**Assignment **

Name: Umme Rubaiyat Chowdhury ID: 20216021, Batch:6th

#Task Select the R dataset with your ID: modulo (12, 5) = 2 in R 12 %% 5 = 2 ■ Students with ID_last_two_digits %% 6 = 0 use nottem ■ Students with ID_last_two_digits %% 6 = 1 use USAccDeaths ■ Students with ID_last_two_digits %% 6 = 2 use austres ■ Students with ID_last_two_digits %% 6 = 3 use UKgas ■ Students with ID_last_two_digits %% 6 = 4 use AirPassengers ■ Students with ID_last_two_digits %% 6 = 5 use UKDriverDeaths

21%%6 == 3 Hence, As per the assignment instruction my selected Dataset is UKgas

According to the instruction my modulo is 3. So I will work with UKgas Dataset

Loading Dataset

In [1]:

```
data(UKgas)
```

Data Description

UKgas - A quarterly time series of length 108.

Source

Durbin, J. and Koopman, S. J. (2001) Time Series Analysis by State Space Methods.

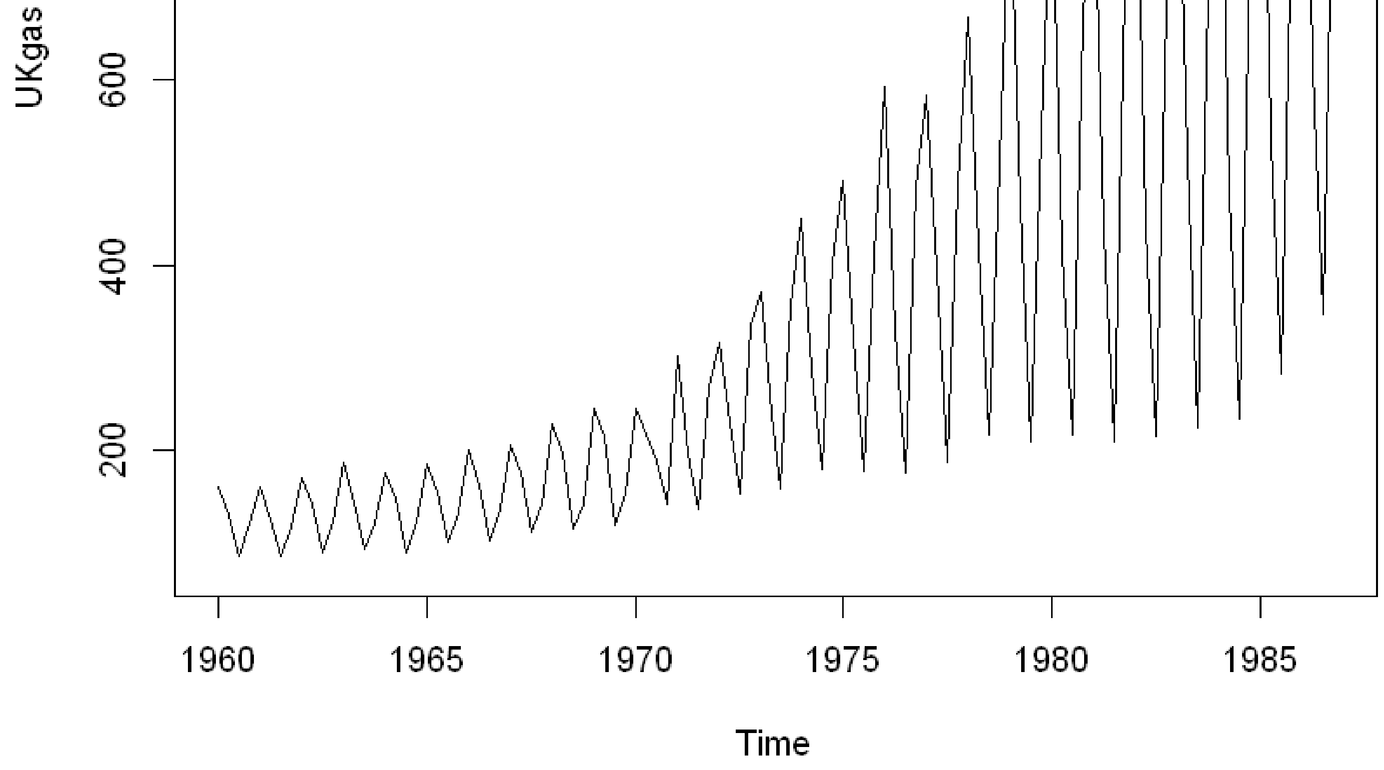
Answer to the Question No 1

Plotting The Dataset

Time Plot

In [2]:

```
plot(UKgas)
```

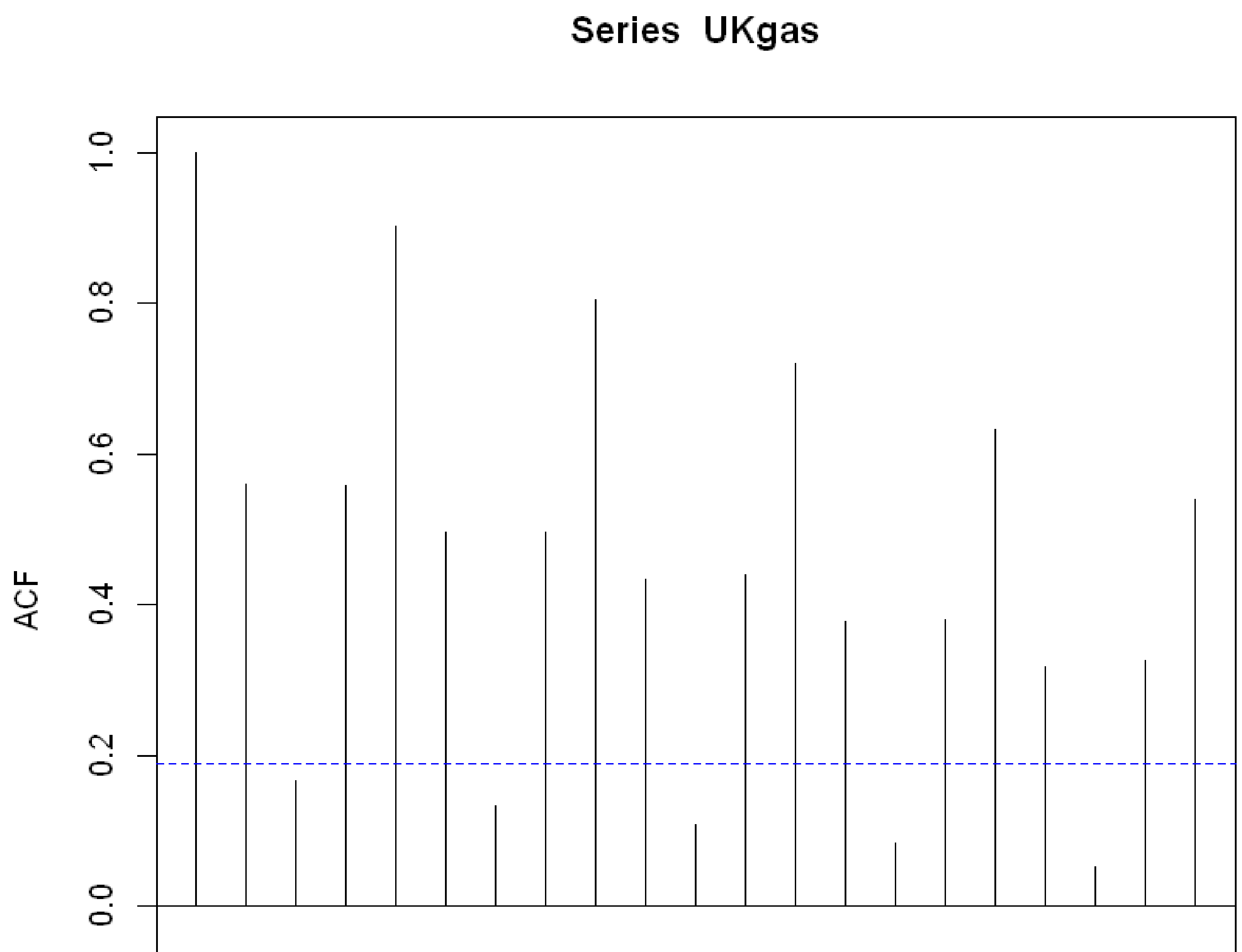


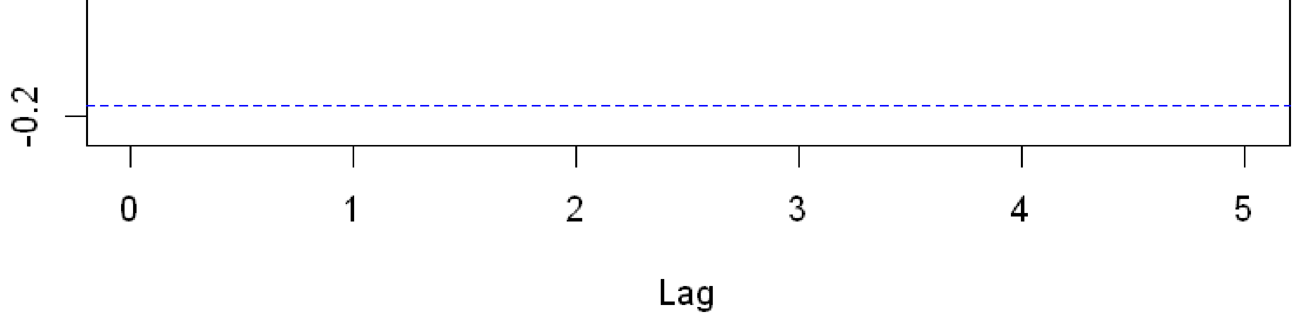
According to the Time plot above it can be said that, it has a Trend & Multiplicative Seasonality

ACF Plot

In [3]:

```
acf(UKgas)
```



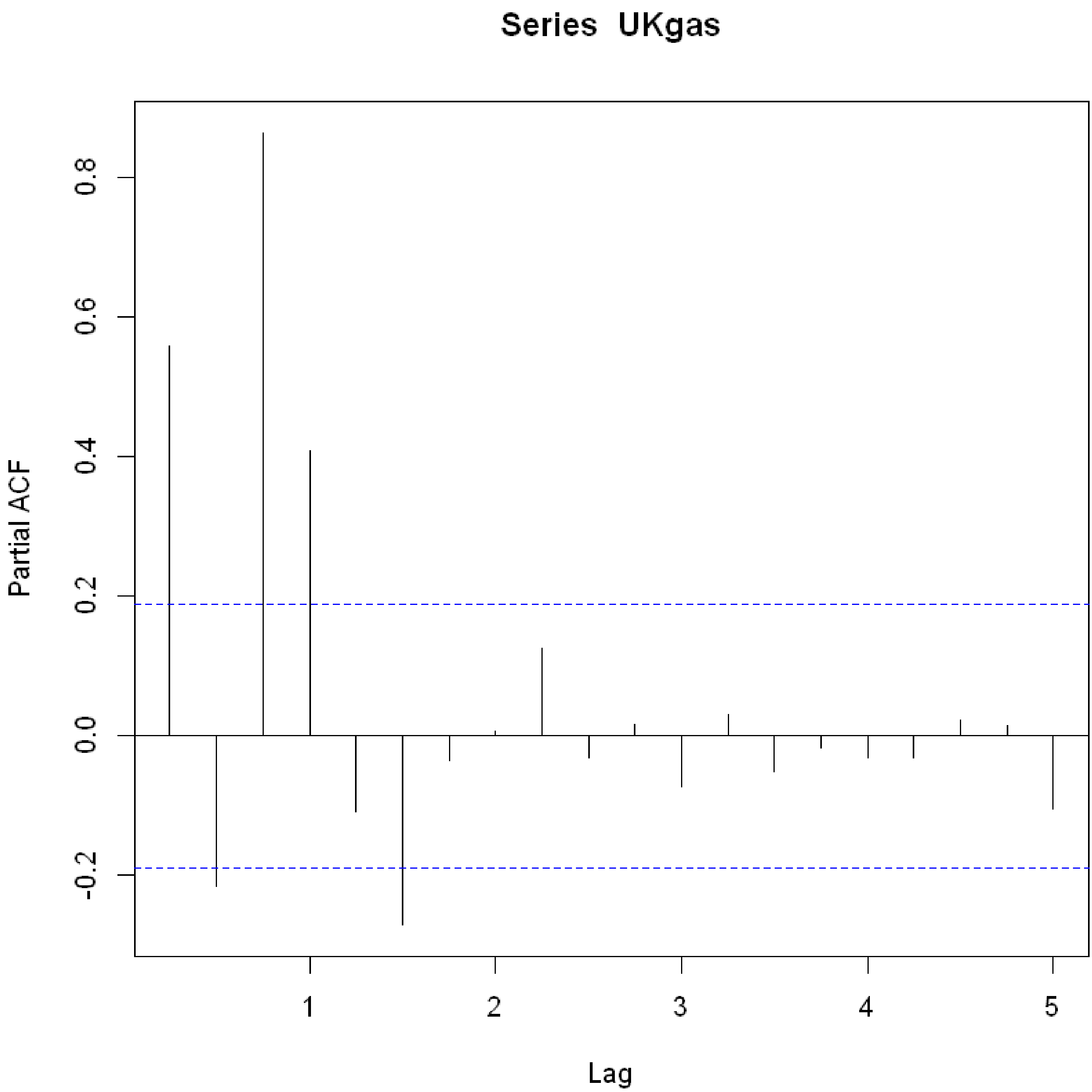


The above ACF Plot describes about trend & seasonality.

