Progressive Education Society's
Modern College of Arts, Science and Commerce (Autonomous)
Shivajinagar, Pune - 411005

Multivariate Time Series Analysis of GDP and It's Components

Acknowledgement:-

We hereby express our heartfelt gratitude and token of appreciation to all those associated with our project. We would like to thank Prof. Dixit and Prof. Kapre, our mentors who guided us in our endeavour during this time.

Group Members:-

Name	Roll no.
Akshay Deshmukh	2102726
Srishti Batra	2102420
Vaishnavi Ingle	2102425

Motivation :-

As statistics students we love to play with numbers. And when these numbers can be used to measure something or to know something important about our country's growth, we felt that there's nothing more fascinating than this.

We have close relations with other subjects which provide us numbers to play and one such subject is economics.
We've always been fond of how we can study about our country's economic growth and one thing people want to know about an economy of a country is whether it's total output of goods and services is growing or shrinking.

Then, we searched how we can measure the total output of goods and services of the country.

The answer is GDP.

Economists consider GDP as an important factor that determines whether an economy is growing or experiencing a recession.

This motivated us to study more about GDP.

Hence, we decided to do this project to know more about our country's economic growth.

Now, let us understand '**What is GDP ?'** before going through our analysis.

<u>Introduction :-</u>

GDP is the final value of the goods and services produced within the geographic boundaries of a country during a specified period of time, normally a year. GDP growth rate is an important indicator of the economic performance of a country.

KEY TAKEAWAYS

- Gross domestic product tracks the health of a country's economy.
- It represents the value of all goods and services produced over a specific time period within a country's borders.
- Economists can use GDP to determine whether an economy is growing or experiencing a recession.
- Investors can use GDP to make investments decisions—a bad economy often means lower earnings and stock prices.

<u>There are three different methods to measure a country's GDP</u>

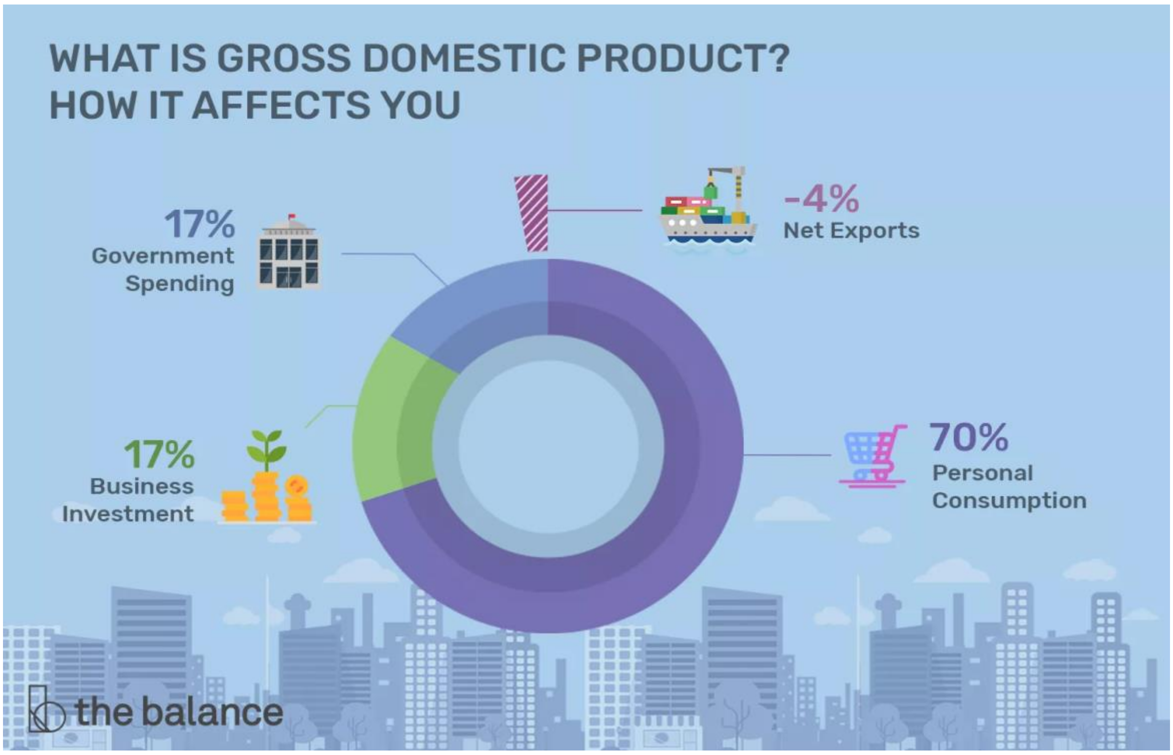
1. *<u>Output Method:</u>* This measures the monetary or market value of all the goods and services produced within the borders of the country. In order to avoid a distorted measure of GDP due to price level changes, GDP at constant prices o real GDP is computed. GDP (as per output method) = Real GDP (GDP at constant prices) – Taxes + Subsidies.

2. *<u>Expenditure Method:</u>* This measures the total expenditure incurred by all entities on goods and services within the domestic boundaries of a country. GDP (as per expenditure method) = C + I + G + (X-IM) C: Consumption expenditure, I: Investment expenditure, G: Government spending and (X-IM): Exports minus imports, that is, net exports.

3. *<u>Income Method</u>:* It measures the total income earned by the factors of production, that is, labour and capital within the domestic boundaries of a country. GDP (as per income method) = GDP at factor cost + Taxes – Subsidies.

Out of these three methods, we used the 2nd method for our analysis.
That is, the **<u>Expenditure method.</u>**

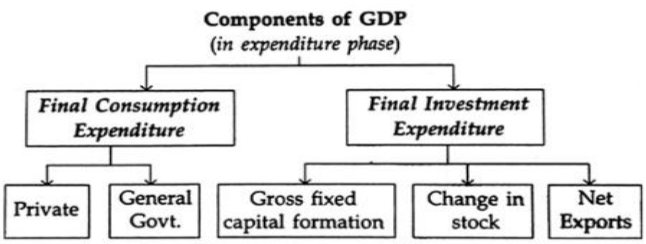
Because India uses expenditure method how different areas of the economy perform.
And, in theory, which method we use-the end result should be the same value.



This expenditure method depends on following four factors:-
1.Personal consumption
2.Business investment (Gross Capital Formation)
3.Government Spending
4.Net exports

Further we studied that Indian GDP as a developing country is influenced by Consumption (Government Final Expenditure and Household Final Expenditure) and Gross Capital formation.

GDP_{MP} = Private final consumption expenditure + Government final consumption expenditure + Gross fixed capital formation + Change in stocks + Net exports.



Methodology and data collection:-

We collected secondary data from the site of world bank.

We got the data of three variables from the year 1960 to 2021.

1.GDP at constant LCU (Dependent Variable)

2.Consumption at constant LCU(Independent variable)

3.Gross Capital formation at constant LCU (Independent variable)

(Here LCU means Local Currency Unit)

We have dropped the data of the pandemic years to be specific, 2020 and 2021 as it could have given us erroneous results.

