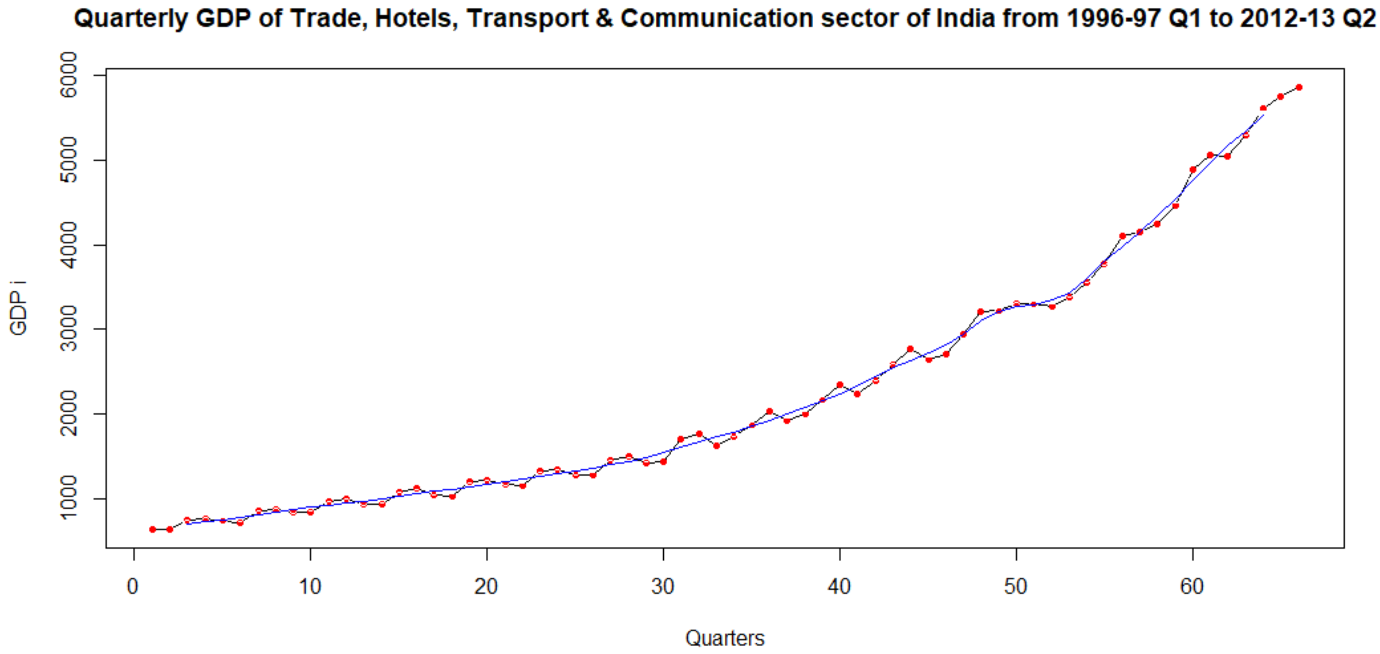
MG222-Assignment #4

Kumar Prerak

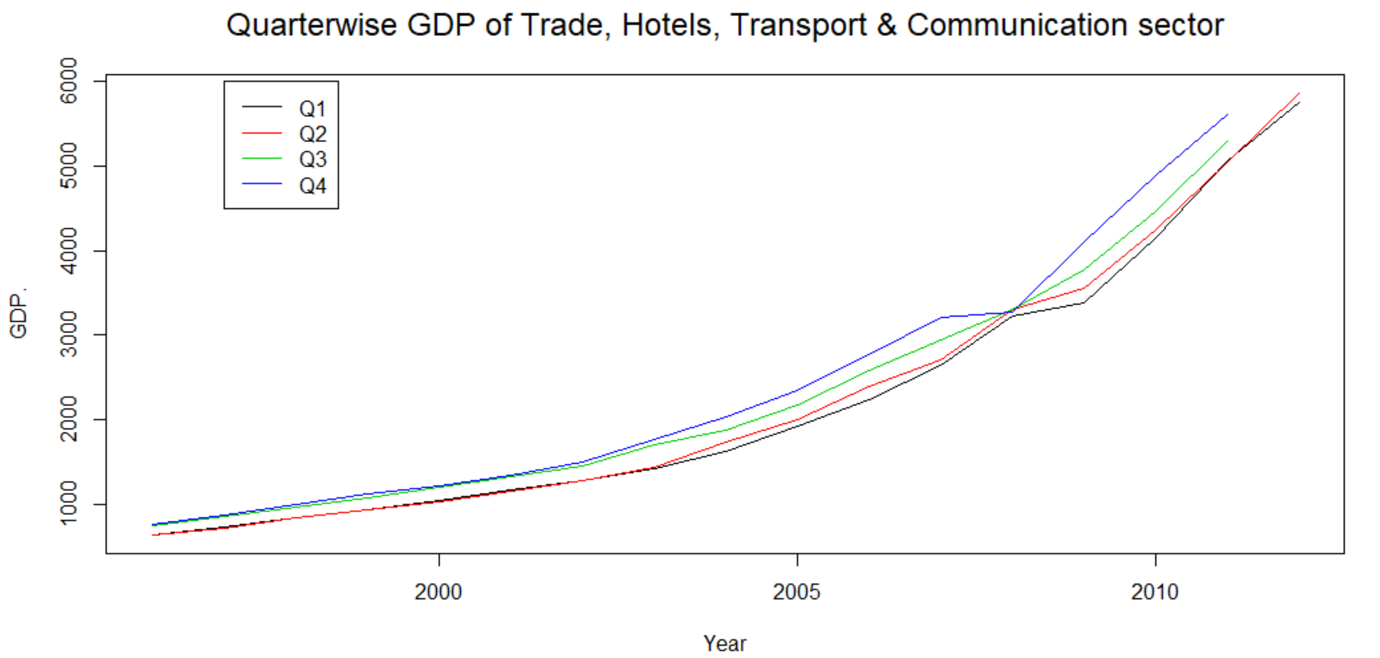
Sector: Trade, Hotels, Transport & Communication

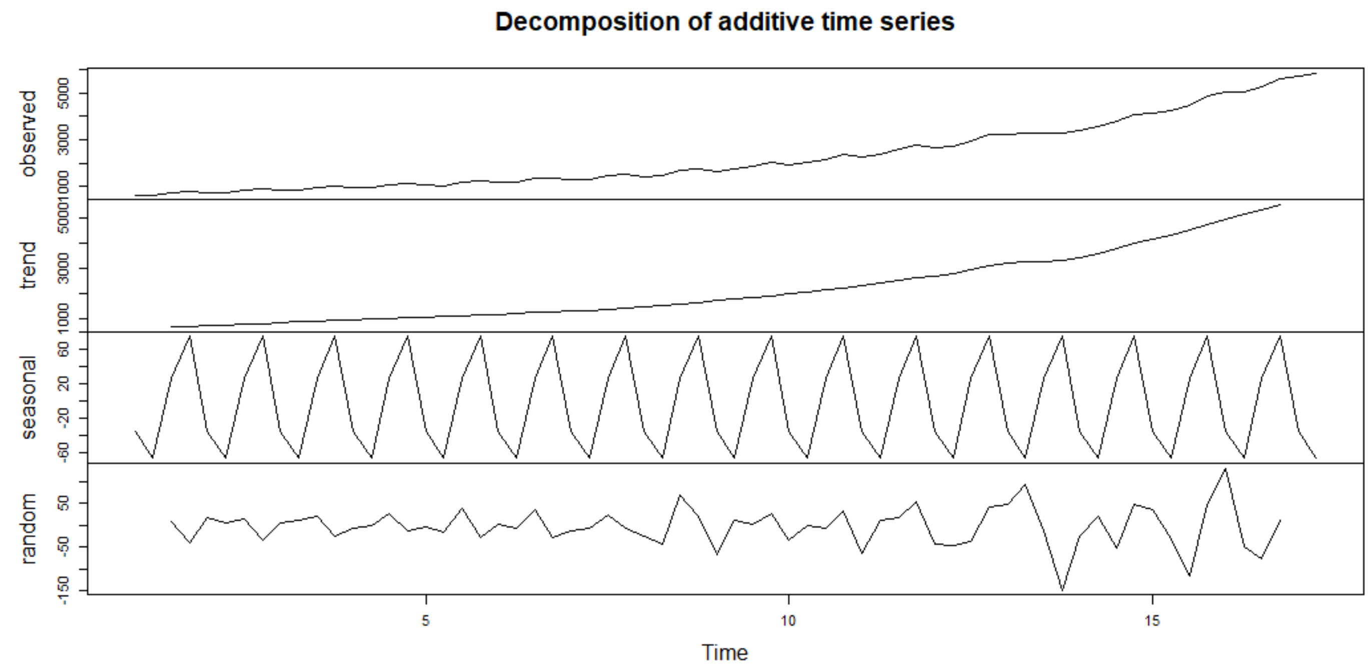
Q.1) Build a regression model that you deem to be most appropriate for modeling the trend and seasonality present in your chosen time series. Do the residuals of this regression model satisfy all the required assumptions?

Ans) The time series plot for GDP in Trade, Hotels, Transport & Communication sector :



The red line shows the actual time series plot whereas the blue line shows the trend.





Naive Modeling I: Linear Model of log-GDP with Seasonal Dummies

```
> lgdp<-log(gdptr)
> w2<-pi
> w1<-pi/2
> s1<-cos(w1*t)
> s2<-cos(w2*t)
> s3<-sin(w1*t)
> model1<-lm(lgdp~t+s1+s2+s3)
> summary(model1)

Call:
lm(formula = lgdp ~ t + s1 + s2 + s3)

Residuals:
      Min         1Q    Median         3Q        Max
-0.098415 -0.037470 -0.001464  0.042006  0.102957

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.4104354  0.0135146 474.334  < 2e-16 ***
t            0.0334398  0.0003507  95.359  < 2e-16 ***
s1           0.0391784  0.0094487   4.146 0.000106 ***
s2          -0.0010033  0.0066835  -0.150 0.881173
s3          -0.0237586  0.0094487  -2.514 0.014574 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05425 on 61 degrees of freedom
Multiple R-squared:  0.9934,	Adjusted R-squared:  0.9929
F-statistic:  2281 on 4 and 61 DF,  p-value: < 2.2e-16
```

Inference: We observe that the time series of Trade, Hotels, Transportation & Comm. has seasonality and trend components present in it. s2 seasonality component is not significant.

```
> model2<-lm(lgdp~t+s1+s3)
> summary(model2)


Call:
lm(formula = lgdp ~ t + s1 + s3)

Residuals:
     Min       1Q   Median       3Q      Max
-0.09939 -0.03705 -0.00149  0.04142  0.10397
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.4104817  0.0134042 478.245  < 2e-16 ***
t            0.0334384  0.0003478  96.149  < 2e-16 ***
s1           0.0392087  0.0093718   4.184 9.19e-05 ***
s3          -0.0237282  0.0093718  -2.532   0.0139 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05382 on 62 degrees of freedom
Multiple R-squared:  0.9934,	Adjusted R-squared:  0.993 
F-statistic:  3089 on 3 and 62 DF,  p-value: < 2.2e-16
```

Inference: All the terms are significant in this model.

