

BIDHANNAGAR COLLEGE

DEPARTMENT OF STATISTICS

ANALYSIS ON GROSS DOMESTIC PRODUCT



REGISTRATION NO.: 1082011400136

ROLL: 6231125

NO.: 14761

COLLEGE ROLL NO.: 20287

SESSION: 2020-23

ACKNOWLEDGEMENT

“Luck is what happens when preparation meets opportunity”

(Seneca)

Yes, I'm very lucky to have the opportunity to understand the educational theories of my undergraduate level through real practical examples. All my preparation, my thoughts, my concepts throughout my 3 years journey have been very useful to complete this project. I would like to express my gratitude to several individuals for supporting me and giving me this opportunity throughout my final year project.

First, it has been a tremendous honour and privilege to work under Professor **Dr. Kiranmoy Chatterjee** and Professor **Soumyadeep Das** for their exemplary guidance, enthusiasm, helpful information, practical advice and unceasing ideas that have helped me tremendously in my research and doing this project.

I am also extremely grateful to all my other Professors in Department of Statistics at Bidhannagar College, Bidhannagar, Kolkata for help me to finish this project in the best possible way. Their immense knowledge, profound experience and professional expertise have enabled me to complete this research and project successfully.

I also wish to express my sincere thanks to the West Bengal State University and Bidhannagar College for giving me this opportunity for doing this project. I would also like to express my gratitude for our Principal sir for providing the means and support to complete the project.

Last but not the least; I want to thank my family and my classmates for their immense support, care, help, constant encouragement and for exchanging interesting ideas throughout the course of this project.

Thanks for all the encouragement!

CONTENTS

Topic	Page Number
Abstract	4
Introduction	4 – 5
Data Description	6 – 7
Objectives	7
Methodology	8
Analysis of Data	8 – 42
• Objective 1 Analysis	8 – 21
1. Data	8
2. Study the Correlations & Some Scatterplots	9 – 13
3. Study the effect on GDP by Multiple Linear Regression	13 – 21
• Objective 2 Analysis	22 – 35
1. Data	22
2. Study the graph, Year vs GDP of India	22 – 24
3. Model Identification	24 – 33
4. Forecasting	34 – 35
• Special Analysis	36 – 42
Conclusion	43 – 45
Appendix	46 – 56
1. Theoretical Framework	46 – 50
2. R – Code	51 – 56
Bibliography	56

ABSTRACT

This project aims to investigate the *effects of various factors on the Gross Domestic Product (GDP)* of selected 14 countries on the year 2020. The study focuses on analysing the social and economic factors, such as goods exports, goods imports, gross national product (GNP), consumer spending, foreign direct investment (FDI), external debt, GDP per capita, inflation rate, unemployment rate, that affect GDP. The whole sample is subjected to regression analysis as GDP is being the dependent variable and the rest of the factors are being the covariates. By examining these factors and their impact on GDP, this project seeks to contribute to a deeper understanding of the economic dynamics within these countries. Also, this project aims to *develop a forecast model for predicting the total Gross Domestic Product (GDP) of selected countries for the years 2023 to 2030*. The study focuses on utilizing historical GDP data of India (1960-2022), economic indicators, and statistical techniques to build an accurate and reliable forecasting model and forecast for upcoming years. As a result, we found significant and less significant factors and also find an approximate forecast which will be discussed below in this work.

Key Words: *Gross Domestic Product (GDP), different factors, India, Bangladesh, Indonesia, Vietnam, Pakistan, Philippines, Egypt, Malaysia, Brazil, Iran, Mexico, Turkey, Ukraine, Nigeria, time series data.*

INTRODUCTION

GDP (Gross Domestic product) is a monetary measure of the market value of all the final goods and services produced and sold (not resold) in a specific time period by countries. Gross Domestic Product (GDP) is crucial indicator of a country's economic performance and is commonly used to measure the overall economic health and growth. It represents the total value if goods & services produced within a country's borders over a specific period. There are various factors which affect GDP of a country and understanding these factors is vital for forecasting future economic growth.

Here we are taking mainly the developing countries more precisely similar to India (comparing with the year of our interest 2020). 14 countries are *India, Bangladesh, Indonesia, Vietnam, Pakistan, Philippines, Egypt, Malaysia, Brazil, Iran, Mexico, Turkey, Ukraine, Nigeria*.

Here our study variable is GDP of the countries & the explanatory variables of interest are respectively given below. *Goods exports of a country* that is the goods & services that are produced in one country and sold it to buyers in another. *Goods Imports of a country*, a goods or services bought in one country that was produced in another. *Consumer spending*, a significant driver of economic growth. When a consumer purchases goods and services, it

stimulates the demand and encourages business to produce more. *Gross National product (GNP)* is the value of all products & services produced by the citizen of the study country both domestically and internationally minus income earned by a foreign resident. *Foreign direct investments (FDI)*, refers to purchase of an asset such that it gives the control to the purchaser over the asset. *Gross Govt. external debt* is the portion of a country's debt that is borrowed from foreign lenders, including commercial banks, governments or international financial institutions. If a country can't repay its external debt it is said to be in sovereign debt and faces a debt crisis. Another important factor is *GDP per capita* which is the sum of gross value added by all resident producers in the economy plus any product taxes (less subsidies) not included in the valuation of output, divided by mid-year population. *Inflation rate* measures how much more expensive a set of good & services has become over a certain period, usually a year. Lastly, we take *unemployment rate*, the percentage of the labour force that is looking for a job.

As one of the world's fastest-growing economies, India has witnessed remarkable economic progress for the upcoming years. Understanding & forecasting India's GDP growth is of great importance for policymakers, businesses and investors to make informed decisions and plan for the future.

The project utilizes a comprehensive methodology that involves analysing historical GDP data, examining the trends and patterns and identifying the key drivers of economic growth. Additionally, we employ statistical techniques and time series analyse to construct a robust forecast model that captures the underlying dynamics of India's GDP growth.

Forecasting GDP involves utilizing historical data, economic models & various indicators to estimate future economic growth. However, forecasting with accuracy is quite challenging due to complexity and dynamic nature of economies. Economists and analysts employ different forecasting techniques like, econometric models, time series analysis etc. It's essential to note that unforeseen events such as natural disasters, geopolitical tensions or pandemics can significantly impact GDP and make accurate forecasts more challenging.

Understanding the factors influencing GDP provides insights into economic growth. So, we shall try to examine which factors gave more impact on GDP in the pandemic year 2020 & forecasting GDP involves analysing these factors and employing some models and data to estimate future economic performance.

Lastly, of this work we're trying to give a comparison between a newspaper article and my own analysis.

DATA DESCRIPTION

Here in our project, we are considering total 10 variables, they are divided into two groups study variable and explanatory variables. Study variable in which we're trying to explain or predict & sometimes it is called dependent variable as it depends on another variable(s). on the other hand, the explanatory variables are used to explain or predict the response variable.

Study Variable:

Gross domestic Product (GDP) of a country is a measure of the total value of all final goods and services produced within its borders during a specific period, typically a year. It provides an overall snapshot of the economic activity and output within a country's economy.

Explanatory Variable:

There are various socio-economic factors which have impact on GDP of a country. Working with all is little bit laborious so, we selected some factors as explanatory variables which are given below –

- i. **Total goods export by a country:** It means the value of physical goods that are produced within its borders and sold to other countries. It represents the monetary worth of goods shipped out of the country to international markets. Goods exports encompass various products, including manufactured goods, raw materials, agricultural products and commodities.
- ii. **Total import of goods by a country:** It means the value of physical goods that are purchased from other countries & brought into the country for domestic use. It represents the monetary worth of goods that a country buys from international markets. The value of import goods is typically measured in the country's national currency.
- iii. **Consumer spending of a country:** It refers to the total amount of money spent by households on goods & services within a specific time period. It represents the consumption expenditure by individuals & households in the economy.
- iv. **Gross National Product (GNP):** GNP of a country is a measure of the total value of all final goods and services produced by the residents of a country, regardless of their location, during a specific period typically a year.
- v. **Foreign Direct Investment (FDI):** FDI of a country refers to the investment made by individuals, business or entities from one country into another country with the intention of establishing a lasting interest and control in the foreign economy. FDI is a key component of global integration and plays a significant role in stimulating economic growth, forecasting technological transfer, creating employment opportunities and enhancing international trade.

- vi. **External debt of a country:** External debt of a country means total amount of money that a country owes to foreign creditors. It includes both public debt (which is incurred by the govt.) and private debt (which is incurred by private sector entities such as corporations or individuals).
- vii. **GDP per capita for a country:** GDP per capita refers to the GDP divided by the total number of populations of that country. It is a very important measure which is used by economists, policymakers & analysts to assess the economic wellbeing and development of a country.
- viii. **Inflation rate of a country:** The inflation rate of a country refers to the percentage change in the avg. price level of goods and services is increasing and the purchasing power of a currency is declining. It is typically expressed as an annualized percentage and can be calculated using various price indices. It is influenced by economic development, income distribution, population size, natural resources etc.
- ix. **Unemployment rate of a country:** The unemployment rate of a country measures that the percentage of the labour force that is currently unemployed and actively seeking employment. It is typically calculated by dividing the number of unemployed individuals by the total labour force (which includes both employed and unemployed individual actively seeking work).

Also, we have a time series dataset from the year 1960 to 2022 of India on GDP and various factors. This is a historical data of India. We're interested in successive years from 1960 and net GDP for a particular year to form our time series analyse and forecast through the time series model.

At first, we plot the series and focus on whether there is any pattern, followed by the data. Then the model identification for forecast plays an important role to forecast. Here we made our forecast using a suitable ARIMA model to predict the net GDP for the years 2023-'30.

OBJECTIVES

- In this study we are going to investigate the **effect of different selected factors** as explanatory variables on Gross Domestic Product (GDP) in selected fourteen countries.
- Develop a **forecast** for the total GDP of India for the years 2023-'30

METHODOLOGY

Data Collection: Throughout my project, I am working with secondary data which are collected from different websites on Google platform e.g.- <http://www.macrotrends.net>, <http://www.worldbank.org>, <http://www.kaggle.com> etc.

Procedure: To explore the first objective we analyse the effect on GDP with the help of linear models more precisely, we're going to use the Regression technique. We should fit a multiple linear regression equation between study variable and explanatory variables. The whole procedure was carried out in R software.

For the second objective, that is for forecasting we use a time series model. Also, the forecasting process was carried out in R software.

Necessary Software: R software.

ANALYSIS OF THE DATA

Objective 1:

In this section we are going to investigate the effect of different selected factors as explanatory variables on Gross Domestic Product (GDP) in selected fourteen countries.

Data

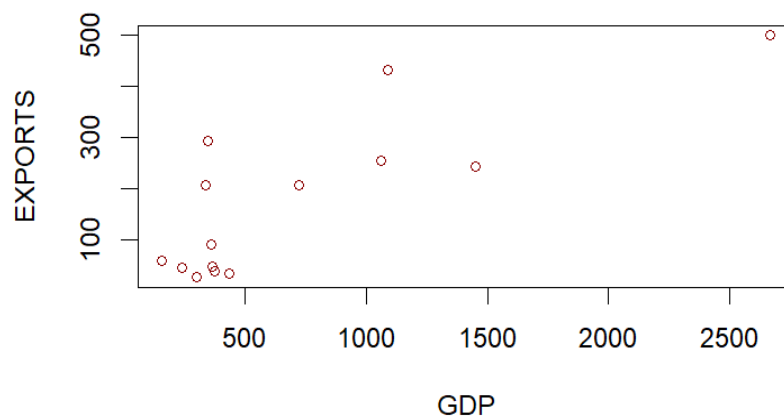
Here is our data for analysis and the unit is in billion USD –

Country	GDP	Exporsts	Imports	Con Sp	GNP	FDI	Ext debt	GDP /cap	Inf rate	Jnem rate
India	2667.69	499.1	509.43	1621.9	2641.49	64.36	565.0527	1910	6.62	8
Bangladesh	373.9	39.05	59.18	250.33	385.22	1.53	73.50719	2233	5.69	5.41
Indonesia	1058.69	255.73	166.26	624.14	1059.88	19.18	417.0462	3894	1.92	4.28
Vietnam	346.62	292.48	273.36	193.99	333.83	15.8	125.065	3586	3.22	2.39
Pakistan	300.43	27.94	52.33	244.89	321.9	2.06	115.6953	1322	9.74	4.3
Philippines	361.75	91.17	119.26	271.56	375.81	6.82	98.49369	3224	2.39	2.52
Egypt	365.25	47.87	75.43	313.49	307.45	5.85	129.7918	3399	5.04	9.17
Malaysia	337.34	207.8	186.31	205.29	342.56	4.06	209.25	10161	-1.14	4.5
Brazil	1448.56	243.29	233.15	911.01	1666.97	37.79	549.2632	6794	3.21	13.69
Iran	239.74	46.57	58.46	110.92	287.58	1.34	5.433188	2746	30.59	12.17
Mexico	1090.52	430.5	408.11	687.55	1102.89	31.38	605.0517	8655	3.4	4.45
Turkey	720.29	206.96	232.1	408.89	762.75	7.83	441.1503	8561	12.28	13.11
Ukraine	156.62	60.8	63.17	114.78	149.17	0.3	132.3128	3752	2.73	9.13
Nigeria	432.2	35.09	35.59	276.22	421.34	2.39	70.5249	2075	13.25	9.71

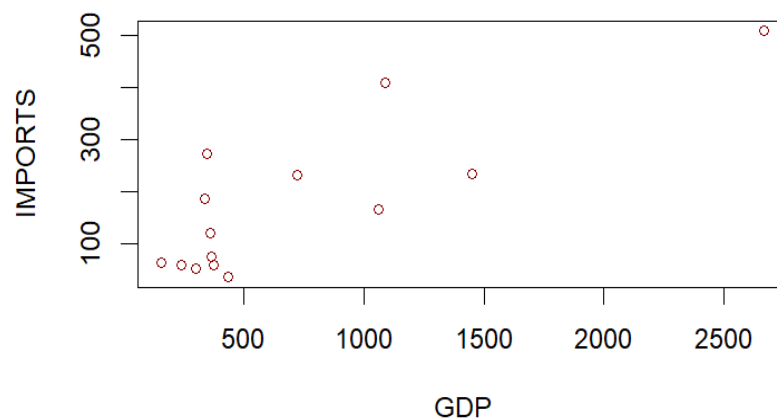
Study the Correlation & Some Scatterplots

At first, we take a look of some scatterplots between the study variables and the other variables. Also, their correlation is specified in braces. We observed that the correlation between GDP and *consumer spending* is maximum correlation (0.9962289). Also, correlation between GDP and *GNP & FDI* is also high (0.9959903, 0.9661529, respectively). The minimum correlation (-0.1585357, 0.04161905, 0.1671824) occurs for the factor *inflation rate*, *GDP per capita* and *unemployment rate* respectively. So, intuitively we can say that for our further analysis that the factors having max correlation would be among the significant factors that explain our study variable.