Final consumption expenditure (formerly total consumption) is the sum of household final consumption expenditure (formerly private consumption) and general government final consumption expenditure (formerly general government consumption). Final consumption expenditure shows an exponential trend.

Gross capital formation (formerly gross domestic investment) consists of outlays on additions to the fixed assets of the economy plus net changes in the level of inventories. Gross capital formation shows an exponential trend.

GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. GDP shows an exponential trend.

Objectives :-

1. To forecast the values of GDP for upcoming years.
2. To learn about the practical application of multivariate time series analysis.

Statistical Tools :-

1.Time Series Analysis:-

Time series analysis is a specific way of analysing a sequence of data points collected over an interval of time.

Time series analysis helped us to understand the underlying causes of trends or systemic patterns over time, using data visualizations.

2.Econometrics

Econometrics analyses data using statistical methods in order to test or develop economic theory. These methods rely on statistical inferences to quantify and analyse economic theories by leveraging tools such as frequency distributions, probability, and probability distributions, statistical inference, correlation analysis, simple and multiple regression analysis, simultaneous equations models, and time series methods.

Softwares used :-

For the analysis of data in our project we used the following packages

1. MS-EXCEL
2. R-STUDIO
3. E-VIEWS

Exploratory Data Analysis:-

❖ Plots

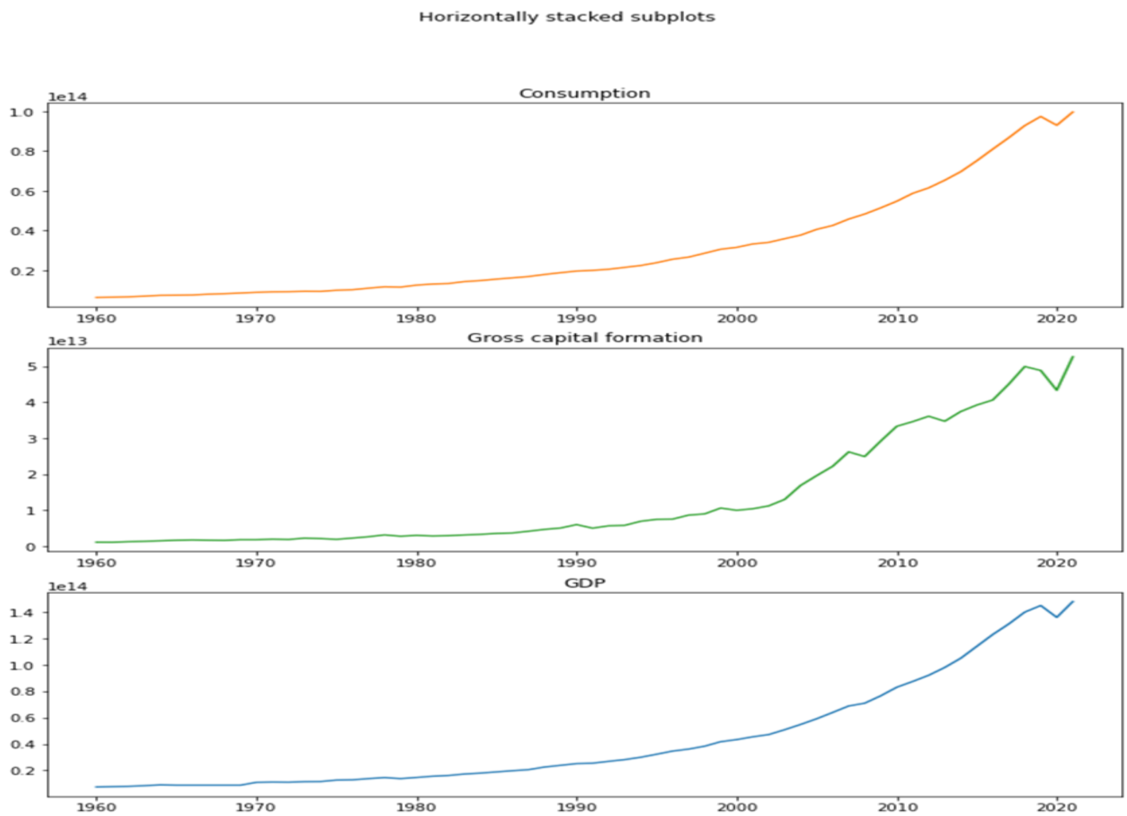
We have the following three variables for the analysis.

1.GDP at constant LCU (Dependent Variable)

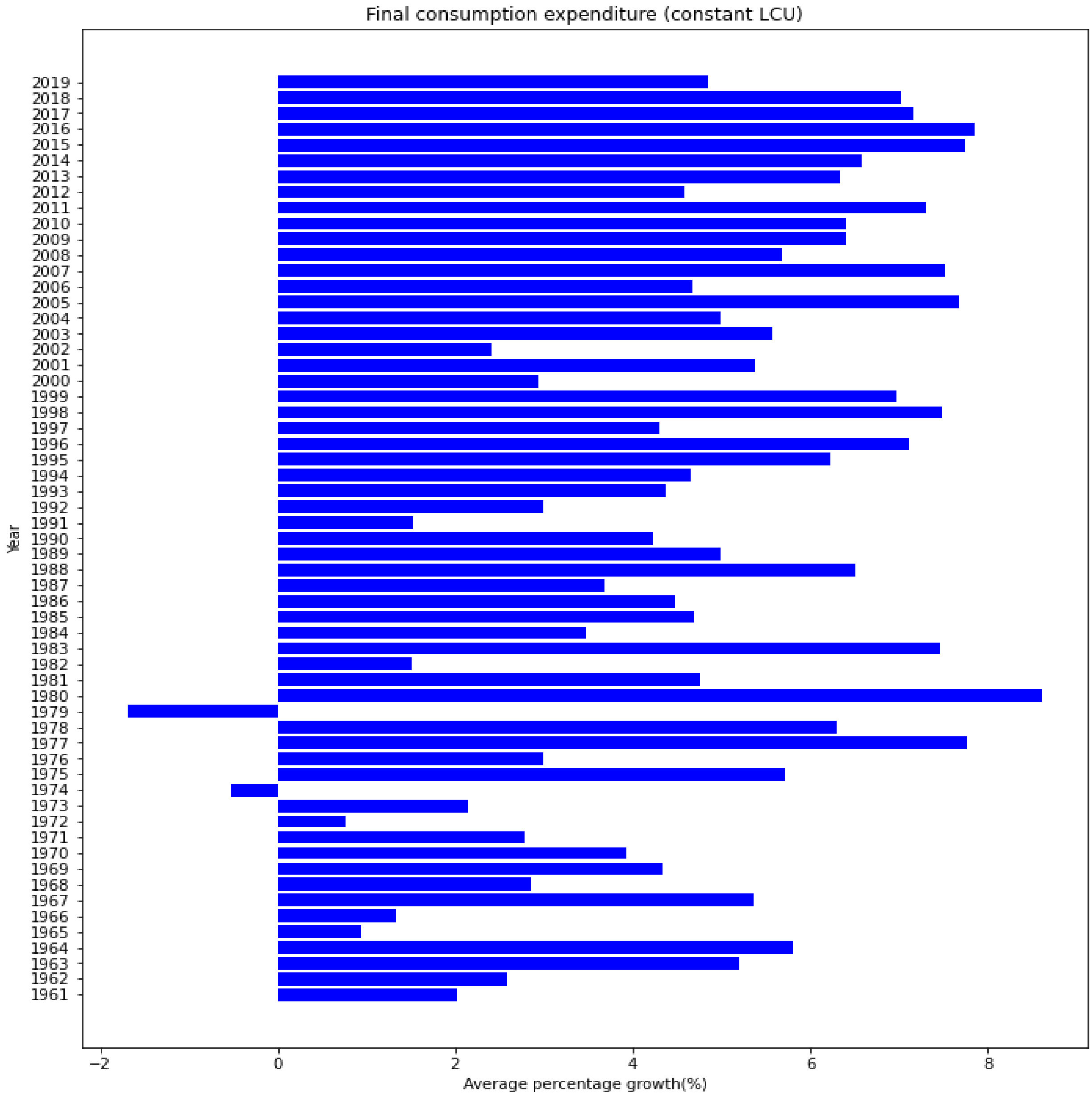
2.Consumption at constant LCU (Independent variable)