PACF Plot

In [4]:

```
pacf(UKgas)
```



Splitting Dataset - Train-70% Test-30%

Splitting Dataset - Train 70% Test 30%

In [5]:

```
print(UKgas)
```

```
        Qtr1   Qtr2   Qtr3   Qtr4
1960  160.1  129.7   84.8  120.1
1961  160.1  124.9   84.8  116.9
1962  169.7  140.9   89.7  123.3
1963  187.3  144.1   92.9  120.1
1964  176.1  147.3   89.7  123.3
1965  185.7  155.3   99.3  131.3
1966  200.1  161.7  102.5  136.1
1967  204.9  176.1  112.1  140.9
1968  227.3  195.3  115.3  142.5
1969  244.9  214.5  118.5  153.7
1970  244.9  216.1  188.9  142.5
1971  301.0  196.9  136.1  267.3
1972  317.0  230.5  152.1  336.2
1973  371.4  240.1  158.5  355.4
1974  449.9  286.6  179.3  403.4
1975  491.5  321.8  177.7  409.8
1976  593.9  329.8  176.1  483.5
1977  584.3  395.4  187.3  485.1
1978  669.2  421.0  216.1  509.1
1979  827.7  467.5  209.7  542.7
1980  840.5  414.6  217.7  670.8
1981  848.5  437.0  209.7  701.2
1982  925.3  443.4  214.5  683.6
1983  917.3  515.5  224.1  694.8
1984  989.4  477.1  233.7  730.0
1985 1087.0  534.7  281.8  787.6
1986 1163.9  613.1  347.4  782.8
```

In [6]:

```
library(forecast) # Necessary Library for Splitting the data
```

```
Warning message:
"package 'forecast' was built under R version 3.6.3"Registered S3 methods overwritten by 'g
gplot2':
  method         from
  [.quosures     rlang
  c.quosures     rlang
  print.quosures rlang
Registered S3 method overwritten by 'xts':
  method     from
  as.zoo.xts zoo
Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo
```

In [7]:

```
train <- head(UKgas, round(length(UKgas) * 0.7))
```

In [8]:

```
print(train)
```

```
      Qtr1  Qtr2  Qtr3  Qtr4
1960 160.1 129.7  84.8 120.1
1961 160.1 124.9  84.8 116.9
1962 169.7 140.9  89.7 123.3
1963 187.3 144.1  92.9 120.1
1964 176.1 147.3  89.7 123.3
1965 185.7 155.3  99.3 131.3
1966 200.1 161.7 102.5 136.1
1967 204.9 176.1 112.1 140.9
1968 227.3 195.3 115.3 142.5
1969 244.9 214.5 118.5 153.7
1970 244.9 216.1 188.9 142.5
1971 301.0 196.9 136.1 267.3
```

```
1972 317.0 230.5 152.1 336.2
1973 371.4 240.1 158.5 355.4
1974 449.9 286.6 179.3 403.4
1975 491.5 321.8 177.7 409.8
1976 593.9 329.8 176.1 483.5
1977 584.3 395.4 187.3 485.1
1978 669.2 421.0 216.1 509.1
```

In [9]:

```
h <- length(UKgas) - length(train)
```

In [10]:

```
test <- tail(UKgas, h)
```

In [11]:

```
print(test)
```

```
       Qtr1   Qtr2   Qtr3   Qtr4
1979  827.7  467.5  209.7  542.7
1980  840.5  414.6  217.7  670.8
1981  848.5  437.0  209.7  701.2
1982  925.3  443.4  214.5  683.6
1983  917.3  515.5  224.1  694.8
1984  989.4  477.1  233.7  730.0
1985 1087.0  534.7  281.8  787.6
1986 1163.9  613.1  347.4  782.8
```

As per Assignment Instruction the Train data will be used to answer Question No 2,3,4,5 & The Test data will be used to answer Question No 6

Answer to the Question No 2

Stationarity Checking

In [12]:

```
library(tseries) #Required Library for checking Stationarity
```

```
Warning message:
"package 'tseries' was built under R version 3.6.3"
```

Augmented Dickey-Fuller Test (Original Data)

In [13]:

```
#Augmented Dickey-Fuller Test
adf.test(train)
```

```
 Augmented Dickey-Fuller Test

data:  train
Dickey-Fuller = -0.35195, Lag order = 4, p-value = 0.9859
alternative hypothesis: stationary
```

For the Augmented Dickey-Fuller Test

Null Hypothesis: Not Stationary

Alternate Hypothesis: Stationary

As per the significant p value .9859 , we do not reject the null hypothesis & Hence, the Series is not

Stationary

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test (Original Data)

In [14]:

```
kpss.test(train)
```

```
Warning message in kpss.test(train):
"p-value smaller than printed p-value"
```

```
 KPSS Test for Level Stationarity

data:  train
KPSS Level = 1.7999, Truncation lag parameter = 3, p-value = 0.01
```

For the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

Null Hypothesis: Stationary

Alternate Hypothesis: Not Stationary

As per the small p value .01 , we reject the null hypothesis & Hence, the Series is not Stationary

Now we will try take differences to Make The Data Stationary

For Difference = 1

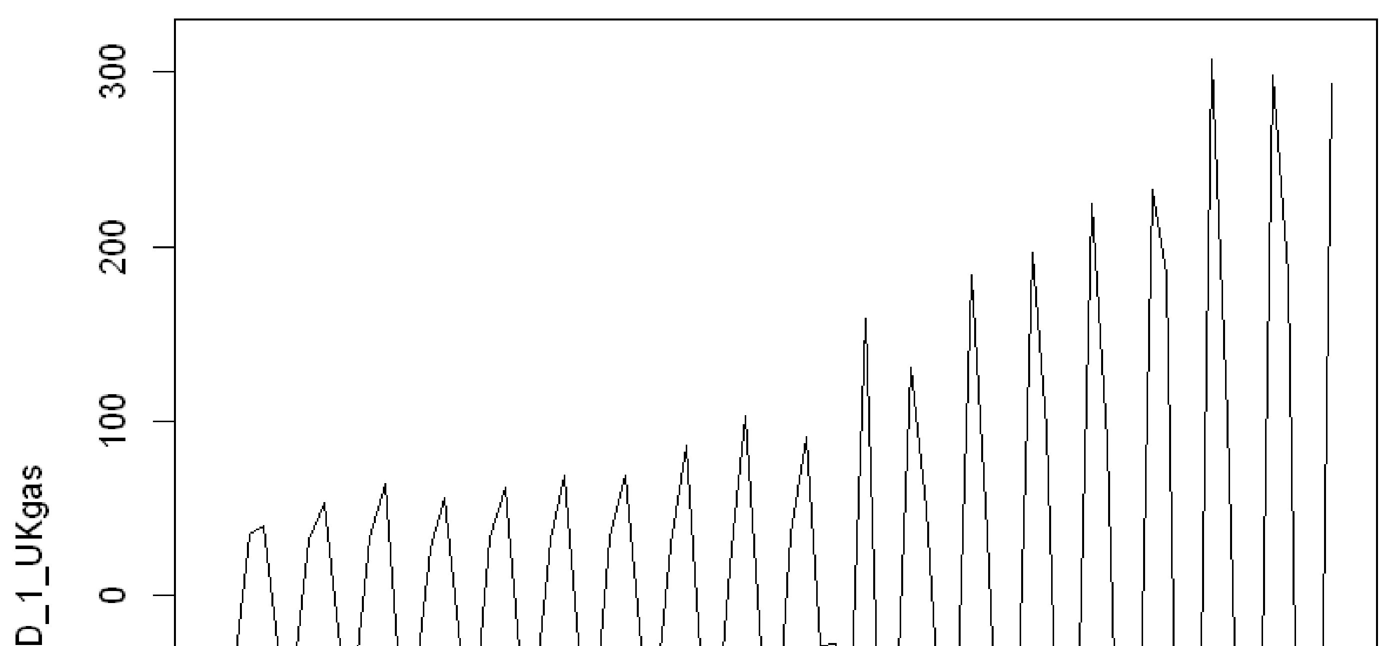
In [15]:

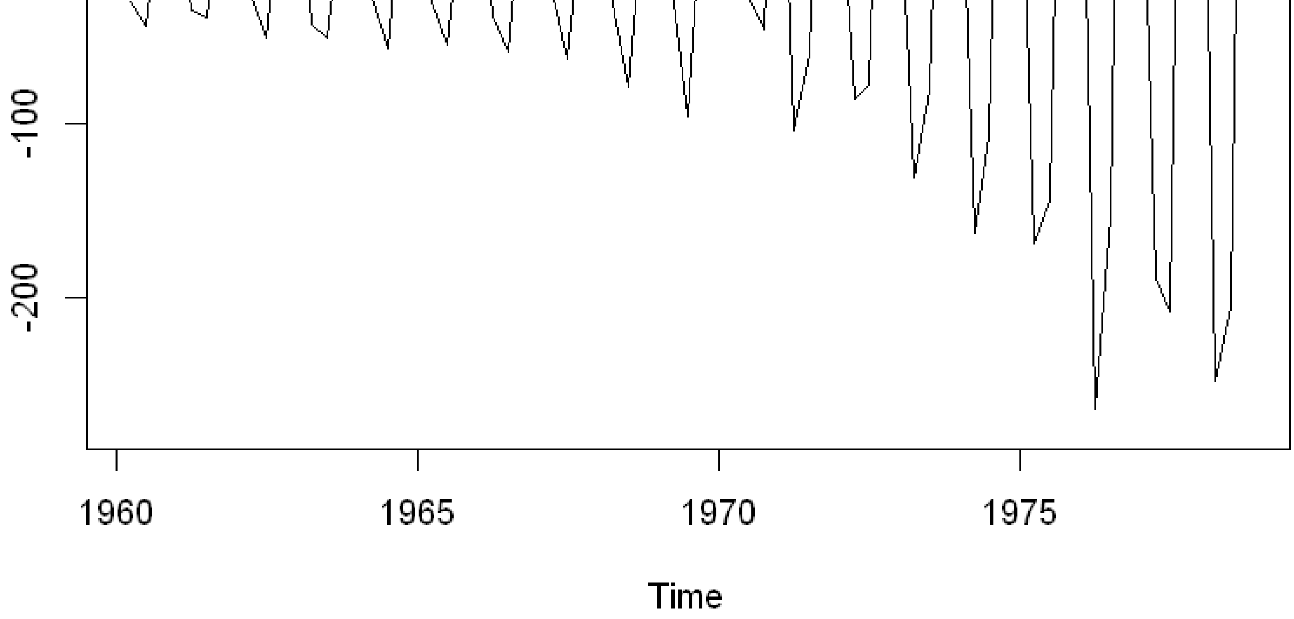
```
D_1_UKgas=diff(train, differences =1)
```