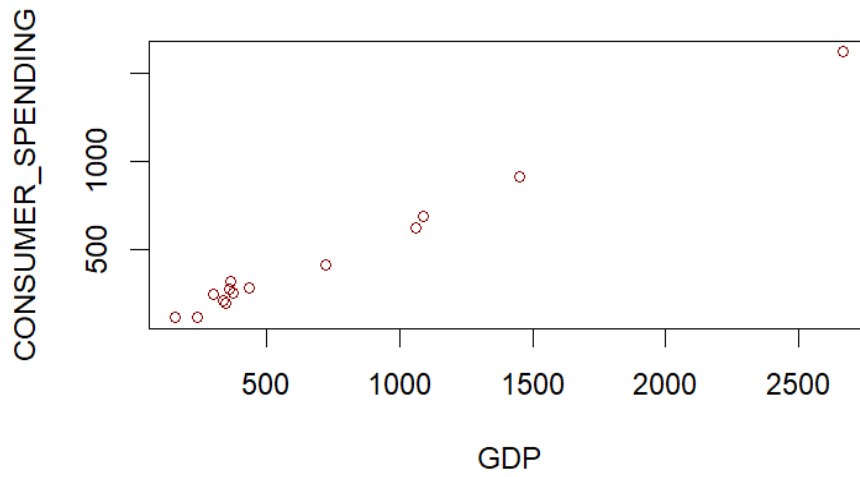
- **Diagram 1.1: Scatter plot between GDP & Exports of countries (corr = 0.8021576)**



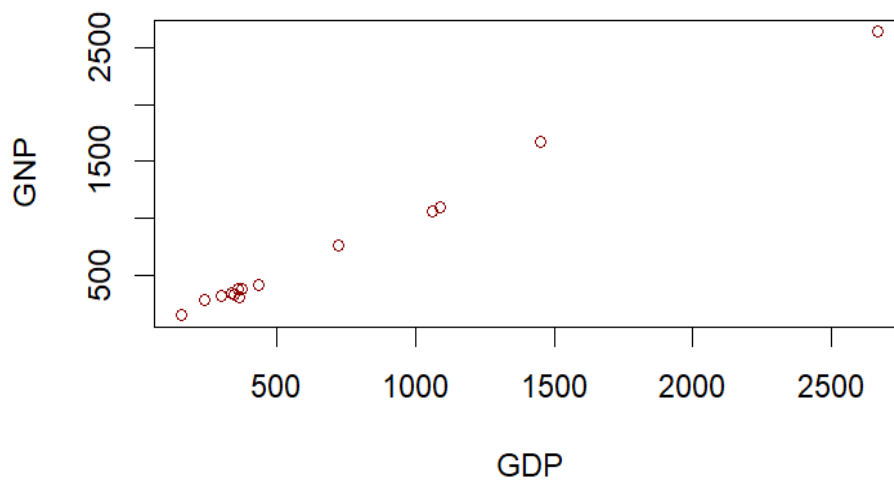
- **Diagram 1.2: Scatter plot between GDP & Imports of countries (corr. = 0.8166605)**



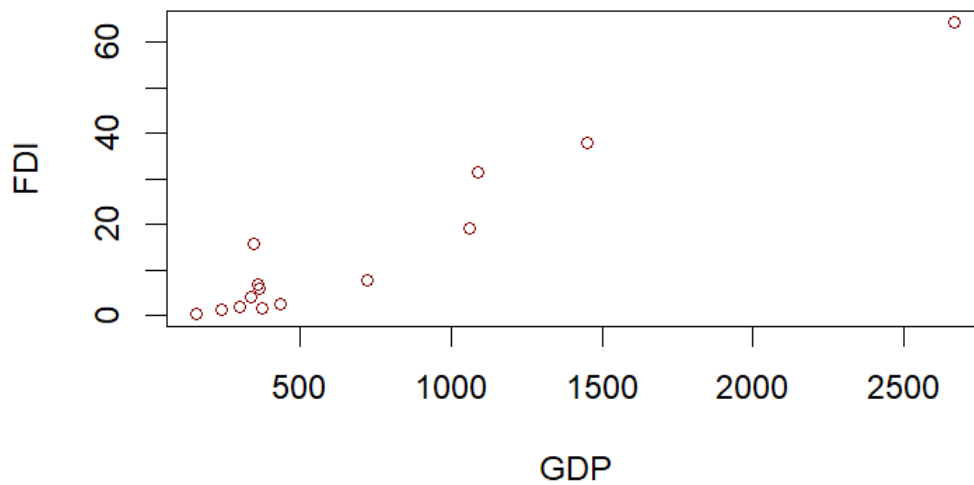
- **Diagram 1.3: Scatter plot between GDP & consumer spending of countries (corr. = 0.9962289)**



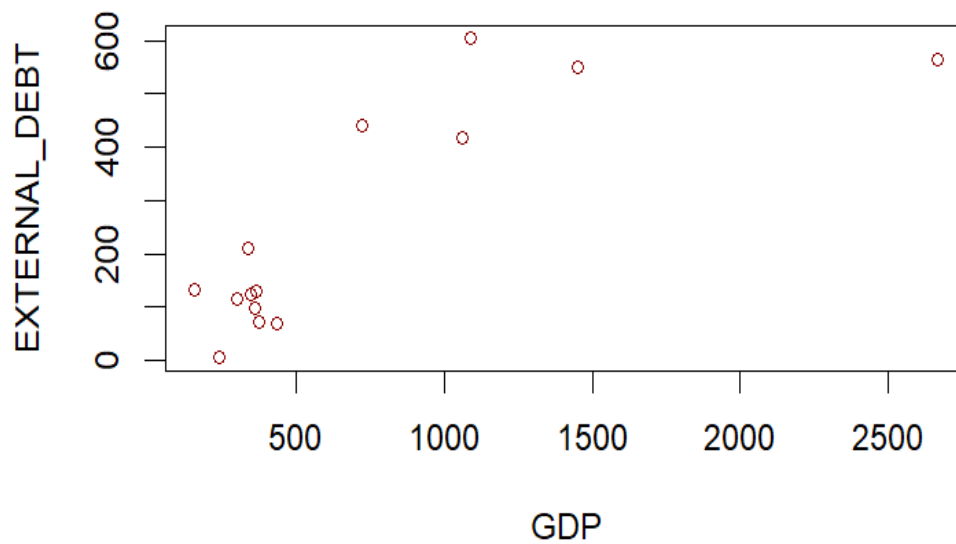
- **Diagram 1.4: Scatter plot between GDP & GNP of countries (corr. = 0.9959903)**



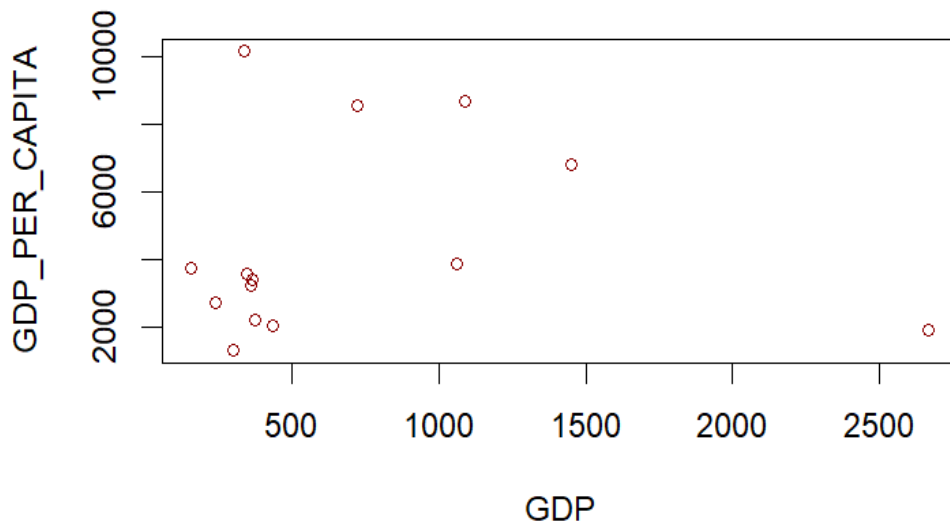
• Diagram 1.5: Scatter plot between GDP & FDI of countries (corr. = 0.9661529)



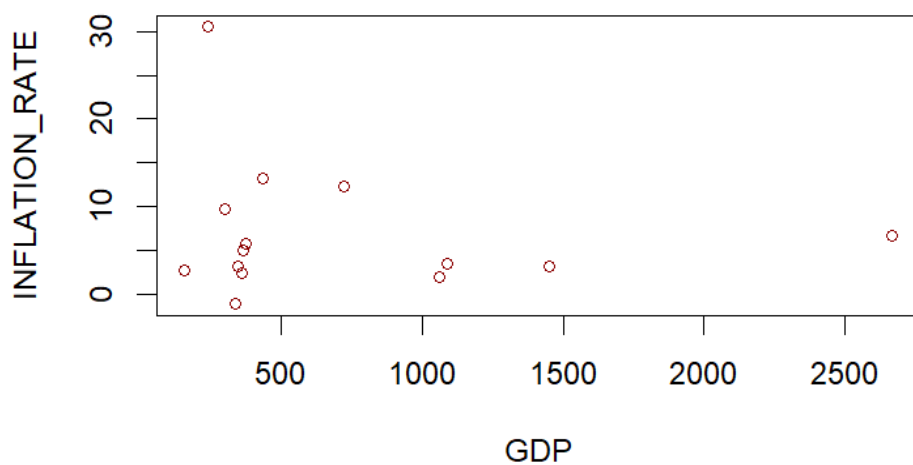
• Diagram 1.6: Scatter plot between GDP & External debt of countries (corr. = 0.8123983)



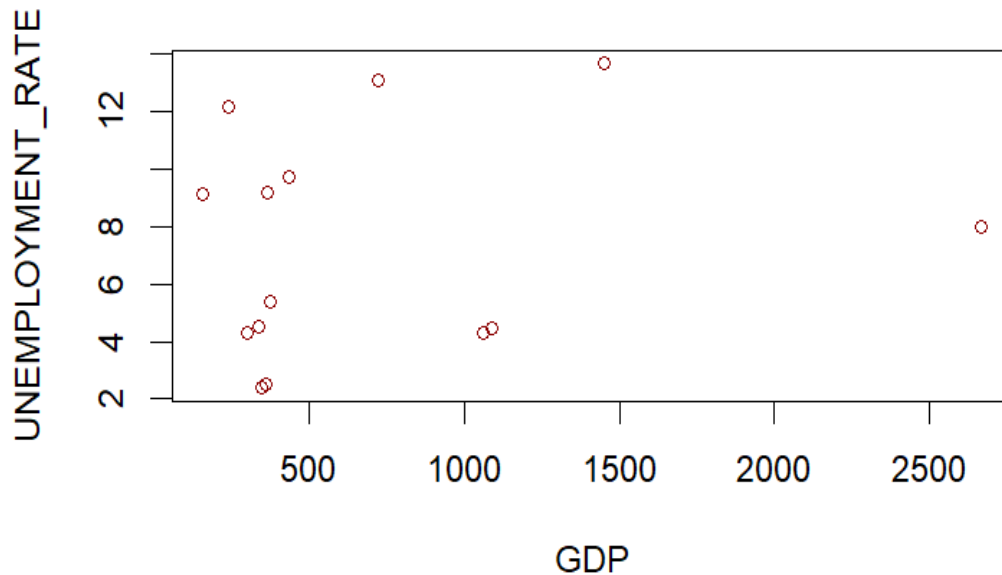
- **Diagram 1.7: Scatter plot between GDP & GDP per capita of countries**
(corr. = 0.04161905)



- **Diagram 1.8: Scatter plot between GDP & Inflation rate of countries**
(corr. = -0.1585357)



- **Diagram 1.9: Scatter plot between GDP & Unemployment rate of countries**
(corr. = 0.1671824)



Study the effect on GDP by Multiple Linear Regression

Now we're interested to study the effect on GDP of 9 different selected factors of 14 differently chosen countries. We brought all units are in same in billion USD (\$) or population figure brought in billion forms for our easier calculation purpose.

So, number of total factors (i) = 9

Number of total countries (j) = 14

[N.B.- Related theories are given in Appendix]

Notations:

1. Y_j : Y_j be the GDP of jth country (Study variable)
2. X_{ij} : X_{ij} be the data of ith factor of jth country (Explanatory Variables),

$$i = 1, 2, \dots, 9 ; j = 1, 2, \dots, 14$$

Factors -

- a. X_1 : Goods export (In billion \$)
- b. X_2 : Goods import (In billion \$)
- c. X_3 : Consumer spending (In billion \$)
- d. X_4 : GNP (In billion \$)
- e. X_5 : FDI (In billion \$)
- f. X_6 : External debt (In billion \$)
- g. X_7 : GDP per capita (In billion \$)
- h. X_8 : Inflation rate (In percentage)
- i. X_9 : Unemployment rate (In percentage)

Model:

To explore the systematic variation in Y values we have introduced the 9 factors discussed above and each having 14 different levels (countries). Corresponding to each level combination we administer the test at random to a single country. To analyse the data, consider multiple linear regression model of interest is given below,

$$Y_j = \mu + \beta_1 X_{1j} + \beta_2 X_{2j} + \beta_3 X_{3j} + \beta_4 X_{4j} + \beta_5 X_{5j} + \beta_6 X_{6j} + \beta_7 X_{7j} + \beta_8 X_{8j} + \beta_9 X_{9j} + \varepsilon_j$$

$j = 1(1)14$

Named the above equation by model 1.

Where, μ be the general effect
 β_i be the coefficient of the i th regressor variable of the regression model 1, $i = 1(1)9$
 ε_j be the random error

Hypothesis:

- We're interested to check whether the error part follows Gaussian distribution or not.
 $H_{01} : \varepsilon_j \sim N(0, \sigma^2)$ against,
 $H_{11} : \text{Not } H_{01}$
- We are interested to test whether all the regressors variables i.e., explanatory variables $X_1, X_2, X_3, \dots, X_9$ have significant presence in the regression model. To address this question, we construct a statistical test for
 $H_{02} : \beta_i = 0$ (Insignificant) against,
 $H_{12} : \text{at least one inequality holds in } H_{02}$ for all $i = 1(1)9$

First Hypothesis Testing:

For the first hypothesis testing we completely use R software & we use Shapiro-Wilk normality test where null hypothesis is assumption of the model error follows Gaussian distribution with mean is zero and with finite variance (σ^2).

[N.B.- Related theory is in Appendix]

Analysis:

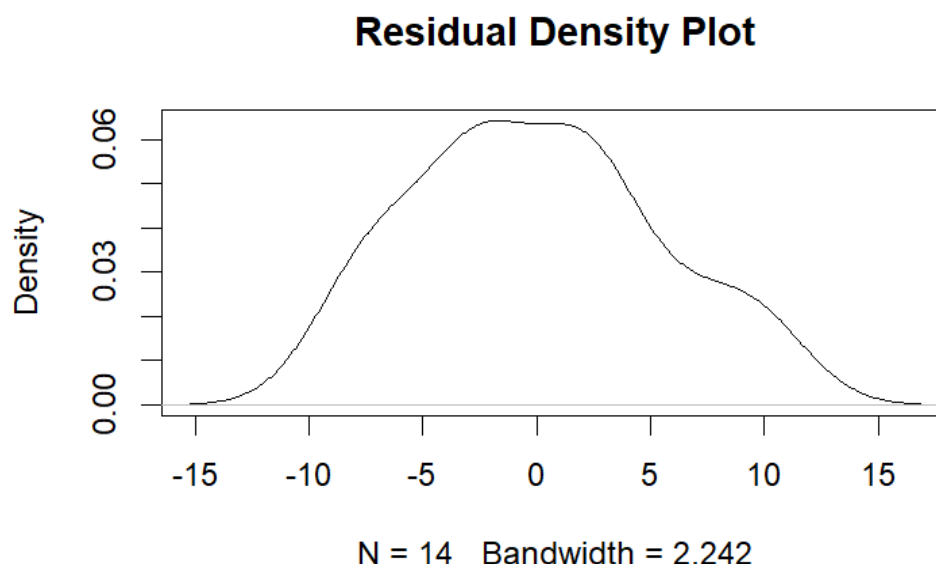
We get the p-value for model 1 of the test is 0.8542 (>0.05 , preassigned level of the test) it says that our assumption is correct that is the error follows Gaussian distribution with parameters mean=zero and variance = $\sigma^2(>0)$.

We're going to provide the analysed result of Shapiro-Wilk test;

```
> shapiro.test(e)
      Shapiro-Wilk normality test
data:  e
W = 0.96838, p-value = 0.8542
```

Here is a graphical nature to understand the errors are being normal;

- **Diagram 1.10: Density plot of Residuals of our dataset (observed – fitted)**



Here from the above plot, we observe that max density is around the zero and about to symmetrical in nature so we can conclude our desirable result.

Second Hypothesis Testing:

We should reject H_{02} against the alternative hypothesis H_{12} for an approximate level α of the test (here we set 0.05 level) by p-value of the test.

Observed p-value of the test for a suitable t-statistic is measured by, $P_{H_0}(T > t_{obs})$.

Then we should reject H_{02} if p-value, $P_{H_0}(T > t_{obs}) \leq \alpha (=0.05)$

Here is the analysed result for our model1 from our software;

```
> summary(lm(data$GDP~data$Goods.Exp.In.B.USD.+data$Goods.imp.In.B.USD.+
+ data$Consumer.spending+data$GNP+data$FDI+data$External.debt+
+ data$Inf.rate+data$Unemp.rate+data$GDP.per.capita))

Call:
lm(formula = data$GDP ~ data$Goods.Exp.In.B.USD. + data$Goods.imp.In.B.USD. +
  data$Consumer.spending + data$GNP + data$FDI + data$External.debt +
  data$Inf.rate + data$Unemp.rate + data$GDP.per.capita)

Residuals:
    1     2     3     4     5     6     7     8     9    10    11    12    13
 1.588  1.649  1.731 -6.709 -8.480  7.976 -2.461 -1.932 -3.130  5.014  3.017 -2.336 10.145
    14
-6.076

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -94.542130   11.569975  -8.171 0.001221 **
data$Goods.Exp.In.B.USD.    1.185171    0.126186   9.392 0.000716 ***
data$Goods.imp.In.B.USD.   -0.255414    0.123542  -2.067 0.107544
data$Consumer.spending    1.069355    0.075734  14.120 0.000146 ***
data$GNP              0.485171    0.047149  10.290 0.000503 ***
data$FDI             -9.537608    1.015594  -9.391 0.000716 ***
data$External.debt     -0.239811    0.042650  -5.623 0.004920 **
data$Inf.rate         -0.297971    0.623124  -0.478 0.657477
data$Unemp.rate        6.466488    1.430625   4.520 0.010658 *
data$GDP.per.capita    -0.008970    0.002283  -3.929 0.017123 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.848 on 4 degrees of freedom
Multiple R-squared:  0.9999,    Adjusted R-squared:  0.9998
F-statistic: 6921 on 9 and 4 DF, p-value: 5.102e-08
```

From the above picture we say that from p-value that *exports, consumer spending, GNP, FDI, external debt, unemployment rate, inflation rate for a country* are **significant** factors. Also remaining *imports, inflation rate for a country* are **less significant** factors.