3.Gross Capital formation at constant LCU (Independent variable)

(Here LCU means Local Currency Unit)



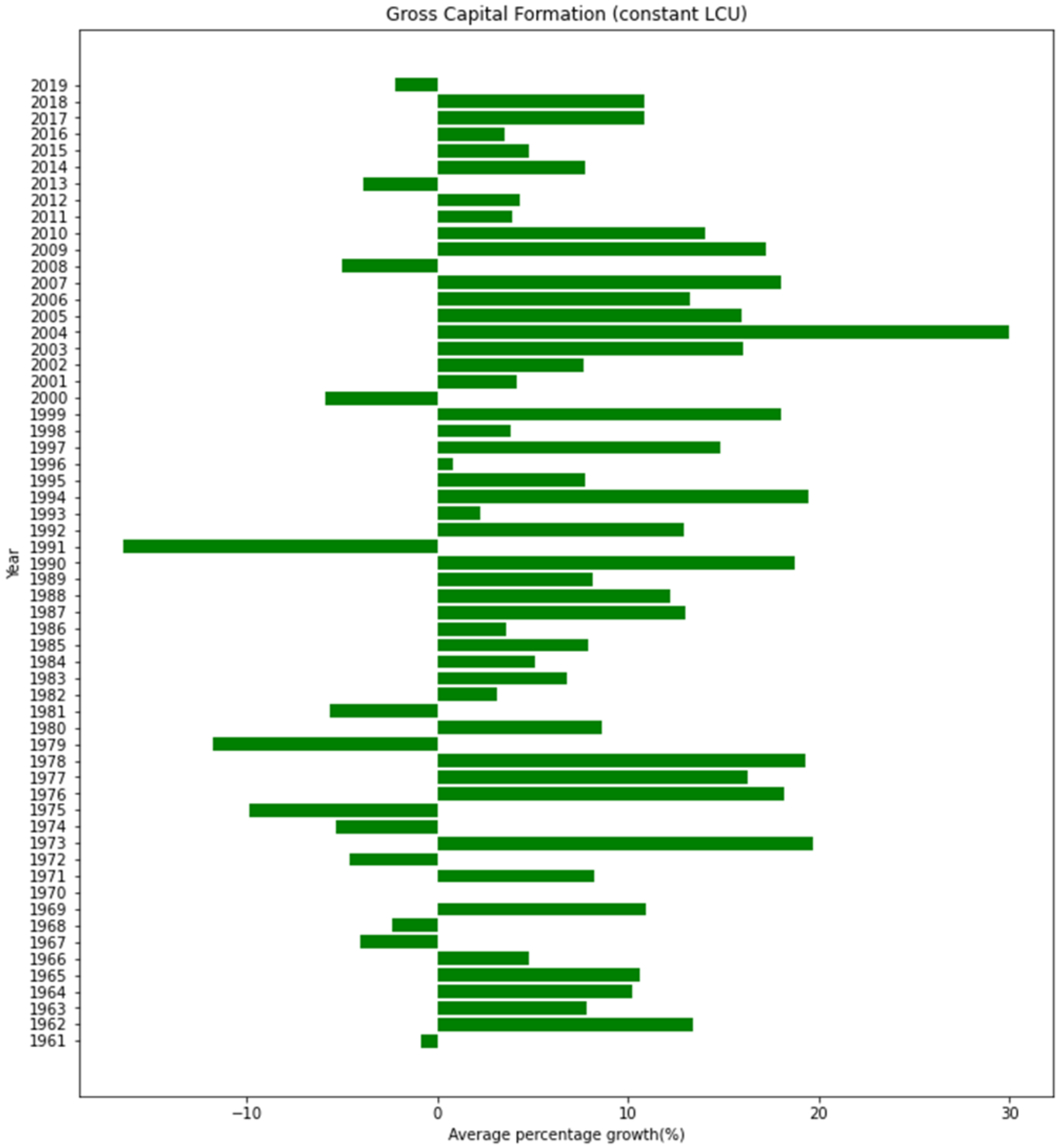
❖ Final Consumption Expenditure



Here we observe that, for the first 30 years, average percentage contribution is around 4 % but as we go further, the average for the next 30 years is around 6 % which concludes average growth rate for consumption is increased over the years.

We also observe negative growth in consumption in the year 1979-80 because the preceding year under the Janata Party government had led to the strongest recession in the history of modern India.

