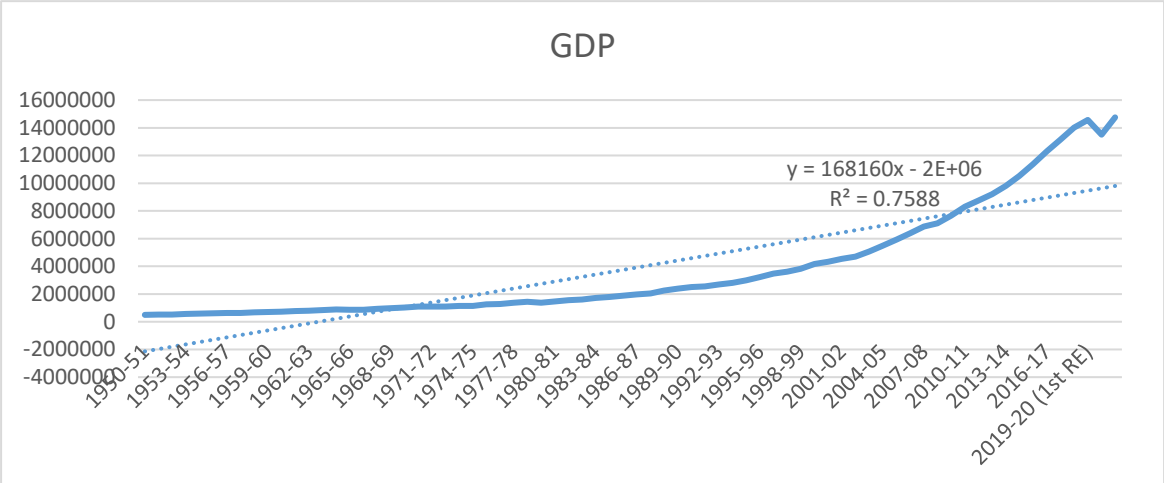
4. Results And Discussion:

Time series analysis starts with following steps:

1) Trend of time series:

Basically Trend is the pattern or forms in data which shows the movements of a process or series to respectively variant to very long period of time. We can say, a trend is observable when there is any increasing or decreasing patterns or slopes are present in the series. And it is noted that trends usually observable for some time and then disappears, it does not repeat all the time.

For GDP data the trend line is:



2) Stationarity check:

Stationarity is the most important concept in every time series analysis. Stationarity reffers to the concept that the statistical properties of a a time series are constant over time. And it is very useful because our results rely on it. Here we have two test to do this task:

1.Augmented Dickey-Fuller Test

In ADF test our p-value = 0.99 which is greater than the 5% level of significance.
hence we do not reject our null hypothesis i.e. GDP time series is not stationary

2.KPSS Test for Trend Stationarity

In KPSS test our P_value =0.01 which is less than the 5% level of significance.
hence we reject our null hypothesis ie. GDP time series is trend stationary

3) **Plot of the results to see non stationary behaviour of GDP data:**

Plot: Non stationary behaviour of GDP data

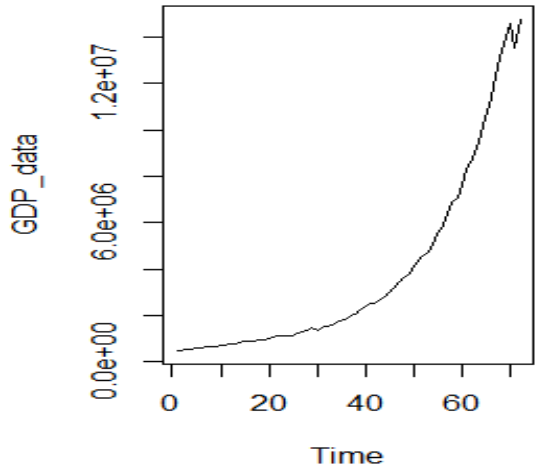


Fig.1 Non stationary behaviour of GDP data

From above fig. we conclude that time series has increasing trend with no seasonal variation.

4) **Making non stationary GDP data as stationary by method of log and differencing:**
Log transformations is used because logarithms helps to stabilise the variance of a time series. Then we take first difference of series and again check for the stationarity of the data. If stationarity is present then we will move further else we will perform second difference of the series to make process stationary. After we came to the result that series is stationary we need to check whether the variance is constant or not.
5) **Variance Ratio test to check for the constant variance:**

The hypothesis of variance ratio test are as follows:

Ho : variance is constant

H1: variance is not constant

Here for the series stats value=-2.929738, which is significant hence we accept our H0 ie. var iance of the series constant

Stationarity plot :