Analysis:

Here are some tabulated values of our slope and coefficients;

Table 1: Values of coefficients of regression & corresponding p-value

Coefficients	Estimated value	p-value
Intercept(μ)	-94.542130	0.001221
β_1	1.185171	0.000716
β_2	-0.255414	0.107544

β_3	1.069355	0.000146
β_4	0.485171	0.000503
β_5	-9.537608	0.000716
β_6	-0.239811	0.004920
β_7	-0.297971	0.657477
β_8	6.466488	0.010658
β_9	-0.008970	0.017123

So, from the above table we get a fitted model with the estimates;

$$Y_j = (-94.542130) + (1.185171)*X_{1j} + (-0.255414)*X_{2j} + (1.069355)*X_{3j} + (0.485171)*X_{4j} + (-9.537608)*X_{5j} + (-0.239811)*X_{6j} + (-0.297971)*X_{7j} + (6.466488)*X_{8j} + (-0.008970)*X_{9j}$$

$j = 1(1)14$

The above equation is named by fitted model 1

Here is another tabulated value of real value and estimated value;

Table 2: Fitted values of GDP and Error of 14 countries from fitted model 1

Country	GDP	Estimated GDP	Error of estimation
India	2667.69	2666.1020	1.588024
Bangladesh	373.9	372.2507	1.649288
Indonesia	1058.69	1056.9588	1.731192
Vietnam	346.62	353.3288	-6.708827
Pakistan	300.43	308.9096	-8.479571
Philippines	361.75	353.7736	7.976430
Egypt	365.25	367.7109	-2.460932
Malaysia	337.34	339.2715	-1.931542
Brazil	1448.56	1451.6896	-3.129554
Iran	239.74	234.7257	5.014257
Mexico	1090.52	1087.5026	3.017380
Turkey	720.29	722.6255	-2.335507
Ukraine	156.62	146.4747	10.145296
Nigeria	432.2	438.2759	-6.075934

From our estimation we get;

Value of Multiple R-squared: 0.9999

Value of Adjusted R-squared: 0.9998

More Analysis:

From previous model fitting by regression technique, there found two less significant variables. If we use the command “step()” in R then we get the value of Akaike Information Criterion (AIC).

[N.B.- Related theory is in Appendix]

For the model 1 the value if AIC is 66.5. For the case of insignificance further we should check the best fitted model with least AIC.

For model 1;

```
> step(model1)
Start: AIC=66.5
data$GDP ~ data$Goods.Exp.In.B.USD. + data$Consumer.spending +
  data$GNP + data$FDI + data$External.debt + data$GDP.per.capita +
  data$Unemp.rate + data$Goods.imp.In.B.USD. + data$Inf.rate
```

Now we exclude one of the insignificant factors, import of a country then the model is named by model 2;

$$\text{Model 2: } Y_j = \mu + \beta_1 X_{1j} + \beta_3 X_{3j} + \beta_4 X_{4j} + \beta_5 X_{5j} + \beta_6 X_{6j} + \beta_7 X_{7j} + \beta_8 X_{8j} + \beta_9 X_{9j} + \epsilon_j$$

$$j = 1(1)14$$

For model 2;

```
> step(model2)
Start: AIC=74.68
data$GDP ~ data$Goods.Exp.In.B.USD. + data$Consumer.spending +
  data$GNP + data$FDI + data$External.debt + data$GDP.per.capita +
  data$Unemp.rate + data$Inf.rate
```

	Df	Sum of Sq	RSS	AIC
<none>			802.4	74.680
- data\$Inf.rate	1	127.8	930.2	74.749
- data\$Unemp.rate	1	1829.9	2632.3	89.312
- data\$GDP.per.capita	1	2377.9	3180.3	91.959
- data\$External.debt	1	2701.3	3503.7	93.315
- data\$FDI	1	10171.2	10973.6	109.299
- data\$Goods.Exp.In.B.USD.	1	15824.2	16626.6	115.116
- data\$GNP	1	15922.1	16724.5	115.198
- data\$Consumer.spending	1	21774.0	22576.4	119.398

AIC of model 2 is 74.68

Now we exclude the factor inflation rate of a country then the model of our interest is named by model 3;

$$\text{Model 3: } Y_j = \mu + \beta_1 X_{1j} + \beta_2 X_{2j} + \beta_3 X_{3j} + \beta_4 X_{4j} + \beta_5 X_{5j} + \beta_6 X_{6j} + \beta_7 X_{7j} + \beta_9 X_{9j} + \epsilon_j$$

$$j = 1(1)14$$

For model 3;

```
> step(model3)
Start: AIC=65.28
data$GDP ~ data$Goods.Exp.In.B.USD. + data$Consumer.spending +
  data$GNP + data$FDI + data$External.debt + data$GDP.per.capita +
  data$Unemp.rate + data$Goods.imp.In.B.USD.

      Df Sum of Sq    RSS    AIC
<none>                410.1  65.282
- data$Goods.imp.In.B.USD.  1     520.1   930.2  74.749
- data$GDP.per.capita      1    1814.3  2224.4  86.954
- data$Unemp.rate          1    2693.0  3103.0  91.615
- data$External.debt      1    3070.6  3480.7  93.223
- data$Goods.Exp.In.B.USD.  1    8532.7  8942.8 106.434
- data$FDI                  1    9778.2 10188.3 108.259
- data$GNP                  1   11930.8 12340.9 110.943
- data$Consumer.spending   1   22796.4 23206.4 119.784
```

AIC of our model 3 is 65.28

Now we exclude both the insignificant factors from the model and make model 4;

$$\text{Model 4: } Y_j = \mu + \beta_1 X_{1j} + \beta_3 X_{3j} + \beta_4 X_{4j} + \beta_5 X_{5j} + \beta_6 X_{6j} + \beta_7 X_{7j} + \beta_9 X_{9j} + \varepsilon_j$$

$$j = 1(1)14$$

For model 4;

```
> step(model4)
Start: AIC=74.75
data$GDP ~ data$Goods.Exp.In.B.USD. + data$Consumer.spending +
  data$GNP + data$FDI + data$External.debt + data$GDP.per.capita +
  data$Unemp.rate

      Df Sum of Sq    RSS    AIC
<none>                930.2  74.749
- data$Unemp.rate        1    2194.8  3125.0  89.714
- data$GDP.per.capita    1    2427.8  3358.0  90.721
- data$External.debt     1    2581.6  3511.8  91.348
- data$FDI                1   10715.2 11645.4 108.131
- data$GNP                1   16782.2 17712.5 114.002
- data$Goods.Exp.In.B.USD.  1   18999.6 19929.9 115.653
- data$Consumer.spending  1   24300.1 25230.3 118.954
```

AIC of our model 4 is 74.75

So, from the above data we conclude that;

$$\text{AIC}_{(\text{Model 3})} < \text{AIC}_{(\text{model 1})} < \text{AIC}_{(\text{model 2})} < \text{AIC}_{(\text{model 4})}$$

Therefore, the **best model** with least AIC is **model 3**.

Model 3 Analysis:

Here is the analysed result for our model1 from our software;

```
Call:
lm(formula = data$GDP ~ data$Goods.Exp.In.B.USD. + data$Consumer.spending +
    data$GNP + data$FDI + data$External.debt + data$GDP.per.capita +
    data$Unemp.rate + data$Goods.imp.In.B.USD.)

Residuals:
    1     2     3     4     5     6     7     8     9    10    11    12    13
 1.841  2.609  1.617 -5.255 -9.583  7.889 -2.737 -1.954 -2.500  2.215  1.055 -1.436 12.352
 14
-6.113

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -95.368276  10.520890  -9.065  0.000273 ***
data$Goods.Exp.In.B.USD.  1.182433  0.115926  10.200  0.000155 ***
data$Consumer.spending  1.082461  0.064927  16.672  1.42e-05 ***
data$GNP        0.475814  0.039450  12.061  6.92e-05 ***
data$FDI       -9.342937  0.855658 -10.919  0.000112 ***
data$External.debt -0.236194  0.038601  -6.119  0.001691 **
data$GDP.per.capita -0.008395  0.001785  -4.703  0.005321 **
data$Unemp.rate   6.059471  1.057466  5.730  0.002266 **
data$Goods.imp.In.B.USD. -0.273058  0.108429  -2.518  0.053284 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.056 on 5 degrees of freedom
Multiple R-squared:  0.9999,    Adjusted R-squared:  0.9998
F-statistic: 9206 on 8 and 5 DF,  p-value: 5.481e-10
```

From the above picture we say that from p-value that *exports, consumer spending, GNP, FDI* are **most significant** factors. Also, *external debt, GDP per capita, unemployment rate imports* are **more significant** factors and remaining *imports* for a country is **significant** factor.

Tabulated value of parameters estimation;

Table 3: Values of coefficients of regression & corresponding p-value

Coefficients	Estimated value	p-value
Intercept(μ)	-95.368276	0.000273
β_1	1.182433	0.000155
β_2	1.082461	1.42e-05
β_3	0.475814	6.92e-05
β_4	-9.342937	0.000112
β_5	-0.236194	0.001691
β_6	-0.008395	0.005321
β_7	6.059471	0.00226
β_9	-0.273058	0.053284

So, the fitted model 3 is given below;

$$Y_j = (-95.368276) + 1.182433 * X_{1j} + 1.082461 * X_{2j} + 0.475814 * X_{3j} + -9.342937 * X_{4j} + 0.236194 * X_{5j} + -0.008395 * X_{6j} + 6.059471 * X_{7j} + -0.273058 * X_{9j}$$

Tabulated value of real value and fitted values.

Table 4: Fitted values of GDP and Error of 14 countries from fitted model 3

Country	GDP	Estimated GDP	Error of estimation
India	2667.69	2665.849	1.840706
Bangladesh	373.9	371.2912	2.608781
Indonesia	1058.69	1057.074	1.616529
Vietnam	346.62	351.8747	-5.2547
Pakistan	300.43	310.0129	-9.58295
Philippines	361.75	353.8607	7.889331
Egypt	365.25	367.987	-2.737
Malaysia	337.34	339.2937	-1.95373
Brazil	1448.56	1451.061	-2.50047
Iran	239.74	237.5248	2.215244
Mexico	1090.52	1089.465	1.055115
Turkey	720.29	721.7257	-1.43567
Ukraine	156.62	144.2681	12.35192
Nigeria	432.2	438.3131	-6.11311

From our estimation we get;

Value of Multiple R-squared: 0.9999

Value of Adjusted R-squared: 0.9998

Objective 2:

In this section we are going to develop a *forecast* for the total GDP of India for the years 2023- '30.

Data

Here is our data and GDP is in USD –

Year	GDP (IND)
1960	3.7E+10
1961	3.92E+10
1962	4.22E+10
1963	4.84E+10
1964	5.65E+10
1965	5.96E+10
1966	4.59E+10
1967	5.01E+10
1968	5.31E+10
1969	5.84E+10
1970	6.24E+10

For more information visit the site where we have collected the data:

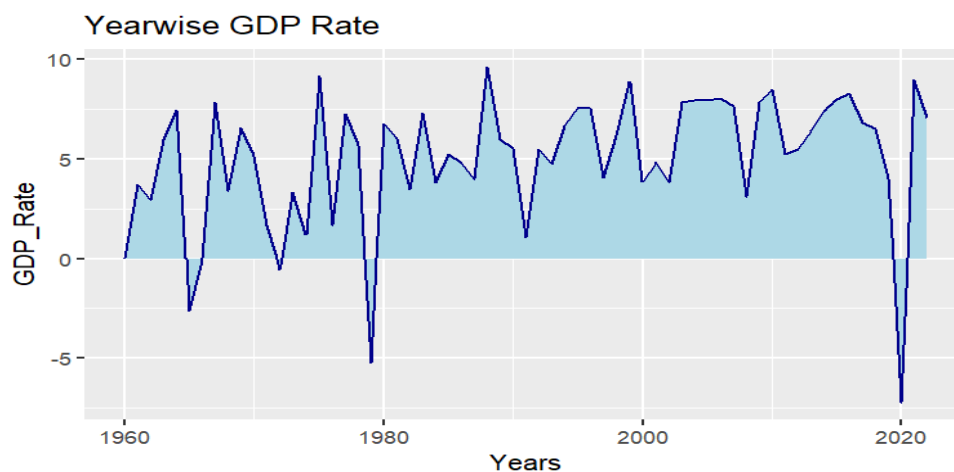
<https://www.macrotrends.net/countries/IND/india/gdp-gross-domestic-product>

Study the graph, Year vs GDP of India

Here in our hand has a time series data of India's GDP from the year 1960 to the year 2022. Here we have taken the unit of GDP in USD (\$). First, we see the GDP rate diagram, is given below by taking the successive years on our X-axis and corresponding year wise rate of that particular year on Y-axis.

[N.B. – Related theory about on GDP is in Appendix]

• Diagram 2.1: GDP growth rate structure since 1960



From the diagram 2.1, we observe the GDP growth rate is decreased on 1965-'66, 1972, 1979 and 2020.

India faced issues at its borders with wars in 1962 & 1965 and this led to a repaid increase in defence expenditure and massive import of arms which dented the fight against poverty. Also, the terrible drought in East India in the mid 1960's setback India's development and aggravated poverty. It led to the country's greater dependence on the Western powers for aid.

(The wire)

Fifty years back in 1971, there was two landmark events for India. The Bangladesh war and the quick victory achieved by the Indian forces. Indira Gandhi's thumping victory in the general elections due to the evocated slogan of 'Garibi Hatao' which brought poverty alleviation to centre stage in Indian politics. Undoubtedly, each of the decades since Independence has brought new challenges due to the fast-changing global situation and India trying to find its feet, internally, externally. 1971 reflects the global and the internal challenges in a continuum with the events shaping India since the mid 1960's.

(The wire)

In the year 1979-'80 the Indian economy faced a double whammy with a severe drought gripping most of the country and crude oil prices almost doubling due to supply disruptions because of the Iranian revolution against the state's monarch.

(India Today)

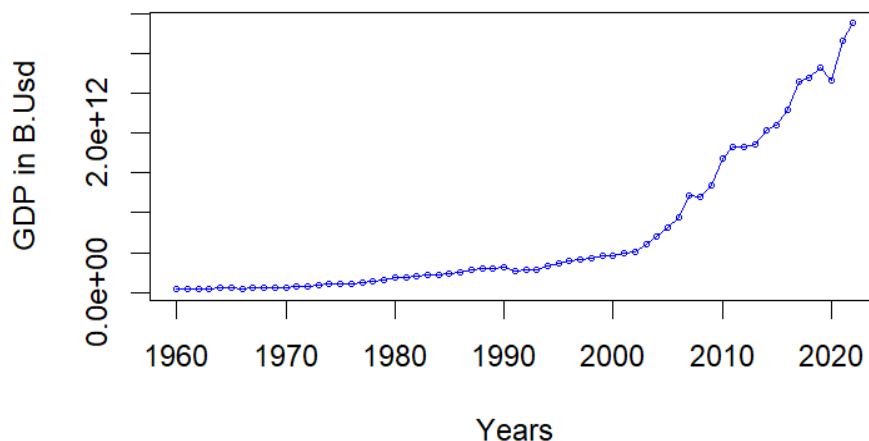
From April to June 2020, India's GDP dropped by a massive 24.4% due to Covid pandemic. According to the latest national income estimates, in the second quarter of the 2020-'21 financial year, July-September 2020, the economy contracted by a further 7.4%.

(TOI)

Now we are going to observe the graphical nature of net GDP of India from 1960 to 2022 by taking net GDP on Y-axis and corresponding years on X-axis.

• **Diagram 2.2: Net GDP growth structure since 1960 of India.**

GDP Growth of IND since 1960



From diagram 2.2, we say that if we study the nature of the growth curve that it has a direction of upward. So, our objective is to make a forecast for 2023-2030. Intuitively if there are no conditions that make economic recession, it will remain upward. Though during the year 2021 or after effects of pandemic that is in 2021 many start-ups open their wings so fast that India is now in third position of making of Unicorn Companies across the world and reached a milestone of 100 unicorns. Sectors and services activities expansion was reinforced by increases in trade, financial, real estate and professional services (4.6%), hotels, transportation and communication (6.1%); and public management, defence and other services.

Now we are concentrating over our procedure of forecasting.