Time Plot

In [16]:

```
plot(D_1_UKgas)
```

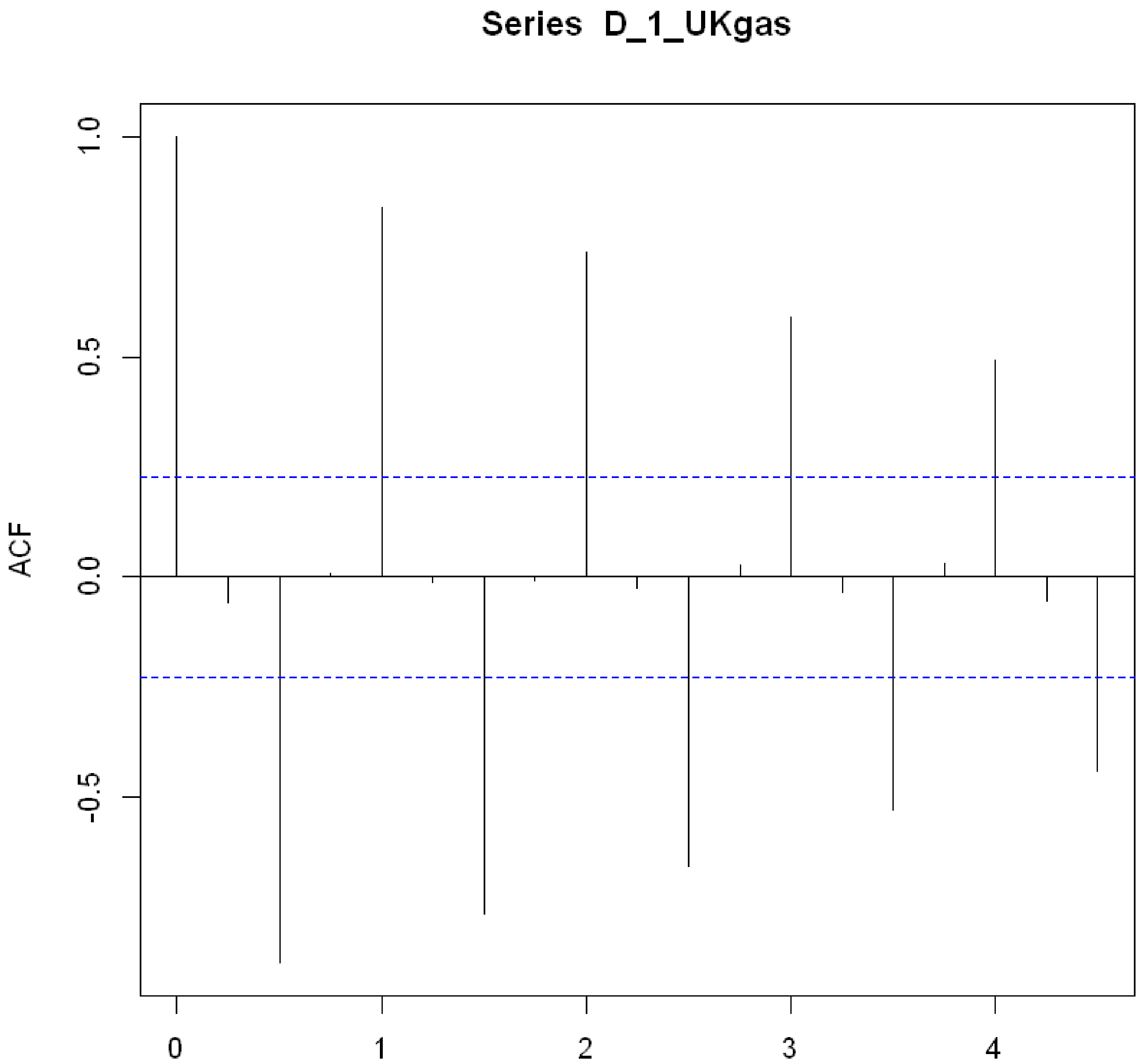




ACF Plot

In [17]:

```
acf(D_1_UKgas)
```

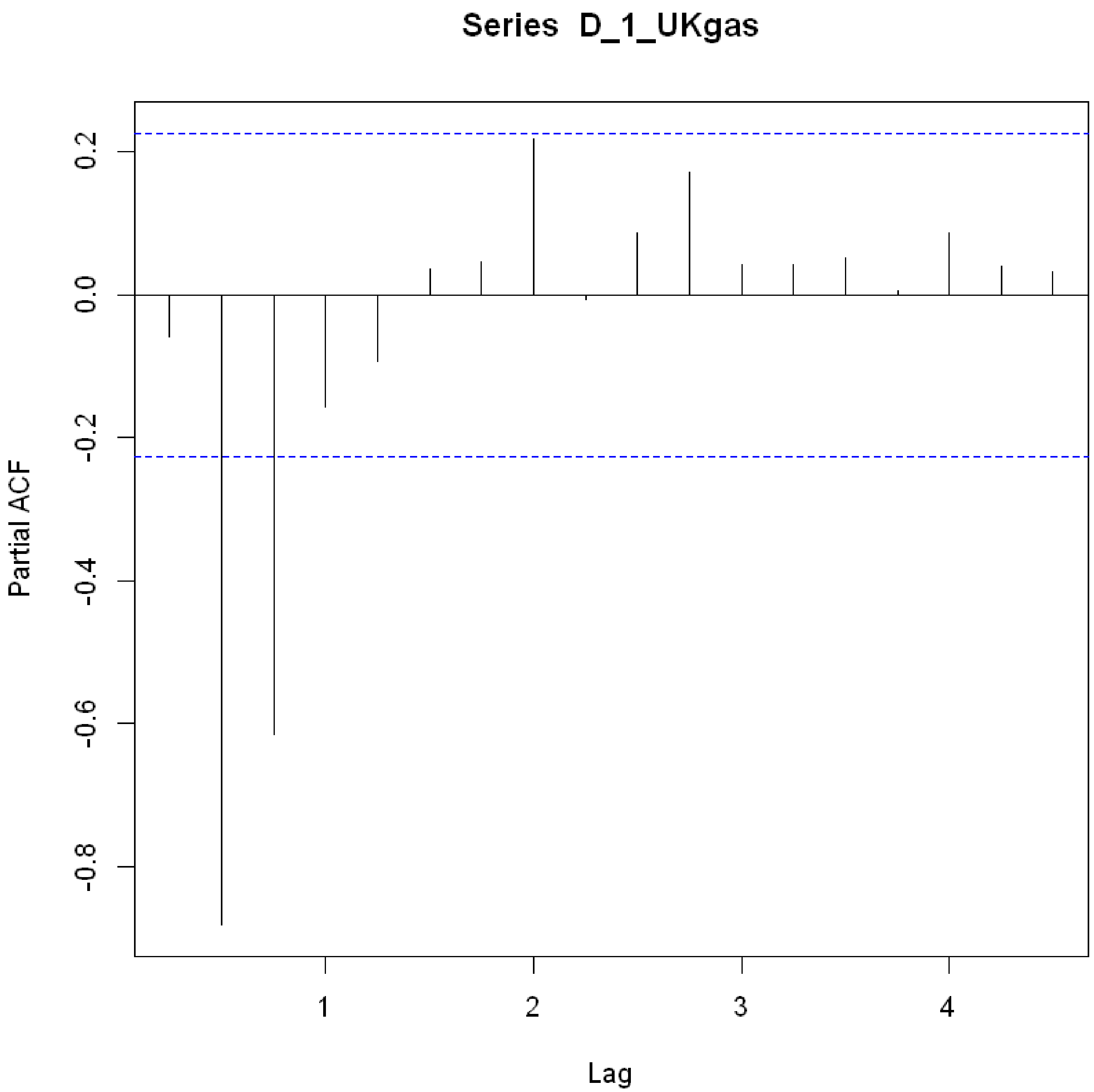


Lag

PACF Plot

In [18]:

```
pacf(D_1_UKgas)
```



Augmented Dickey-Fuller Test (Difference = 1)

In [19]:

```
adf.test(D_1_UKgas)
```

```
Warning message in adf.test(D_1_UKgas):
"p-value smaller than printed p-value"
```

```
	Augmented Dickey-Fuller Test
```

```
data:  D_1_UKgas
Dickey-Fuller = -7.0544, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

For the Augmented Dickey-Fuller Test with Difference = 1

Null Hypothesis: Not Stationary

Alternate Hypothesis: Stationary

As per the small p value .01 , we reject the null hypothesis & Hence, the Series is Stationary

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test (Difference = 1)

In [20]:

```
kpss.test(D_1_UKgas)
```

```
 KPSS Test for Level Stationarity

data:  D_1_UKgas
KPSS Level = 0.41295, Truncation lag parameter = 3, p-value = 0.07157
```

For the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test with Difference = 1

Null Hypothesis: Stationary

Alternate Hypothesis: Not Stationary

As per the significant p value .07 , we do not reject the null hypothesis & Hence, the Series is Stationary

Model Fitting (M1)

Now we will fit Model M1 using ACF & PACF plots of the stationary series. When using CSS (conditional sum of squares), it is possible for the autoregressive coefficients to be non-stationary (i.e., they fall outside the region for stationary processes).

In [21]:

```
M1_css = arima(D_1_UKgas, order=c(2,1,2))
```

```
Error in arima(D_1_UKgas, order = c(2, 1, 2)): non-stationary AR part from CSS
Traceback:

1. arima(D_1_UKgas, order = c(2, 1, 2))
2. stop("non-stationary AR part from CSS")
```

We also force R to use MLE (maximum likelihood estimation) instead by using the argument method="ML". This is slower but gives better estimates and always returns a stationary model.

In [22]:

```
M1_mle = arima(D_1_UKgas, order=c(2,1,2), method="ML")
```

In [23]:

```
summary(M1_mle)
```

```
Call:
arima(x = D_1_UKgas, order = c(2, 1, 2), method = "ML")

Coefficients:
```

```
         ar1      ar2      ma1     ma2
      -0.0242  -0.9708  -1.6283  0.6628
s.e.   0.0347   0.0231   0.1091  0.1143

sigma^2 estimated as 923.1:  log likelihood = -363.67,  aic = 737.34

Training set error measures:
                   ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
Training set 5.803502 30.17859 19.78969 1.759453 22.14559 0.1598835 -0.5765526
```

In [24]:

```
summary(M1_css)
```

```
Error in summary(M1_css): object 'M1_css' not found
Traceback:

1. summary(M1_css)
```

We find an error in implementing the CSS method and hence fit the model M_1 with the MLE method.

Residual Checking for model M1

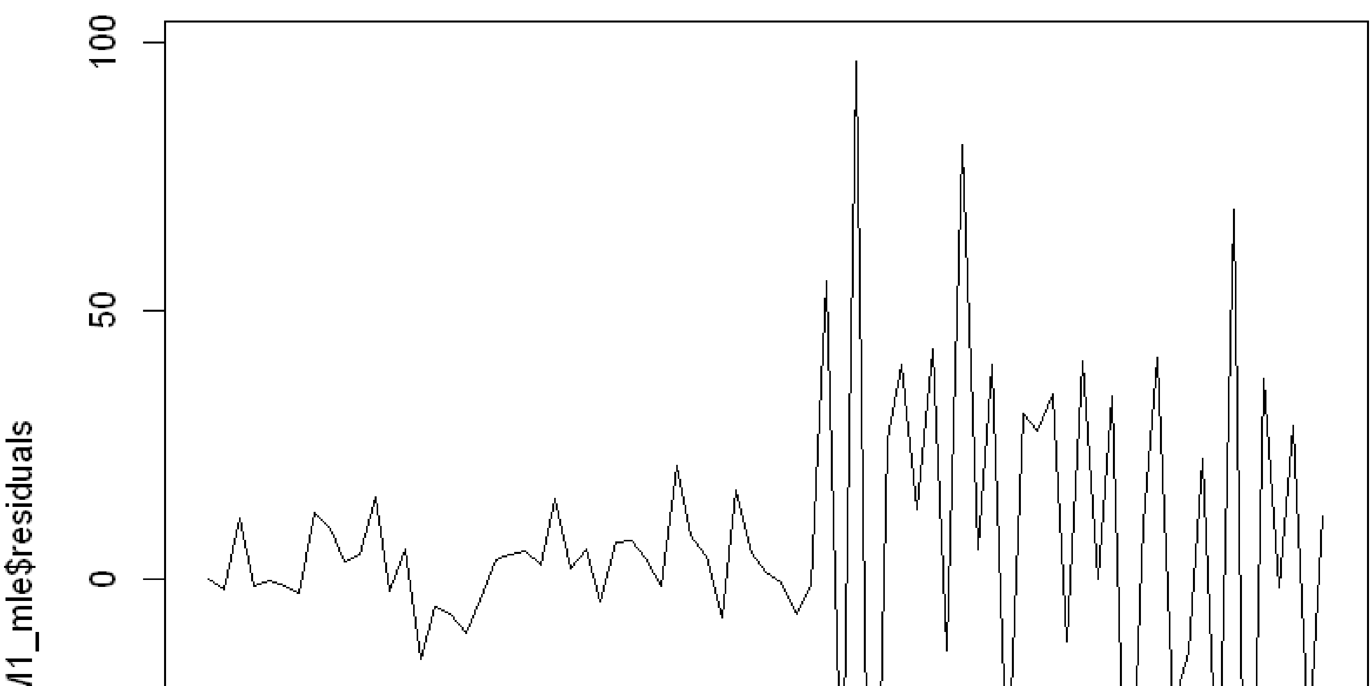
In [25]:

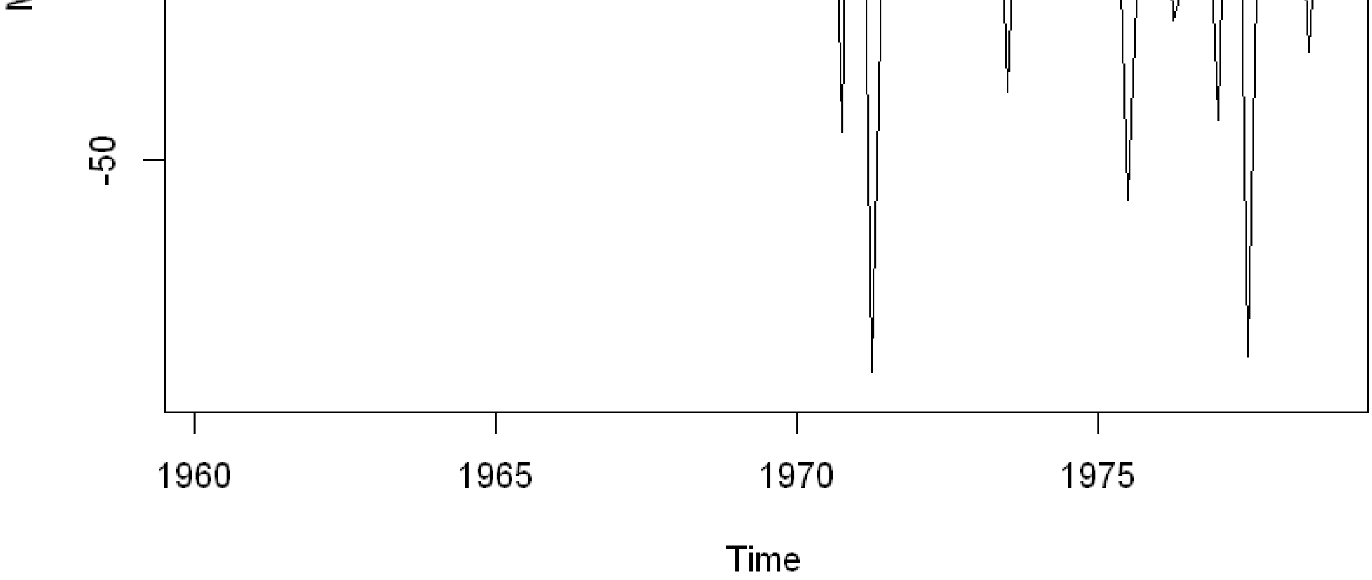
```
library(itsmr) # Required Library for Residual Checking
```

```
Attaching package: 'itsmr'

The following object is masked from 'package:tseries':

    arma

The following object is masked from 'package:forecast':

    forecast
```

In [26]:

```
plot(M1_mle$residuals)
```

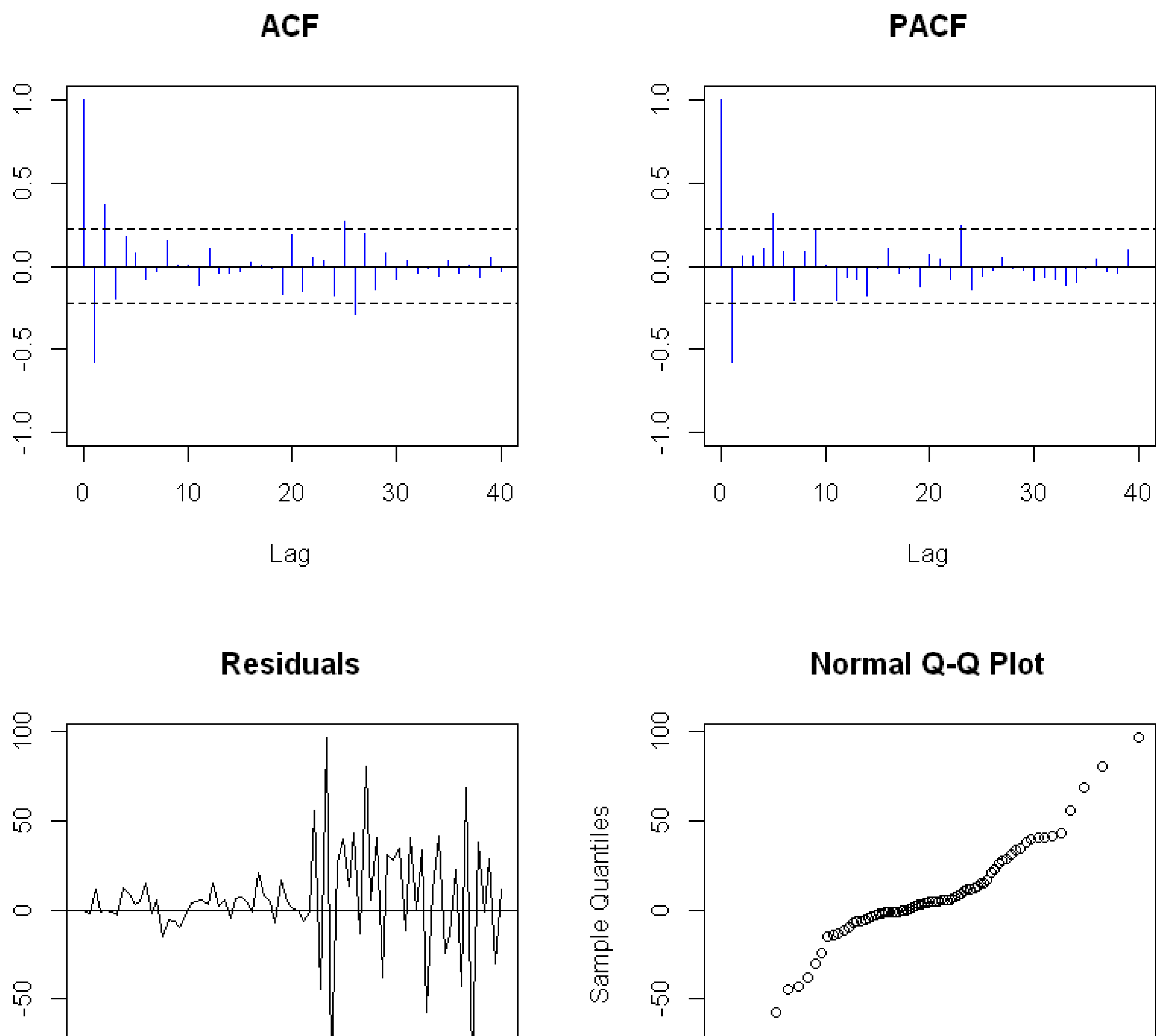


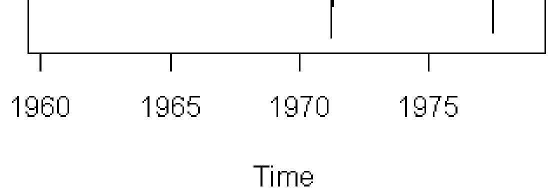


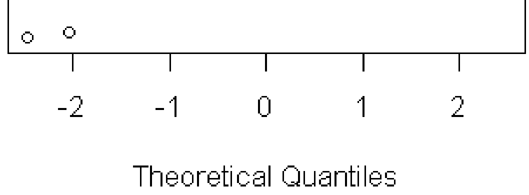
In [27]:

```
test(M1_mle$residuals)
```

```
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                Q ~ chisq(20)     54.88         0 *
McLeod-Li Q                Q ~ chisq(20)     31.11    0.0537
Turning points T    (T-48.7)/3.6 ~ N(0,1)       55    0.0791
Diff signs S          (S-37)/2.5 ~ N(0,1)       37         1
Rank P          (P-1387.5)/109.3 ~ N(0,1)     1531    0.1892
```







According to the Ljung-Box Q test & their p value we reject the null hypothesis & hence reject the null hypothesis, Hence the Residual is not iid moise Test Distribution Statistic p-value Ljung-Box Q Q ~ chisq(20) 54.88 0 * According to othert results we found significant p value & hence do not reject the Null hypothesis (Residuals are iid noise.) Null hypothesis: Residuals are iid noise. Test Distribution Statistic p-value McLeod-Li Q Q ~ chisq(20) 31.11 0.0537 Turning points T (T-48.7)/3.6 ~ N(0,1) 55 0.0791 Diff signs S (S-37)/2.5 ~ N(0,1) 37 1 Rank P (P-1387.5)/109.3 ~ N(0,1) 1531 0.1892

As we notice unstabilized variance, a transformation might be useful. So we will apply different Transformations

Answer to the Question No 3

Original Data

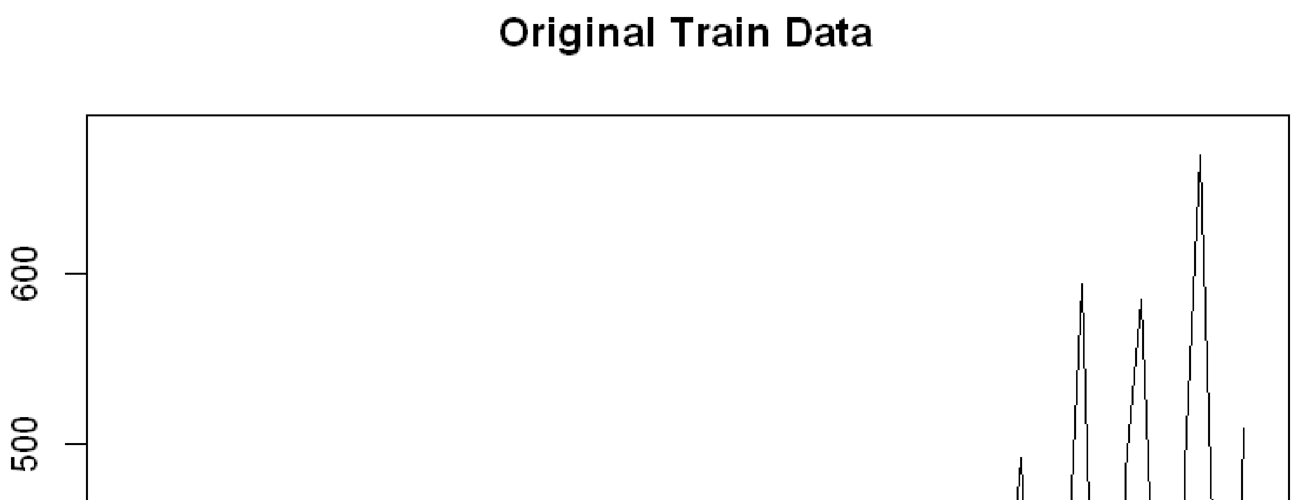
In [28]:

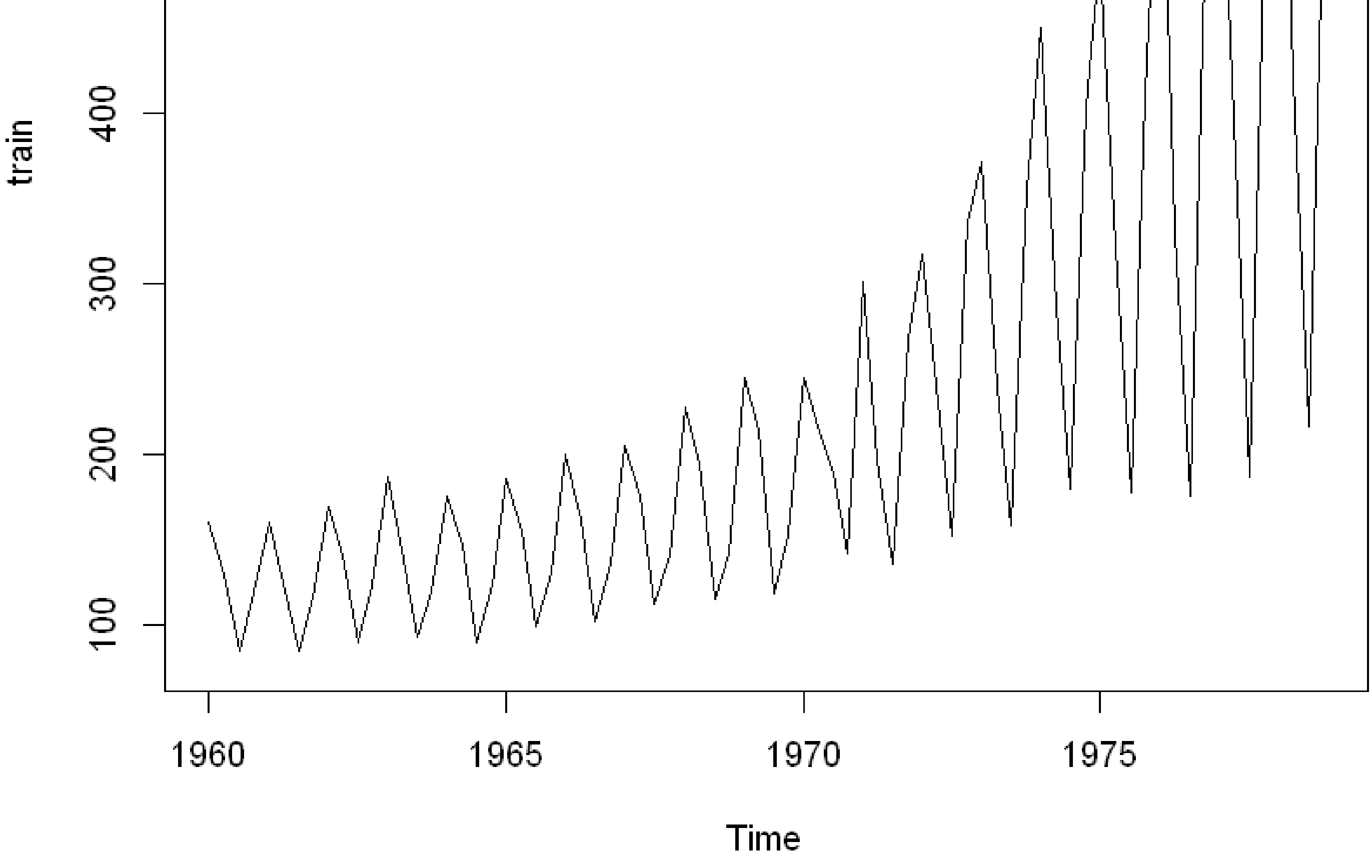
```
print(train)
```

```
       Qtr1  Qtr2  Qtr3  Qtr4
1960 160.1 129.7  84.8 120.1
1961 160.1 124.9  84.8 116.9
1962 169.7 140.9  89.7 123.3
1963 187.3 144.1  92.9 120.1
1964 176.1 147.3  89.7 123.3
1965 185.7 155.3  99.3 131.3
1966 200.1 161.7 102.5 136.1
1967 204.9 176.1 112.1 140.9
1968 227.3 195.3 115.3 142.5
1969 244.9 214.5 118.5 153.7
1970 244.9 216.1 188.9 142.5
1971 301.0 196.9 136.1 267.3
1972 317.0 230.5 152.1 336.2
1973 371.4 240.1 158.5 355.4
1974 449.9 286.6 179.3 403.4
1975 491.5 321.8 177.7 409.8
1976 593.9 329.8 176.1 483.5
1977 584.3 395.4 187.3 485.1
1978 669.2 421.0 216.1 509.1
```

In [29]:

```
plot(train, main="Original Train Data")
```





Square Root Transformation

In [30]:

```
sq_transform <- sqrt(train)
```

In [31]:

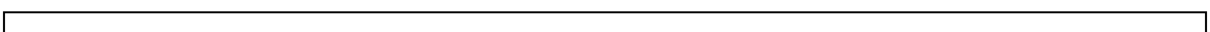
```
print(sq_transform)
```

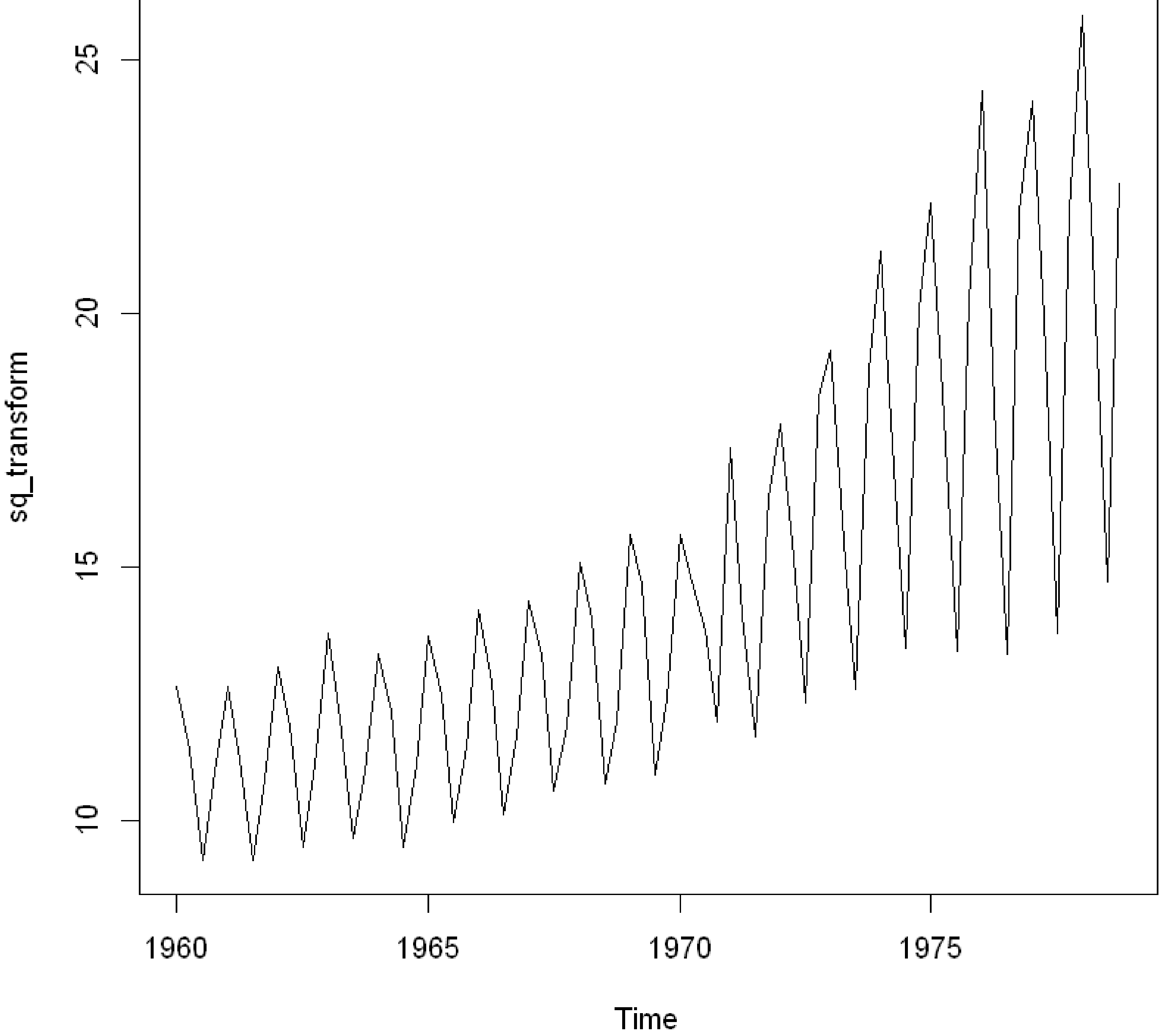
```
          Qtr1      Qtr2      Qtr3      Qtr4
1960 12.653063 11.388591  9.208692 10.959015
1961 12.653063 11.175867  9.208692 10.812030
1962 13.026895 11.870131  9.471008 11.104053
1963 13.685759 12.004166  9.638465 10.959015
1964 13.270268 12.136721  9.471008 11.104053
1965 13.627179 12.461942  9.964939 11.458621
1966 14.145671 12.716131 10.124228 11.666190
1967 14.314328 13.270268 10.587729 11.870131
1968 15.076472 13.974978 10.737784 11.937336
1969 15.649281 14.645819 10.885771 12.397580
1970 15.649281 14.700340 13.744090 11.937336
1971 17.349352 14.032106 11.666190 16.349312
1972 17.804494 15.182226 12.332883 18.335757
1973 19.271741 15.495161 12.589678 18.852056
1974 21.210846 16.929265 13.390295 20.084820
1975 22.169799 17.938785 13.330416 20.243517
1976 24.370064 18.160396 13.270268 21.988633
1977 24.172298 19.884667 13.685759 22.024986
1978 25.868900 20.518285 14.700340 22.563244
```

In [32]:

```
plot(sq_transform, main="Square Root Transformation - UKgas Train")
```

Square Root Transformation - UKgas Train





Cube Root Transformation

In [33]:

```
cube_transform <- ((train)^(1/3))
```

In [34]:

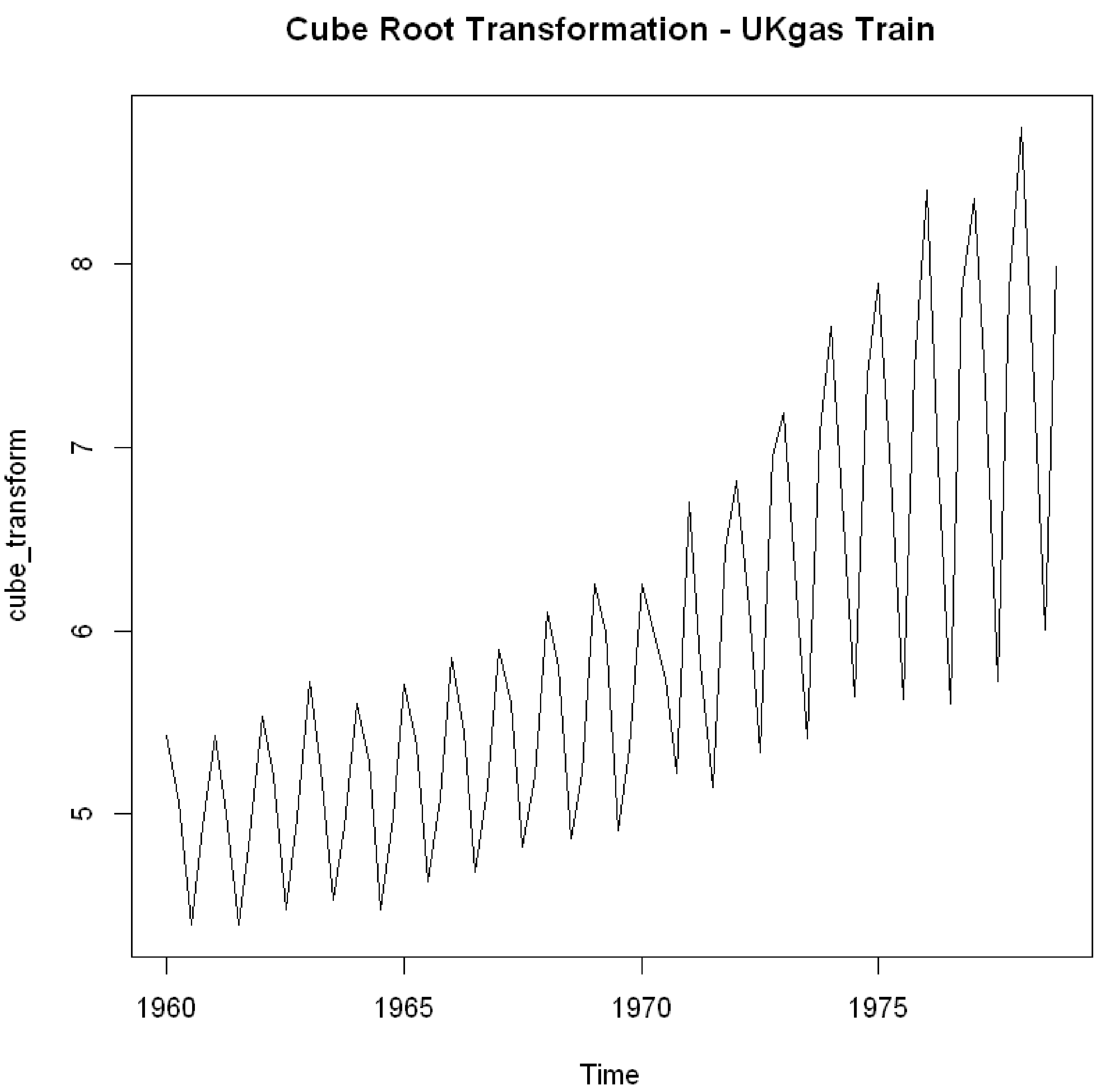
```
print(cube_transform)
```

```
         Qtr1     Qtr2     Qtr3     Qtr4
1960 5.429966 5.061897 4.393378 4.933794
1961 5.429966 4.998666 4.393378 4.889579
1962 5.536398 5.203597 4.476420 4.977230
1963 5.721535 5.242696 4.529030 4.933794
1964 5.605140 5.281220 4.476420 4.977230
1965 5.705197 5.375149 4.630733 5.082627
1966 5.849010 5.447995 4.679951 5.143823
1967 5.895410 5.605140 4.821719 5.203597
1968 6.102856 5.801862 4.867169 5.223220
1969 6.256473 5.986079 4.911786 5.356626
1970 6.256473 6.000926 5.737781 5.223220
1971 6.701759 5.817663 5.143823 6.441688
1972 6.818462 6.131362 5.337973 6.953432
1973 7.188098 6.215328 5.411817 7.083357
1974 7.662527 6.593136 5.638887 7.388880
1975 7.891772 6.852705 5.622064 7.427751
1976 8.405646 6.909027 5.605140 7.848720
1977 8.360109 7.339710 5.721535 7.857368
```

```
1978 8.746856 7.494811 6.000926 7.984867
```

In [35]:

```
plot(cube_transform, main="Cube Root Transformation - UKgas Train")
```



Logarithm Transformation

In [36]:

```
log_transform <- log10(train)
```

In [37]:

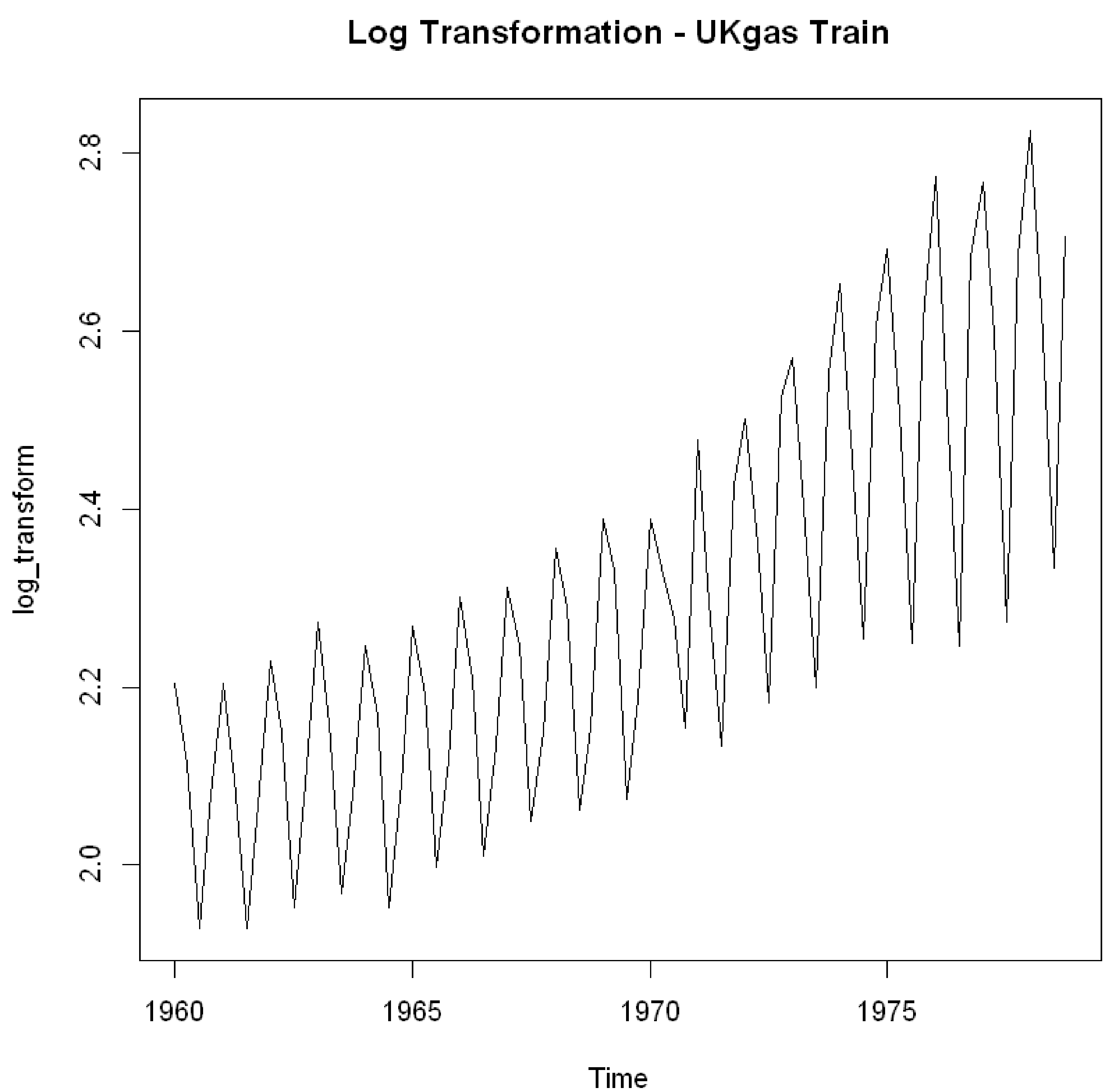
```
print(log_transform)
```

```
         Qtr1     Qtr2     Qtr3     Qtr4
1960 2.204391 2.112940 1.928396 2.079543
1961 2.204391 2.096562 1.928396 2.067815
1962 2.229682 2.148911 1.952792 2.090963
1963 2.272538 2.158664 1.968016 2.079543
1964 2.245759 2.168203 1.952792 2.090963
1965 2.268812 2.191171 1.996949 2.118265
1966 2.301247 2.208710 2.010724 2.133858
```

```
1966 2.301247 2.208710 2.010724 2.153858
1967 2.311542 2.245759 2.049606 2.148911
1968 2.356599 2.290702 2.061829 2.153815
1969 2.388989 2.331427 2.073718 2.186674
1970 2.388989 2.334655 2.276232 2.153815
1971 2.478566 2.294246 2.133858 2.426999
1972 2.501059 2.362671 2.182129 2.526598
1973 2.569842 2.380392 2.200029 2.550717
1974 2.653116 2.457276 2.253580 2.605736
1975 2.691524 2.507586 2.249687 2.612572
1976 2.773713 2.518251 2.245759 2.684396
1977 2.766636 2.597037 2.272538 2.685831
1978 2.825556 2.624282 2.334655 2.706803
```

In [38]:

```
plot(log_transform, main="Log Transformation - UKgas Train")
```



Augmented Dickey-Fuller Test (For Square Root Transformed Data)

In [39]:

```
adf.test(sq_transform)
```

```
 Augmented Dickey-Fuller Test

data:  sq_transform
Dickey-Fuller = -1.3172, Lag order = 4, p-value = 0.8542
alternative hypothesis: stationary
```

For the Augmented Dickey-Fuller Test

Null Hypothesis: Not Stationary

Alternate Hypothesis: Stationary

As per the significant p value .85 , we do not reject the null hypothesis & Hence, the Series is not Stationary

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test (For Square Root Transformed Data)

In [40]:

```
kpss.test(sq_transform)
```

```
Warning message in kpss.test(sq_transform):
"p-value smaller than printed p-value"
```

```
 KPSS Test for Level Stationarity

data:  sq_transform
KPSS Level = 1.8664, Truncation lag parameter = 3, p-value = 0.01
```

For the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

Null Hypothesis: Stationary

Alternate Hypothesis: Not Stationary

As per the small p value .01 , we reject the null hypothesis & Hence, the Series is not Stationary

Now we will try take differences to Make The Square Root Transformed Data Stationary

For Difference = 1

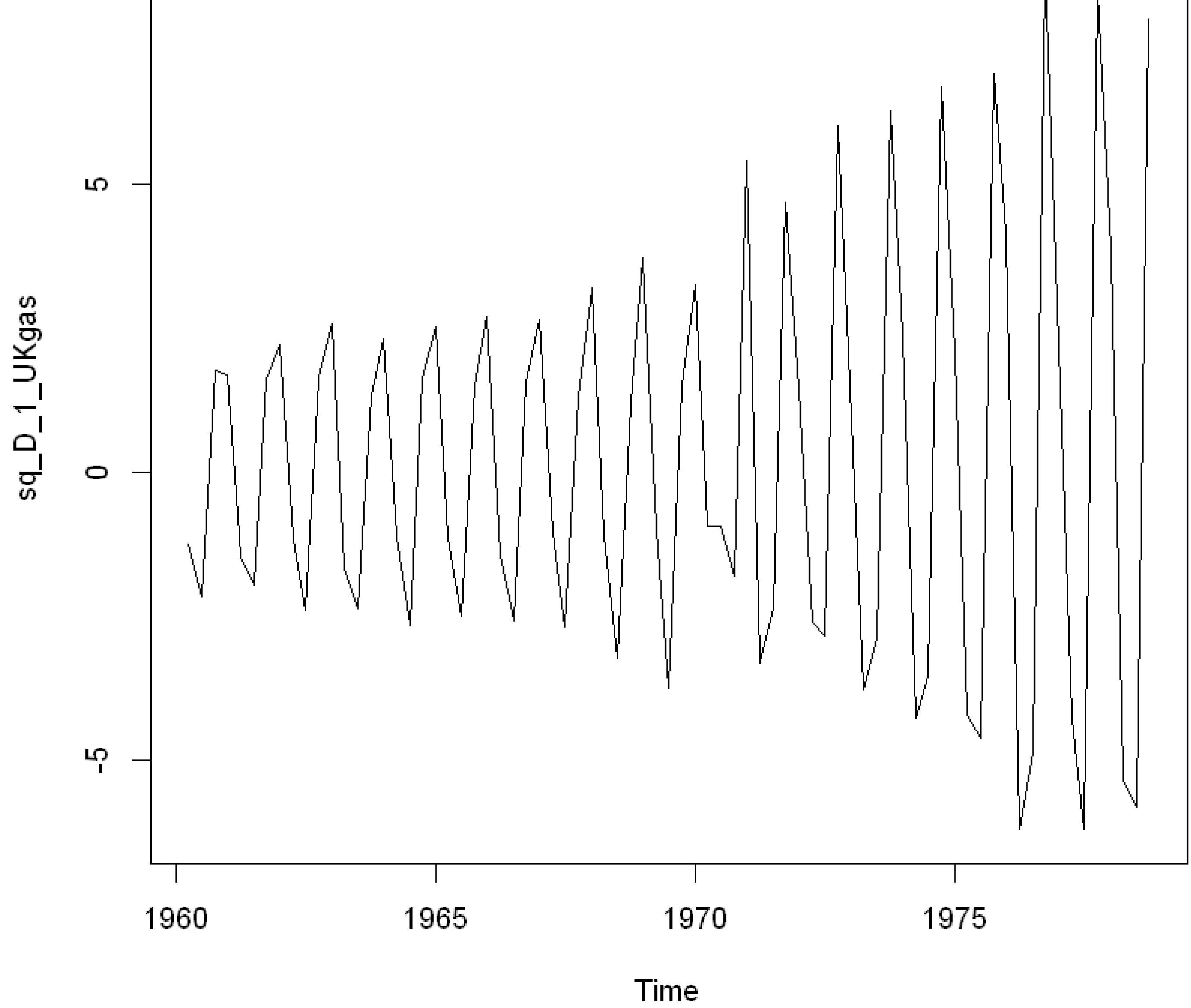
In [41]:

```
sq_D_1_UKgas=diff(sq_transform, differences =1)
```

Time Plot

In [42]:

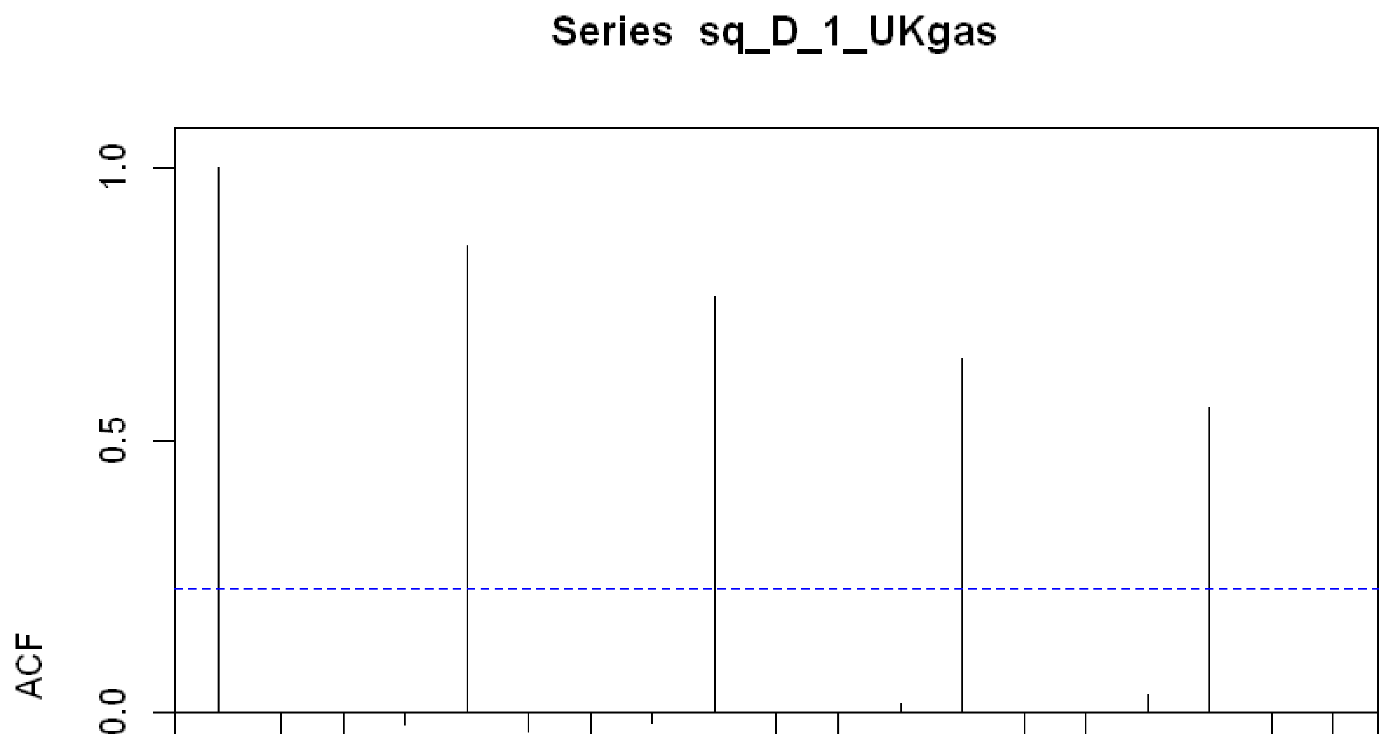
```
plot(sq_D_1_UKgas)
```

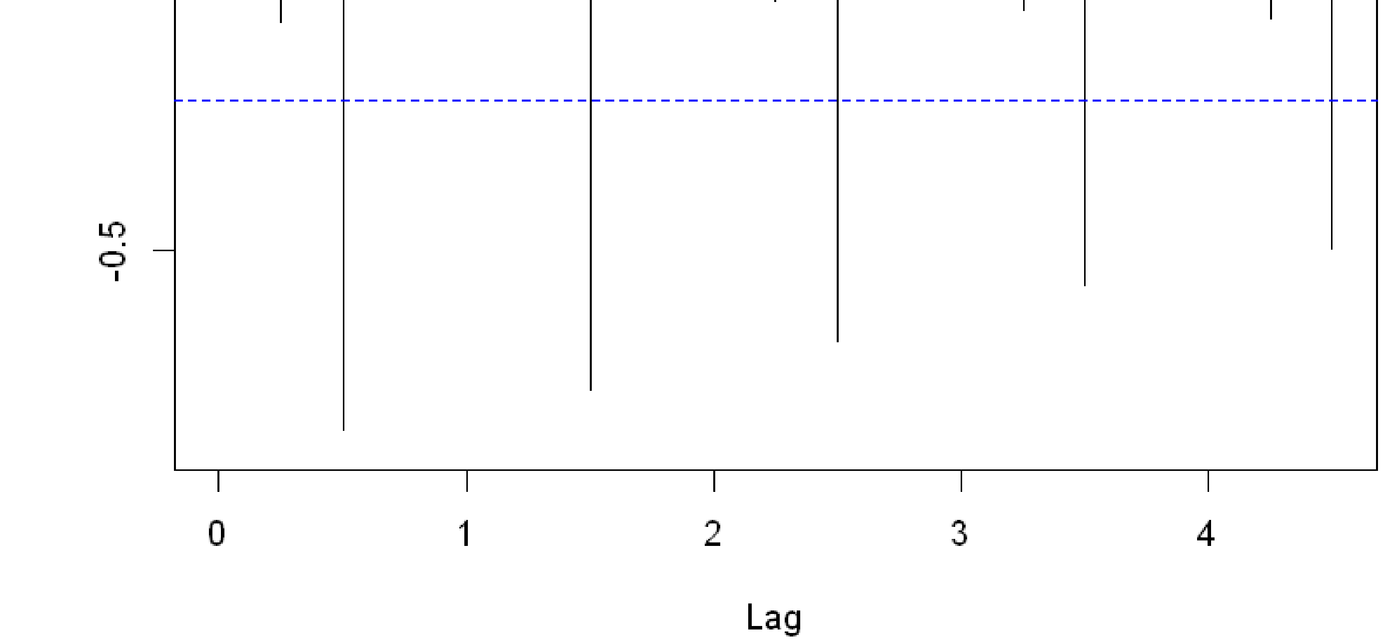


ACF Plot

In [43]:

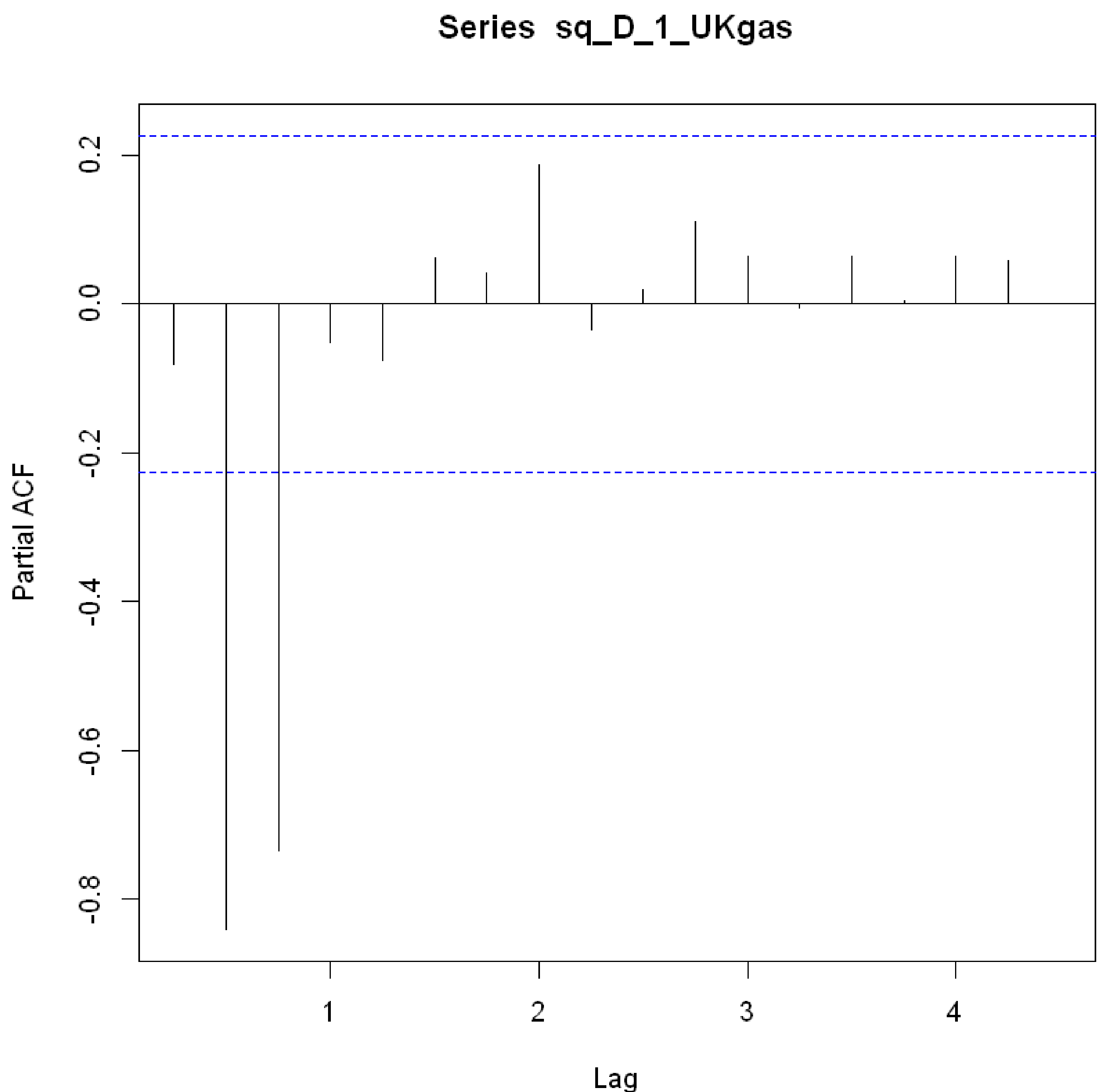
```
acf(sq_D_1_UKgas)
```





In [44]:

```
pacf(sq_D_1_UKgas)
```



Augmented Dickey-Fuller Test for Square Root Transformed Data (Difference = 1)

In [45]:

```
adf.test(sq_D_1_UKgas)
```

```
Warning message in adf.test(sq_D_1_UKgas):
"p-value smaller than printed p-value"
```

```
	Augmented Dickey-Fuller Test

data:  sq_D_1_UKgas
Dickey-Fuller = -7.0177, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

For the Augmented Dickey-Fuller Test for Square Root Transformed Data with Difference = 1

Null Hypothesis: Not Stationary

Alternate Hypothesis: Stationary

As per the small p value .01 , we reject the null hypothesis & Hence, the Series is Stationary

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test for the Square Root Transformed Data (Difference = 1)

In [46]:

```
kpss.test(sq_D_1_UKgas)
```

```
	KPSS Test for Level Stationarity

data:  sq_D_1_UKgas
KPSS Level = 0.4328, Truncation lag parameter = 3, p-value = 0.06302
```

For the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test for the Square Root Transformed Data with Difference = 1

Null Hypothesis: Stationary

Alternate Hypothesis: Not Stationary

As per the significant p value .06 , we do not reject the null hypothesis & Hence, the Series is Stationary

Augmented Dickey-Fuller Test (For Cube Root Transformed Data)

In [47]:

```
adf.test(cube_transform)
```

```
	Augmented Dickey-Fuller Test

data:  cube_transform
Dickey-Fuller = -1.5648, Lag order = 4, p-value = 0.7531
alternative hypothesis: stationary
```

For the Augmented Dickey-Fuller Test

Null Hypothesis: Not Stationary

Alternate Hypothesis: Stationary

As per the significant p value .75 , we do not reject the null hypothesis & Hence, the Series is not Stationary

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test (For Cube Root Transformed Data)

In [48]:

```
kpss.test(cube_transform)
```

```
Warning message in kpss.test(cube_transform):
"p-value smaller than printed p-value"
```

```
 KPSS Test for Level Stationarity

data:  cube_transform
KPSS Level = 1.8846, Truncation lag parameter = 3, p-value = 0.01
```

For the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

Null Hypothesis: Stationary

Alternate Hypothesis: Not Stationary

As per the small p value .01 , we reject the null hypothesis & Hence, the Series is not Stationary

Now we will try take differences to Make The Cube Root Transformed Data Stationary

For Difference = 1

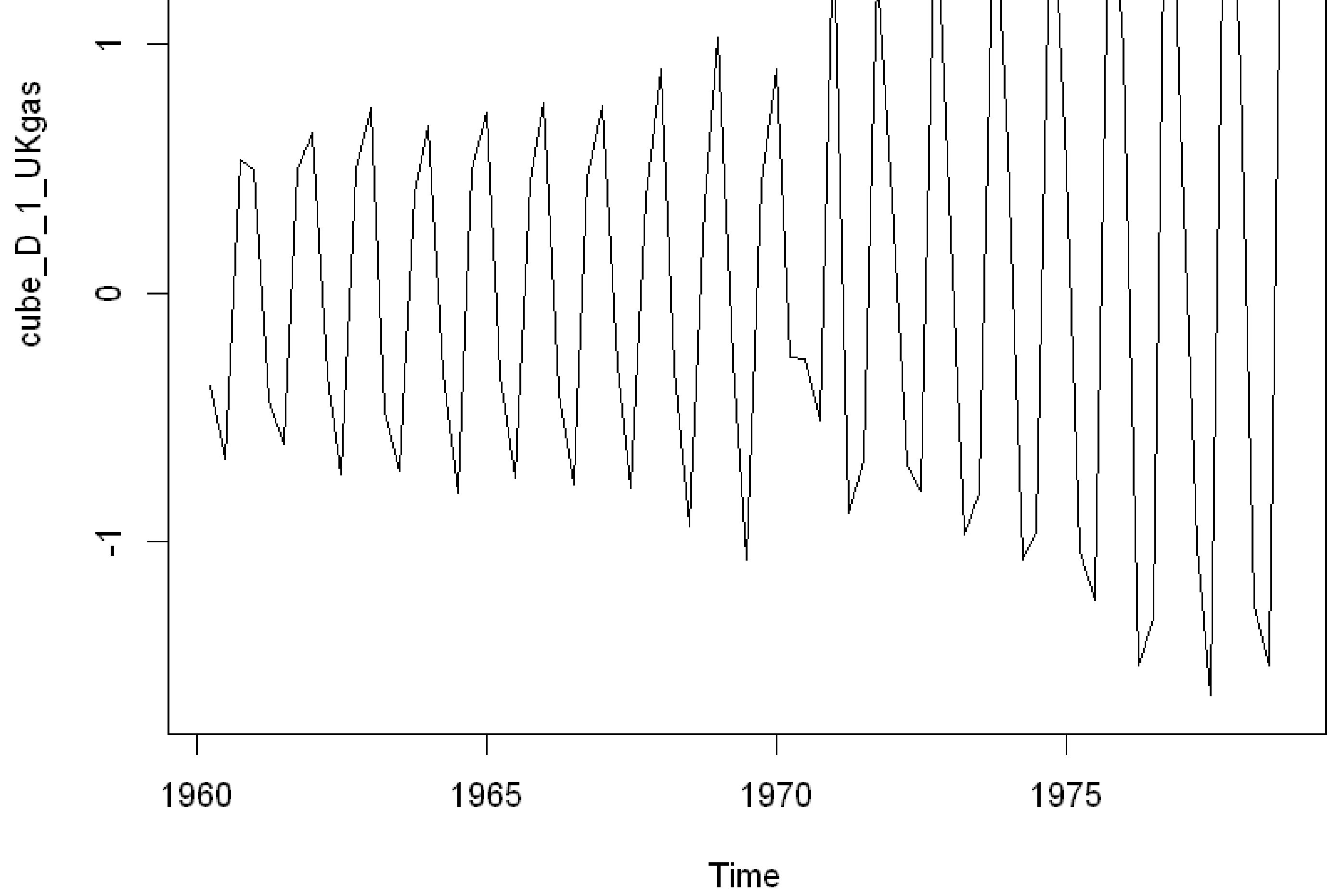
In [49]:

```
cube_D_1_UKgas=diff(cube_transform, differences =1)
```