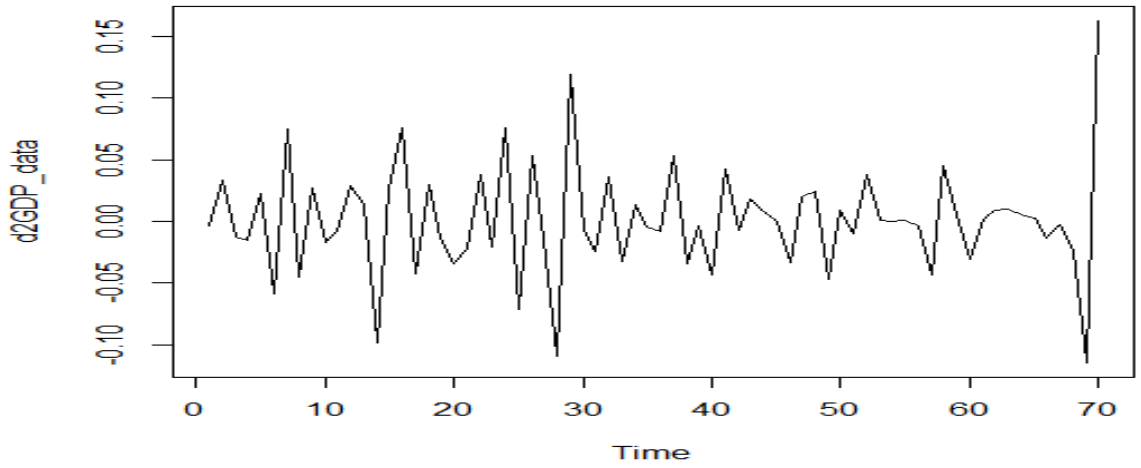


Fig.2 Stationary behaviour of stationary data

6) Identifying AR and MA using ACF and PACF Plots:

ACF plot:

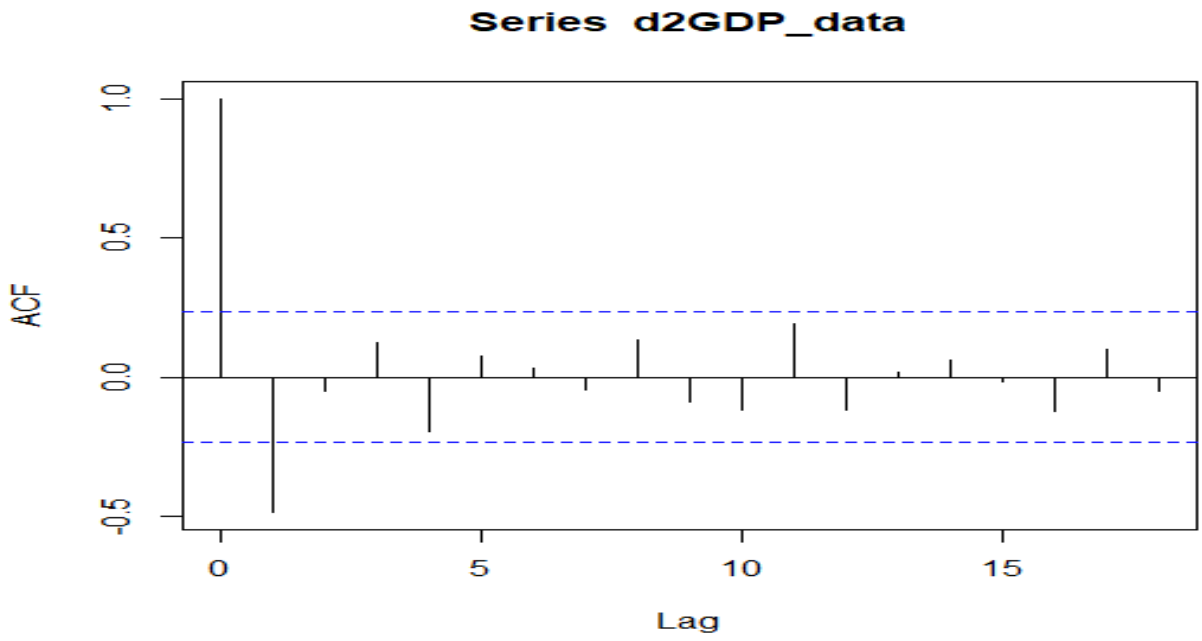


Fig.3 ACF correlogram of stationary GDP series with lags

PACF plot:

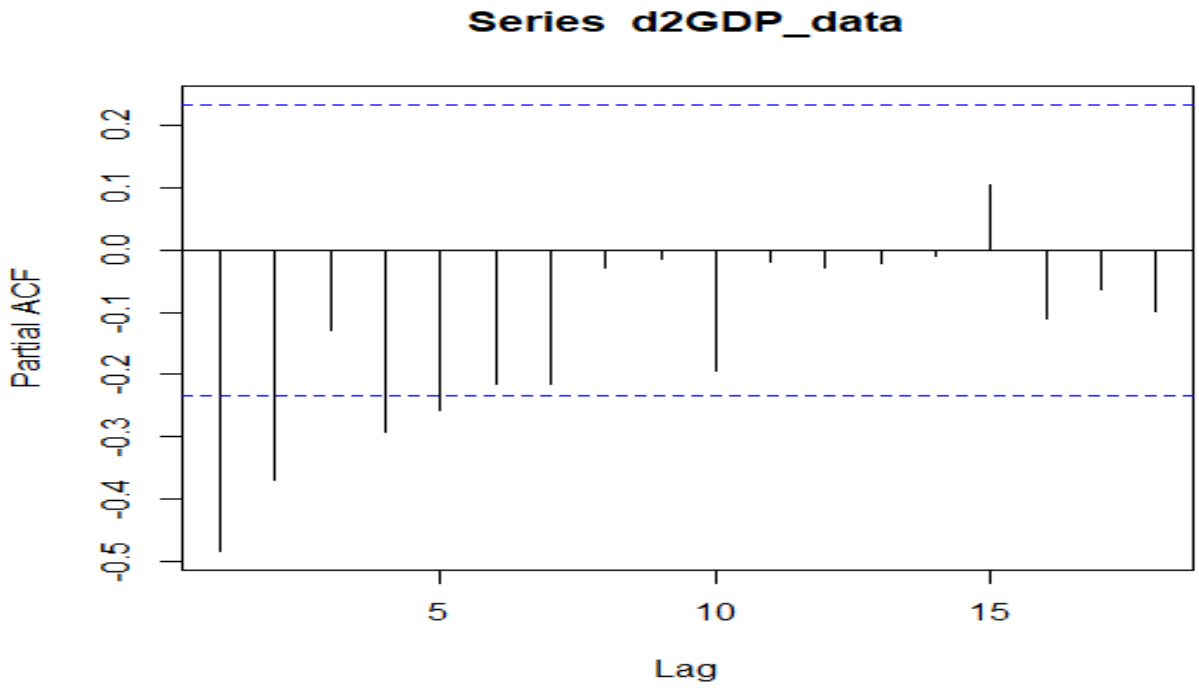


Fig.3 PACF correlogram of stationary GDP series with lags

7) ARIMA model:

Table 1. Different ARIMA(p,d,q) models for Forecasting comparisons for GDP parameters with minimum AIC measures

Model	AIC
ARIMA(0,1,0)	-158.92
ARIMA(1,1,0)	-193.08
ARIMA(2,1,1)	-253.38
ARIMA(1,0,0)	-253.49
ARIMA(0,0,1)	-281.46
ARIMA(1,2,1)	-181.95
ARIMA(1,1,2)	-267.31
ARIMA(1,0,2)	-279.23
ARIMA(2,2,2)	-236.84

Result of best fit model:

Series	Best Model	MAPE	RMSE
dGDP_data	ARIMA(0,0,1)	5490.81	0.0255666

Interpretation of ARIMA(0,0,1) :

1. Here the PACF value is 0 i.e. p=0
   Basically ir means that the lags value in the PACF graph are not to be significant and it can be interpreted as trend components does not play a significant role for the given time series data.

2. d=0

   It indicates that the given time series data is stationary and it does not need check nonseasnal differences.

3. q=1

   It indicates that ACF value is 1. Which is nothing but the lag after which the ACF value will becomes a zero in the graph is 1. It suggest that our time series has a randomness component and the ACF after the lag 2 becomes 0. It is because of the reason that the first small pick in ACF plot is because of the correlation of the variable with itself.

5. **Conclusion** :

From the above tables, the lowest AIC values are considered to be the best model for modelling properties and GDP data. Hence, it can be concluded that ARIMA (0, 0, 1) is the best model for GDP data where the AIC, MAPE and RMSE values are -281.46, 5490.81 and 0.0255666 respectively. And the equation for ARIMA(0,0,1) is given as:

GDP=0.0001-0.9289et-1

6. Future scope of this study:

From above result we got the ARIMA(0,0,1) model and later after performing the ArchTest to see the ARCH effect we have got the presence of ARCH in the GDP dataset.

Hence the study can continued to fit ARCH and GARCH models on the GDP data for forecasting the series of next 10 years.

7. References:

1. GDP modelling and forecasting using ARIMA: an empirical study from India
   By Varun Agrawal; Budapest, Hungary (2018)

2. Determining the best ARCH/GARCH model and comparing JKSE with stock index in developed countries Kharisya Ayu Effendi Faculty of Business and Management in Widyatama University, Bandung, Indonesia(2014)

3. Comparative Performance of ARIMA and GARCH Models in Modelling and Forecasting
   Volatility of Malaysia Market Properties and Shares Nor Hamizah Miswan , Nor Azazi Ngatiman, Khairum Hamzah and Zaminor Zamzamir Zamzamin(2014)
4. Modeling the Volatility of Real GDP Growth: The Case of Japan Revisited WenShwo Fang University of Connecticut (2008)
5. An Emperical Study of GDP in Indian Economy on Inflation and Population Article, Palvi Jearth ,Shikhil Munjal, Patiala 7 Gurcharan Singh,Punjabi University (2019)
6. Analysis and comparison of overall GDP depending only on three major sectors in Indian economy
   Dr. Konda Hari Prasad Reddy Associate Professor, Department of Management Centre for Management Studies, Jain University Lalbagh Road, Bengaluru – 560027
7. Forecasting of demand using ARIMA model Jamal Fattah , Latifa Ezzine , Zineb Aman , Haj El Moussami , and Abdeslam Lachhab(2018)
8. A Comparative Study of Indian Economy in Pre and Post Reform period: An Econometric Analysis, Shivani Gupta(2014)