Model Identification

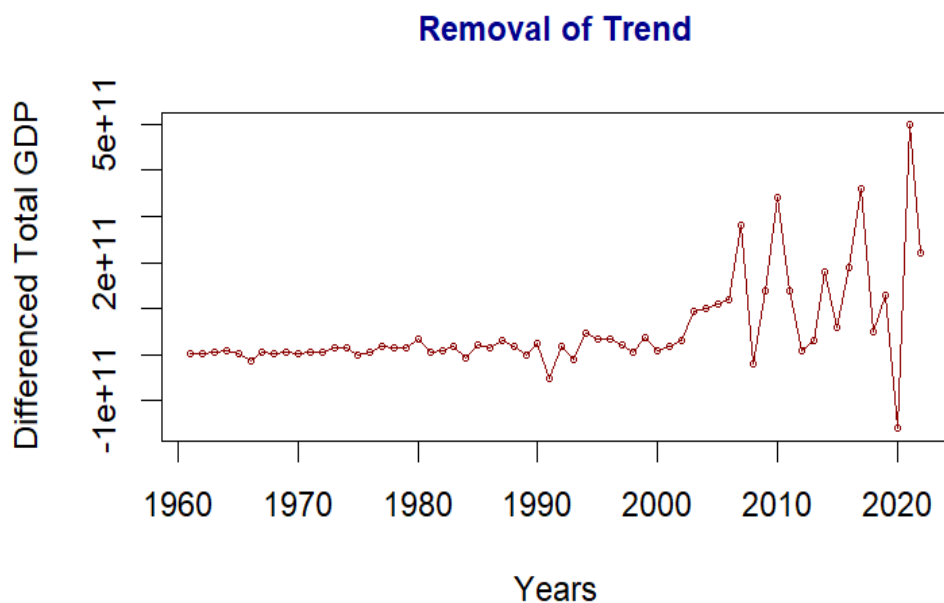
- **Time series plot -**

From diagram 2.2 we see an upward trend in the graph and variation changes with respect to time.

- **Removing the trend & seasonal fluctuation –**

In this step firstly we are differencing (1st order difference) our data on net GDP to remove the upward trend and seasonal fluctuations. Then we plot it on Y-axis, the 1st order differenced data on replace of GDP and X-axis is same on years.

- **Diagram 2.3: Removing the trend**



From the above diagram we can say clearly that, the above series is NOT stationary on variance i.e., variation in the plot is increasing as we move towards the right of the graphical plot. We need to make the series stationary on variance to produce reliable forecasts through ARIMA models. In order to make the series stationary on variance, one of the popular approaches is transformation of the original series into log transform data.

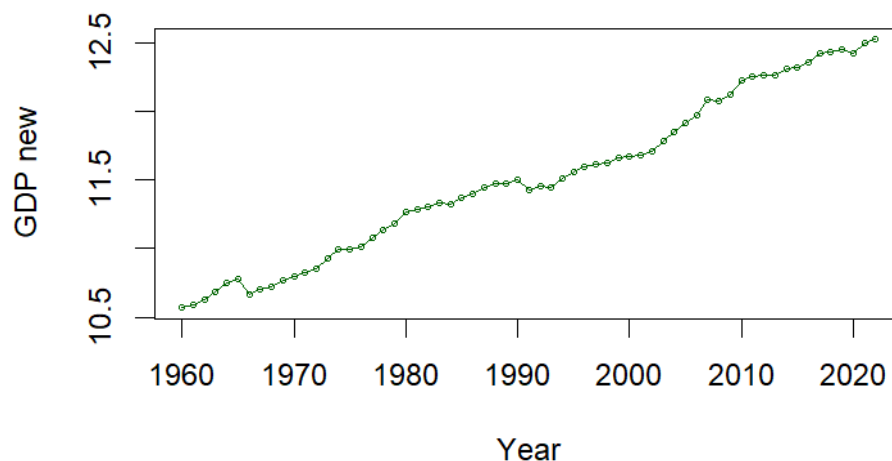
Log transformation –

In this case, it appears that an additive model is not appropriate for describing the original time series, since the spike of the fluctuations which are mainly seasonal fluctuations and random fluctuations seem to increase with the level of the time series. Thus, we may need to transform our time series data in order to get a transformed time series data that can be described using an additive model. We can transform the time series by taking the log (base 10) of the original data i.e., we assume that a multiplicative model is appropriate for the original data. By taking \log_{10} the multiplicative model becomes additive model. The reason behind we're interested for assuming the multiplicative model is an additive model, is appropriate if there is no interaction among the different components. More specifically if the magnitude of the seasonal fluctuation doesn't vary with the level of the series. But the multiplicative model is appropriate if there is any interaction among the different components. So, if the seasonal fluctuation increases and decreases proportionally with the increases and decreases of the level of the series then a multiplicative model is appropriate. Also, if we taking logarithmic transformation then variation of the series or graph decreases but the pattern (i.e., upward or downward) will be same. So, the transformed data is

$$GDP_{\text{new}} = \log_{10} GDP$$

• Diagram 2.4: Removing the trend

Multiplicative Model to Additive Model, GDP Growth

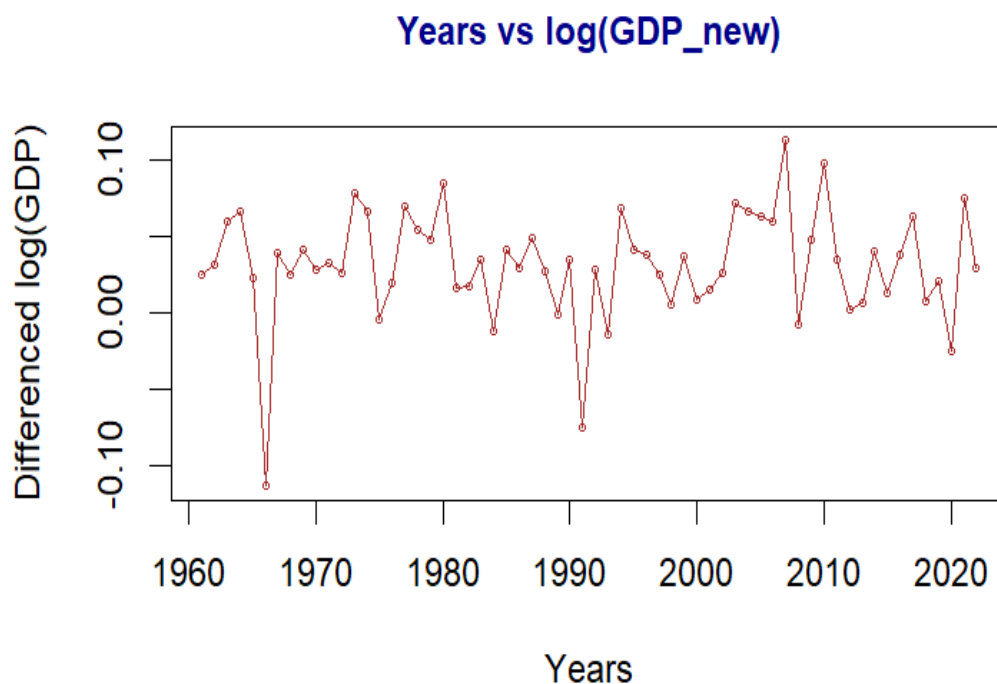


By taking log now we can say that it can help to stabilize the variance and reduce amplitude of the seasonal patterns. Also, by taking logarithm on GDP we may find that transform data exhibits a more linear trend. If we plot in graph our original data then we have exponential growth pattern, but the transformation can make trend linear, enabling simpler modelling and analysis.

First order differenced on GDP_{new} -

Now taking the first order differenced again on transformed GDP that is on GDP_{new} to eliminate the trend. We plot differenced GDP_{new} on Y-axis and successive years on X-axis. Here the plot is given below,

- **Diagram 2.5: Removal of trend with first order differenced on GDP_{new}**



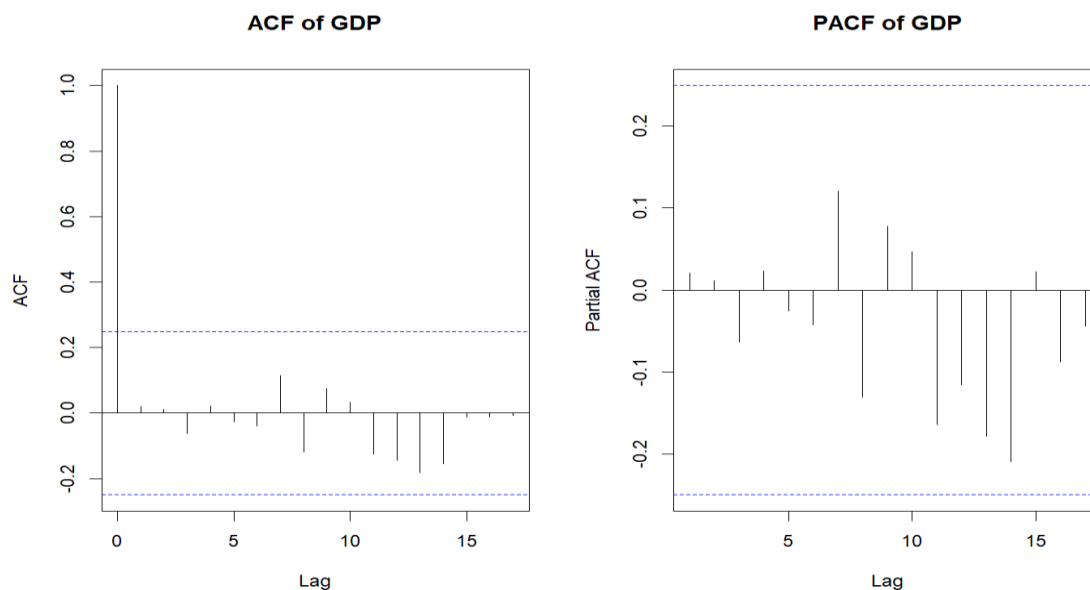
Now this series looks stationary on both mean and variance by removing the trend component. This also gives us the clue that integrated part (d in ARIMA model notation) of our ARIMA model will be equal to 1 as 1st difference is making the log data series stationary.

ACF & PACF of our new data -

Now let us make the plots of auto correlation factor (ACF) and partial autocorrelation factor (PACF) to study the pattern of the above transformed data which is stationary on both mean and variance. The work will be made on our first order differenced of GDP_{new} . With the help of this idea, we come to know whether any presence of AR and MA process in residuals.

[N.B. – Related theory is in Appendix]

• Diagram 2.6: ACF & PACF of diff (GDP_{new})



The horizontal two lines in both plot of ACF and PACF are called significant lines and the place between two lines is called significant zone. So, we can see that no spikes are exceed the significant zone. Hence, we added that our residuals are random. This implies that there is no information available in residuals to be extracted by AR and MA models.

Model identification & fitting –

Fitting a model is a very important things in analysis. For this purpose, we are going take the help of R software entirely. To get the exact model we have gone through trial-and-error process. That is, we take a model and keep tracking it's AIC value and then we take that model which has lowest AIC value. Whole procedure we have done on R.

[N.B. – Related theory is in Appendix]

For ARIMA (1,1,0), we get AIC value is **-210.41**

```

> G1=arima(gdp_new, order = c(1,1,0))
> G1

Call:
arima(x = gdp_new, order = c(1, 1, 0))

Coefficients:
      ar1
    0.4407
s.e. 0.1129

sigma^2 estimated as 0.001837:  log likelihood = 107.2,  aic = -210.41

```

For ARIMA (0,1,1), we get AIC value is **-205.23**

```

> G2=arima(gdp_new, order = c(0,1,1))
> G2

Call:
arima(x = gdp_new, order = c(0, 1, 1))

Coefficients:
      ma1
    0.2957
s.e. 0.0980

sigma^2 estimated as 0.002001:  log likelihood = 104.62,  aic = -205.23

```

For ARIMA (1,1,1), we get AIC value is **-225.6**

```

> G3=arima(gdp_new, order = c(1,1,1))
> G3

Call:
arima(x = gdp_new, order = c(1, 1, 1))

Coefficients:
      ar1      ma1
    0.9999  -0.9897
s.e. 0.0004   0.0260

sigma^2 estimated as 0.001326:  log likelihood = 115.8,  aic = -225.6

```

Model identification through R -

For ARIMA (1,1,1) we see the lowest AIC value (-225.6) but we do not say this so surely. To ensure we use the package in R “auto.arima()” to get the appropriate ARIMA model for forecast.

Now the question is what is basic reason or how can anyone find to get the appropriate ARIMA model, that is by observing the AIC value of the model. Least AIC value of a model is going to be the best and appropriate. As we discussed before in theoretical framework section about the AIC

So, here is our output;

```

> G = auto.arima(gdp_new)
> summary(G)
Series: gdp_new
ARIMA(0,1,0) with drift

Coefficients:
    drift
    0.0316
s.e.    0.0046

sigma^2 = 0.001314:  log likelihood = 118.25
AIC=-232.5   AICc=-232.3   BIC=-228.24

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 0.0001672525 0.035666818 0.02488124 0.0002422384 0.2160903 0.6251348 0.01949474

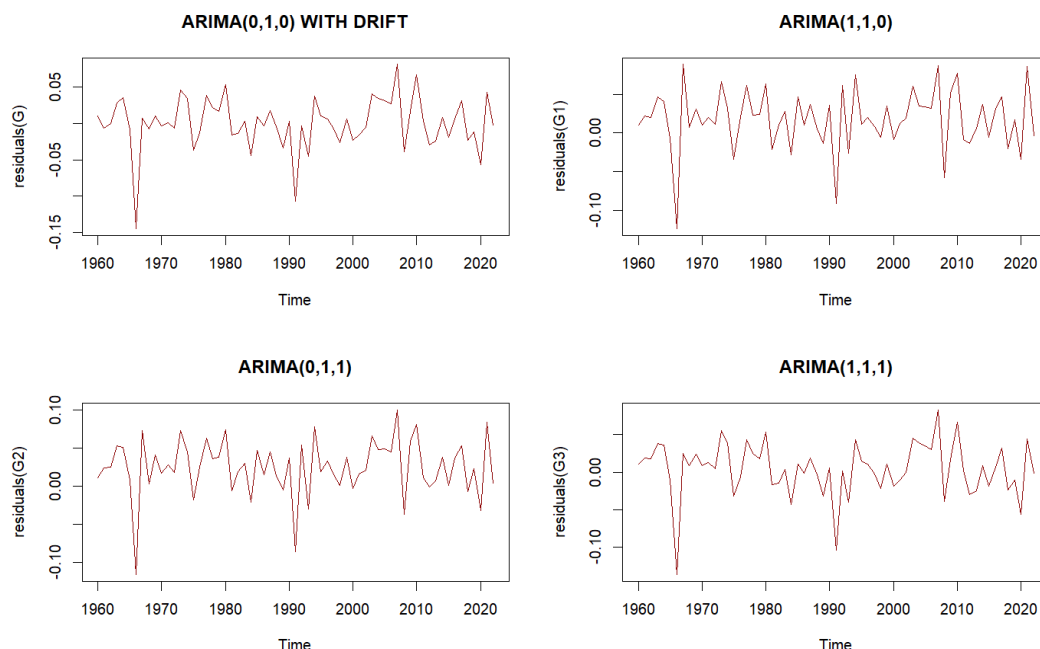
```

So, we get ARIMA (0,1,0) with drift model for forecasting which is appropriate for our work and with least AIC value (-232.5). As expected, the integrated part of the above model is 1 as we took first differencing. Above model has AR & MA of order 0.