Time Plot

In [50]:

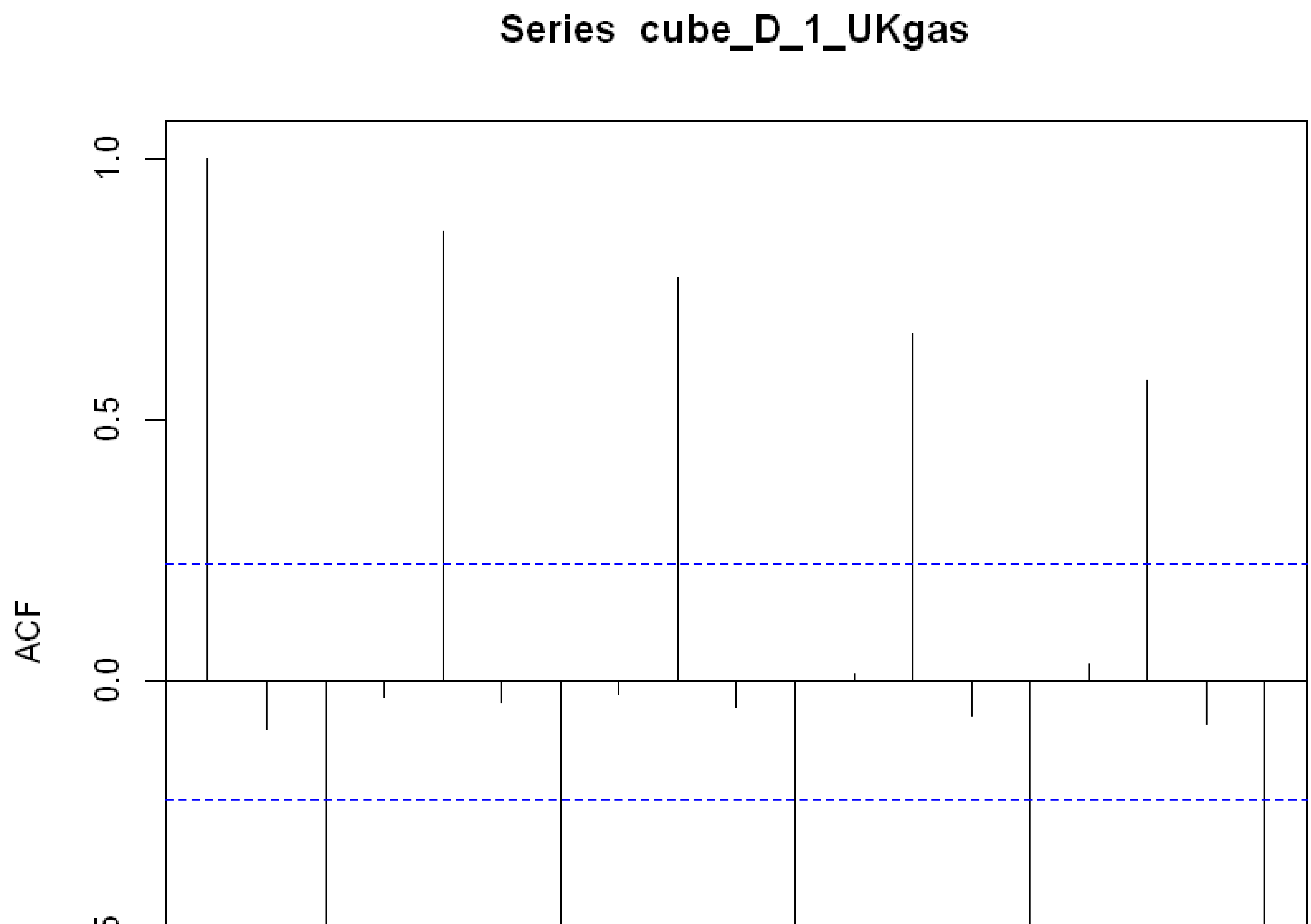
```
plot(cube_D_1_UKgas)
```

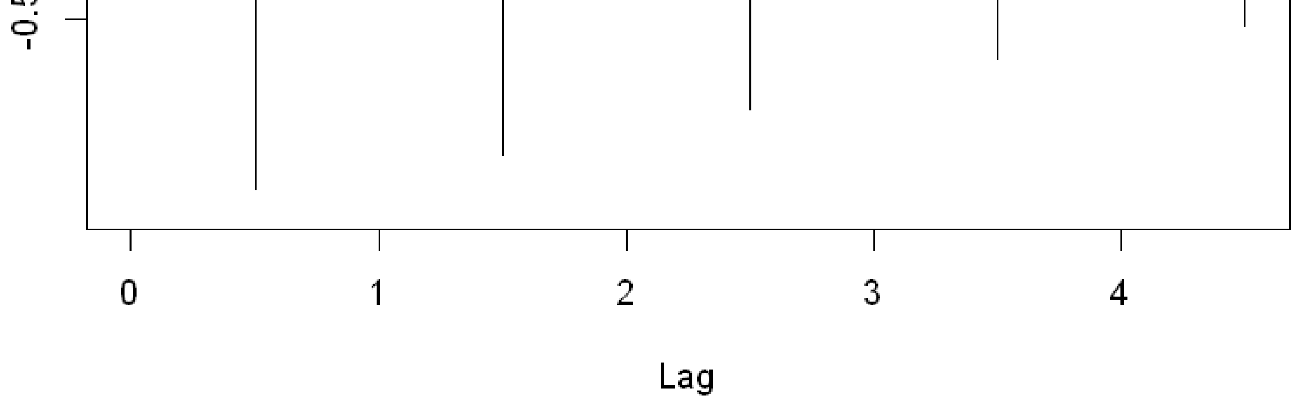


ACF Plot

In [51]:

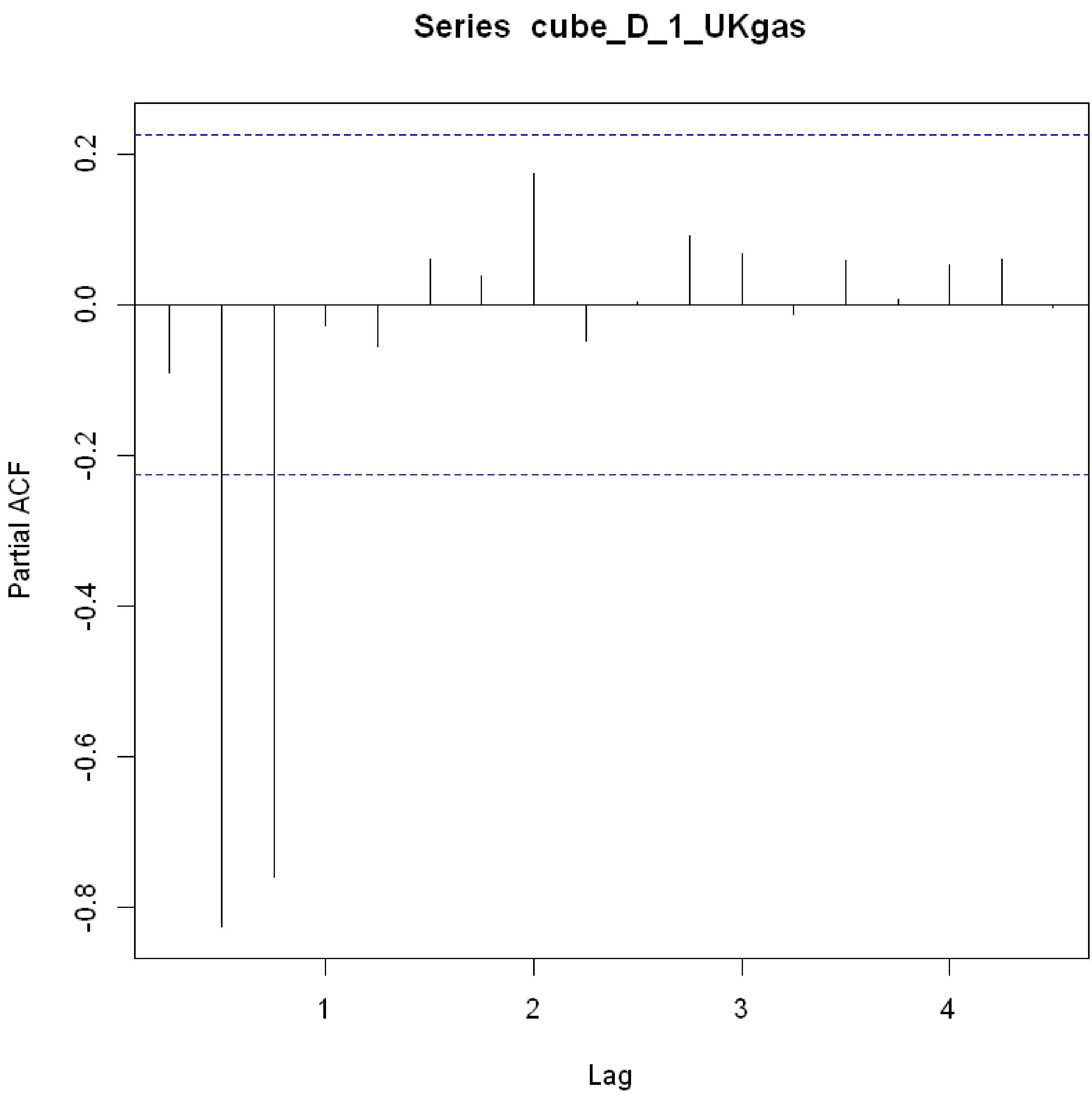
```
acf(cube_D_1_UKgas)
```





In [52]:

```
pacf(cube_D_1_UKgas)
```



Augmented Dickey-Fuller Test for Cube Root Transformed Data (Difference = 1)

In [53]:

```
adf.test(cube_D_1_UKgas)
```

```
Warning message in adf.test(cube_D_1_UKgas):
"p-value smaller than printed p-value"
```

```
	Augmented Dickey-Fuller Test

data:  cube_D_1_UKgas
Dickey-Fuller = -6.9341, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

For the Augmented Dickey-Fuller Test for Cube Root Transformed Data with Difference = 1

Null Hypothesis: Not Stationary

Alternate Hypothesis: Stationary

As per the small p value .01 , we reject the null hypothesis & Hence, the Series is Stationary

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test for the Cube Root Transformed Data (Difference = 1)

In [54]:

```
kpss.test(cube_D_1_UKgas)
```

```
	KPSS Test for Level Stationarity

data:  cube_D_1_UKgas
KPSS Level = 0.43036, Truncation lag parameter = 3, p-value = 0.06407
```

For the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test for the Cube Root Transformed Data with Difference = 1

Null Hypothesis: Stationary

Alternate Hypothesis: Not Stationary

As per the significant p value .06 , we do not reject the null hypothesis & Hence, the Series is Stationary

Augmented Dickey-Fuller Test (For Log Transformed Data)

In [55]:

```
adf.test(log_transform)
```

```
	Augmented Dickey-Fuller Test

data:  log_transform
Dickey-Fuller = -1.9792, Lag order = 4, p-value = 0.584
alternative hypothesis: stationary
```

For the Augmented Dickey-Fuller Test

Null Hypothesis: Not Stationary

Alternate Hypothesis: Stationary

As per the significant p value .58 , we do not reject the null hypothesis & Hence, the Series is not Stationary

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test (For Log Transformed Data)

In [56]:

```
kpss.test(log_transform)
```

```
Warning message in kpss.test(log_transform):
"p-value smaller than printed p-value"
```