Checking if Residuals are homoscedastic or not:

1.Breusch-Pagan Test

```
> bptest(model2)

	studentized Breusch-Pagan test

data:  model2
BP = 2.7979, df = 3, p-value = 0.4238
```

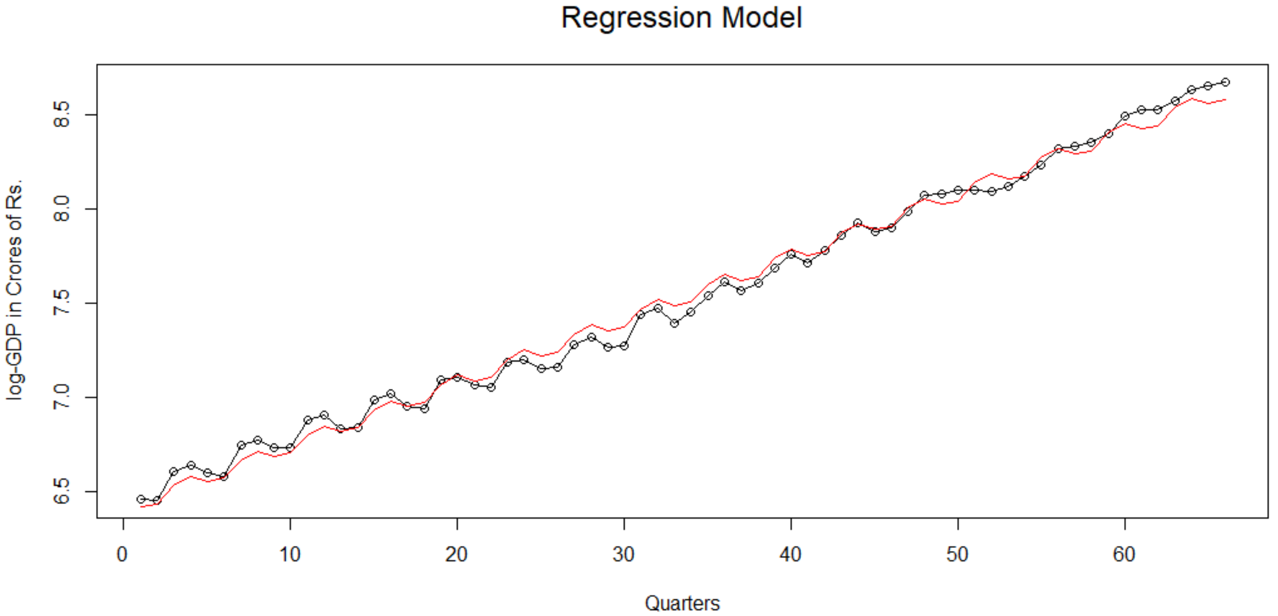
Here, we fail to reject the null hypothesis that the residual is homoscedastic. Hence, the assumption of homoscedasticity is satisfied.

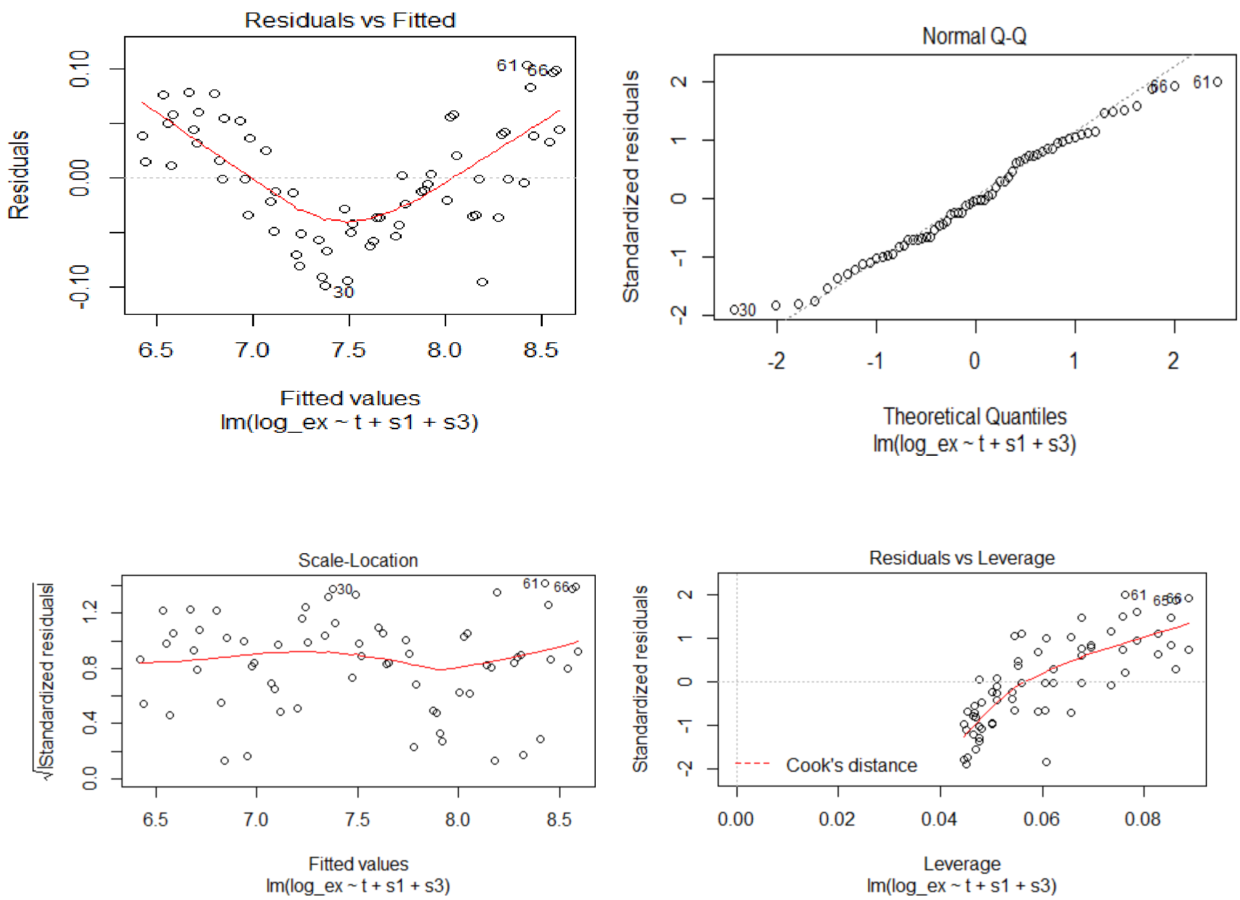
Checking if Residuals are normally distributed or not:

```
> normtest(residuals(model2))
                                     Method   P.value
1                  Shapiro-Wilk normality test 0.2803130
2              Anderson-Darling normality test 0.4189847
3              Cramer-von Mises normality test 0.3680612
4 Lilliefors (Kolmogorov-Smirnov) normality test 0.4058885
5               Shapiro-Francia normality test 0.5187843
```

Here, we fail to reject the null hypothesis that the residuals are normally distributed. Hence, the assumption of normal distribution is satisfied.

First order Fitted model vs the actual time series on a log scale:





Inferences:

- From the plots, we infer that there was some non-linear relationship between the log of gdp ,the outcome variable and the predictor variables that was not explained by our model.
- From the Residuals vs Leverage plot, we see that few of the residuals are having high leverage
- Normally distributed and Homoscedasticity assumptions seem to be valid.

Naive Modeling II: Quadratic Model of GDP with Seasonal Dummies

```
> tsq<-t^2
> model3<-lm(lgdp~t+tsq+s1+s2+s3)
> summary(model3)

Call:
lm(formula = lgdp ~ t + tsq + s1 + s2 + s3)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.094677 -0.022041  0.001579  0.018420  0.072801 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.505e+00  1.334e-02 487.670  < 2e-16 ***
t            2.514e-02  9.188e-04  27.358  < 2e-16 ***
tsq          1.239e-04  1.329e-05   9.322 2.85e-13 ***
s1           4.190e-02  6.096e-03   6.874 4.10e-09 ***
s2          -1.003e-03  4.307e-03  -0.233    0.817    
s3          -2.648e-02  6.096e-03  -4.345 5.47e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03496 on 60 degrees of freedom
```

```
Multiple R-squared:  0.9973,	Adjusted R-squared:  0.9971 
F-statistic:  4411 on 5 and 60 DF,  p-value: < 2.2e-16
```

Since, s2 is insignificant, let us remove it and refit the model.

```
> model3a<-lm(lgdp~t+tsq+s1+s3)
> summary(model3a)

Call:
lm(formula = lgdp ~ t + tsq + s1 + s3)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.095685 -0.021939  0.001557  0.019413  0.073184 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.505e+00  1.323e-02 491.553  < 2e-16 ***
t            2.514e-02  9.117e-04  27.572  < 2e-16 ***
tsq          1.239e-04  1.319e-05   9.395 1.84e-13 ***
s1           4.193e-02  6.047e-03   6.935 3.00e-09 ***
s3          -2.645e-02  6.047e-03  -4.375 4.83e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03469 on 61 degrees of freedom
Multiple R-squared:  0.9973,	Adjusted R-squared:  0.9971 
F-statistic:  5600 on 4 and 61 DF,  p-value: < 2.2e-16
```

Inference: All the terms are significant in this new model.

Checking if Residuals are homoscedastic or not:

1.Breusch-Pagan Test

```
> bptest(model3a)

	studentized Breusch-Pagan test

data:  model3a
BP = 1.1342, df = 4, p-value = 0.8888
```

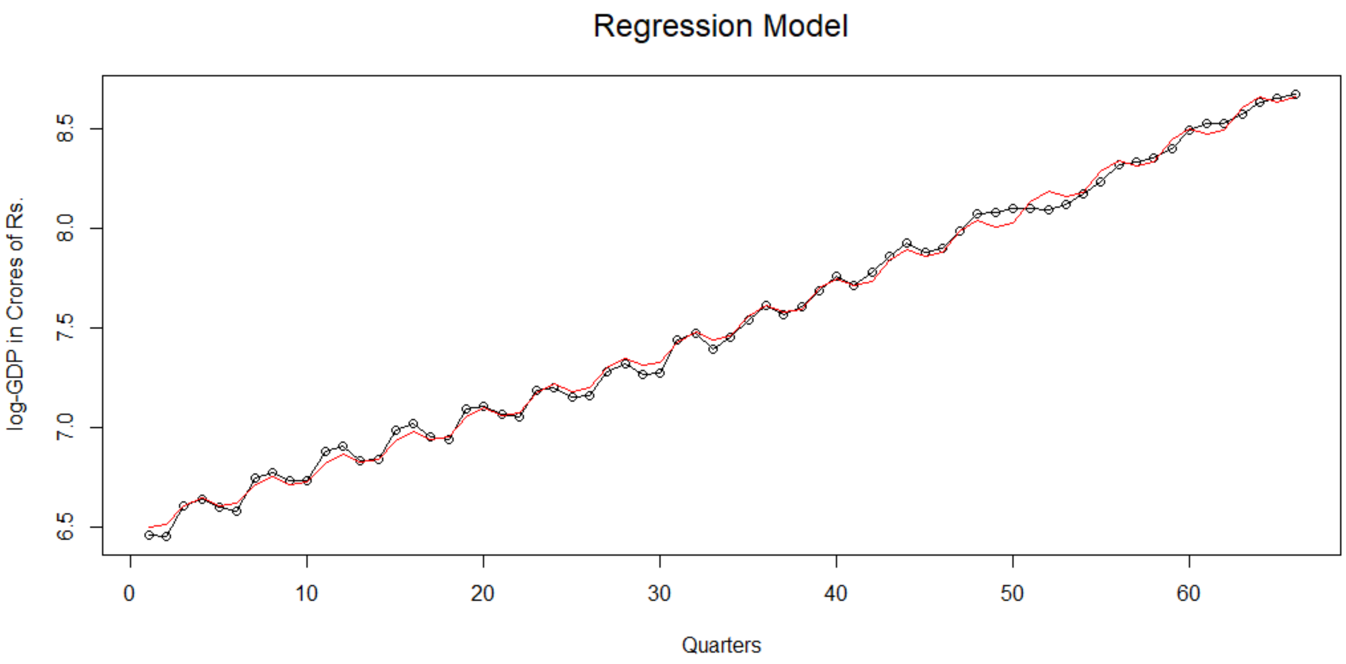
Here, we fail to reject the null hypothesis that the residual is homoscedastic. Hence, the assumption of homoscedasticity is satisfied.