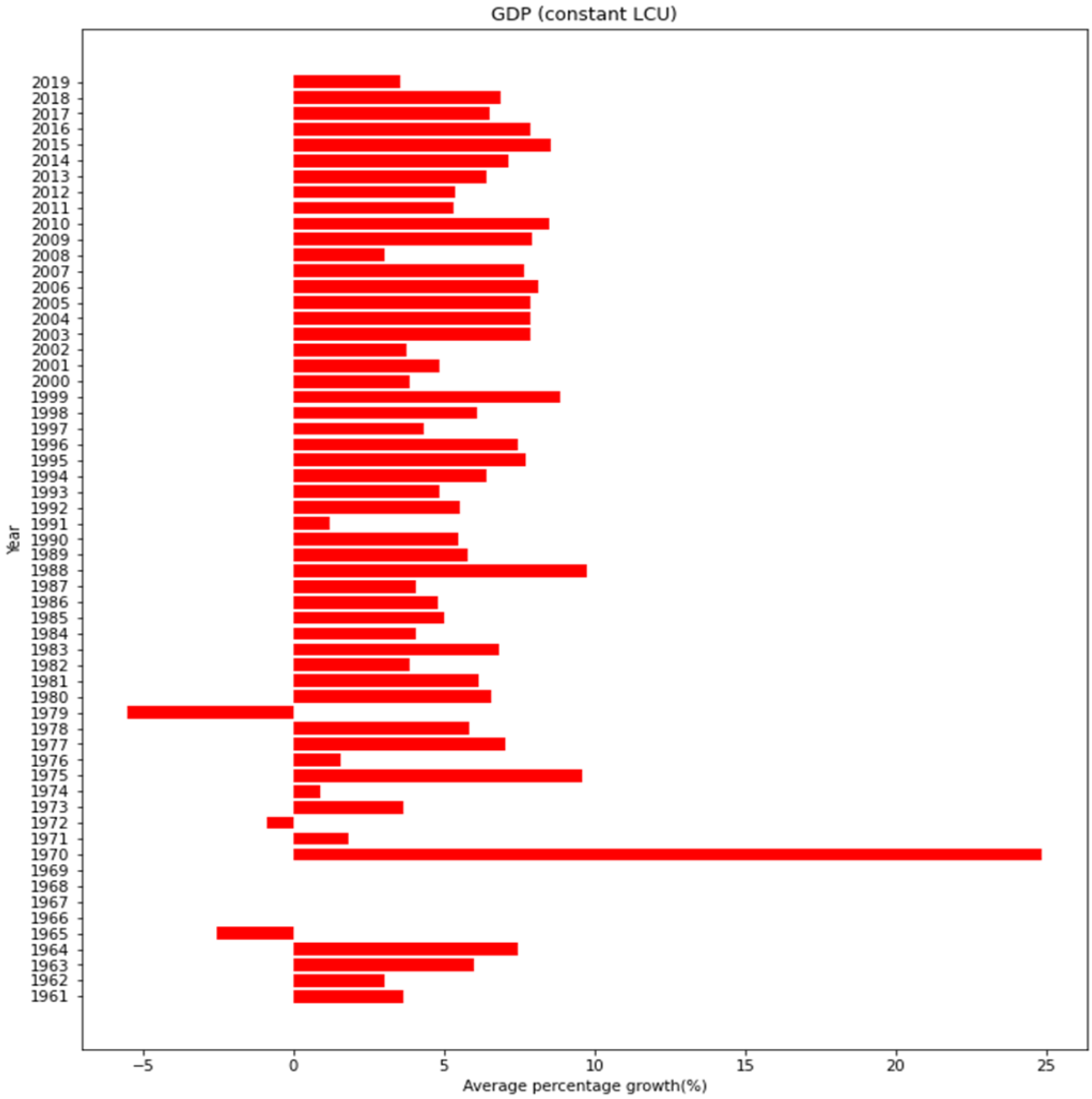
❖ **Gross Capital Formation**



Here we observe that, there are considerable negative downfalls in the graph but for the first 30 years, average is around 6% and as we go further, the average for next 30 years is around 8 % which concludes average growth rate for Gross capital formation is increased over the years

We also observe that the maximum negative growth in the gross capital formation is in the year 1991-1992 because of the 1991 Indian economic crisis.

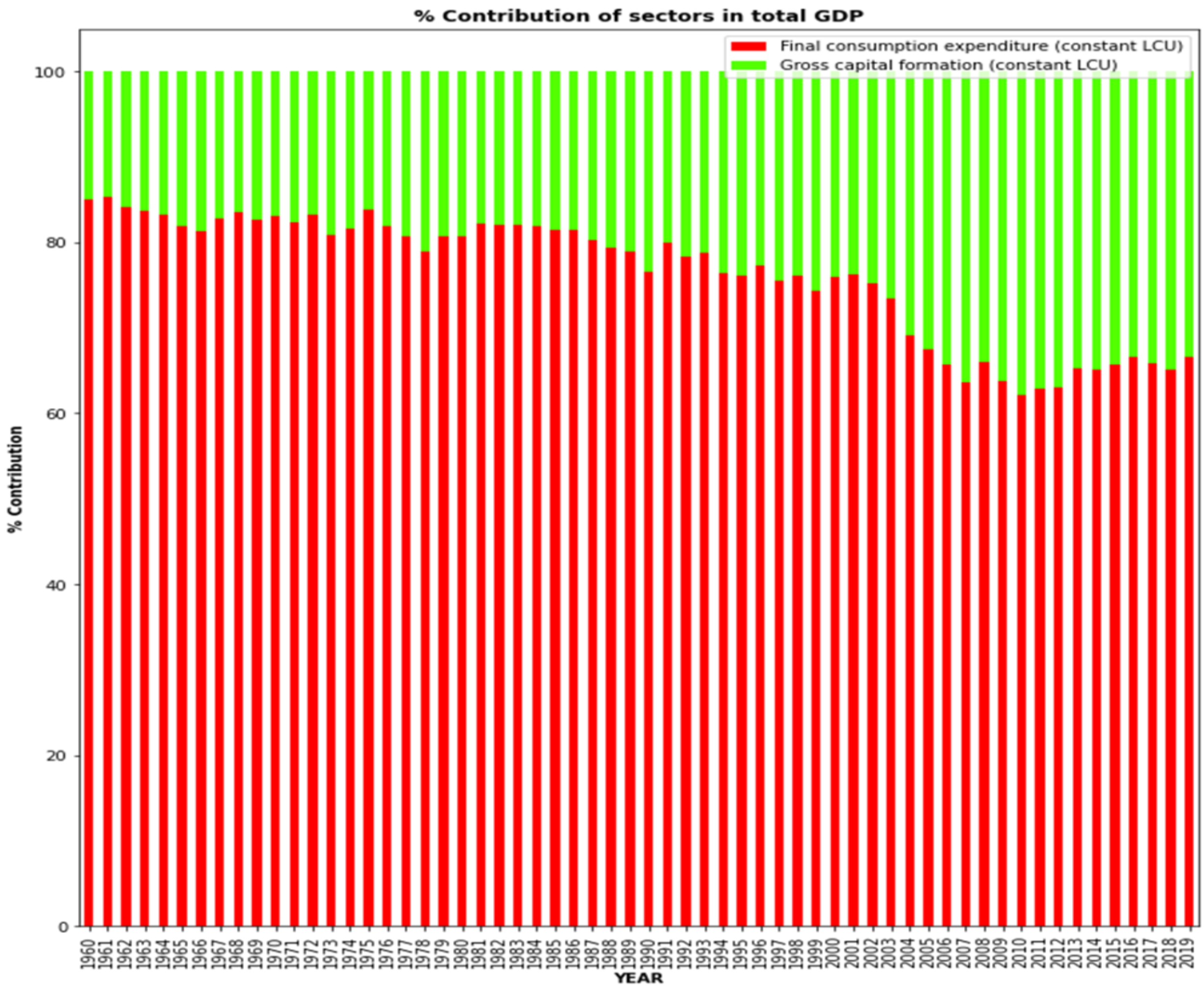
- ## **GDP :-**



Here we observe that, there are few negative downfalls and some zero growth rate in the graph, while for the first 30 years, the average is around 4 % but as we go further, the average GDP for the next 30 years is around 6 % which concludes average growth rate for GDP is increased over the years.

We also observe that the highest growth in GDP is seen in the year 1970-1971 because the then prime minister Indira Gandhi announced to the nation that 14 major commercial banks which between them controlled 85 percent of bank deposits in the country, had been nationalized.