Residuals plot -

Now we plot the residuals of our generated four models. Here it's given below;

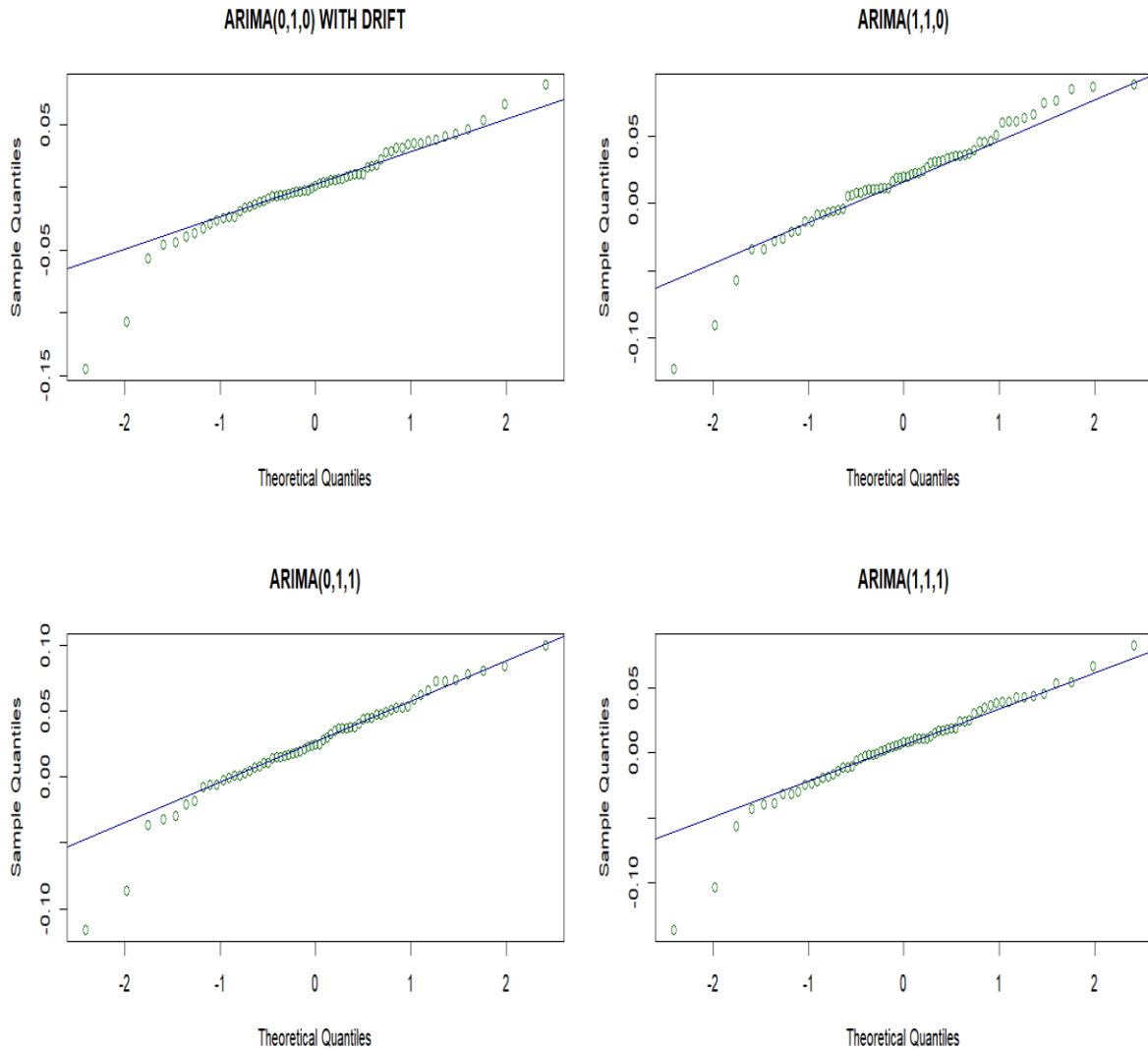
- **Diagram 2.7: Plots of Residuals of different models.**



By conclusion we can say that from diagram 2.7, graphical nature for all models is more or less same but we take ARIMA (0,1,0) with drift as it has least AIC value.

Now we are going to plot “qq-plot” and also draw “qq-line” for four models in diagram 2.8, is given below;

• **Diagram 2.8: QQ-plot of residuals**

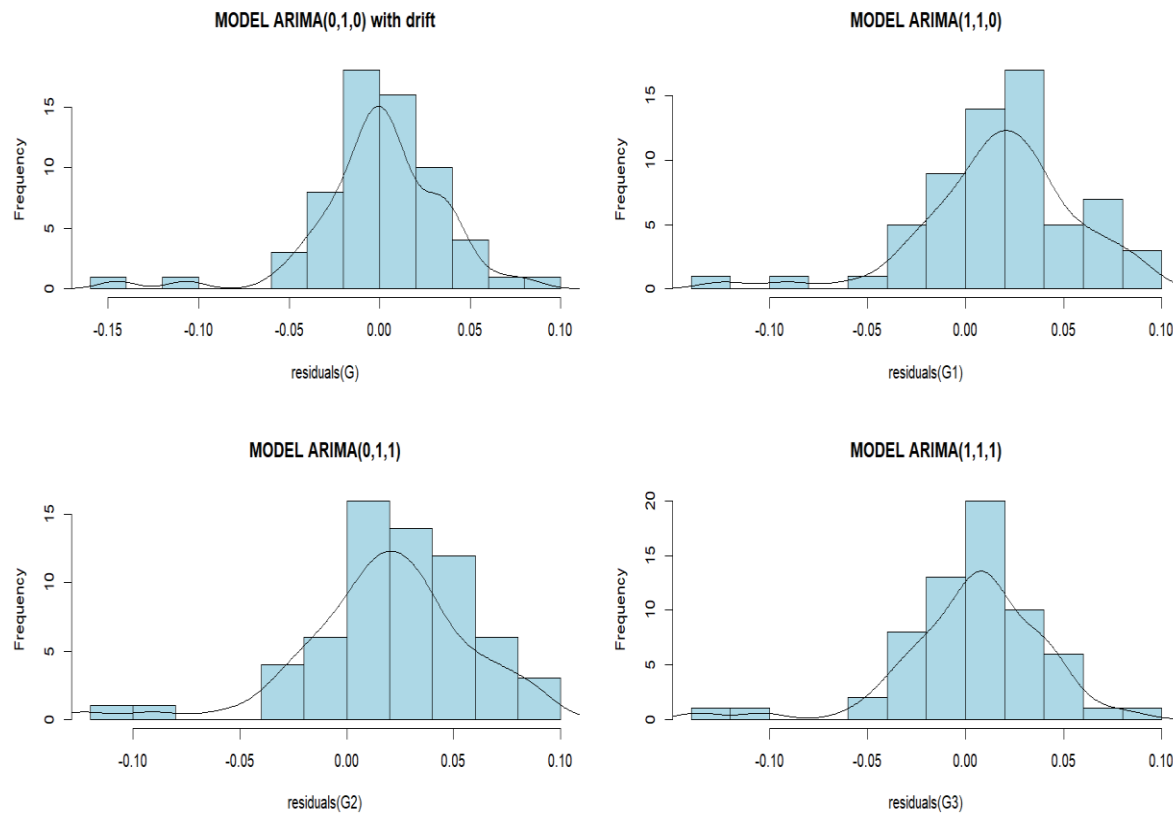


From above graphical representation we observe that the nature or structure of residual plot of the models, ARIMA (1,1,1) and ARIMA (0,1,0) with drift almost same. The residuals are distributed Gaussian distribution.

But we assume to take the model ARIMA (0,1,0) with drift for the value of lowest AIC value.

Also, we plot the histogram of residuals and along with their density plot. We have to observe in which model we have the most of the observations (residuals) are clustered around 0.

• **Diagram 2.9: Histogram plot of residuals**



From diagram 2.9 we can say that for ARIMA (0,1,0) with drift, the number of observations of residuals are clustered around 0 is more than the other three.

Now, we assume that the error of our model ARIMA (0,1,0) with drift is from Gaussian distribution with mean 0 and standard deviation 1.

Working model –

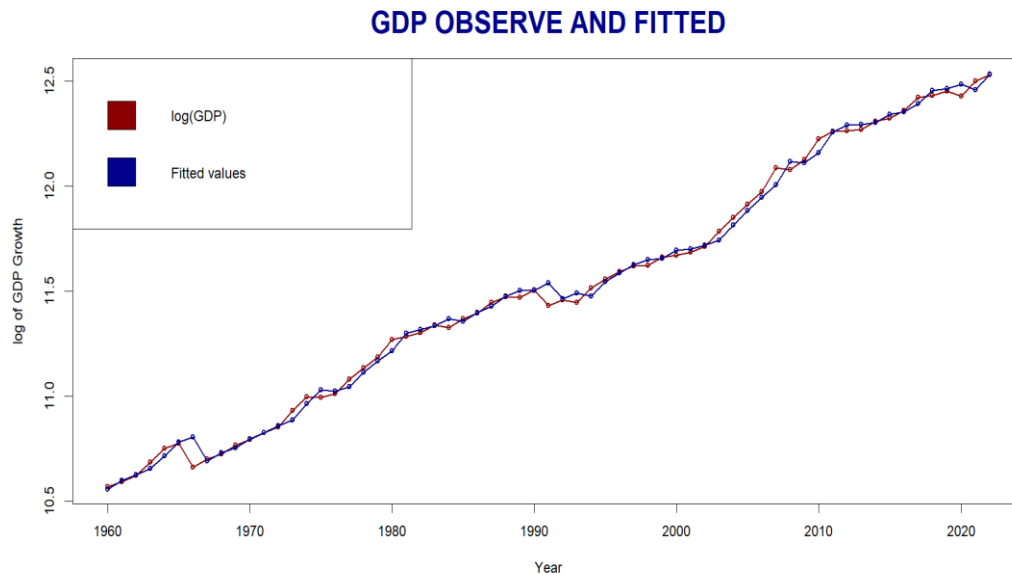
From above some graphical representation and moreover from AIC value we conclude that our working model for forecasting is ARIMA (0,1,0) with drift. Value of the coefficient (drift) is 0.03163947.

[N.B. – Related theory is in Appendix]

Graphical presentation of fitted values –

Now we are going to plot the fitted value along with our data on GDP, given below;

- **Diagram 2.10: Fitted values and given observations**



In the diagram 2.10 the red line is our observations which we got by logarithmic transformation and the red line is our fitted observations.

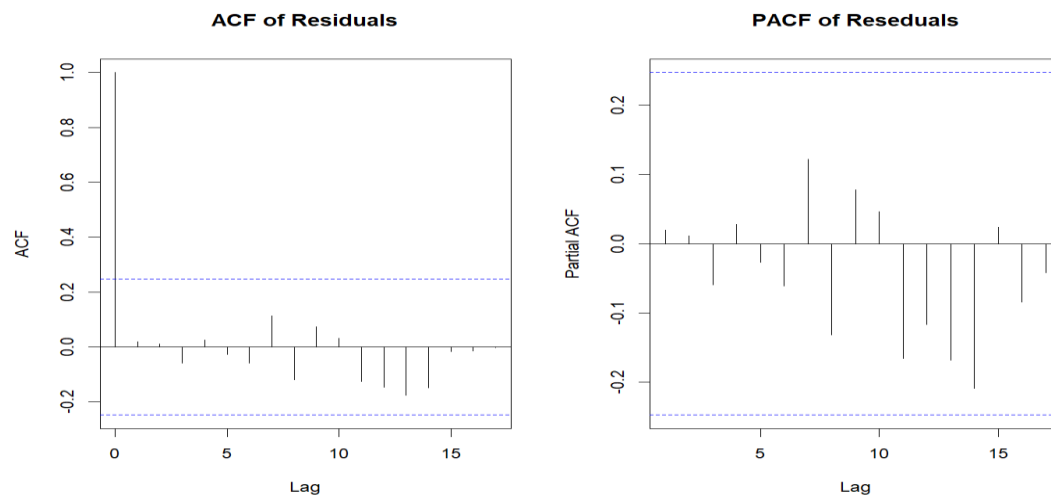
Here are errors of our estimations;

```
> residuals(g)
Time Series:
Start = 1960
End = 2022
Frequency = 1
 [1]  0.0105369076 -0.0065465509 -0.0003688792  0.0284866968  0.0352154073 -0.0086192224
 [7] -0.1450708690  0.0070152725 -0.0068044408  0.0101545933 -0.0030680511  0.0013634576
[13] -0.0059010227  0.0463217740  0.0342529672 -0.0362593117 -0.0121185200  0.0383086723
[19]  0.0222957235  0.0163313903  0.0531820400 -0.0155951086 -0.0140007249  0.0036209628
[25] -0.0437601060  0.0093805867 -0.0027960473  0.0177653828 -0.0044872273 -0.0331042116
[31]  0.0035738480 -0.1067807416 -0.0036107498 -0.0454277578  0.0373040760  0.0101152748
[37]  0.0064505763 -0.0069386931 -0.0264507081  0.0058911163 -0.0232063058 -0.0161435878
[43] -0.0055739829  0.0404568769  0.0351031826  0.0315281438  0.0276745279  0.0815925037
[49] -0.0388180580  0.0162840790  0.0665650100  0.0031226329 -0.0292597716 -0.0245776189
[55]  0.0084777499 -0.0190503460  0.0059767142  0.0317709182 -0.0235215831 -0.0112168020
[61] -0.0569146475  0.0429085275 -0.0024990374
```


ACF & PACF of residuals –

Let us create the ACF & PACF plot for the residuals of our working model to check any pattern in the residuals.

- **Diagram 2.11: ACF & PACF of residuals in ARIMA (0,1,0) with drift**



So, from the above plot we can easily observe that there are no spikes which is crossed the significant line. Hence, we can say that there is no information left in our residuals and it is random in nature. Hence, our ARIMA model is working properly.

Checking presence of auto-correlation in the residuals -

To check any presence of autocorrelation in our residuals we go through a test is called Ljung-Box test. Here the null hypothesis is there is no autocorrelation. In conclusion we can say that the p-value of the test is less than our approximate level α (i.e., take 0.05) then we should reject our null hypothesis against our alternative (i.e., presence of autocorrelation). Here is the value of p-value is 0.9782 (>0.05). Hence, we should accept our null hypothesis indicating absence of auto-correlation in the residuals.

[N.B. – Related theory is in Appendix]

```
> Box.test(ts(residuals(G)), lag=10, type="Lj")  
  
Box-Ljung test  
data: ts(residuals(G))  
X-squared = 3.1295, df = 10, p-value = 0.9782
```

Forecasting

Forecasted values -

Here we are going to begin our main part of analysis that is to forecast for upcoming some years. As we can see the graph of GDP of India is tending to upward growth. Here in this section, we shall forecast from 2023 to 2030. For this purpose, we have to install “forecast” package on R.

```
> require(forecast)
> #Forecast= predict(G,n.ahead = 3)
> G_new=forecast(G,8)
> G_new
  Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2023   12.56184 12.51539 12.60829 12.49079 12.63288
2024   12.59348 12.52778 12.65917 12.49301 12.69395
2025   12.62512 12.54466 12.70558 12.50206 12.74817
2026   12.65676 12.56385 12.74967 12.51467 12.79885
2027   12.68840 12.58452 12.79227 12.52954 12.84726
2028   12.72004 12.60625 12.83382 12.54601 12.89406
2029   12.75168 12.62877 12.87458 12.56371 12.93964
2030   12.78332 12.65192 12.91471 12.58237 12.98426
```

Here we present our forecasting through tabulated version.

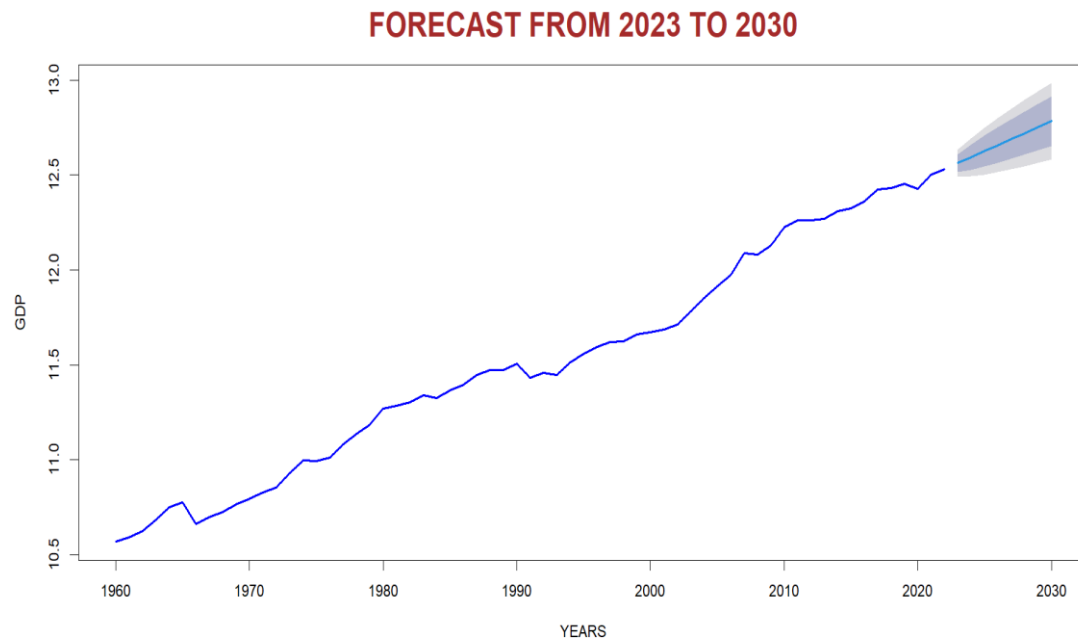
Table 5: Forecasted Observations from the year 2023 to 2030

Years	Forecasted value (in log base 10 form) = x	Forecasted value (10^x)
2023	12.56184	3.646196e+12
2024	12.59383	3.924913e+12
2025	12.62512	4.218130e+12
2026	12.65676	4.536908e+12
2027	12.68840	4.879777e+12
2028	12.72004	5.248558e+12
2029	12.75168	5.645209e+12
2030	12.78332	6.071836e+12

Plot of forecasted values -

Now we plot the forecasted values, here it is given below;

- **Diagram 2.12: Plot of forecasting**



So, here we give 8 points forecast that is from 2023 to 2030. From above graphical plot we can say that the net GDP will be in increasing nature. It will be 6071.836 billion USD for the year 2030 where on 2022, it was just 3390 billion USD.