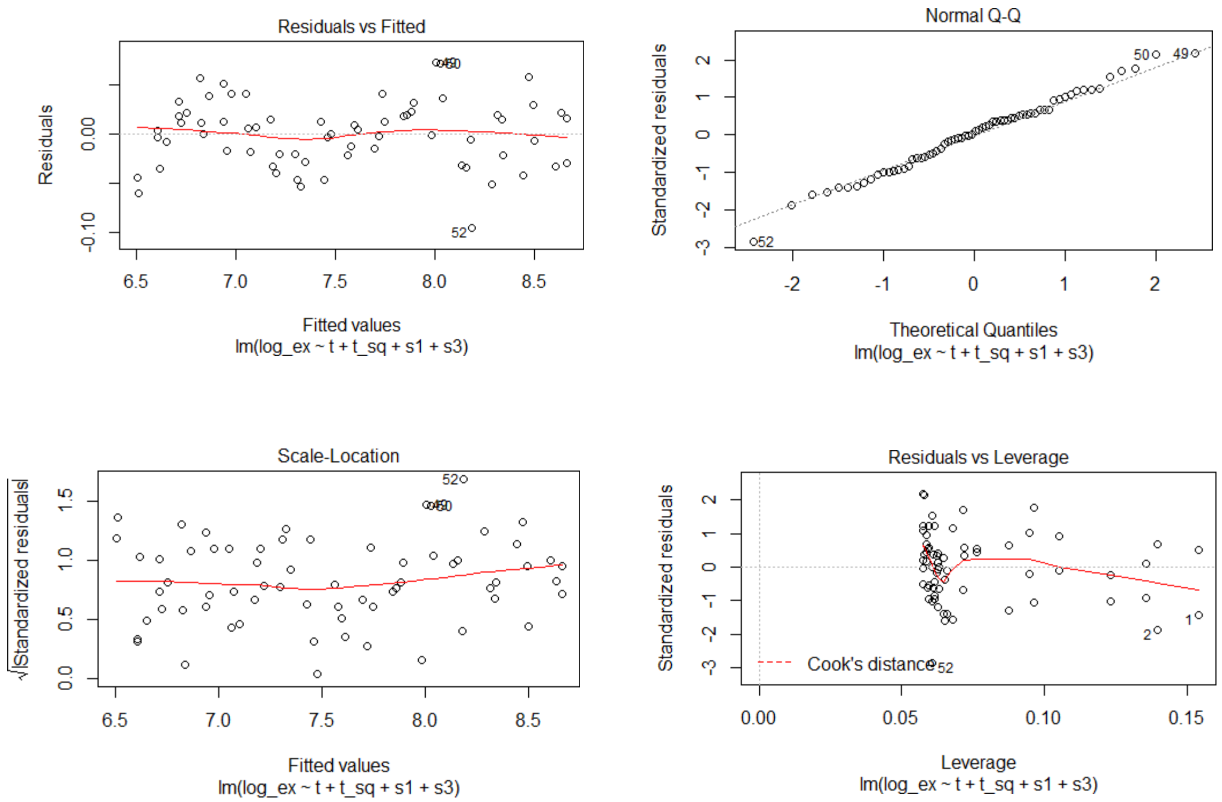
Checking if Residuals are normally distributed or not:

```
> normtest(residuals(model3a))
                                     Method   P.value
1                   Shapiro-Wilk normality test 0.8949353
2                Anderson-Darling normality test 0.8754890
3                  Cramer-von Mises normality test 0.8038212
4 Lilliefors (Kolmogorov-Smirnov) normality test 0.8681600
5                 Shapiro-Francia normality test 0.7844464
```

Here, we fail to reject the null hypothesis that the residuals are normally distributed. Hence, the assumption of normal distribution is satisfied.

Second order Fitted model vs the actual time series on a log scale:





Inferences:

1. This second order model seems to take care of the non-linearity between dependent variable and the predictor variables.

2. The assumptions of normality and homoscedasticity of residuals seem to be valid.

Checking if Residuals are white noise or not:

1. Box Pierce test:

```
> Box.test(model3a$residuals,lag=20)

        Box-Pierce test

data:  model3a$residuals
X-squared = 61.38, df = 20, p-value = 4.346e-06
```

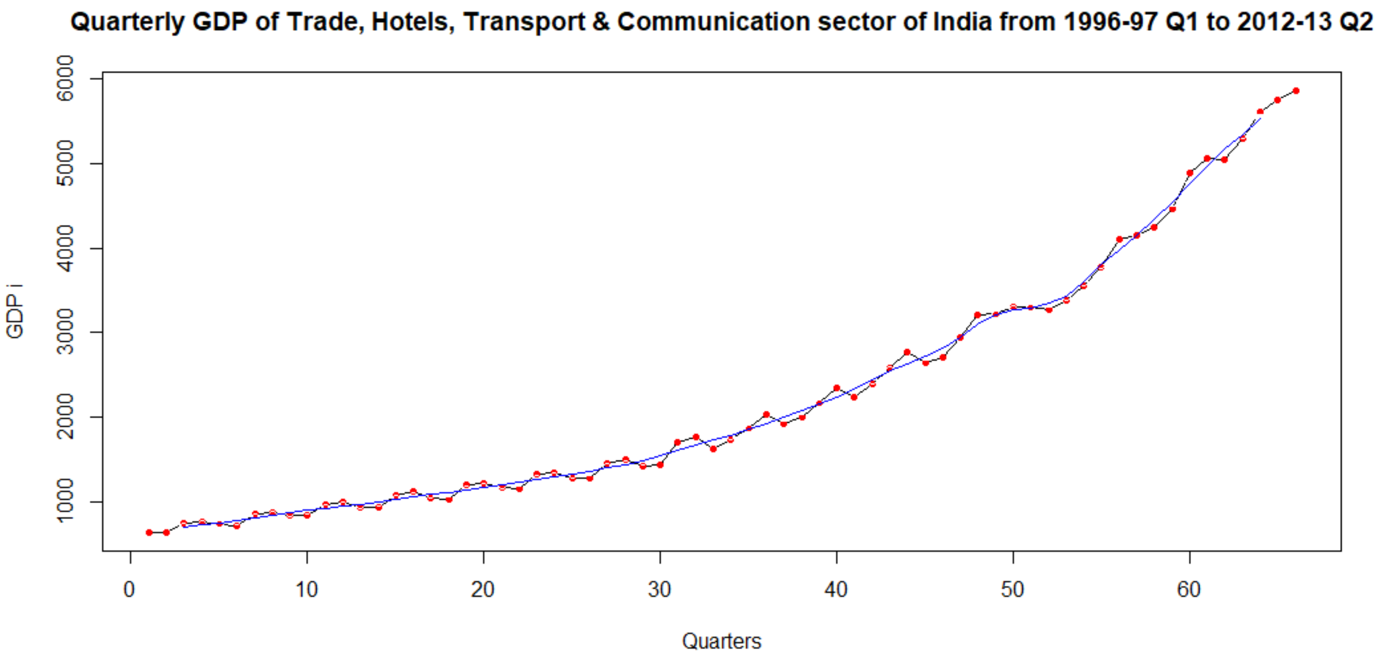
2.Box Ljung test:

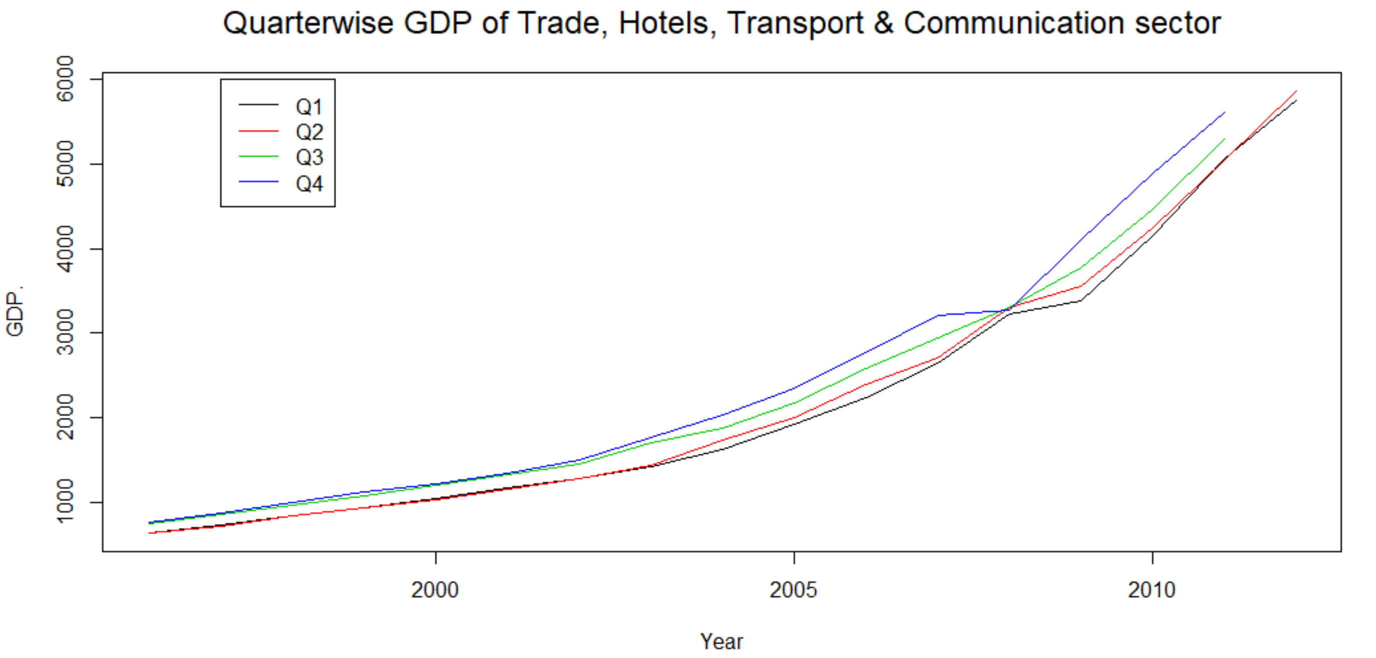
```
> Box.test(model3a$residuals,lag=20,type='L')

        Box-Ljung test

data:  model3a$residuals
X-squared = 73.983, df = 20, p-value = 4.022e-08
```

Inference: We reject the null hypothesis that residuals are white noise. So, we can say that the residual of the regression model does not satisfy the White noise assumption.

Q.2) Build an appropriate SARIMA model for the same and interpret the fitted model in terms of its IRF, ACF, PACF and η^2_n's. Do the residuals of this SARIMA model satisfy all the required assumptions?

Ans)
The time series plot of quarterly gdp:





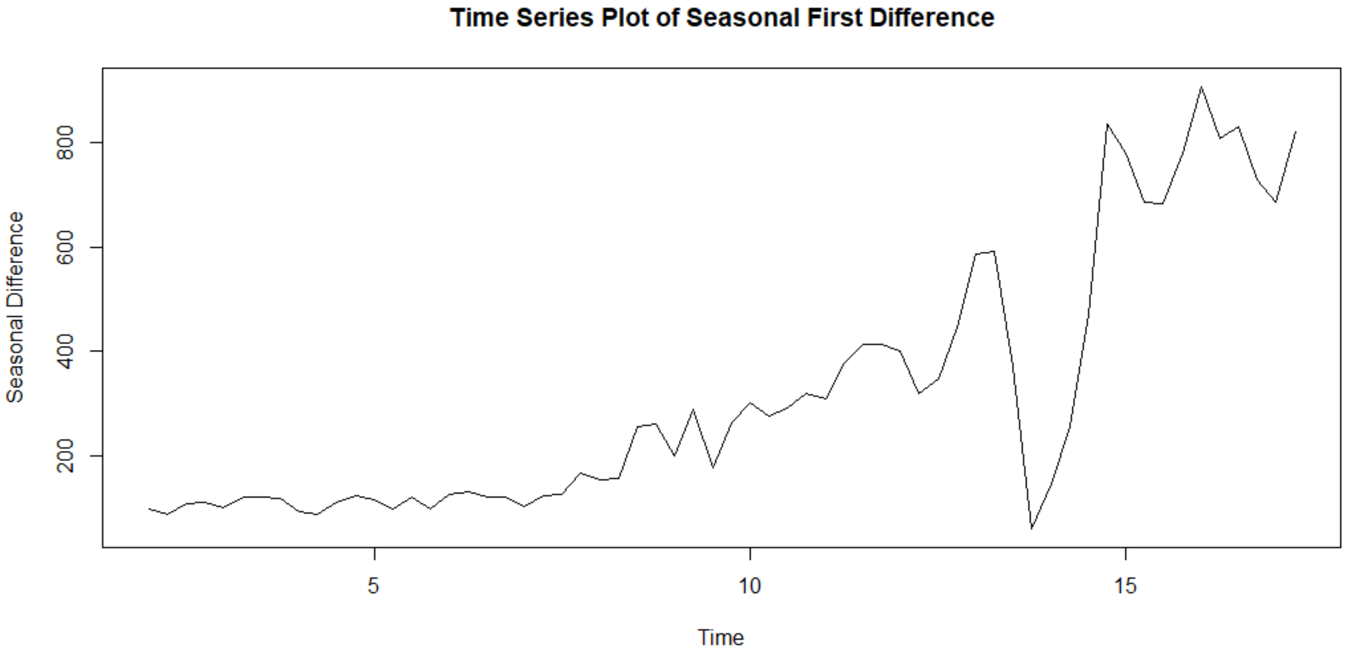
Clearly, the process does not seem to be stationary as there is a clear upward trend. Also, it looks like there is some seasonality in the series.
Let us check for stationarity using statistical tests for each quarter.

```
> adf.test(q1)$p.value: 0.99
> pp.test(q1)$p.value:  0.99
> kpss.test(q1)$p.value: 0.02010307

> adf.test(q2)$p.value: 0.99
> pp.test(q2)$p.value:  0.99
> kpss.test(q2)$p.value: 0.01955969
```

```
> adf.test(q3)$p.value: 0.99
> pp.test(q3)$p.value:  0.99
> kpss.test(q3)$p.value: 0.0214582

> adf.test(q4)$p.value: 0.99
> pp.test(q4)$p.value:  0.99
> kpss.test(q4)$p.value:0.02163848
```

The statistical tests confirm that the time series is non-stationary for each quarter.
There is a seasonal unit root for each quarter. So, let's take first seasonal difference of the series.