❖ Percentage Contribution (Final Consumption and Gross Capital formation towards GDP)



- Final consumption expenditure has contributed more over the years as compared to Gross Capital Formation.
- With time there is increase in contribution of Gross Capital formation, increasing up to 33% in 2019, with decrease in percentage contribution of Final Consumption expenditure.

❖ Correlation Analysis

	Final consumption expenditure (constant LCU)	Gross capital formation (current LCU)	GDP (constant LCU)
Final consumption expenditure (constant LCU)	1.000000	0.977498	0.999485
Gross capital formation (current LCU)	0.977498	1.000000	0.979901
GDP (constant LCU)	0.999485	0.979901	1.000000

We observe that, Final consumption expenditure (constant LCU) and Gross

capital formation (current LCU) of a country have certain effects on GDP change. The interpretation of above table comprising GDP and Final consumption expenditure (constant LCU) or Gross capital formation (current LCU) highlights a highly positive correlation. In this context we can conclude that values of Final consumption expenditure and Gross capital formation reflects the change of the GDP. Linear relation

Statistical Data Analysis:-

ARDL Model:-

- ARDL stands for Autoregressive distributed lag which are often expressed in relation to the number of lags, ARDL(p, q) where:
    - p is the number of lags of the dependent variable.
    - q is the number of lags of the independent variables.

Mathematically we represent the ARDL model as:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \ldots\ldots.+ \beta p\ y_{t-p} + \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots\ldots\ldots + \alpha_q x_{t-q} + \varepsilon_t$$

- It means this model includes lag of dependent variable and independent variables. According to the theory GDP(constant LCU) is the dependent variable and Gross Capital Formation(constant LCU) and Final consumption expenditure(constant LCU) are the independent variables.

- ARDL model will work on stationary as well as non-stationary data.

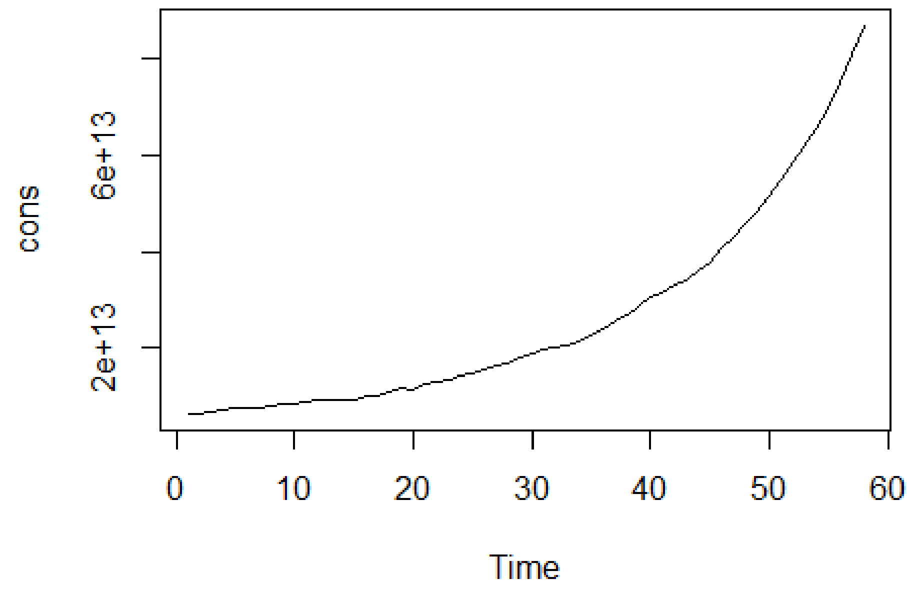
Steps:-

- Check the stationarity of data: The ARDL model can be specified for a combination of variables with I(1) and I(0), but not I(2) or higher. It can also be specified if all the variables are of order I(1).
- Perform the test for cointegration.
- Determine the optimal lag structure.
- Apply ARDL model.
- Check the goodness of fit.
- Obtain the forecasts.

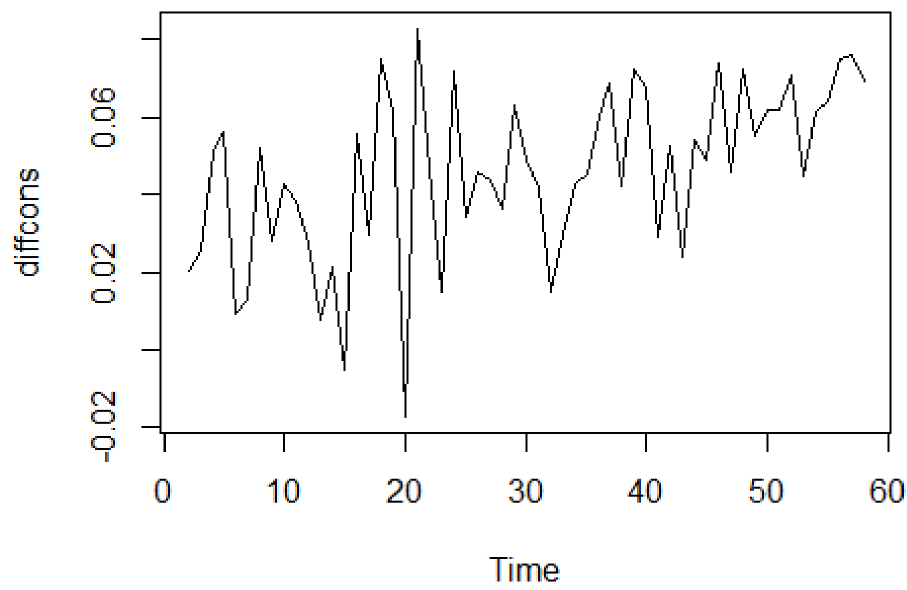
NOTE: Here, using the log transformation we observe that log of all the variables i.e Final consumption expenditure, Gross capital formation and GDP are stationary I(1).

R- Work:-

```
library(tsDyn)

library(readxl)

library(dynamac)

library(forecast)

library(tidyverse)

library(tseries)

library(urca)

library(TSstudio)

library(vars)

library(dLagM)

##STEPS  FOR ARDL MODEL

##1)Check the stationarity of variable(Variables should be stationary at I(0) and I(1)
##2)select the optimum lag
##3)Apply ARDL model

###1)###CONSUMPTION
data=read.csv("G:\\M.Sc\\M.Sc Project\\consumption.csv")
d=data[1:58,2]
cons=ts(d)
plot(cons)
```