Special Analysis

Data Source:

<https://www.financialexpress.com/economy/india-may-overtake-germany-japan-too-after-beating-uk-to-become-worlds-third-largest-economy/2655867/>

Description of the Source:

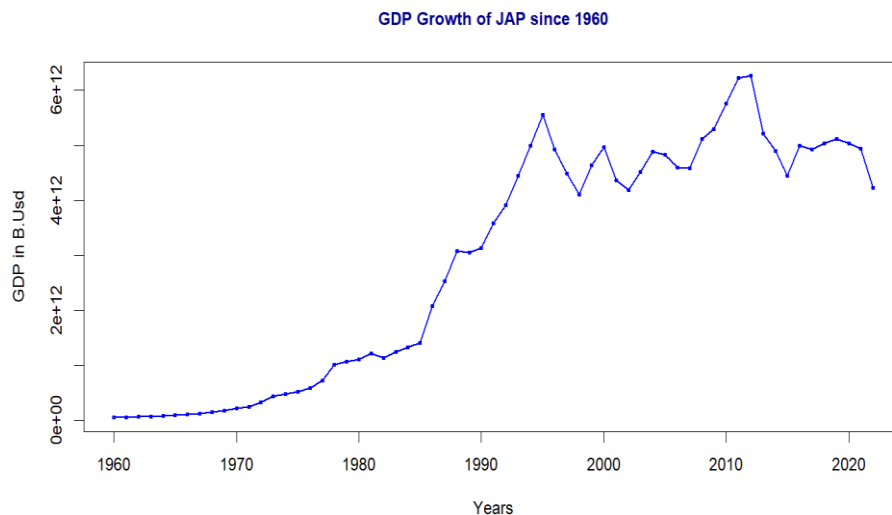
According to analysts, “India should surpass Germany in 2027 and most likely Japan by 2029 at the current rate of growth, a remarkable achievement by any standards.”

Own Analysis:

For Japan

First of all, we take the data on GDP of Japan from 1960 to 2022. We got the data from the site <https://www.macrotrends.net/countries/JPN/japan/gdp-gross-domestic-product>. Then we fit a suitable model for the dataset as we have done before for India. For better visualisation we provide the graphical nature of the GDP growth over the years, given below;

• Diagram 2.13: GDP growth of Japan since 1960 to 2022



Now we get ARIMA (0,2,2) model by R. Here is the result given below;

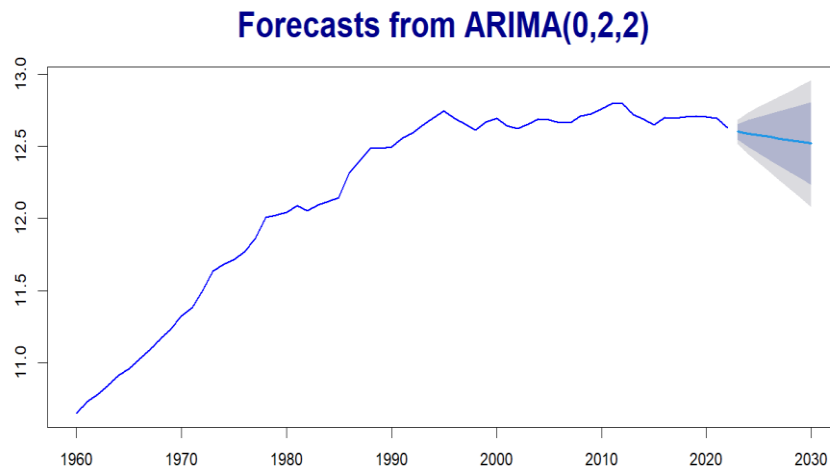
```
Series: g_new
ARIMA(0,2,2)

Coefficients:
      ma1      ma2
    -0.5967  -0.2330
s.e.    0.1255   0.1172

sigma^2 = 0.001838:  log likelihood = 106.12
AIC=-206.23  AICc=-205.81  BIC=-199.9
```

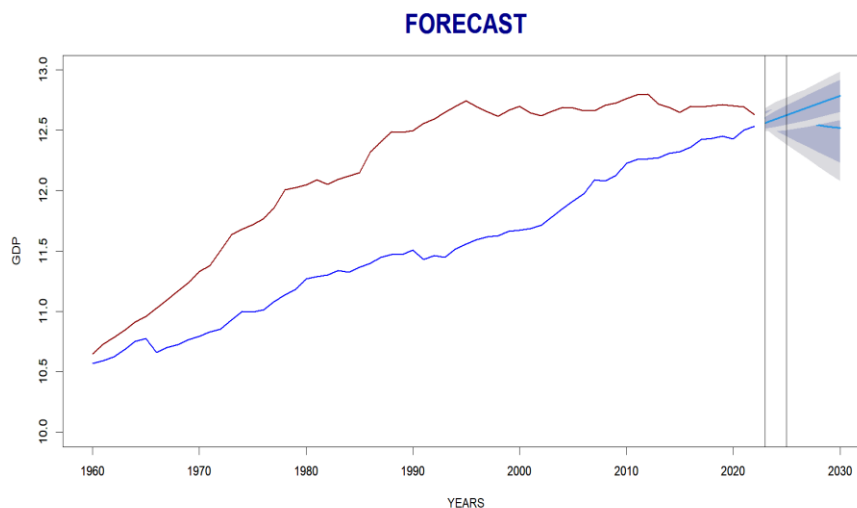
Here we get the model where MA of order 2 and differencing based on second order differencing. Now we plot the forecasted value from 2023 to 2030 where in X axis we take successive years & on Y axis logarithmic transformation on GDP values.

• **Diagram 2.14: Forecasted values of Japan 2023-2030**



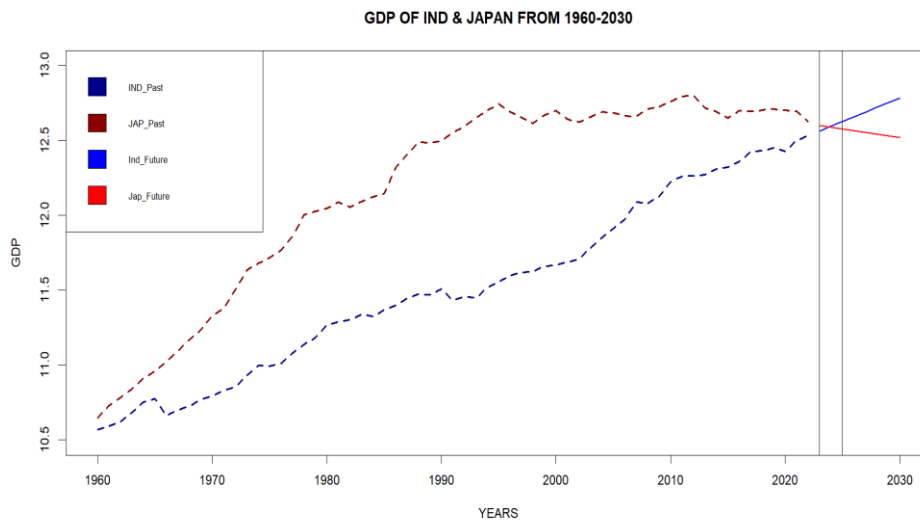
Now we are interested to compare this plot with India. We overlap two graphical plots on forecasting and observe the superimposed area of the intersection area from which we can make sense about on which year India overtake Japan.

• **Diagram 2.15: Forecast comparison of India & Japan**



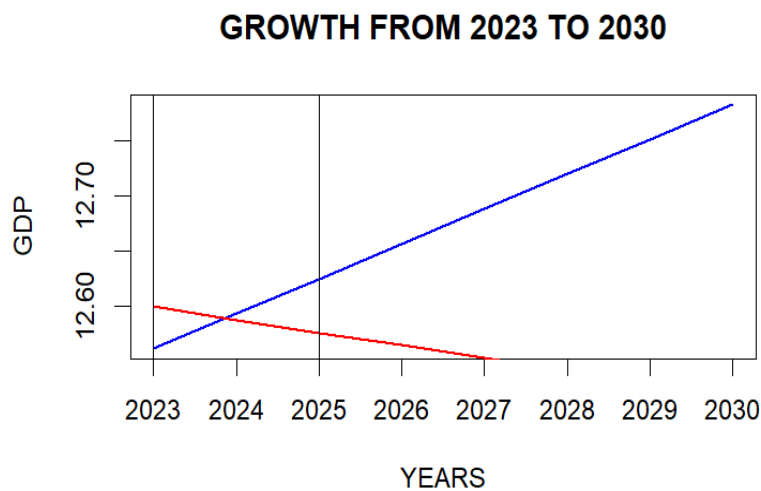
Here in the above model, we draw two vertical lines for the years 2023 and 2025. But we can't visualise its properly. So, our simplicity of understanding we plot a curve from 1960 to 2030 for both countries India and Japan.

• **Diagram 2.16: GDP from 1960 to 2030 for both Japan & India**



Here red dotted line is past observation (1960-2022) of Japan, blue dotted line is past observation of India, and continuous lines are in blue and red is for forecasted years for India and Japan respectively. Again, for simplicity in our understanding we only focus on the plot from 2023 to 2030.

• **Diagram 2.17: Magnification of the diag. 2.16 near the cutting points**



From the above graphical representation, it's clear that GDP growth curve of India will exceed the GDP growth curve of Japan about the interval between 2023 and 2025.

But as we can see in the data source that India will overtake Japan's GDP around 2029 but based on our analysis, we see that this will be occur more earlier that is somewhere in between the years 2023 and 2025.

Hence there is a contradiction in years but the nature of the growth comparison is same that is India's GDP will overtake Japan's GDP growth curve.

For Germany

First of all, we take the data on GDP of Germany from 1960 to 2022. We got the data from the site <https://www.macrotrends.net/countries/DEU/germany/gdp-gross-domestic-product>. Then we fit a suitable model for the dataset as we have done before for India. For better visualisation we provide the graphical nature of the GDP growth over the years, given below;

- **Diagram 2.18: GDP growth of Germany since 1960 to 2022**



Now we get ARIMA (0,2,2) model by R. Here is the result given below;

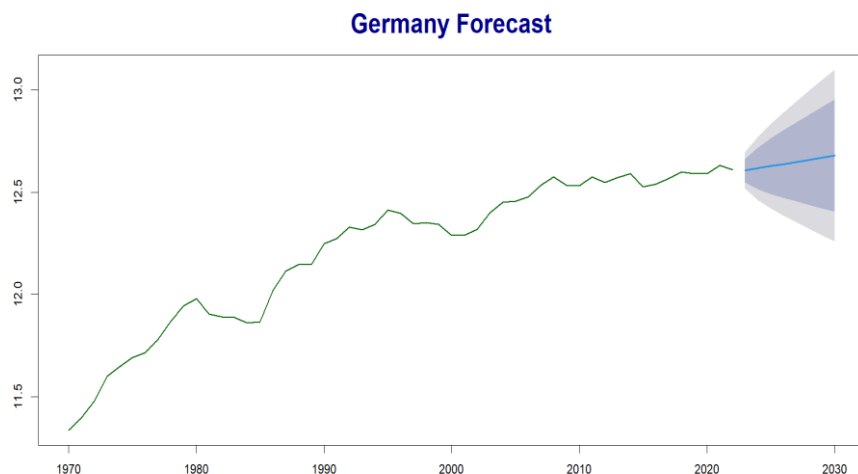
```
Series: g_new2
ARIMA(0,2,2)

Coefficients:
      ma1      ma2
    -0.5697  -0.3270
s.e.    0.1348   0.1295

sigma^2 = 0.002047: log likelihood = 85.85
AIC=-165.7   AICc=-165.19   BIC=-159.9
```

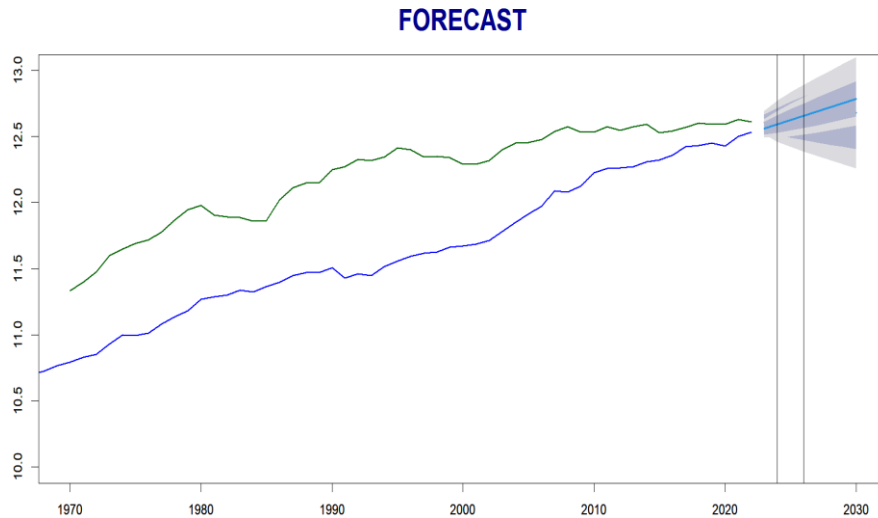
Here we get the model where MA of order 2 and differencing based on second order differencing. Now we plot the forecasted value from 2023 to 2030 where in X axis we take successive years & on Y axis logarithmic transformation on GDP values.

- **Diagram 2.19: Forecasted values of Germany 2023-2030**



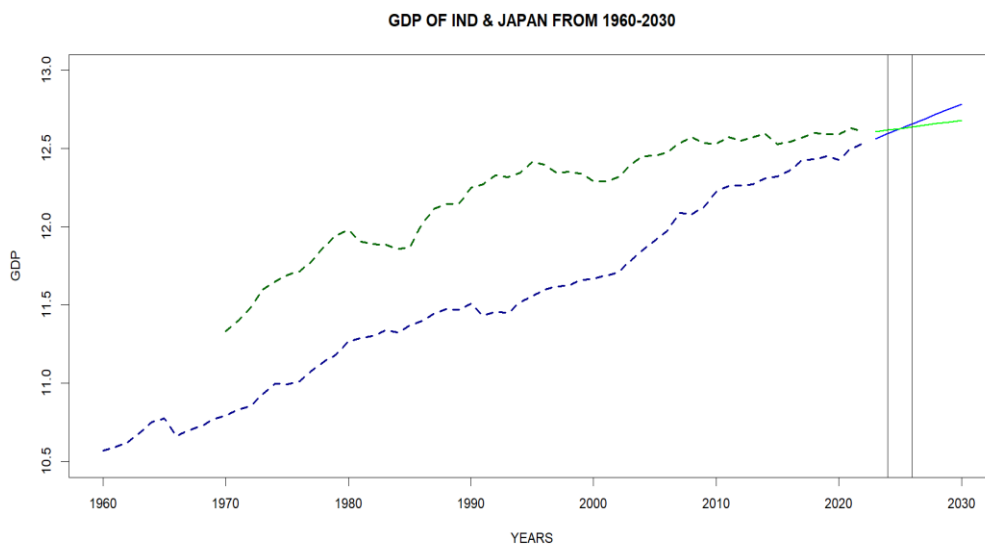
Now we are interested to compare this plot with India. We overlap two graphical plots on forecasting and observe the superimposed area of the intersection area from which we can make sense about on which year India overtake Germany.

• **Diagram 2.20: Forecast comparison of India & Japan**



But we can't visualise in which year India overtakes properly from the above graphical diagram. So, for our simplicity of understanding we plot a curve from 1960 to 2030 for India and 1970 (as we have data of Germany from 1970) to 2030 for Germany.

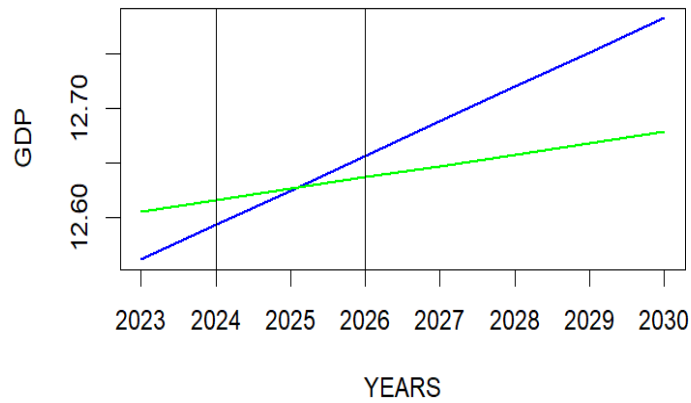
• **Diagram 2.21: GDP from 1960 to 2030 for both India & Germany**



Here green dotted line is past observations (1970-2022) of Germany, blue dotted line is past observation of India, and continuous lines in blue and green is for forecasted years for India and Germany respectively. Again, for simplicity in our understanding we only focus on the plot from 2023 to 2030.

