Time Series Analysis on Indian GDP RATE

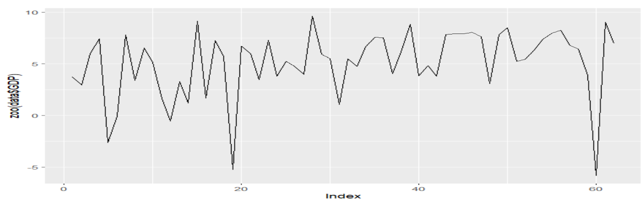
Name : Devendra Sanjay Patil - 2234

Abstract: Gross Domestic Product is one of the most important economic indicators of the country and its positive or negative growth indicates the economic development of the country. It is calculated quarterly and yearly at the end of the financial year. The GDP growth of India has seen fluctuations from last few decades after independence and reached as high as 10.25 in 2010 and declined to low of -5.23 in 1979. The GDP growth has witnessed a continuous decline in the past five years, taking it from 8.15 in 2015 to 1.87 in 2020.The lockdown imposed in the country to curb the spread of COVID-19 has caused massive slowdown in the economy of the country by affecting all major contributing sectors of the GDP except agricultural sector. To keep on track on the GDP growth is one of the parameters for deciding the economic policies of the country. In this study, we are analyzing and forecasting the GDP growth using the time series forecasting techniques Arima model. This model can assist policy makers in framing policies or making decisions

Objectives

Analysis and Forecasting of GDP growth trend in India from 1951 to 2020

Data Description: In data there are 62 rows means data contain 62 years of GDP data.We divide the data into train and test part train data is from 1961 to 2010 and test data has from 2011to 2020 the following plot show how the GDP rate fluctuating in 62 years,



This are some values of data:

```
Year      GDP
1  1961   3.72
2  1962   2.93
3  1963   5.99
4  1964   7.45
5  1965  -2.64
6  1966  -0.0553
```

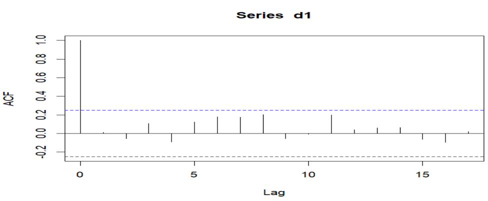
Now converting the data into time series component with frequency 1 because we have yearly data.

Checking For Stationarity :

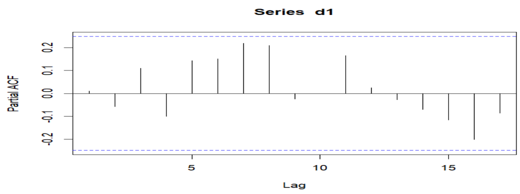
```
adf.test(d1)
Augmented Dickey-Fuller Test
data:  d1
Dickey-Fuller = -5.4847, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

From the above output the data is stationary because p-value lower than 0.05 which we reject Ho:non-stationary.

ACF/PACF plot:



The ACF plot for "Series d1" shows a strong positive autocorrelation at lag 0, which drops significantly afterwards. This suggests little correlation between the time series and its lagged values. The correlations are within the bounds of the blue region, indicating they may not be statistically significant.



From the above plot there is no significant partial correlation between the GDP rate

Model Building

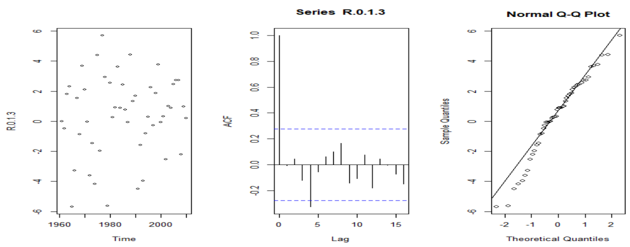
1.ARIMA model:

Fitting the ARIMA model to train data by using the auto.arima() function which automatically calculate the order of AR, I and MA and fit the model, here are the results which are from the above function,

```
Series: train_data
ARIMA(0,1,3)
Coefficients:
        ma1      ma2     ma3
     -1.2263  -0.1000  0.5110
s.e.  0.1463   0.2164  0.1572
sigma^2 = 7.635:  log likelihood = -119.93
AIC=247.86   AICc=248.77   BIC=255.43
Training set error measures:
ME          RMSE       MAPE
0.3987795  2.650246  180.9956
```

The order of AR,I and MA has 0,1,3.

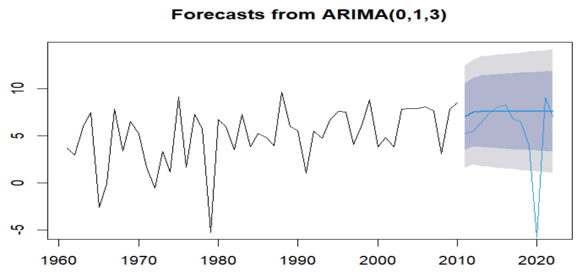
Residual Analysis of ARIMA model:



From the above plot1 we can say that assumption of constant variance is satisfy, from plot 2 residuals are independent and from plot 3 residuals are normally distributed. There is one test which gives that the residuals are normally distributed are not, which shown below:

```
shapiro.test(R.0.1.3)
              Shapiro-Wilk normality test
data:  R.0.1.3
W = 0.96907, p-value = 0.2121
```

p-value is greater than 0.05 i.e accept Ho residuals are normally distributed.

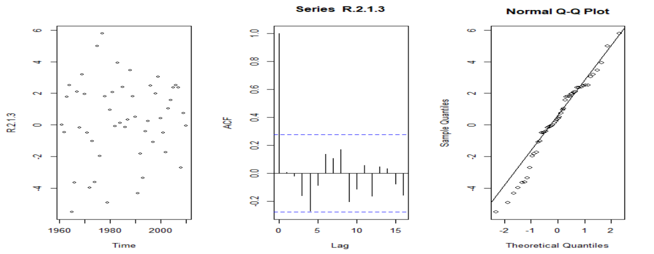


From the above Forecasted plot some values are forecasted well but in 2020 there sharp decreasing rate in GDP this because of COVID-19 otherwise the predicted model is quite good.

2.ARIMA model with order (2,1,3)

```
arima(x = train_data, order = c(2, 1, 3))
Coefficients:
         ar1     ar2      ma1      ma2     ma3
     -0.3218  0.3736  -0.8476  -0.8874  0.9602
s.e.  0.1628  0.1617   0.2347   0.1324  0.2569
sigma^2 estimated as 6.534:  log likelihood = -118.9,
aic = 249.8
Training set error measures:
ME            RMSE       MAPE
0.3293676   2.530433   192.4487
```

Residual Analysis:

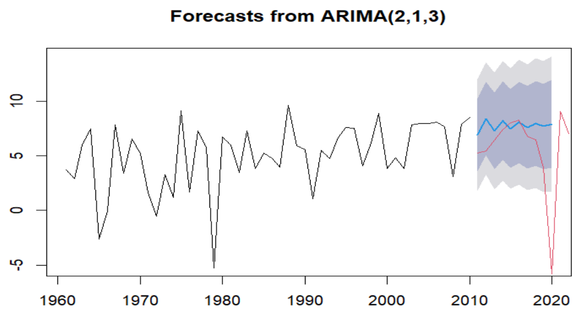



```
> shapiro.test(R.2.1.3)

        Shapiro-Wilk normality test

data:  R.2.1.3
W = 0.9721, p-value = 0.2811
```

ARIMA(2,1,3) residuals are also follow all the assumptions.



Conclusion:
Both Models ARIMA(0,1,3) and ARIMA(2,1,3) are fitted good but from **ME and RMSE** ARIMA(2,1,3) is slightly better than ARIMA(0,1,3)

```
Code:
rm(list=ls(all=T))
library(readxl)
library(readxl)
library(tseries)
# Library Simple Moving Average
library(TTR)
# Library to forecast
library(forecast)
# Data visualisation
library(plotly)
library(xts)
library(TSstudio)
library(ggplot2)
data=read_xls("C:\\Users\\Shiv\\Documents\\Time Series\\Time_projects\\Indian_GDP_data.xls")
data$Year=as.numeric(data$Year)
autoplot(zoo(data$GDP))
dim(data)
class(data)
head(data)

#### convert the data into time series component
d1=ts(data$GDP,start=c(1961,1),frequency = 1)

#### stationarity test
adf.test(d1)


#### Acf and PCAF
acf(d1)
pacf(d1)

#### train and test data
length(d1)
train_data=head(d1,50)
test_data=tail(d1,12)
train_data
test_data

#### model fitting
library(forecast)
library(seasonal)
ARIMA=auto.arima(train_data)
summary(ARIMA)

plot(forecast(ARIMA,h=12))
lines(test_data,col=4)
summary(ARIMA)
##### residual analysis
R.0.1.3=resid(ARIMA)
mean(R.0.1.3^2)
par(mfrow=c(1,3))
plot(R.0.1.3,type="p")
jarque.bera.test(R.0.1.3)
acf(R.0.1.3)
qqnorm(R.0.1.3)
qqline(R.0.1.3)
shapiro.test(R.0.1.3)

ARIMA1=arima(train_data,order=c(2,1,3))
summary(ARIMA1)
plot(forecast(ARIMA1))
lines(test_data,col=2)
R.2.1.3=resid(ARIMA1)
mean(R.2.1.3^2)
plot(R.2.1.3,type="p")
jarque.bera.test(R.2.1.3)
acf(R.2.1.3)
qqnorm(R.2.1.3)
qqline(R.2.1.3)
shapiro.test(R.2.1.3)
summary(ARIMA1)
```