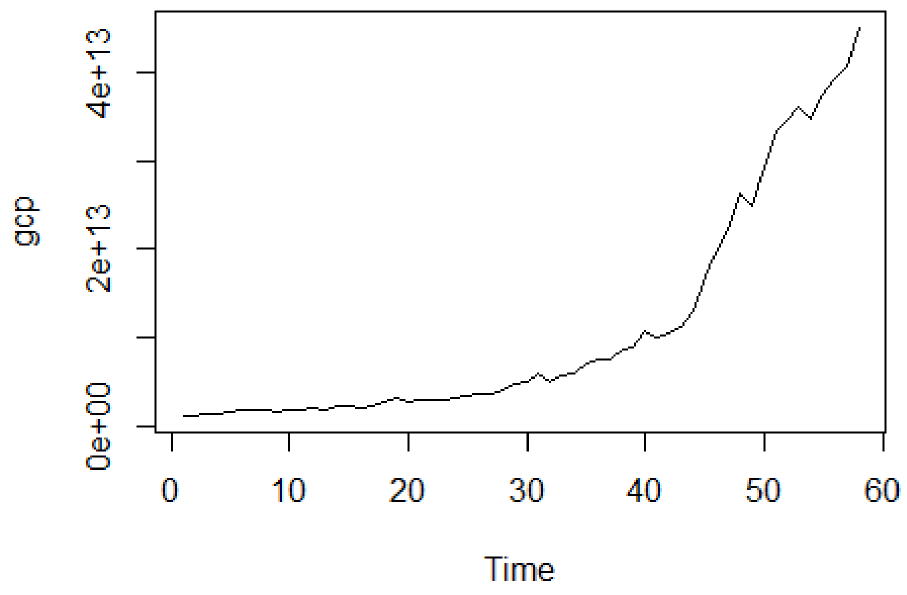

```
diffcons=diff(log(cons),1)
plot(diffcons)
```



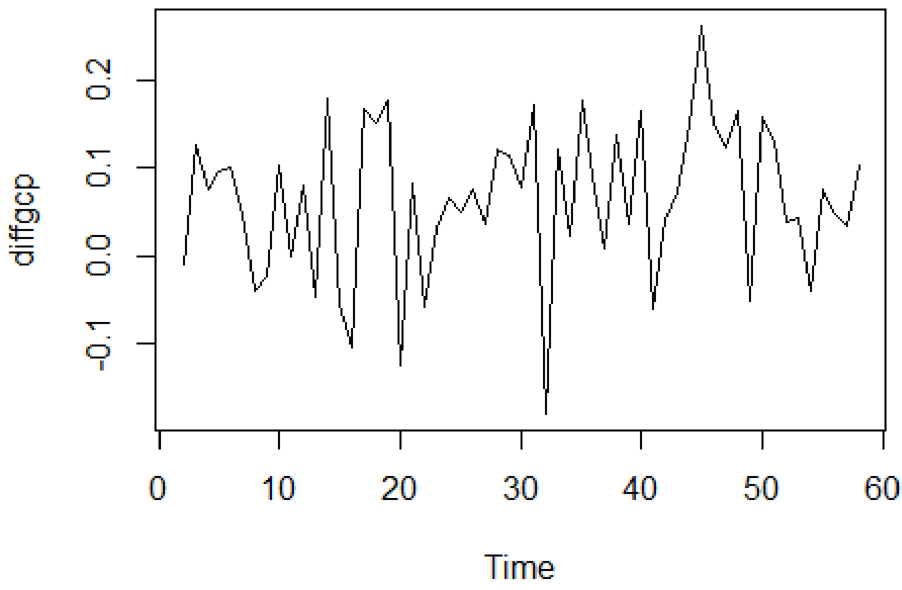

```
adf.test(diffcons)

## Warning in adf.test(diffcons): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diffcons
## Dickey-Fuller = -5.0861, Lag order = 3, p-value = 0.01
## alternative hypothesis: stationary
```

```
####The order of integration is 1

####GCP####
data2<- read.csv("G:\\M.Sc\\M.Sc Project\\gcp.csv")
d2=data2[1:58,2]
gcp=ts(d2)
plot(gcp)
```

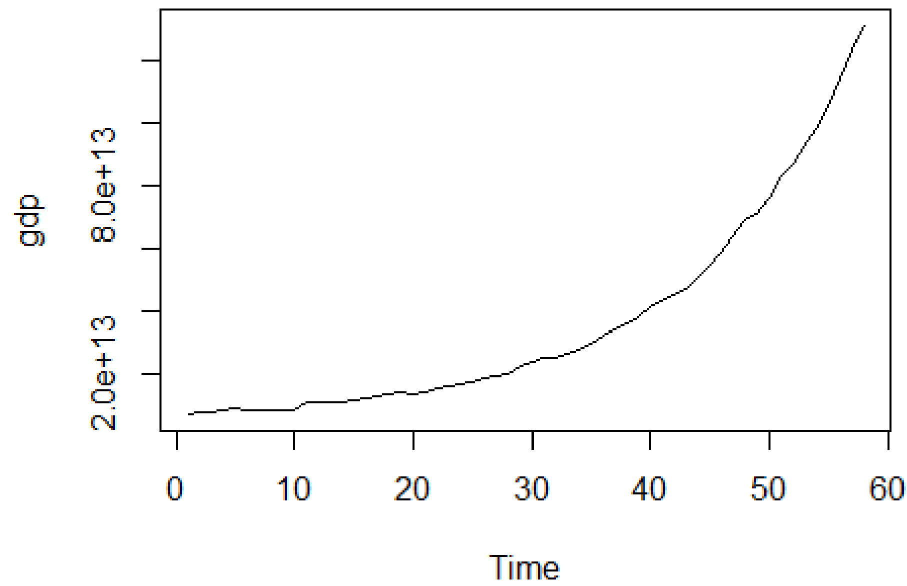



```
diffgcp=diff(log(gcp),1)
plot(diffgcp)
```

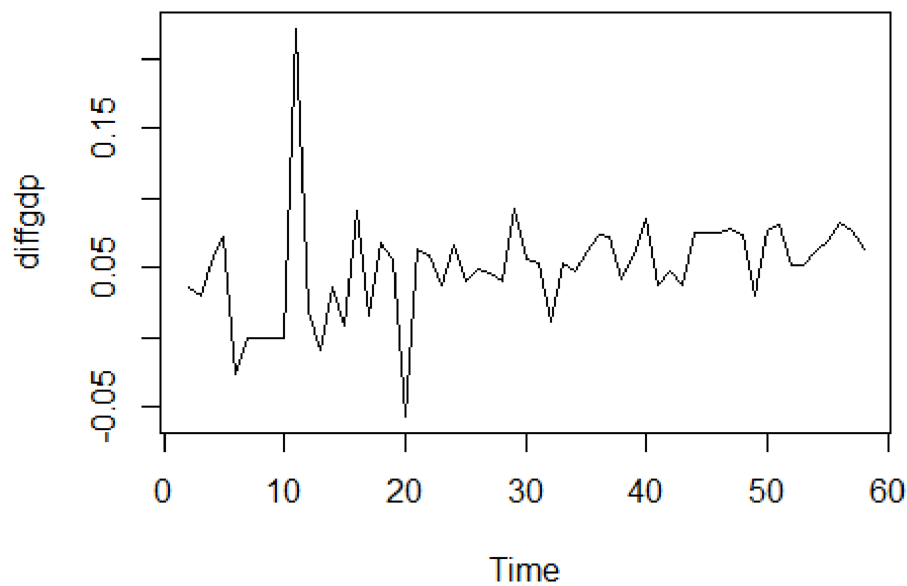



```
adf.test(diffgcp)

## Warning in adf.test(diffgcp): p-value smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  diffgcp
## Dickey-Fuller = -4.6603, Lag order = 3, p-value = 0.01
## alternative hypothesis: stationary
```

```
####The order of integration is 1


####GDP###
data3=read.csv("G:/M.Sc/M.Sc Project/gdp time.csv")
d3=data3[1:58,2]
```




```
diffgdp=diff(log(gdp),1)
plot(diffgdp)
```




```
adf.test(diffgdp)

## Warning in adf.test(diffgdp): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diffgdp
## Dickey-Fuller = -6.7629, Lag order = 3, p-value = 0.01
## alternative hypothesis: stationary
```