Seasonal first difference of lag=4:



Even after taking first seasonal difference of lag 4, the trend still seems to persist. The time series plot does not look like stationary
Let us examine this using statistical tests.

```
> sdgdptr<-diff(gdptr,lag=4)
```

1. Augmented Dickey-Fuller Test

```
> adf.test(sdgdptr)

        Augmented Dickey-Fuller Test


data:  sdgdptr
Dickey-Fuller = -2.507, Lag order = 3, p-value = 0.3697
alternative hypothesis: stationary
```

2. Phillips-Perron Unit Root Test

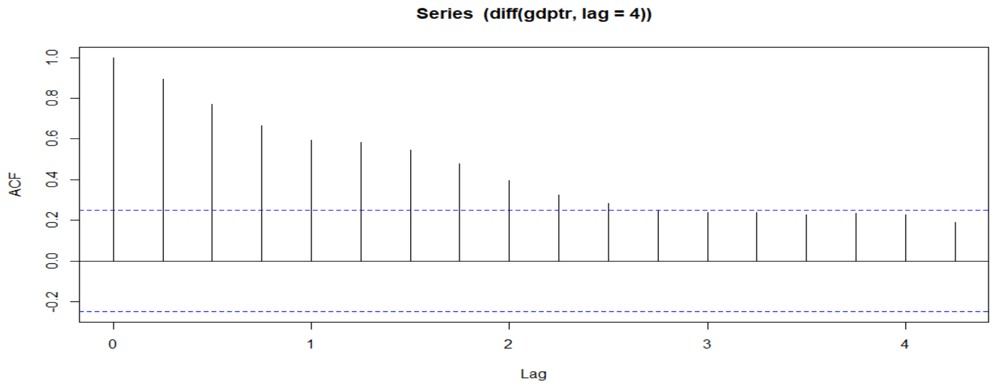
```
> pp.test(sdgdptr)

        Phillips-Perron Unit Root Test

data:  sdgdptr
Dickey-Fuller Z(alpha) = -17.006, Truncation lag parameter = 3, p-value = 0.09915
alternative hypothesis: stationary
```

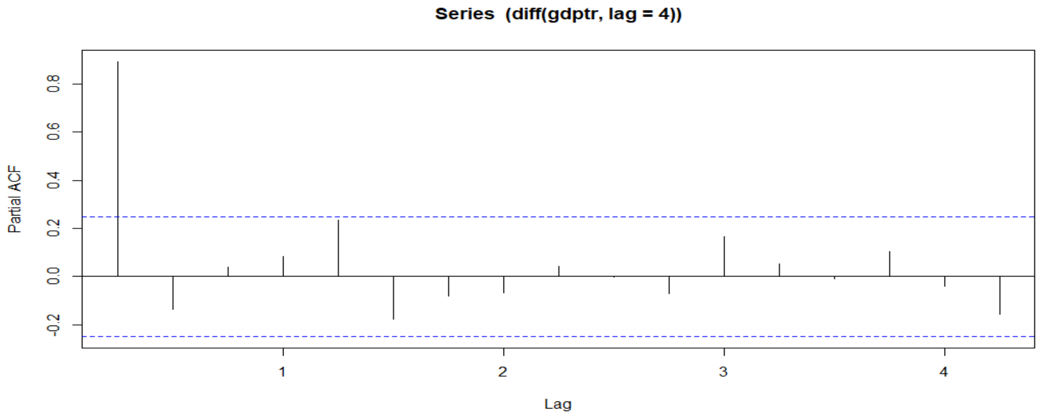
3. KPSS Test for Level Stationarity

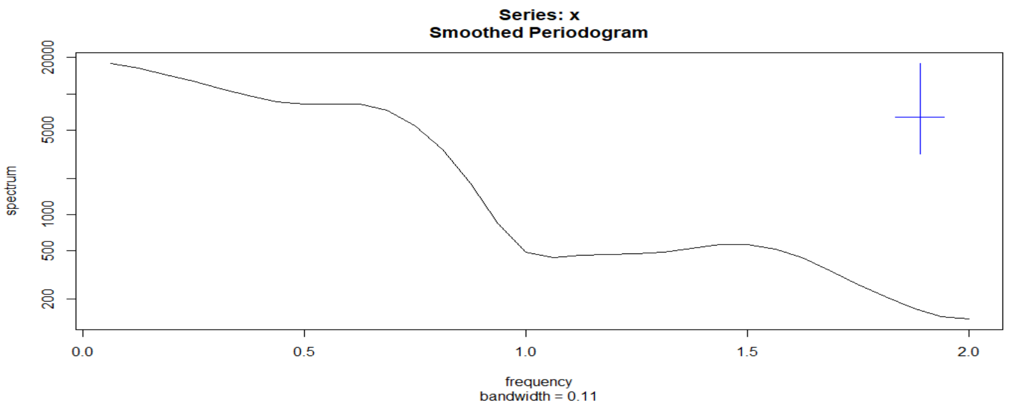
```
> kpss.test(sdgdptr)

        KPSS Test for Level Stationarity

data:  sdgdptr
KPSS Level = 1.333, Truncation lag parameter = 3, p-value = 0.01
```

Inference: The statistical tests suggest that the first seasonal differenced time series is non-stationary and there is seasonal unit root present in the series.

Plotting ACF, PACF and smoothed periodogram:

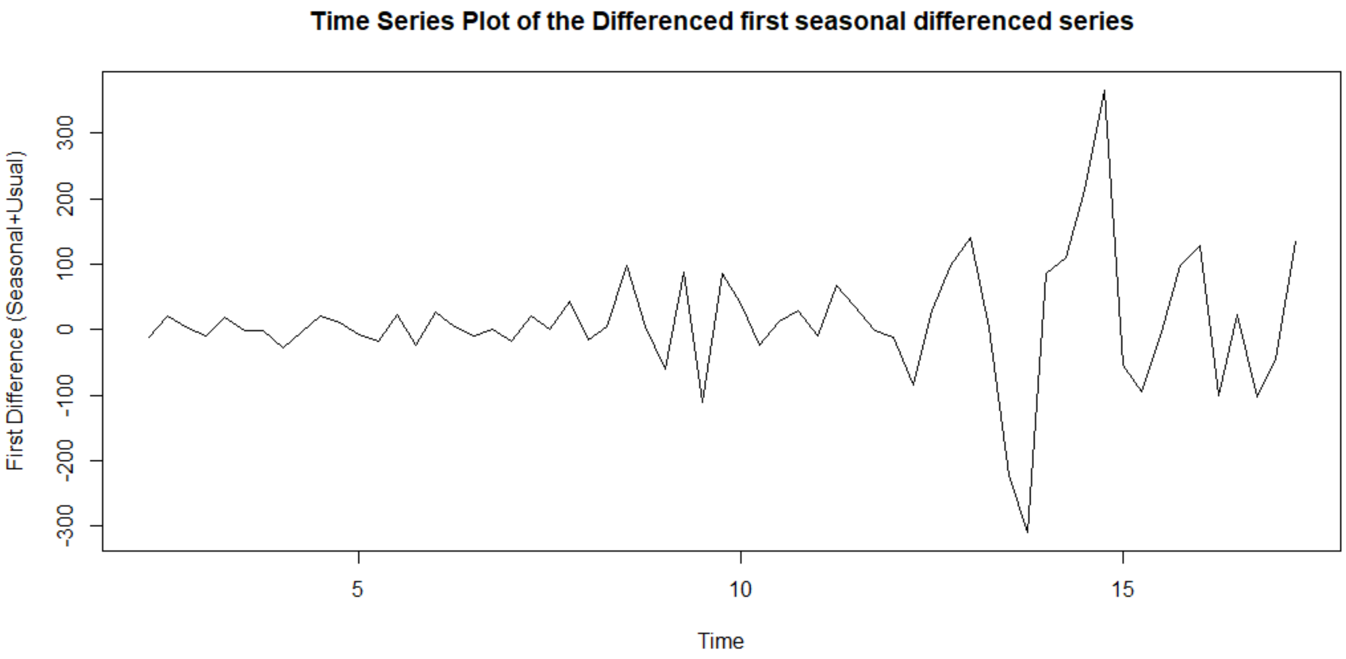






- Clearly, we can observe that there is a peak at zero in the periodogram, which implies that the time series still has a trend component.

Differenced series of the first seasonal differenced time series of lag 4:



From the plot, the time series looks like stationary.Let us carry out statistical tests to check for stationarity.

1. **Augmented Dickey-Fuller Test:**

```
> adf.test(diff(diff(gdptr,lag=4)))$p.value
[1] 0.01
```

2. **Phillips-Perron Unit Root Test**

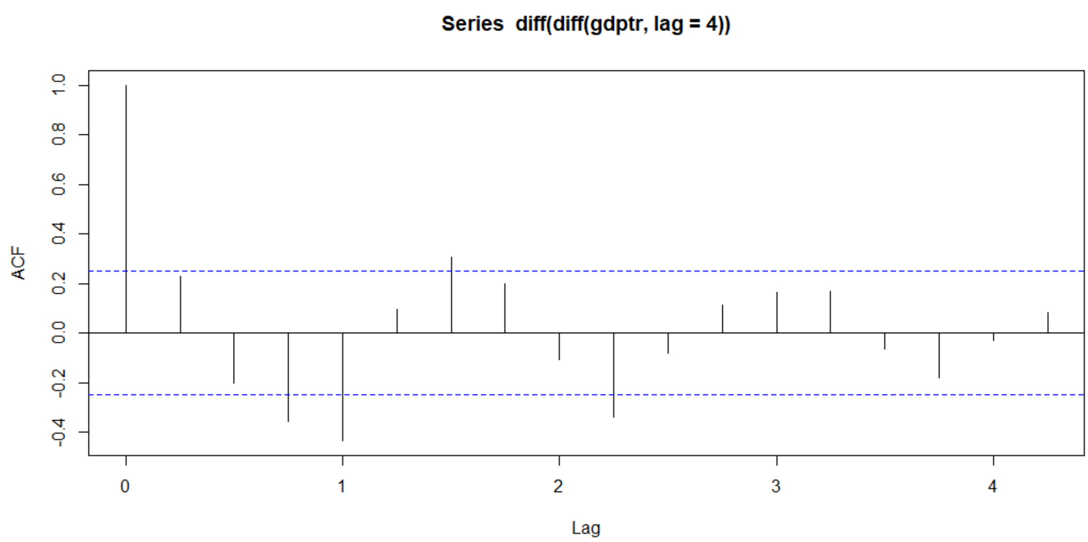
```
> pp.test(diff(diff(gdptr,lag=4)))$p.value
[1] 0.01
```