• **Diagram 2.22: Magnification of the diag. 2.21 near the cutting points**

GROWTH FROM 2023 TO 2030



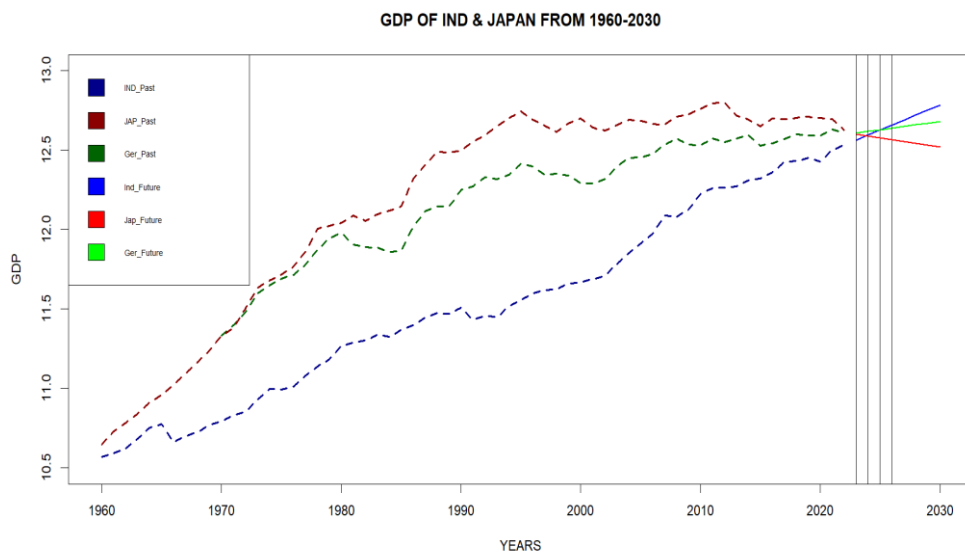
From here we can say that GDP growth curve of India will exceed the GDP growth curve of Germany at somewhere in between the years 2024 and 2026.

But as we can see in the data source that India will overtake Germany's GDP around 2027 but based on our analysis, we see that this will occur more earlier that is in between the year 2024 and 2026.

Hence, there is a contradiction in years but the nature of the GDP growth comparison is same that is India's GDP will overtake Germany's GDP curve.

Now we plot 3 countries together and observe the change-

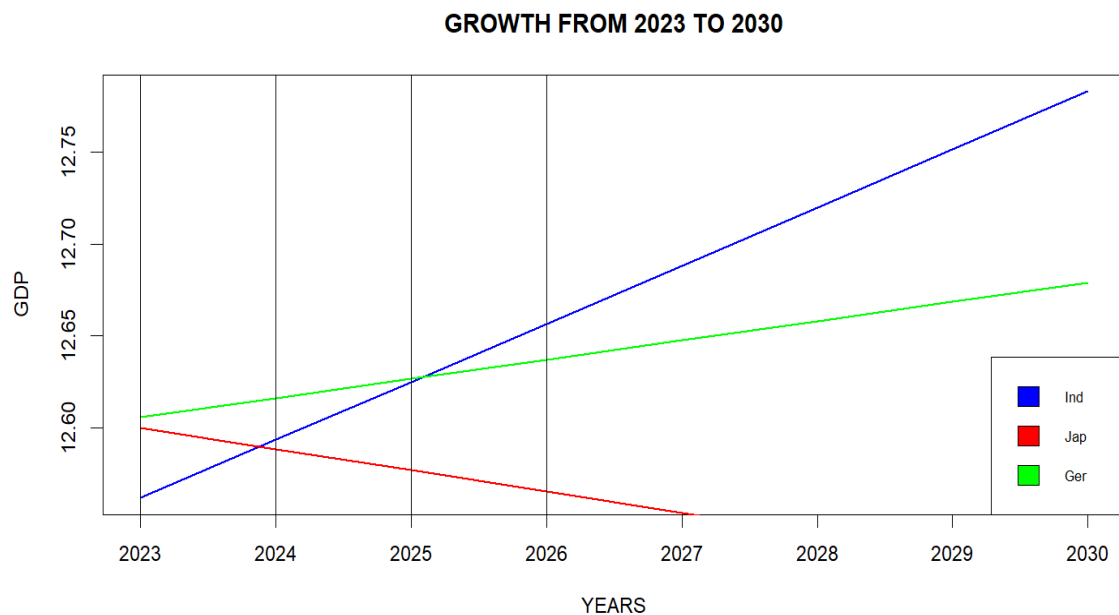
• **Diagram 2.23: GDP from 1960 to 2030 for India, Japan & Germany**



Here, red dotted line is past observation (1960-2022) of Japan, green dotted line is past observation (1970-2022) of Germany, blue dotted line is past observation of India, and continuous lines in blue, red and green is for forecasted years for India, Japan & Germany respectively.

Again, for simplicity in our understanding we only focus on the plot from 2023 to 2030

• **Diagram 2.24: Magnification of the diag. 2.23 near the cutting points**



As we can see, GDP of India will overtake Japan and Germany between the years 2023-2025 and 2024-2026 respectively.

Conclusion:

For Objective 1

Out of nine factors in model 1 we've found that seven factors are found to be significant as their p-value come less than 0.05. These significant factors are *export of the country, consumer spending, GNP, FDI, external debt, GDP per capita, unemployment rate* of a country. Among them consumer spending, *GNP, FDI* are highly significant for our model. Correlation between these three with GDP is 0.9962289, 0.9959903, 0.9661529 respectively. Which shows high correlation between them. Also, two factors are less significant in our model 1 that are, *imports of a country* and *inflation rate of a country*. As we said in our earlier intuition at the starting of our analysis to observe the correlated values on significant factors and less significant factors.

This study found that export of a country, GNP, consumer spending, unemployment rate, GDP per capita has a **positive impact** on our study variable GDP of a country in model 1. On the other hand, it is also found that there is a **negative impact** of FDI, external debt on our study variable GDP of a country in model 1. On the model 1 summary above R square closer to 1 means that **99.99%** of the variations on the study variable are explained by the variability of the explanatory variables. R which is the square root of R square shows the correlation between study variable and explanatory variables. 99.99% means that there is high correlation between study and explanatory variables.

The other two less significant factors, *inflation rate* and *import of a country* have been weak in explaining.

In the model 3 out of eight factors in model 1 we've found that eight factors are found to be significant as their p-value come less than 0.05. These significant factors are *export of the country, import of a country, consumer spending, GNP, FDI, external debt, GDP per capita, unemployment rate* of a country.

This study found that export of a country, GNP, consumer spending, unemployment rate has a **positive impact** on our study variable GDP of a country in model 3. On the other hand, it is also found that there is a **negative impact** of import of a country, FDI, external debt, GDP per capita on our study variable GDP of a country in model 3.

In the model 3 summary above R square closer to 1 means that **99.99%** of the variations on the study variable are explained by the variability of the explanatory variables. R which is the square root of R square shows the correlation between study variable and explanatory variables. 99.99% means that there is high correlation between study and explanatory variables.

Also, we get the p-value for model 3 of the Shapiro-Wilk test is 0.7344 (>0.05, preassigned level of the test) it says that our assumption is correct that is the error follows Gaussian distribution with parameters mean=zero and variance = $\sigma^2(>0)$.

```
> shapiro.test(e1)

Shapiro-Wilk normality test

data:  e1
W = 0.96069, p-value = 0.7344
```

As least AIC and best fitting we should take model 3 for obtaining the effect on GDP.

So, the factors; *export of a country, consumer spending of a country, GNP, FDI, external debt, GDP per capita, unemployment rate* and *import of a country* are the **significant** factors that affect GDP growth of our selected developing countries.

For Objective 2

As we said that by forecasting, we can't predict appropriately and it's so much unpractical to predict exactly. But we can say something about changes about future through some theoretical steps or guess this.

So, here we give 8 points forecast that is from 2023 to 2030. From graphical nature we conclude that the net GDP will be in increasing nature. It will be 6071.836 billion US Dollar for the year 2030 where on 2022, it was just 3390 billion US Dollar. That is, we can say according our analysis that GDP will be about to be doubled after 7 years.

Also, we can say that India's economic health will be got more stronger than before as per our prediction. But if there will be some natural calamities or any pandemic situation just like the year 2020 then it will be deviate from our assumption.

According to the World GDP ranking 2023 list, India is the 5th largest economy in the world. Other countries before India are United States of America, China, Japan, Germany etc. As per some articles or newspaper report India will be the 3rd largest after some years and will overtake Japan and Germany. Now at the end of our study we would give a conclusion on comparison between my analysis and an article report that whether there is a chance of overtaking both countries or not.

For Special Analysis

We have done in our entire analysis on time series data of three countries. Basically, we are interested on India's forecasting. There, we noticed that economic growth has positive trend. Throughout our comprehensive analysis of the Indian economy and rigorous forecasting methodologies, this project aimed to project the GDP growth trajectory of India for the period 2023 to 2030. The findings of the analysis suggest a positive outlook, with India's GDP growth expected to be upward during the forecasted period. During the historical period considered in the analysis, India has displayed growth potential, even in the face of global challenges. The economy has demonstrated the ability to recover from downturns and achieve higher growth rates.

The forecasted upward GDP growth presents an opportunity for policymakers to leverage India's economic potential and further foster an enabling environment for sustainable growth. Targeted policy interventions in areas like education, skill, development, innovation and infrastructure can support the growth momentum. While the forecast indicates positive growth, it is essential to be aware of potential risks that could impact the trajectory. These risks may include global economic uncertainties, geopolitical developments, and domestic challenges. Mitigation strategies and contingency plans can help mitigate adverse impacts.

This project opens avenues for future research and analysis to delve deeper into specific sectors, regional disparities, and the impact of external factors on India's GDP growth. Additionally, exploring the implications of environmental sustainability and social inclusiveness on economic growth can be areas of further investigation.

As we can see, GDP of India will overtake Japan and Germany between the years 2023-2025 and 2024-2026 respectively and become *3rd largest economy in the world* after USA and China.

GDP of India on 2023 will be 3646.196 billion USD comparatively in Japan it will be 3981.988 billion USD and in Germany it will be 4033.759 billion USD.

GDP of India on 2024 will be 3921.751 billion USD comparatively in Japan it will be 3877.305 billion USD. Hence, India will *overtake* GDP of Japan between the mentioned year's interval.

GDP of India on 2025 will be 4218.13 billion USD comparatively in Germany it will be 4235.351 billion USD. GDP of India on 2026 will be 4536.908 billion USD comparatively in Germany it will be 4336.806 billion USD.

Hence, India will *overtake* GDP of Germany between the mentioned year's interval.

In conclusion, the forecasted upward trajectory of India's GDP growth bodes well for the country's economic prospects in the coming years. As India positions itself as a major player in the global economy, strategic policies and collaborative efforts will be pivotal in unlocking the full potential of its growth opportunities. This project provides valuable insights into India's economic outlook, and it is hoped that its findings will contribute to informed decision-making and policy formulation to foster sustainable and inclusive economic growth in India.

APPENDIX

● THEORETICAL FRAMEWORK

1. How GDP of a country is formulated?

For encounter this question we said that, GDP of a country is a measure of the total value of all goods and services produced within its border over a specific period of time, typically a year. It's often used as an indicator of a country's economic performance.

The formula for calculating GDP of a country is given by,

$$\mathbf{GDP = C+I+G+(X-M);}$$

where,

C represents, consumer spending or private consumption expenditure,

I represents, investment spending (including business and residual investment),

G represents, government spending on goods & services,

X represents, exports of goods & services,

M represents, imports of goods & services.

In other words, GDP is the sum of consumer spending, govt. spending and net exports (exports minus imports). It's important to note that there are different methods for calculating GDP, such as the expenditure approach, income approach and production approach. The formula provided above is based on the expenditure approach, which is commonly used.

It's also worth mentioning that GDP can be calculated on a nominal basis (current prices) or a real basis (constant prices adjusted for inflation). The formula provided above calculates GDP on a nominal basis. To calculate real GDP, the values of goods and services are adjusted for inflation using a price index, such as the Consumer Price Index (CPI) or the GDP deflator.

2. What is Regression Analysis?

Regression analysis is conceptually simple method for investigating functional relationships among variables. The relationship is expressed in the form of an equation or a model connecting the response or dependent variable and one or more explanatory variables or predictor variables. In our analysis we have already specified our study and explanatory variables in data description section. Now we're going to discuss about what is multiple linear regression.

3. What is Multiple Linear Regression?

Consider a dataset which consists of n observations on a dependent or response variable Y ($y_1, y_2, y_3, \dots, y_n$) and p predictor or explanatory variables, X_1, X_2, \dots, X_p . The relationship is formulated as a linear model;

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Where, $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ are constants referred to as the model *regression coefficients* and ε is a *random error*.

The above equation may be written as,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

So, based on any n-tuple data our first motto is to estimate the parameters that is coefficients. Here we use the method of least squares parameters estimation. In this method we minimise the sum of squares of the errors. The sum of squares of the error is,

$$S(\beta_0, \beta_1, \beta_2, \dots, \beta_p) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 X_{i1} - \dots - \beta_p X_{ip})^2$$

By direct application of calculus, it can be shown that the LS (least sq.) estimates $\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_p$, which minimise $S(\beta_0, \beta_1, \beta_2, \dots, \beta_p)$, are given the solution of a system of linear equations. The estimate $\widehat{\beta}_0$ is referred to as the intercept or constant, and other estimated β 's is called the estimate of the regression coefficient of the predictor X_j . By least square method estimates are given below;

$$\widehat{\beta}_0 = \bar{y}$$

$$\widehat{\beta}_j = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_i - \bar{y})}{\sum_{i=1}^n (x_{ij} - \bar{x})^2}$$

So, our linear model equation becomes,

$$\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \dots + \widehat{\beta}_p x_{ip}, \quad i=1, 2, \dots, n$$

From the above estimation formula, we can get the estimated value of regression coefficients and the slope of our multiple linear regression model.

4. What is AIC?

Akaike Information Criterion (AIC), which is a commonly used statistical measure for model selection. The AIC is a measure of the relative quality of the statistical model for a given dataset. It was developed by Japanese statistician Hirotugu Akaike. AIC is based on information theory and provides a trade-off between the goodness of fit of the model and its complexity. The principal behind the AIC is to find the model that best balances the fit to the

data with the number of parameters used in the model. The AIC value calculated using the formula:

$$\text{AIC} = (-2) * \log\text{-likelihood} + 2 * (\text{number of parameters})$$

Where log-likelihood represents the goodness of fit of the model to the data and the number of parameters reflects the complexity of the model. *The lower AIC value represents the best fit of the model.*

AIC is commonly used in various statistical techniques, such as linear regression, generalised linear model, time series analysis.

5. What is ARIMA model?

ARIMA stands for Autoregressive Integrated Moving Average. It's a popular time series model for forecasting which combine the components autoregressive (AR), differencing (I), moving average (MA). AR refers to the dependence of the current value of the time series on its previous values. The AR component uses the past values of the series to predict the future values. It assumes that future values are in linear combination of the past values. Differencing is used to make the time series stationary. Stationarity means that the statistical properties of the series such as mean and variance don't change over time. MA refers to the dependency between the current value of the time series and a linear combination of past forecast errors. The MA component helps capture the random shocks or noise in the time series that can't be explained by the AR component. ARIMA models are defined by three parameters; p, d and q. The parameter p refers to the order of AR, which indicates how many past values are considered in the prediction. The parameter d refers to the order of differencing. The parameter q refers to the order of MA component, which signifies the number of past forecast errors that are considered in the prediction.

By notation ARIMA model writes as **ARIMA (p,d,q)**

Necessary equation is given below;

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \dots + \beta_q e_{t-q}$$

6. What is ARIMA (0,1,0) with drift model?

In our second objective we use this model for better forecast. The ARIMA (0,1,0) model with drift specially known as a random walk with drift, is a specific type of series model used for forecasting. This model is often employed when the data exhibits the trend, but the first difference of the data (that is, the difference between two consecutive observations) show no consecutive autocorrelation.

The term drift represents a constant trend component in the model. It allows for a gradual or constant increase or decrease in the forecasted values over time. The drift term captures the long-term trend or growth pattern of the data.