####The order of integration is 1

```
logcons=log(cons)
loggcp=log(gcp)
loggdp=log(gdp)
```

```
#The dependent variable (loggdp) is I(1)
#All the independent variables are either I(0) or I(1)



####Johansen Test for cointegration####
```

Lag selection criteria

```
lagselect=VARselect(ardldata);lagselect
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      6      1      1      6
##
## $criteria
##                     1             2             3             4
5
## AIC(n) -2.037836e+01 -2.010997e+01 -1.993312e+01 -2.002785e+01 -2.04327
8e+01
## HQ(n)  -2.020158e+01 -1.980060e+01 -1.949116e+01 -1.945331e+01 -1.97256
5e+01
## SC(n)  -1.991056e+01 -1.929132e+01 -1.876362e+01 -1.850750e+01 -1.85615
8e+01
## FPE(n)  1.413493e-09  1.858145e-09  2.245024e-09  2.089635e-09  1.44762
9e-09
##                     6             7             8             9
10
## AIC(n) -2.074825e+01 -2.062348e+01 -2.034714e+01 -2.013920e+01 -2.06866
8e+01
## HQ(n)  -1.990854e+01 -1.965118e+01 -1.924225e+01 -1.890172e+01 -1.93166
2e+01
## SC(n)  -1.852620e+01 -1.805058e+01 -1.742339e+01 -1.686460e+01 -1.70612
3e+01
## FPE(n)  1.118741e-09  1.378477e-09  2.046274e-09  2.971515e-09  2.16039
3e-09
```

#####lagselect=6-1=5#####

```
###Johansen Testing(Trace)
ctest1t=ca.jo(ardldata,type="trace",ecdet="const",K=5)
summary(ctest1t)
```

```
######################
# Johansen-Procedure #
######################

Test type: maximal eigenvalue statistic (lambda max) , without linear trend and constant in cointegration

Eigenvalues (lambda):
[1]  4.889147e-01  2.617487e-01  6.623974e-02 -1.496862e-15

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 2 |  3.77  7.52  9.24 12.97
r <= 1 | 16.69 13.75 15.67 20.20
r = 0  | 36.92 19.77 22.00 26.81

Eigenvectors, normalised to first column:
(These are the cointegration relations)

             loggdp.l5 logcons.l5  loggcp.l5   constant
loggdp.l5    1.0000000  1.0000000   1.000000   1.000000
logcons.l5  -0.2976346 -1.0079965   1.802984   2.424732
loggcp.l5   -0.3164352 -0.0899948  -2.081124  -2.017324
constant   -11.3751148  2.6153172 -25.023429 -46.123155

Weights W:
(This is the loading matrix)

           loggdp.l5 logcons.l5    loggcp.l5      constant
loggdp.d  0.11015443 -0.6157616  0.004517786 -8.614879e-13
logcons.d 0.07526449  0.1232181 -0.002838221  6.920989e-13
loggcp.d  0.16682105  0.3133319  0.084239395  4.066745e-12
```

##There is cointegration

#2) Optimal Lag selection

```
VARselect(loggdp)

## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##     10      7      7     10
##
## $criteria
##                   1            2            3            4
5
## AIC(n) -6.566646696 -6.597458773 -6.692425681 -6.730027926 -7.038658924
6
## HQ(n)  -6.537182963 -6.553263173 -6.633498215 -6.656368593 -6.950267725
2
## SC(n)  -6.488679987 -6.480508709 -6.536492264 -6.535111154 -6.804758798
2
## FPE(n)  0.001406574  0.001364052  0.001240751  0.001195406  0.000878456
2
##                    6             7             8             9
10
## AIC(n) -7.0799569441 -7.1685184896 -7.1275055423 -7.0878829324 -7.19428
96422
## HQ(n)  -6.9768338781 -7.0506635571 -6.9949187432 -6.9405642667 -7.03223
91099
## SC(n)  -6.8070734633 -6.8566516545 -6.7766553527 -6.6980493885 -6.76547
27439
## FPE(n)  0.0008435744  0.0007728855  0.0008063312  0.0008403494  0.00075
71089
```

```
##AIC(n)  HQ(n)  SC(n) FPE(n)
##10       7       7     10

###Optimum lag for dependent variable=10

VARselect(data.frame(logcons,loggcp))
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      1      1      1      1
```

```
##
## $criteria
##                     1             2             3             4
5
## AIC(n) -1.272057e+01 -1.261956e+01 -1.250207e+01 -1.252421e+01 -1.26212
7e+01
## HQ(n)  -1.263218e+01 -1.247224e+01 -1.229582e+01 -1.225904e+01 -1.22971
7e+01
## SC(n)  -1.248667e+01 -1.222973e+01 -1.195630e+01 -1.182251e+01 -1.17636
4e+01
## FPE(n)  2.989990e-06  3.311721e-06  3.734558e-06  3.670326e-06  3.35621
3e-06
##                     6             7             8             9
10
## AIC(n) -1.266453e+01 -1.257715e+01 -1.247057e+01 -1.238953e+01 -1.24991
4e+01
## HQ(n)  -1.228150e+01 -1.213520e+01 -1.196969e+01 -1.182972e+01 -1.18804
0e+01
## SC(n)  -1.165096e+01 -1.140765e+01 -1.114514e+01 -1.090816e+01 -1.08618
4e+01
## FPE(n)  3.250149e-06  3.602449e-06  4.092150e-06  4.560534e-06  4.23298
1e-06

#AIC(n)  HQ(n)  SC(n) FPE(n)
#1      1        1        1

###Optimum lag=1

###Optimum lag for dependent variable gdp=10
###Optimum lag for combined independent variables consumption and gcp =1
```