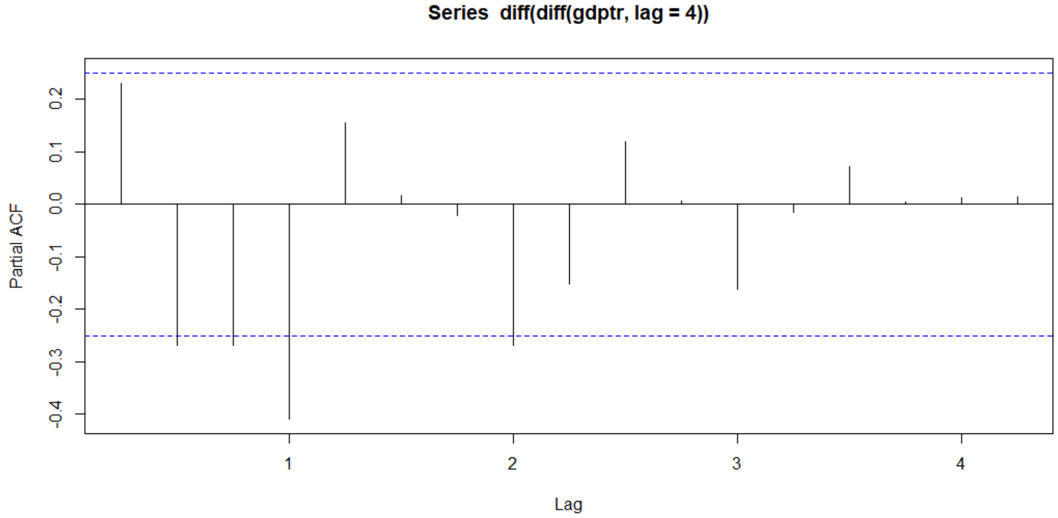
3. KPSS Test for Level Stationarity

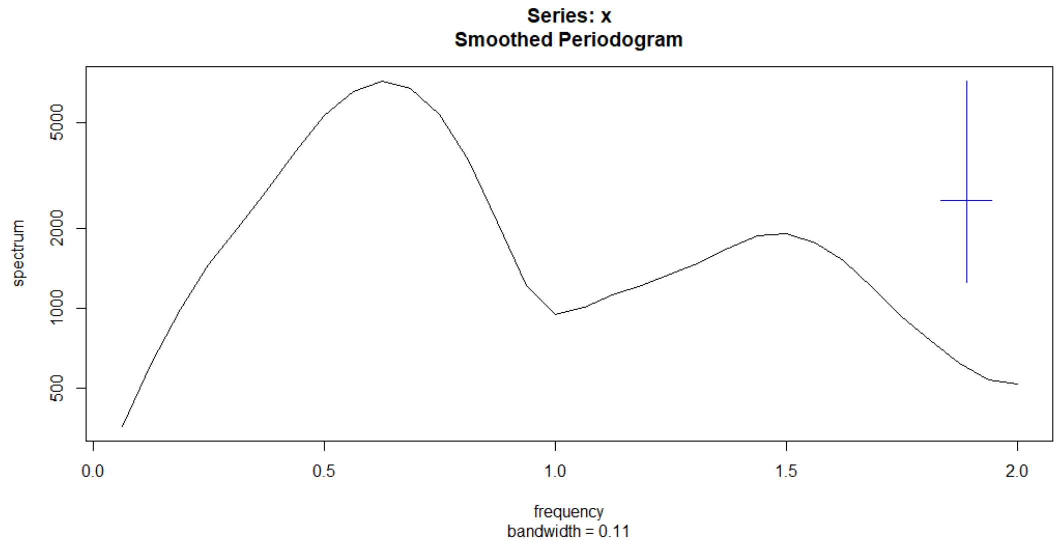
```
> kpss.test(diff(diff(gdptr,lag=4)))$p.value
[1] 0.1
```

Inference: The statistical tests confirm that the series is stationary.

Plotting ACF, PACF and smoothed periodogram:







From the above plots too, we can conclude that this time series is stationary.

Fitting a SARIMA model to this trend and seasonally differenced time series:
We try various models by putting different values for AR, MA, seasonal AR, seasonal MA terms in the model while keeping the value of d and D equal to 1 and period=4.

First 8 Sorted values of AIC in ascending order:

```
> sort(saic_10$AIC)[1:8]
 [1] 710.3564 710.3838 711.4931 711.9035 711.9228 711.9856 712.1342 712.1609
```

```
1.model1<-arima(gdptr,order=c(2,1,2),seasonal=list(order=c(0,1,1),period=4),method="ML
")
Call:
arima(x = gdptr, order = c(2, 1, 2), seasonal = list(order = c(0, 1, 1), period = 4),
    method = "ML")

Coefficients:
         ar1      ar2      ma1     ma2     sma1
      0.8803  -0.9488  -0.7055  0.8557  -0.3117
s.e.  0.0787   0.0582   0.1086  0.1308   0.1357

sigma^2 estimated as 5295:  log likelihood = -349.18,  aic = 710.36
```

2. model2<-arima(gdptr,order=c(2,1,2),seasonal=list(order=c(1,1,0),period=4),method="ML")

```
Call:
arima(x = gdptr, order = c(2, 1, 2), seasonal = list(order = c(1, 1, 0), period = 4),
    method = "ML")

Coefficients:
         ar1      ar2      ma1     ma2     sar1
      0.8879  -0.9530  -0.7108  0.8398  -0.3033
s.e.  0.0784   0.0536   0.1147  0.1351   0.1359

sigma^2 estimated as 5300:  log likelihood = -349.19,  aic = 710.38
```

3. > model3<-arima(gdptr,order=c(4,1,2),seasonal=list(order=c(0,1,0),period=4),method="ML")

```
Call:
arima(x = gdptr, order = c(4, 1, 2), seasonal = list(order = c(0, 1, 0), period = 4),
    method = "ML")

Coefficients:
         ar1      ar2     ar3      ar4      ma1     ma2
      0.4577  -0.8860  0.0309  -0.3617  -0.3104  0.8179
s.e.  0.1566   0.1548  0.1421   0.1360   0.1191  0.1060

sigma^2 estimated as 5225:  log likelihood = -348.75,  aic = 711.49
```

Since, ar3 term is not significant, we will omit that and refit the model:

3a. model3a<-arima(gdptr,order=c(4,1,2),seasonal=list(order=c(0,1,0),period=4),fixed=c(NA,NA,0,NA,NA,NA))

```
Call:
arima(x = gdptr, order = c(4, 1, 2), seasonal = list(order = c(0, 1, 0), period = 4),
    fixed = c(NA, NA, 0, NA, NA, NA))

Coefficients:
         ar1     ar2  ar3      ar4      ma1     ma2
      0.4443  -0.860    0  -0.3485  -0.3134  0.8115
s.e.  0.1443   0.097    0   0.1235   0.1192  0.1046

sigma^2 estimated as 5229:  log likelihood = -348.77,  aic = 709.54
```

4. model4<-arima(gdptr,order=c(3,1,4),seasonal=list(order=c(0,1,0),period=4),method="ML")

```
Call:
arima(x = gdptr, order = c(3, 1, 4), seasonal = list(order = c(0, 1, 0), period = 4),
    method = "ML")

Coefficients:
         ar1      ar2      ar3      ma1     ma2     ma3      ma4
      0.2989  -0.3846  -0.3942  -0.1036  0.2906  0.4011  -0.4168
s.e.  0.2753   0.2369   0.2660   0.2576  0.1781  0.2297   0.1317

sigma^2 estimated as 5076:  log likelihood = -347.95,  aic = 711.9
```

Many terms are not significant in this model.

5. > model5<-arima(gdptr,order=c(2,1,4),seasonal=list(order=c(0,1,0),period=4),method="ML")

```
Call:
arima(x = gdptr, order = c(2, 1, 4), seasonal = list(order = c(0, 1, 0), period = 4),
    method = "ML")
Coefficients:
         ar1      ar2      ma1     ma2     ma3      ma4
      0.6830  -0.7130  -0.4482  0.5160  0.0769  -0.3973
s.e.  0.1499   0.1411   0.1903  0.1929  0.1770   0.1608

sigma^2 estimated as 5245:  log likelihood = -348.96,  aic = 711.92
```

ma3 term is not significant so, we will remove that from our model and refit it.

5a. > model5a<-arima(gdptr,order=c(2,1,4),seasonal=list(order=c(0,1,0),period=4),fixed=c(NA,NA,NA,NA,0,NA))

```
Call:
arima(x = gdptr, order = c(2, 1, 4), seasonal = list(order = c(0, 1, 0), period = 4),
    fixed = c(NA, NA, NA, NA, 0, NA))

Coefficients:
         ar1      ar2      ma1     ma2  ma3      ma4
      0.7117  -0.7292  -0.5141  0.5801    0  -0.3455
s.e.  0.1307   0.1324   0.1171  0.1335    0   0.1135

sigma^2 estimated as 5268:  log likelihood = -349.04,  aic = 710.09
```

6. > model6<-arima(gdptr,order=c(4,1,2),seasonal=list(order=c(1,1,0),period=4),method="ML")

```
Call:
arima(x = gdptr, order = c(4, 1, 2), seasonal = list(order = c(1, 1, 0), period = 4),
    method = "ML")

Coefficients:
         ar1      ar2     ar3      ar4      ma1     ma2     sar1
      0.5369  -0.9829  0.0973  -0.3163  -0.3616  0.8381  -0.2094
s.e.  0.1697   0.1799  0.1603   0.1416   0.1308  0.1117   0.1606

sigma^2 estimated as 5086:  log likelihood = -347.99,  aic = 711.99
```

sar1,ar3 terms are insignificant, so we drop the them and refit the model.

6a. model6a<-arima(gdptr,order=c(4,1,2),seasonal=list(order=c(1,1,0),period=4),fixed=c(NA,NA,0,NA,0,NA,0))

```
Call:
arima(x = gdptr, order = c(4, 1, 2), seasonal = list(order = c(1, 1, 0), period = 4),
    fixed = c(NA, NA, 0, NA, 0, NA, 0))

Coefficients:
         ar1      ar2  ar3      ar4  ma1     ma2  sar1
      0.1868  -0.7201    0  -0.4584    0  0.6100     0
s.e.  0.0979   0.1481    0   0.1070    0  0.1617     0

sigma^2 estimated as 5565:  log likelihood = -350.3,  aic = 710.61
```

7. model7<-arima(gdptr,order=c(4,1,2),seasonal=list(order=c(0,1,1),period=4),method="ML")

```
Call:
arima(x = gdptr, order = c(4, 1, 2), seasonal = list(order = c(0, 1, 1), period = 4),
    method = "ML")

Coefficients:
         ar1      ar2     ar3      ar4      ma1     ma2     sma1
      0.5449  -0.9727  0.0934  -0.3082  -0.3723  0.8356  -0.1999
s.e.  0.1807   0.1810  0.1625   0.1470   0.1404  0.1136   0.1686

sigma^2 estimated as 5100:  log likelihood = -348.07,  aic = 712.13
```

ar3 term is not significant, so we remove that and fit the model

7a. model7a<-arima(gdptr,order=c(4,1,2),seasonal=list(order=c(0,1,1),period=4),fixed=c(NA,NA,0,NA,NA,NA,NA))

```
Call:
arima(x = gdptr, order = c(4, 1, 2), seasonal = list(order = c(0, 1, 1), period = 4),
    fixed = c(NA, NA, 0, NA, NA, NA, NA))
```

```
Coefficients:
         ar1      ar2  ar3      ar4      ma1     ma2     sma1
      0.5016  -0.8872    0  -0.2766  -0.3768  0.8123  -0.1710
s.e.  0.1722   0.0989    0   0.1462   0.1506  0.1118   0.1648

sigma^2 estimated as 5131:  log likelihood = -348.23,  aic = 710.47
```

```
8. model8<-arima(gdptr,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=4),method="M
L")

Call:
arima(x = gdptr, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 4),
    method = "ML")

Coefficients:
         ma1     sma1
      0.2838  -0.4810
s.e.  0.1251   0.1098

sigma^2 estimated as 6123:  log likelihood = -353.08,  aic = 712.16
```