The mathematical structure of this model can be described as follows;

Let, Y_t represent the observed time series data at time t . The model will be expressed as:

$$\Delta Y_t = \mu + \varepsilon_t$$

Where,

- ΔY_t represents the differenced series at time t . It's calculated as the difference between the value at time t and the value at time $(t-1)$.
- μ is the drift term, representing the average change or trend in the series.
- ε_t represents the error term, which is assumed to be a random walk process, with the drift term μ accounting for an overall trend or systematic change in the series.

It's important to note that the ARIMA (0,1,0) model doesn't include autoregressive (AR) or moving average (MA) components. It simply captures the trend or drift in the data through the parameter and assumes that the changes in the series are primarily random or unpredictable as represented by the error term ε_t .

7. What is Shapiro-Wilk test?

Shapiro-Wilk test is a statistical test used to assess the normality of a dataset. It tests the null hypothesis that the data are drawn from a normally distributed population. The test is particularly useful when the sample size is relatively small.

The Shapiro-Wilk statistic is given by,

$$W = \frac{(\sum_{i=1}^n a_i X_{(i)})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Where,

$X_{(i)}$ represents the i -th order statistic of the sample, which is the i -th smallest value after sorting the data in ascending order.

\bar{X} be the sample mean.

a_i be the coefficients provided by Shapiro & Wilk's method, which depend on the sample size and have been precalculated and tabulated.

The Shapiro-Wilk test statistic W measures the discrepancy between the observed data and the expected values under the assumption of normality. A value of W close to 1 indicates that the data are well approximated by a normal distribution, while a value significantly smaller than 1 suggests departure from normality.

8. What is Ljung-Box test?

Ljung-Box test uses for to check whether there is significant auto-correlation between in a time series. It is commonly employed to test the null hypothesis that the autocorrelation of a time series up to a specific lag are zero, indicating no significant residual autocorrelation. The Ljung-Box test is based on following statistic,

$$Q = T(T + 2) \sum_{k=1}^h (T - k)^{-1} r_k^2$$

Where,

h be the maximum lag being considered,

T be the number of observations,

r_k be the sample autocorrelation at lag k

If each residual is close to zero, then Q will be small. If some values are large (positive or negative), then Q will be large. We suggest using h=10 for non-seasonal data and h=2m for seasonal data, where m is the period of seasonality. So, large values of Q suggest that the autocorrelations do not come from a white noise series.

9. What is ACF & PACF?

Let X_t and X_{t+h} be two discrete time points h be the time lag the auto-covariance function at lag h be defined as

$$\gamma(h) = E[X_t - \mu][X_{t+h} - \mu]$$

If mean function be μ and variance function $\gamma(0) = \sigma^2$, then auto correlation function (ACF) defined as;

$$cor(X_t, X_{t+h}) = \rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\gamma(h)}{\sigma^2}$$

In addition to ACF X_t & X_{t+h} we may want to investigate the correlation between X_t & X_{t+h} after their mutual linear dependency on the inter varying variables $X_{t+1}, X_{t+2}, \dots, X_{t+h-1}$ has been removed. Thus, we construct partial auto correlation coefficient (PACF).

$$cor(X_t, X_{t+h} | X_{t+1} X_{t+2} \dots X_{t+h-1})$$

ACF & PACF both are statistical measures and PACF is a extension form of ACF.

• R – Code

For objective 1 –

```
data<-read.csv(file.choose())

data

GDP<-data$GDP

EXPORTS<-data$Exports

IMPORTS<-data$Imports

CONSUMER_SPENDING<-data$Con.Sp

GNP<-data$GNP

FDI<-data$FDI

EXTERNAL_DEBT<-data$Ext.debt

GDP_PER_CAPITA<-data$GDP..cap

UNEMPLOYMENT_RATE<-data$Unp.rate

INFLATION_RATE<-data$Inf.rate

plot(GDP,EXPORTS,col="dark red",cex=0.8)

plot(GDP,IMPORTS,col="dark red",cex=0.8)

plot(GDP,CONSUMER_SPENDING,col="dark red",cex=0.8)

plot(GDP,GNP,col="dark red",cex=0.8)

plot(GDP,FDI,col="dark red",cex=0.8)

plot(GDP,EXTERNAL_DEBT,col="dark red",cex=0.8)

plot(GDP,GDP_PER_CAPITA,col="dark red",cex=0.8)

plot(GDP,UNEMPLOYMENT_RATE,col="dark red",cex=0.8)

plot(GDP,INFLATION_RATE,col="dark red",cex=0.8)

#analysis with MLR

summary(lm(GDP~EXPORTS+IMPORTS+CONSUMER_SPENDING+GNP+FDI+EXTERNAL_DEBT+GDP_PER_CAPI
TA+UNEMPLOYMENT_RATE+INFLATION_RATE))

y_hat<fitted.values(lm(GDP~EXPORTS+IMPORTS+CONSUMER_SPENDING+GNP+FDI+EXTERNAL_DEBT+
GDP_PER_CAPITA+UNEMPLOYMENT_RATE+INFLATION_RATE))

y<-data$GDP

e=y-y_hat

shapiro.test(e)

plot(density(e),main="Residual Density Plot")

model3<lm(GDP~EXPORTS+CONSUMER_SPENDING+GNP+FDI+EXTERNAL_DEBT+GDP_PER_CAPITA+UNEMPL
OYMENT_RATE+IMPORTS)

summary(model3)

step(model3)
```

```

cbind(fitted.values(model3))

e1=data$GDP-fitted.values(model3)

shapiro.test(e1)

model2<-
lm(GDP~EXPORTS+CONSUMER_SPENDING+GNP+FDI+EXTERNAL_DEBT+GDP_PER_CAPITA+UNEMPLOYMENT_
RATE+INFLATION_RATE)

summary(model2)

step(model2)

model4<-lm(GDP~EXPORTS+CONSUMER_SPENDING+GNP+FDI+EXTERNAL_DEBT+GDP_PER_CAPITA+
UNEMPLOYMENT_RATE)

summary(model4)

step(model4)

```

For objective 2 –

```

data2<-read.csv(file.choose())

data2

Years<-data2$Year

GDP_Rate<-data2$GDP.growth..annual...

library(ggplot2)

ggplot(data2,aes(Years,GDP_Rate),col="red")+geom_area(fill="light blue",col="dark
blue")+ggtitle("Yearwise GDP Rate")

#gdp vs years

gdp=ts(data2$GDP..current.US., start = 1960, end = 2022)

plot(gdp,type="o", xlab="Years",ylab="GDP in B.Usd", main="GDP Growth of IND since
1960",cex=.5,col="Blue",col.main="Dark Blue", cex.main=1,cex.lab=1.7)

# remove trend & seasonal component

plot(diff(gdp),type="o", xlab="Years",ylab="Differenced Total GDP", main="Removal
of Trend",cex=.5,col="Dark Red",col.main="Dark Blue",
col.lab="Purple",cex.main=1,cex.lab=.7)

#multiplicative model as additive model

gdp_new=log10(gdp)

plot(gdp_new,type="o",cex=0.5,xlab="Year",ylab="GDP new",main="Multiplicative Model
to Additive Model, GDP Growth",col="dark green",col.main="Dark
Blue",col.lab="Purple",cex.main=1,cex.lab=.7)

plot(diff(gdp_new),xlab="Years", ylab="Differenced log(GDP)",
col="Brown",type="o",cex=0.5, main="Years vs log(GDP_new)",col.main="dark
blue",col.lab="dark green", cex.main=1,cex.lab=0.7)

#Model identification

par(mfrow=c(1,2))

acf(ts(diff(gdp_new)),main='ACF of GDP',cex.main=1.2)

```

```

pacf(ts(diff(gdp_new)),main='PACF of GDP')

require(forecast)

G1=arima(gdp_new, order = c(1,1,0))
G2=arima(gdp_new, order = c(0,1,1))
G3=arima(gdp_new, order = c(1,1,1))

G = auto.arima(gdp_new)

summary(G)

par(mfrow=c(1,2))

residuals(G)

par(mfrow=c(2,2))

plot(residuals(G),main="ARIMA(0,1,0) WITH DRIFT" , col= "Dark red")
plot(residuals(G1), main="ARIMA(1,1,0)",col= "Dark red")
plot(residuals(G2),main="ARIMA(0,1,1)",col= "Dark red")
plot(residuals(G3), main="ARIMA(1,1,1)",col= "Dark red")

residuals(G)

par(mfrow=c(2,2))

hist(residuals(G),breaks=10, main = "MODEL ARIMA(0,1,0) with drift",col="light
blue")

lines(density(residuals(G)))

hist(residuals(G1),breaks=10, main="MODEL ARIMA(1,1,0)",col="light blue")
lines(density(residuals(G1)))

hist(residuals(G2),breaks=10, main="MODEL ARIMA(0,1,1)",col="light blue")
lines(density(residuals(G1)))

hist(residuals(G3),breaks=10, main="MODEL ARIMA(1,1,1)",col="light blue")
lines(density(residuals(G3)))

par(mfrow=c(2,2))

qqnorm(residuals(G),col="Dark Green",pch=1,cex=1, main="ARIMA(0,1,0) WITH DRIFT")
qqline(residuals(G),col="dark blue")

qqnorm(residuals(G1),col="Dark Green",pch=1,cex=1, main="ARIMA(1,1,0)")
qqline(residuals(G1),col="dark blue")

qqnorm(residuals(G2),col="Dark Green",pch=1,cex=1, main="ARIMA(0,1,1)")
qqline(residuals(G2),col="dark blue")

qqnorm(residuals(G3),col="Dark Green",pch=1,cex=1, main="ARIMA(1,1,1)")
qqline(residuals(G3),col="dark blue")

acf(ts(residuals(G)),main="ACF of Residuals")

pacf(ts(residuals(G)),main="PACF of Residuals")

Box.test(ts(residuals(G)),lag=10, type="Lj")

```

```

G_fit=gdp_new-residuals(G)

plot(gdp_new,type="o",cex=0.7,xlab="Year",ylab="log of GDP Growth", main="GDP
OBSERVE AND FITTED",col="dark red",col.main="Dark
Blue",col.lab="Purple",cex.main=2,cex.lab=1)

lines(G_fit,type = "o",cex=0.7,col="dark blue")

legend("topleft", c("log(GDP)","Fitted values"),fill = c("Dark Red","Dark
Blue"),cex=1)

G_new=forecast(G,8)

par(mfrow = c(1,2))

plot(gdp_new,type="l", xlab="Years",ylab="GDP in B.Usd",main="GDP Growth of IND
since 1960",cex=.5,col="Blue",col.main="Dark Blue", cex.main=1,cex.lab=1.7)

plot(forecast(G,8),col = "blue",cex.main=2,col.main="dark
blue",lwd=2,main="FORECAST FROM 2023 TO
2030",col.main="brown",xlab="YEARS",ylab="GDP")

#JAPAN

data3<-read.csv(file.choose())

data3

gdp2=ts(data3$GDP, start = 1960, end = 2022)

plot(gdp2,type="o", xlab="Years",ylab="GDP in B.Usd", main="GDP Growth of JAP since
1960",cex=.5,col="Blue",col.main="Dark Blue", cex.main=1,cex.lab=3)

g_new=log10(gdp2)

plot(g_new,type="o",cex=0.5,xlab="Year",ylab="GDP new", main="Multiplicative
Model to Additive Model, GDP Growth",col="dark green",col.main="Dark
Blue",col.lab="Purple",cex.main=1,cex.lab=.7)

require(forecast)

gd=auto.arima(g_new)

G_new=forecast(gd,8)

plot(forecast(gd,8),col = "blue",cex.main=2,col.main="dark blue",lwd=1.5)

#GERMANY

data4<-read.csv(file.choose())

data4

gdp3=ts(data4$GDP.in.USD, start = 1970, end = 2022)

plot(gdp3,type="o", xlab="Years",ylab="GDP in B.Usd", main="GDP Growth of GERMANY
since 1960", cex=.7,col="dark green",col.main="Dark Blue",
cex.main=2,cex.lab=1.4)

g_new2=log10(gdp3)

plot(g_new2,type="o", xlab="Years",ylab="GDP in B.Usd", main="GDP Growth of GERMANY
since 1960",cex=.7,col="dark green",col.main="Dark Blue",
cex.main=2,cex.lab=1.4)

require(forecast)

gd2=auto.arima(g_new2)

G_new2=forecast(gd2,8)

```

```

plot(forecast(gd2,8),col = "dark green",cex.main=2,col.main="dark blue",lwd=1.5,
main="Germany Forecast")

forecast(gd,8)      #Japan
forecast(G,8)      #Ind
forecast(gd2,8)    #Ger

#forecast comparison

plot(forecast(gd,8),col = "dark red",cex.main=2,col.main="dark
blue",lwd=1.5,main="FORECAST",ylim=c(10,13),xlab = "YEARS",ylab = "GDP")

par(new=T)

plot(forecast(G,8),col = "blue",cex.main=2,col.main="dark
blue",lwd=1.5,ylim=c(10,13),main="FORECAST",xlim=c(1970,2030))

par(new=T)

plot(forecast(gd2,8),col = "dark green",cex.main=2,col.main="dark
blue",lwd=1.5,main="FORECAST",ylim = c(10,13),xlim=c(1970,2030))

abline(v=c(2024,2023),col=c("purple","brown"))

legend("topleft",c("IND","JAP"),fill=c("blue","dark red"))

I<-c(12.56184,12.59383,12.62512,12.65676,12.68840,12.72004,12.75168,12.78332)

Ind=cbind(2023:2030,I)

forecast(gd,8)

J<-c(12.60010,12.58853,12.57695,12.56538,12.55380,12.54223,12.53065,12.51908)

Jap=cbind(2023:2030,J)

Ger=cbind(2023:2030,)

#GROWTH FROM 2023 TO 2030 ONLY

plot(Ind,type="l",col="blue",lwd=1.5,main="GROWTH FROM 2023 TO 2030",
xlab="YEARS",ylab="GDP")

lines(Jap,type="l",col="red",lwd=1.5)

lines(Ger_fut,type = "l",lwd=1.5,col="green")

abline(v=c(2023,2024,2025,2026),col=c("black"))

legend("bottomright",c("Ind","Jap","Ger"),fill=c("blue","red","green"),cex=0.7)

#plot from 1960 to 30

plot(gh2,type="l",col="dark blue",xlab="YEARS",ylab="GDP",main="GDP OF IND & JAPAN
FROM 1960-2030",lwd=2,cex=0.7,ylim=c(10.5,13),xlim=c(1960,2030))

lines(2022:2030,gh2[2022:2030],col="red",lty="dashed")

lines(gh3,type = "l",col="dark red",lwd=1.7,cex=0.7,pch=1,lty="dashed")

abline(v=c(2023,2024.2),col=c("purple","brown"))

legend("topleft",c("IND","JAP"),fill=c("dark blue","dark red"))

#GDP OF IND & JAPAN FROM 1960-2030

Ind_past<-data.frame(1960:2022,gdp_new)

Ind_fut<-data.frame(2023:2030,I)

```

```

J<-c(12.60010, 12.58853,12.57695,12.56538,12.55380, 12.54223,12.53065,12.51908)
Jap_past<-data.frame(1960:2022,g_new)
Jap_fut<-data.frame(2023:2030,J)
G<-c( 12.60571,12.61620,12.62668,12.63717,12.64765,12.65814,12.66862,12.67911)
Ger_past<-data.frame(1970:2022,g_new2)
Ger_fut<-data.frame(2023:2030,G)
plot(Ind_past,type="l",col="dark blue",xlab="YEARS",ylab="GDP",lwd=2,cex=0.7,
ylim=c(10.5,13),xlim=c(1960,2030), main="GDP OF IND & JAPAN FROM 1960-
2030",lty="dashed")
lines(Ind_fut,type = "l",lwd=1.5,col="blue")
lines(Ger_past,type="l",col="dark green",xlab="YEARS",ylab="GDP",lwd=2,cex=0.7,
ylim=c(10.5,13),xlim=c(1970,2030),lty="dashed")
lines(Ger_fut,type = "l",lwd=1.5,col="green")
abline(v=c(2025,2026),col=c("black"))
lines(Jap_past,type="l",col="dark red",xlab="YEARS",ylab="GDP",lwd=2,cex=0.7,
ylim=c(10.5,13),xlim=c(1960,2030),lty="dashed")
lines(Jap_fut,type = "l",lwd=1.5,col="red")
abline(v=c(2023,2024),col=c("black"))
legend("topleft",c("IND_Past","JAP_Past","Ger_Past","Ind_Future","Jap_Future","Ger_
Future"),fill=c("dark blue","dark red","dark green","blue","red","green"),cex=0.6)

```

Bibliography

1. Introduction to Linear Regression Analysis by Doglus C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vinning.
2. Regression Analysis by Example by Samprit Chatterjee & Ali S. Hadi
3. <https://www.financialexpress.com/economy/india-may-overtake-germany-japan-too-after-beating-uk-to-become-worlds-third-largest-economy/2655867/>

Thank You