#3) Creating the model

```
> ardldata=data.frame(loggdp,logcons,loggcp)
> model1=ardlDlm(formula=loggdp~logcons+loggcp,data=ardldata,p=1,q=10)
> summary(model1)

Time series regression with "ts" data:
Start = 11, End = 60

Call:
dynlm(formula = as.formula(model.text), data = data)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.050270 -0.007762  0.000331  0.009559  0.052910 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)   
(Intercept) -0.855239   0.487106  -1.756  0.08788 . 
logcons.t    0.523618   0.220535   2.374  0.02320 * 
logcons.1   -0.074342   0.240135  -0.310  0.75871   
loggcp.t     0.115977   0.041050   2.825  0.00775 **
loggcp.1    -0.019738   0.039963  -0.494  0.62446   
loggdp.1     0.227207   0.119188   1.906  0.06485 . 
loggdp.2    -0.053877   0.110279  -0.489  0.62821   
loggdp.3     0.010196   0.112838   0.090  0.92852   
loggdp.4     0.034701   0.118323   0.293  0.77105   
loggdp.5     0.188202   0.121345   1.551  0.12991   
loggdp.6     0.025539   0.109962   0.232  0.81769   
loggdp.7     0.063030   0.110197   0.572  0.57100   
loggdp.8    -0.005255   0.110896  -0.047  0.96247   
loggdp.9    -0.129431   0.113490  -1.140  0.26184   
loggdp.10    0.133309   0.098939   1.347  0.18651   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02043 on 35 degrees of freedom
Multiple R-squared:  0.9995,	Adjusted R-squared:  0.9994 
F-statistic:  5471 on 14 and 35 DF,  p-value: < 2.2e-16

> ####R^2=0.9994
```

```
> ####Removing the insignifican coefficients
> remove=list(p=list(logcons=c(1),loggcp=c(1)),q=(loggdp=c(1,2,3,4,5,6,7,8,9,10)))
> model2=ardlDlm(formula=loggdp~logcons+loggcp,data=ardldata,p=1,q=10,remove=remove)
> summary(model2)

Time series regression with "ts" data:
Start = 1, End = 60

Call:
dynlm(formula = as.formula(model.text), data = data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.146831 -0.015806  0.002415  0.021427  0.076287

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.69976    0.58065  -2.927  0.00490 **
logcons.t    0.94662    0.05337  17.736  < 2e-16 ***
loggcp.t     0.12207    0.03641   3.352  0.00143 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03439 on 57 degrees of freedom
Multiple R-squared:  0.9986,	Adjusted R-squared:  0.9986
F-statistic: 2.039e+04 on 2 and 57 DF,  p-value: < 2.2e-16
```

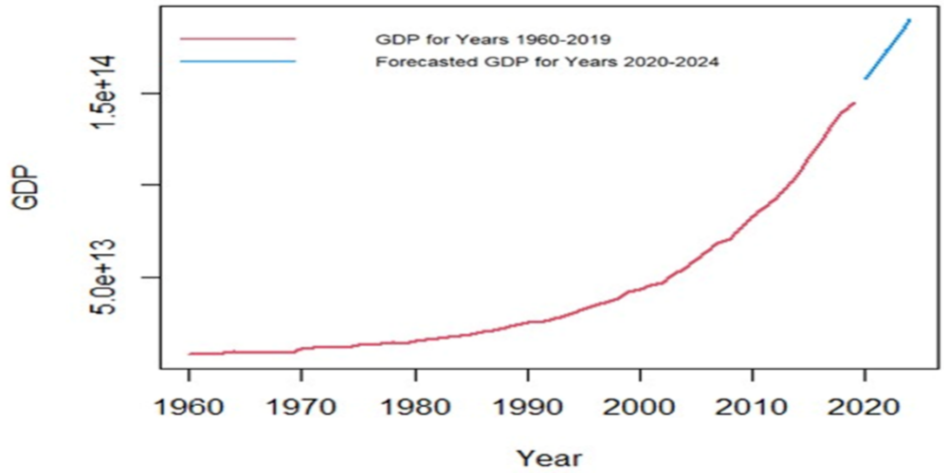
##4.Goodness of fit

```
> GoF(model2)
        n        MAE           MPE         MAPE        sMAPE      MASE         MSE     MRAE
model2 60 0.02330295 -1.303778e-06 0.000764733 0.0007645261 0.4346835 0.001123772 57058236
          GMRAE
model2 1.114653

> #####Forecasting ARDL(10,1)
> X=c(lcons,lgcf)
> a=forecast(model2,h=5,x=matrix(X,nrow=2,byrow=TRUE),interval=TRUE,level=0.95)
> f1=a$forecasts[[2]];f1
[1] 32.69342 32.74213 32.79065 32.83607 32.88013
> fgdp=exp(f1);fgdp
[1] 1.579690e+14 1.658534e+14 1.740992e+14 1.821904e+14 1.903971e+14
> plot(seq(1960,2019),gdp,xlab="Year",ylab="GDP",col=2,lwd=2,type="l",xlim=c(1960,2024),ylim = c(min(gdp),max(fgdp)))
> lines(seq(2020,2024),fgdp,col=4,type="l",lwd=2)
> legend("topleft",c("GDP for Years 1960-2019","Forecasted GDP for Years 2020-2024"),lty=c(1,1),col=c(2,4),cex=1,lwd=2)
```

##5.Forecasts for the GDP

```
> GDFforecasts=data.frame("Years"=seq(2020,2024),"Forecasted Values"=fgdp)

> GDFforecasts

  Years Forecasted.Values

1  2020       1.579690e+14

2  2021       1.658534e+14

3  2022       1.740992e+14

4  2023       1.821904e+14

5  2024       1.903971e+14
```

PLOT :-



Limitations for the calculation of GDP

The GDP is designed to measure the market value for all products and services within a country's borders. Since the measurement hinges on market price, there are many aspects of society including many aspects that factor into economic well-being that aren't included in the GDP numbers.

One of the biggest criticisms of GDP is that it doesn't count environmental costs. For example, the price of plastic is low because it doesn't include the cost of pollution. GDP doesn't measure how these costs impact the well-being of society. A more accurate measurement of a country's standard of living may include environmental conditions.

Another criticism is that GDP doesn't include unpaid services. It leaves out unpaid child care and volunteer work, for example, despite the significant impact they have on the economy and a country's quality of life.

GDP also does not count the shadow or black economy. It underestimates economic output in countries where many people receive their income from illegal activities. These products aren't taxed and don't show up in government records, and although they can estimate, they cannot accurately measure this output. One estimate that is referenced by the Bureau of Labour Statistics pegs the shadow economy's size as 8.8% of the GDP.

Bibliography

- https://www.statsmodels.org/dev/generated/statsmodels.tsa.filters.bk_filter.bkfilter.html
- https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/
- https://en.wikipedia.org/wiki/Band-pass_filter
- https://analyticsindiamag.com/a-complete-tutorial-on-time-series-filters/
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7034805/
- https://www.youtube.com/playlist?list=PLzAfHlPtM1I7mCViv3tZEseWzJb4pQQ_C
- https://www.youtube.com/watch?v=BPnx4ShqUJ8&list=PLzAfHlPtM1I71CIgGhG4-cS8zXWusPEvH&index=8
- https://nomanarshed.wordpress.com/2014/11/16/a-manual-for-ardl-approach-to-cointegration/
- https://cran.r-project.org/web/packages/ARDL/ARDL.pdf
- https://economictimes.indiatimes.com/defaultinterstitial.cms?now=1669757140582
- Macroeconomics 11th Edition is written by N. Gregory Mankiw
- https://www.sciencedirect.com/science/article/pii/S2212567115003743?ref=cra_js_challenge&fr=RR
- https://www.researchgate.net/publication/280324687_Dynamics_of_the_Relation_between_Real_Consumption_Expenditure_and_Economic_Growth_in_India
- https://www.youtube.com/watch?v=qoihlzu7CcM&t=357s
- https://www.youtube.com/@mikleshyadav9830/playlist
- https://www.youtube.com/playlist?list=PL92YnqQQ1gbhUVYZnIzIDzYt3TG4j75Do