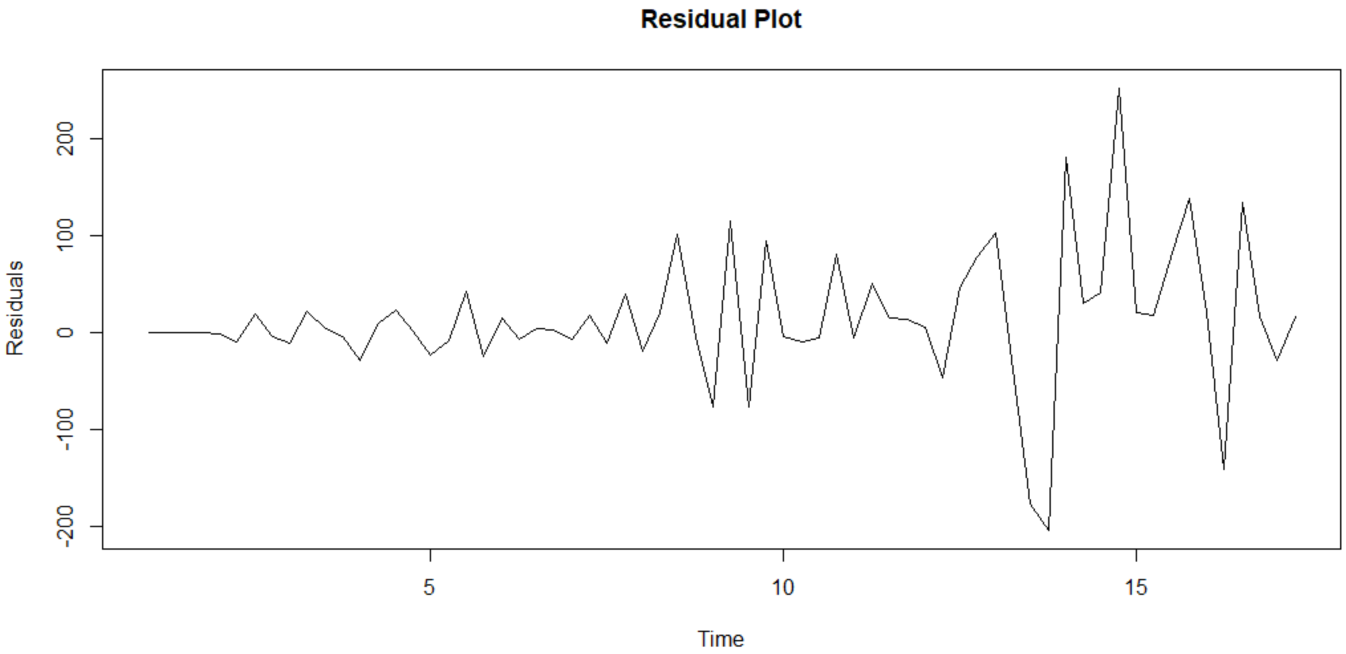
From the above models, we choose the model with least AIC value and all terms being significant.

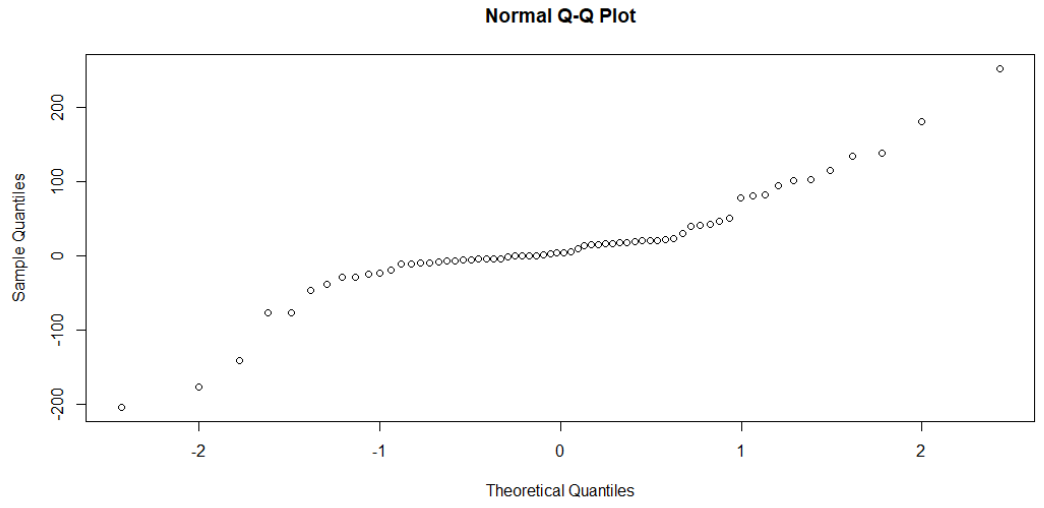
Hence, our selected model is model 1.

```
model1<-arima(gdptr,order=c(2,1,2),seasonal=list(order=c(0,1,1),period=4),method="ML")

Call:
arima(x = gdptr, order = c(2, 1, 2), seasonal = list(order = c(0, 1, 1), period = 4),
    method = "ML")

Coefficients:
         ar1      ar2      ma1     ma2     sma1
      0.8803  -0.9488  -0.7055  0.8557  -0.3117
s.e.  0.0787   0.0582   0.1086  0.1308   0.1357

sigma^2 estimated as 5295:  log likelihood = -349.18,  aic = 710.36
```

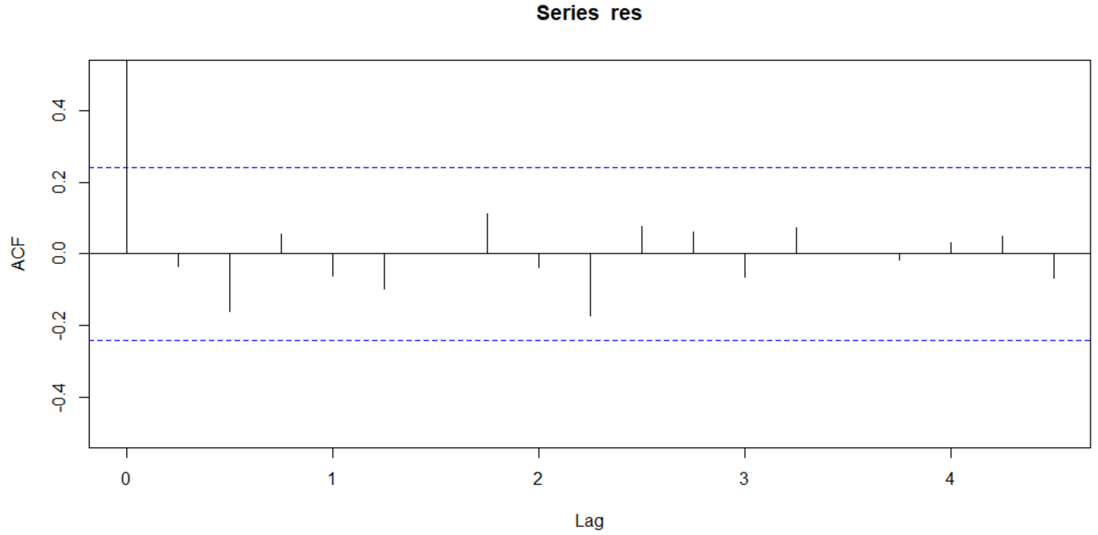
Checking for Model Assumptions

Graphical methods of Residual Analysis

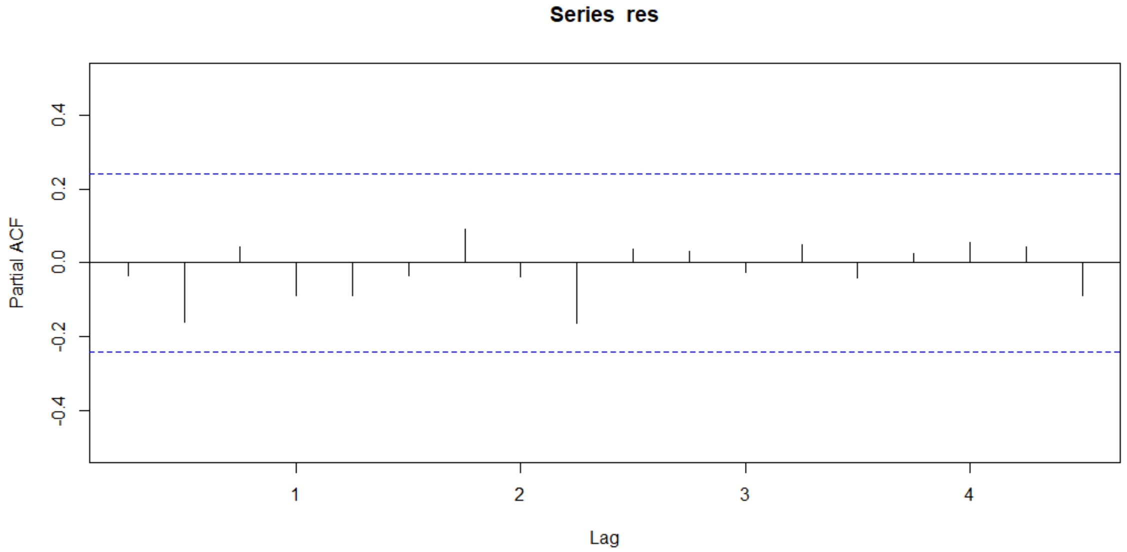




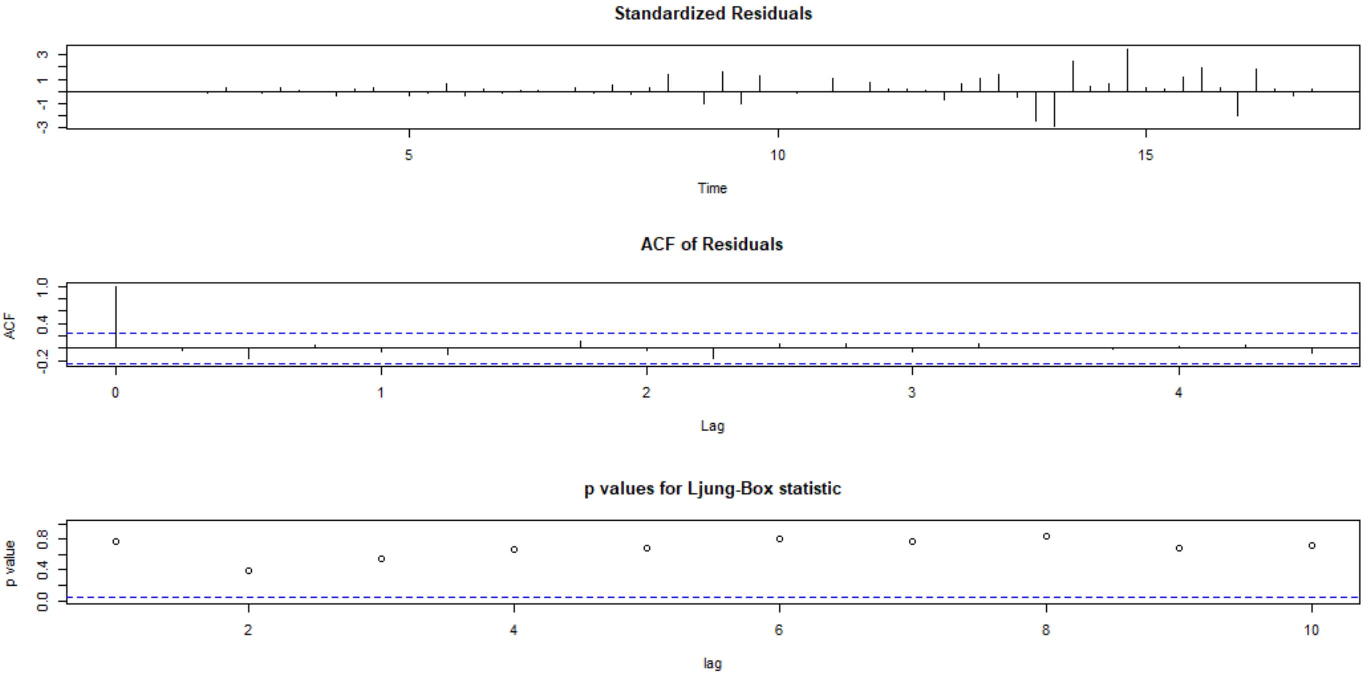
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles

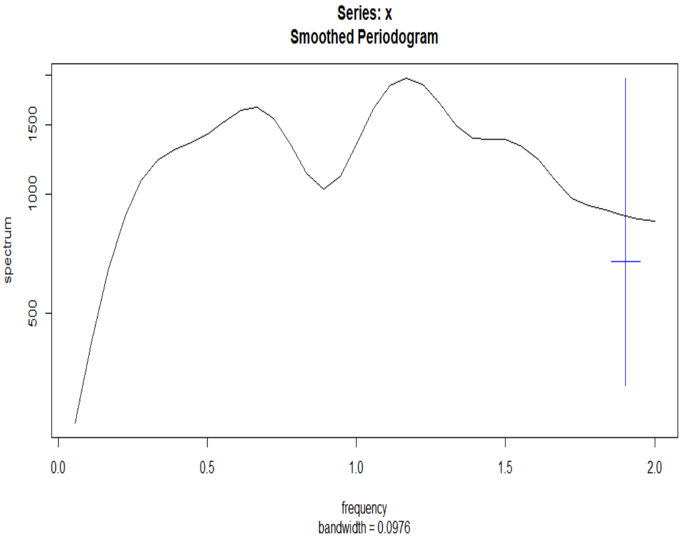


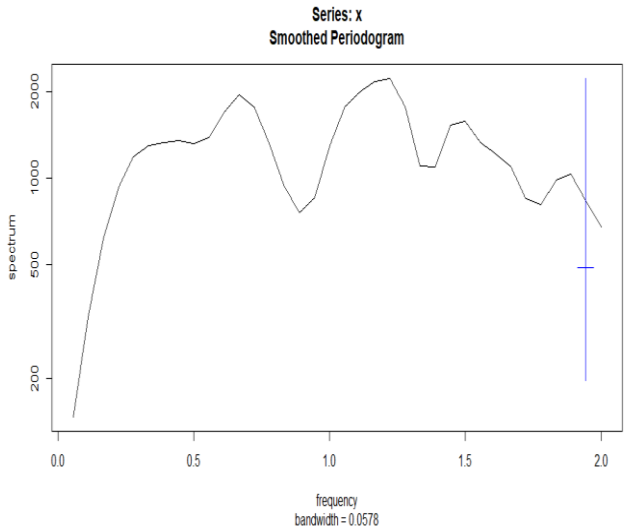
Series res
ACF
Lag

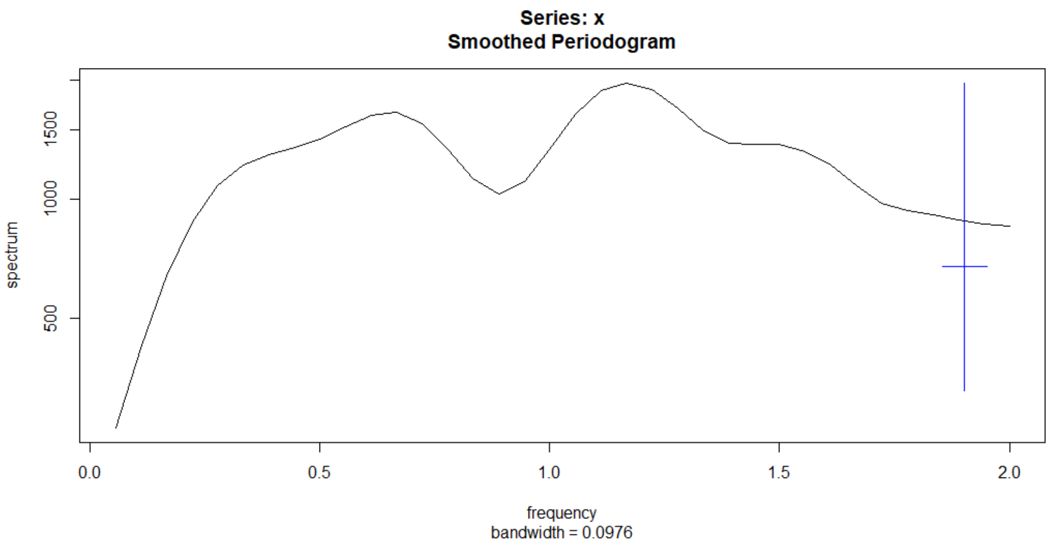


Series res
Partial ACF
Lag


Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag


Series: x
Smoothed Periodogram
spectrum
frequency
bandwidth = 0.0976


Series: x
Smoothed Periodogram
spectrum
frequency
bandwidth = 0.0578


Series: x
Smoothed Periodogram
spectrum
frequency
bandwidth = 0.0976

Inferences:

- From the plot of the residuals we observe that there does not seem to be any pattern in the series.
- From the normal Q-Q plot, we observe that the assumption of normality of residuals seems to be violated.
- The ACF and PACF plots show that there is no autocorrelation or partial autocorrelation at non-zero lags in the residual series.
- From the periodogram, we can observe that there is no trend in the residual series.

Checking for normality of residuals using statistical tests:

```
> normtest(res)
                                       Method      P.value
1                  Shapiro-Wilk normality test 1.510529e-05
2              Anderson-Darling normality test 4.619525e-08
3              Cramer-von Mises normality test 2.109414e-07
4 Lilliefors (Kolmogorov-Smirnov) normality test 4.427708e-06
5               Shapiro-Francia normality test 1.944421e-05
```

We can conclude that the residuals are not normally distributed. However, this is fine if it passes the other model assumptions as the distribution need not be normal always.

Statistical tests to check for White noise:

1. Box-Pierce Test:

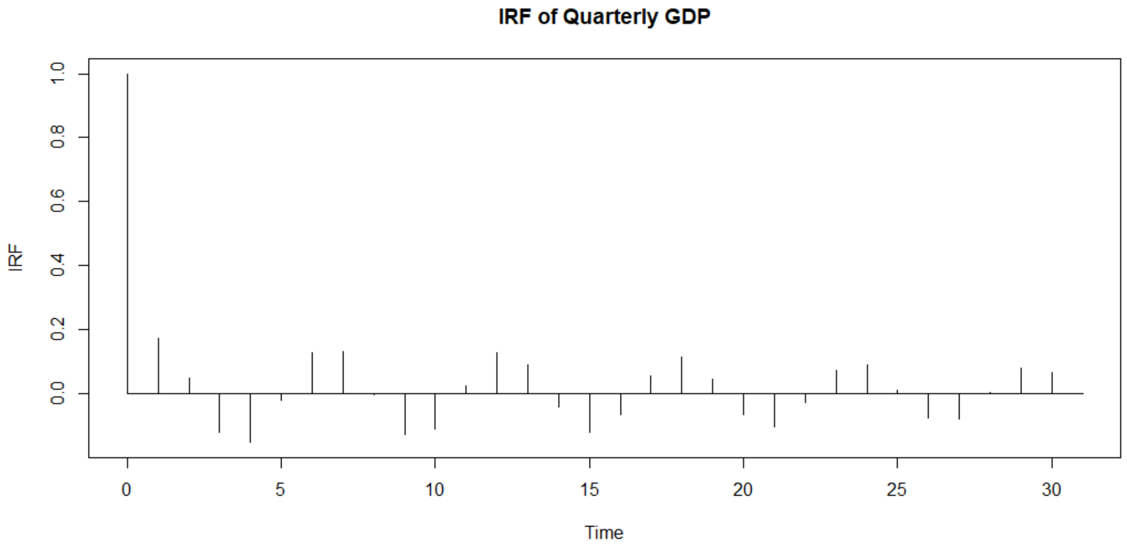
```
> Box.test(res,lag=20)

Box-Pierce test

data:  res
X-squared = 8.6263, df = 20, p-value = 0.9868
```

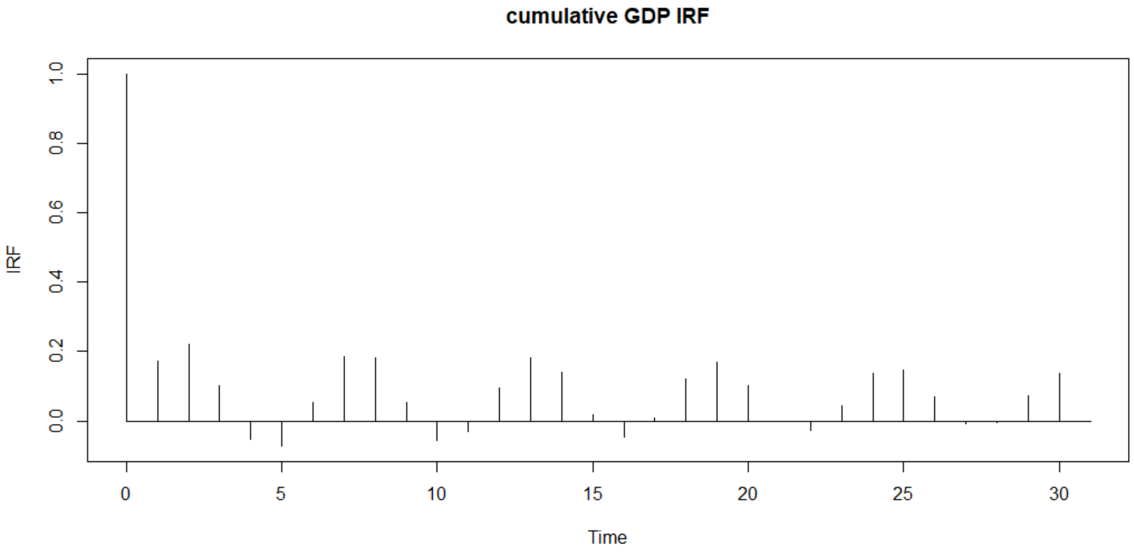
2.Box-Ljung Test:

```
> Box.test(res,lag=20,type='L')

        Box-Ljung test

data:  res
X-squared = 10.378, df = 20, p-value = 0.9608
```

Inference: The residuals are white noise.

IRF Plot of final Model:





The IRF describes the evolution of the variable of interest along a specified time horizon after a shock in a given moment. Since the IRF at lag 0 is high, we can say that the present value is most affected by the current shock. It gradually deceases and then remains constant throughout. Thus the effect of past shocks never become exactly zero ever. Also, the pattern more or less repeats itself after every 4 quarters.

There is at least some permanent effect of past shocks initiated by fiscal policy changes on Current GDP.

%Variance Explained by the model:

Note: Using ARMAtoMA function to approximately find the psi-weights for the series.

```
> psi100<-ARMAtoMA(ar=c(0.8831,-0.9501), ma=c(-0.7107,0.8471),lag.max=100)
> gamma0<-(1+sum(psi100^2))*model1$sigma2
> gamma0
[1] 6942.211
```

```
> pc2<-ARMAacf(ar=c(0.8831,-0.9501), ma=c(-0.7107,0.8471),lag.max=5,pacf=T)^2
> v<-gamma0
> for(i in 2:6) v[i]<-v[i-1]*(1-pc2[i-1])
> ((1-v/gamma0)*100)[-1]

[1]  5.400894  7.906072 14.383032 14.739868 16.840474

> pc2*100

[1] 5.4008945 2.6482038 7.0329936 0.4167824 2.4637604

> cumsum(pc2*100)

[1] 5.400894  8.049098 15.082092 15.498874 17.962635

> ((gamma0-model1$sigma2)/gamma0)*100

[1] 26.729260
```

Inference: If all lags are included, then our model can explain around 26.7% of the total variance in the data.

Q.3) Use the above two fitted models for finding interval forecasts for 2012-13:Q3, 2012- 13:Q4, 2013-14:Q1 and 2013-14:Q2 and check against their realized values. Compare these forecasts and other diagnostics to comment on the aptness of the above two fitted models.

Ans)

1. REGRESSION MODEL

```
lm(formula = lgdp ~ t + tsq + s1 + s3)
```

Quarter	Actual Value	Fitted Value	Lower Bound	Upper Bound
2012-13,Q3	6044.79	6447.742	5979.446	6952.714
2012-13,Q4	6432.43	6828.295	6328.543	7367.511
2013-14,Q1	6253.23	6651.206	6162.940	7178.155
2013-14,Q2	6470.18	6832.412	6326.490	7378.792

MAPE(Mean Absolute Percentage Error)= 0.06195

RMSE(Root Mean Squared Error)= 390.0885

2. SARIMA MODEL

```
arima(gdptr,order=c(2,1,2),seasonal=list(order=c(0,1,1),period=4),method="ML")
```

Year-Quarter	Actual Value	Fitted Value	Lower Bound	Upper Bound
2012-13,Q3	6044.79	6156.460	6010.954	6301.965
2012-13,Q4	6432.43	6474.899	6250.681	6699.117
2013-14,Q1	6253.23	6537.823	6251.687	6823.958
2013-14,Q2	6470.18	6578.923	6250.965	6906.881

MAPE(Mean Absolute Percentage Error)= 0.02184853

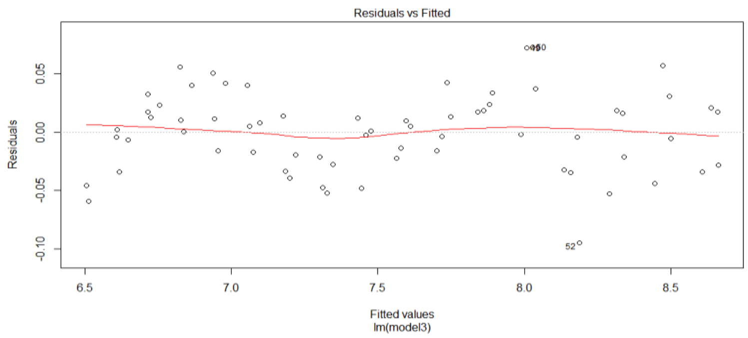
RMSE(Root Mean Squared Error)= 163.6244

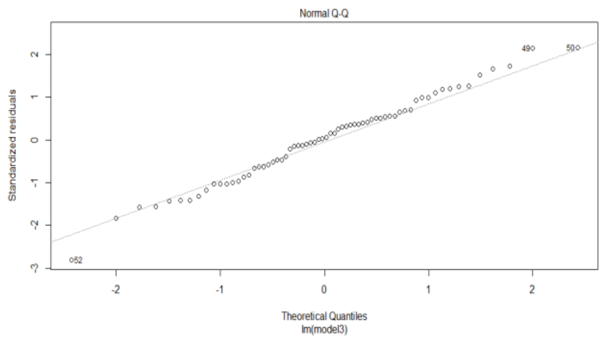
Inference:

- We observe from above tables that the SARIMA model performs better than the Regression model.
- RMSE and MAPE for SARIMA model is considerably lower than that of the Regression model.
- The forecast interval for Regression model is wider than that of the SARIMA model.
- In the SARIMA model, actual values for all the quarters fall within the prediction interval.

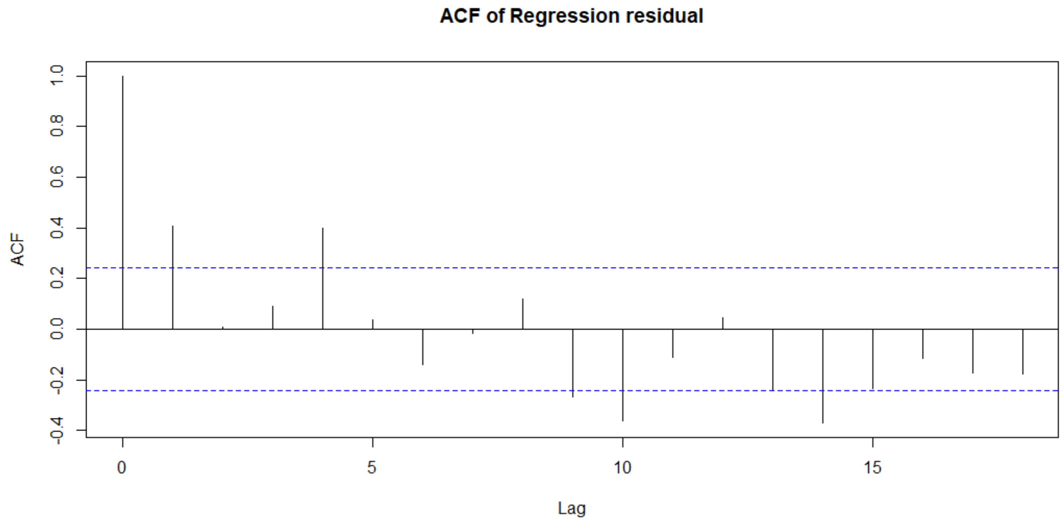
Checking the validity of assumptions of Regression model:

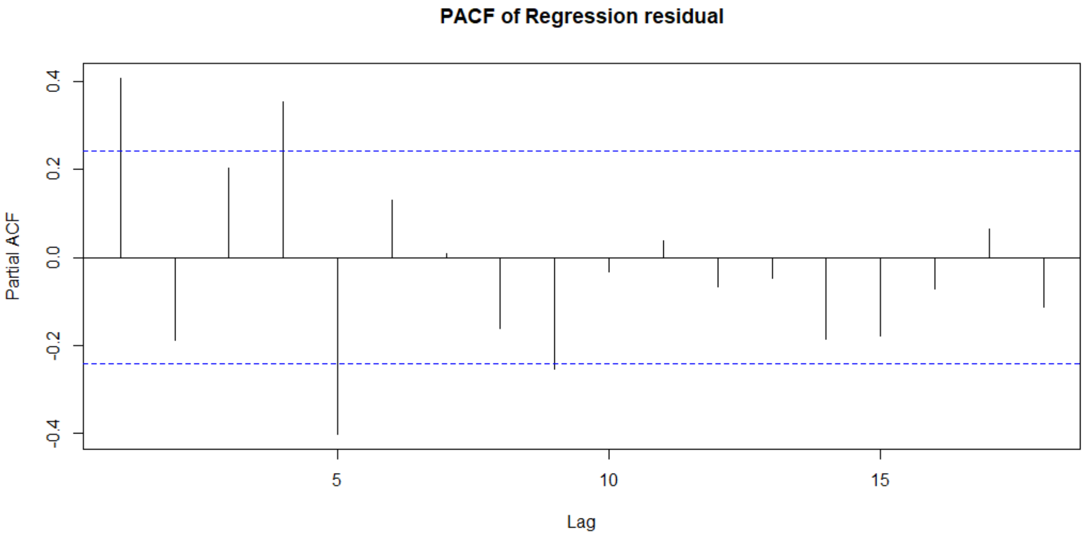
Residual plots:





ACF and PACF plots:





Inference:

- The ACF and PACF plots show that there is some autocorrelation or partial autocorrelation at non-zero lags in the residual series of the regression model.

Checking if Residuals are white noise or not:

1. Box Pierce test:

```
data:  model3a$residuals
X-squared = 61.38, df = 20, p-value = 4.346e-06
```

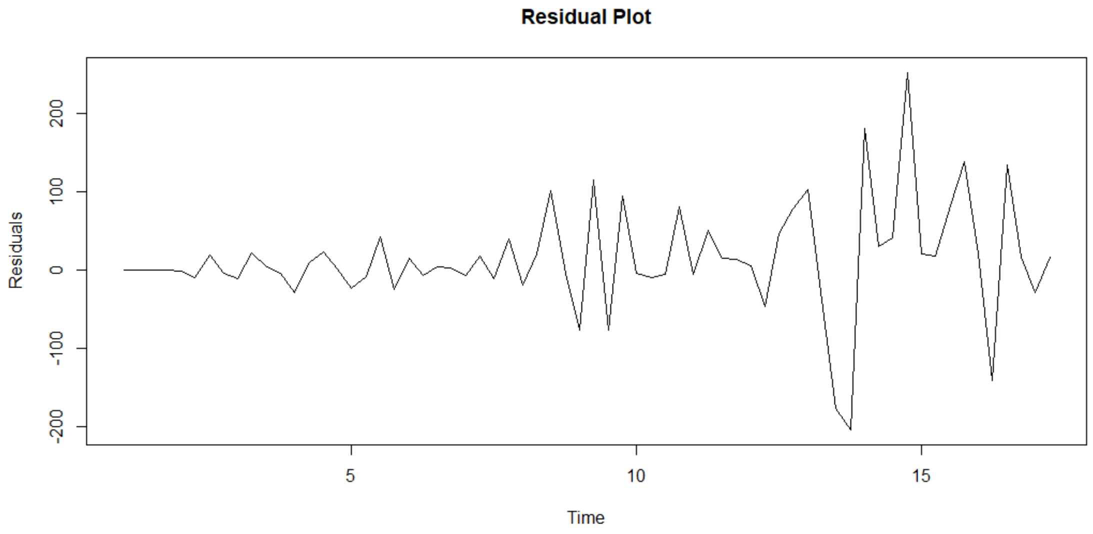
2.Box Ljung test

```
data:  model3a$residuals
X-squared = 73.983, df = 20, p-value = 4.022e-08
```

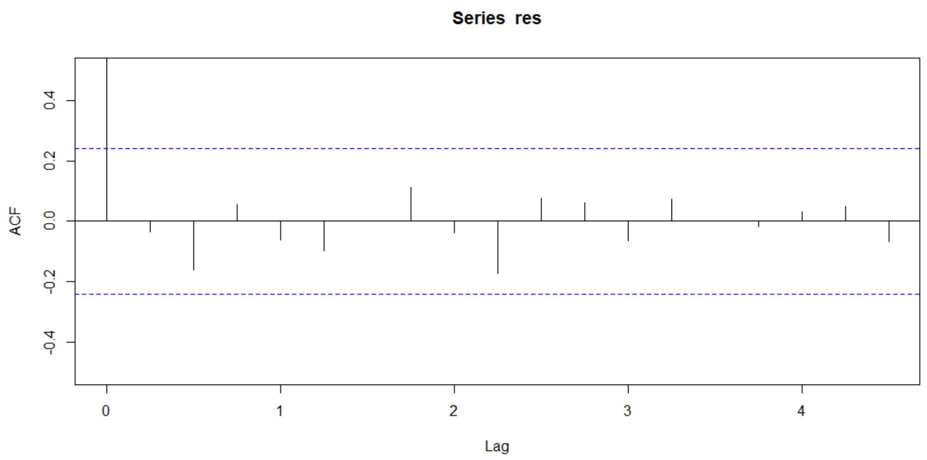
Inference: The residual of the regression model does not satisfy the White noise assumption.

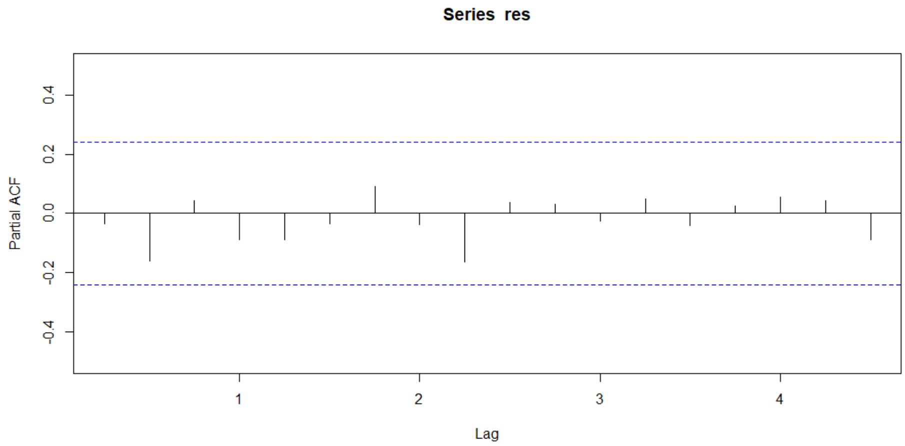
Checking the validity of assumptions of SARIMA model:

Residual Plot:



ACF and PACF of the SARIMA model:





Inference:
The ACF and PACF plots show that there is no autocorrelation or partial autocorrelation at non-zero lags in the residual series of the SARIMA model.

Checking if Residuals are white noise or not:

1. Box-Pierce Test:
```
data:  res
X-squared = 8.6263, df = 20, p-value = 0.9868
```

2.Box-Ljung Test:
```
data:  res
X-squared = 10.378, df = 20, p-value = 0.9608
```

Inference: The residuals of the SARIMA model is white noise.

Conclusion:

The regression approach to modeling time series failed the white noise assumption of the residuals whereas the residuals of the SARIMA model satisfied the white noise assumption, indicating no further pattern that can be modeled. The SARIMA model captures the true value of the data more precisely than the regression model with lower error compared to the Regression model.
Hence, **the SARIMA model is more apt for modeling the quarterly GDP time series data in Trade, Hotels, Transport & Communication sector than the regression model.**