```
 KPSS Test for Level Stationarity

data:  log_transform
KPSS Level = 1.9141, Truncation lag parameter = 3, p-value = 0.01
```

For the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

Null Hypothesis: Stationary

Alternate Hypothesis: Not Stationary

As per the small p value .01 , we reject the null hypothesis & Hence, the Series is not Stationary

Now we will try take differences to Make The Log Transformed Data Stationary

For Difference = 1

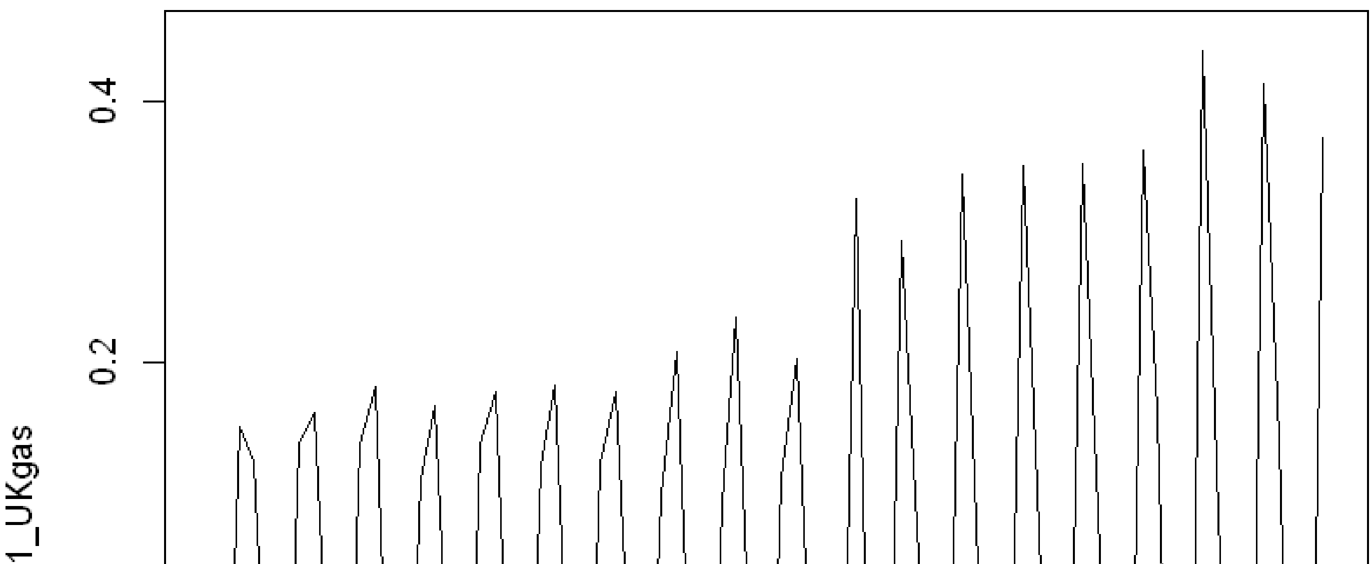
In [57]:

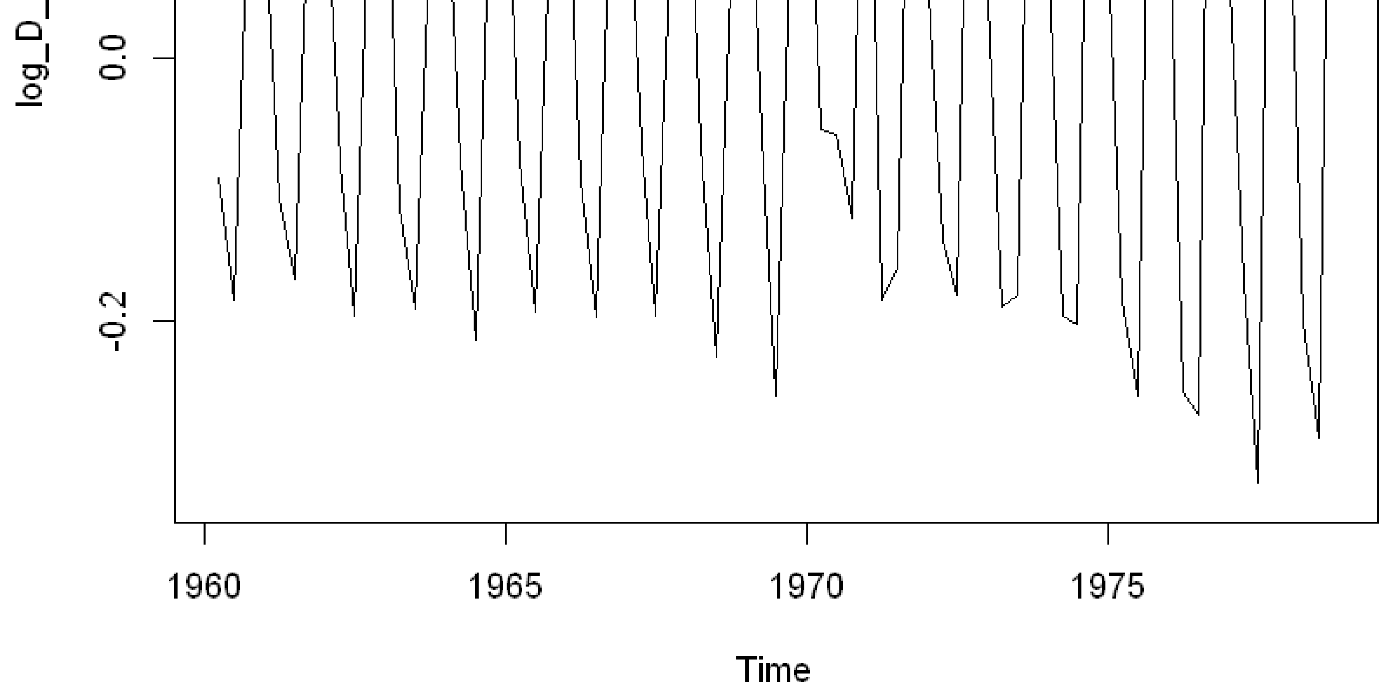
```
log_D_1_UKgas=diff(log_transform, differences =1)
```

Time Plot

In [58]:

```
plot(log_D_1_UKgas)
```

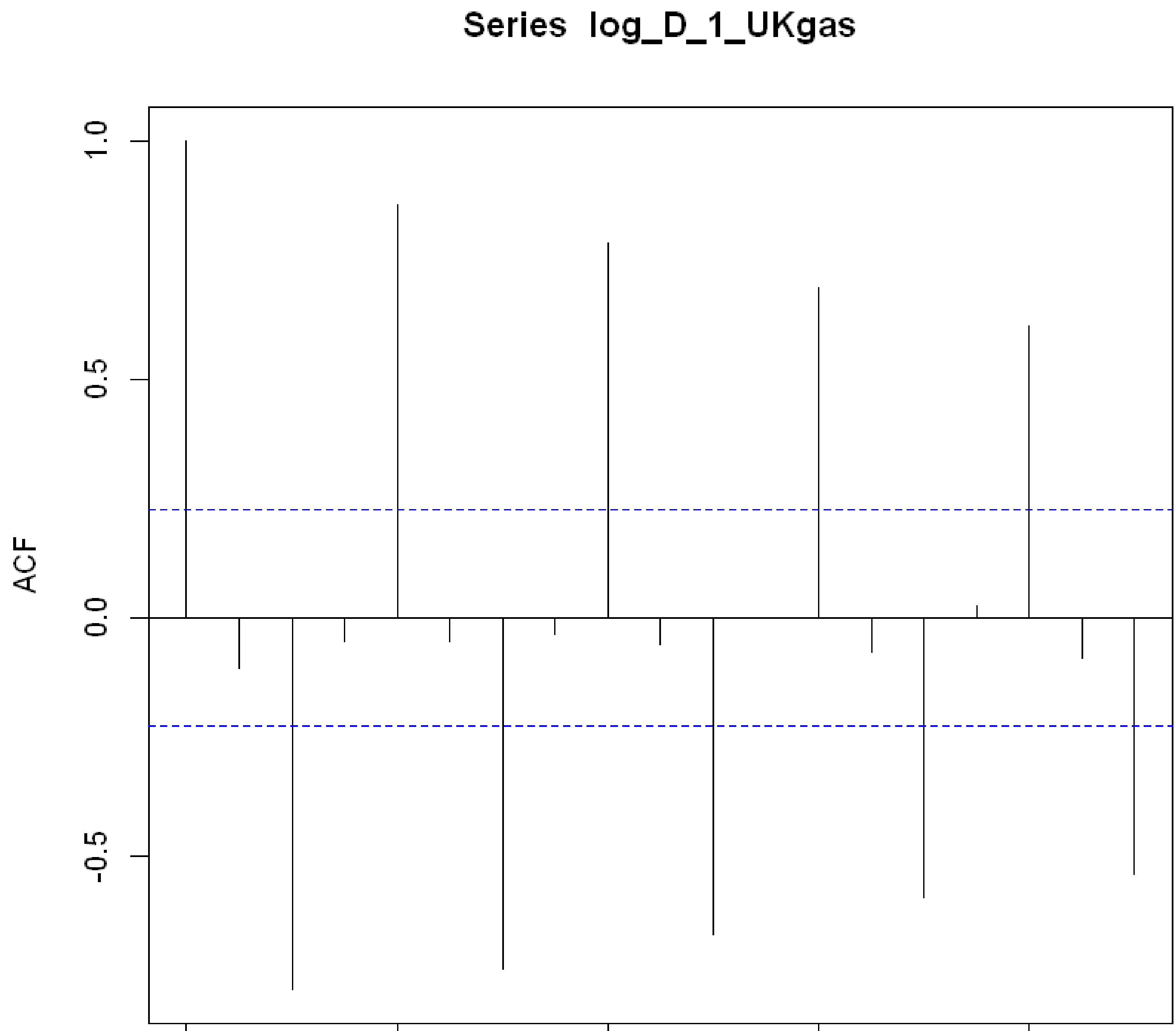


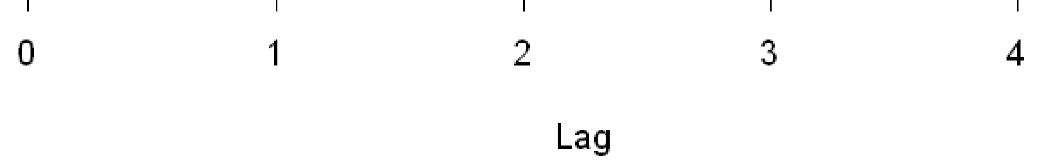


ACF Plot

In [59]:

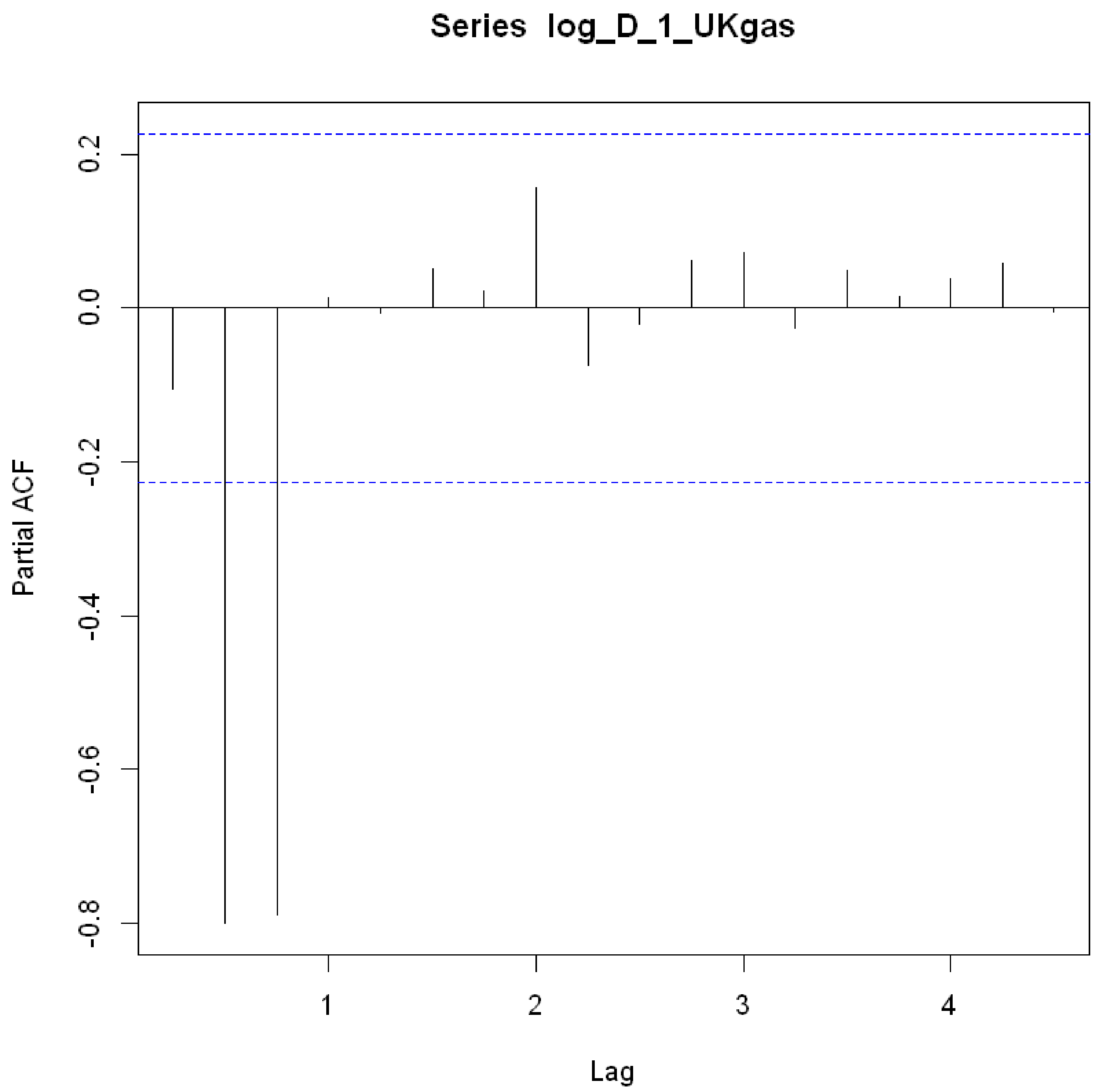
```
acf(log_D_1_UKgas)
```





In [60]:

```
pacf(log_D_1_UKgas)
```



Augmented Dickey-Fuller Test for Log Transformed Data (Difference = 1)

In [61]:

```
adf.test(log_D_1_UKgas)
```

```
Warning message in adf.test(log_D_1_UKgas):
"p-value smaller than printed p-value"
```

```
 Augmented Dickey-Fuller Test
```

```
data:  log_D_1_UKgas
Dickey-Fuller = -6.7559, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

For the Augmented Dickey-Fuller Test for Log Transformed Data with Difference = 1

Null Hypothesis: Not Stationary

Alternate Hypothesis: Stationary

As per the small p value .01 , we reject the null hypothesis & Hence, the Series is Stationary

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test for the Log Transformed Data (Difference = 1)

In [62]:

```
kpss.test(log_D_1_UKgas)
```

```
 KPSS Test for Level Stationarity

data:  log_D_1_UKgas
KPSS Level = 0.41213, Truncation lag parameter = 3, p-value = 0.07193
```

For the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test for the Log Transformed Data with Difference = 1

Null Hypothesis: Stationary

Alternate Hypothesis: Not Stationary

As per the significant p value .07 , we do not reject the null hypothesis & Hence, the Series is Stationary

From the Abobe ADF & KPSS Test Results The Log Transformation seems to be more appropriate in this case.

Model Fitting M2

Now we will fit Model M2 using ACF & PACF plots of the Log Transformed stationary series. When using CSS (conditional sum of squares), it is possible for the autoregressive coefficients to be non-stationary (i.e., they fall outside the region for stationary processes).

In [63]:

```
M2_css = arima(log_D_1_UKgas, order=c(2,1,2))
```

We force R to use MLE (maximum likelihood estimation) instead by using the argument method="ML". This is slower but gives better estimates and always returns a stationary model.

In [64]:

```
M2_mle = arima(log_D_1_UKgas, order=c(2,1,2), method="ML")
```

In [65]:

```
summary(M2_css)
```

```
Call:
arima(x = log_D_1_UKgas, order = c(2, 1, 2))
```

```
Coefficients:
          ar1      ar2      ma1     ma2
      -0.0749  -0.9016  -1.8766  1.0000
s.e.   0.0525   0.0476   0.0602  0.0619

sigma^2 estimated as 0.004443:  log likelihood = 87.82,  aic = -165.63

Training set error measures:
                      ME       RMSE        MAE       MPE     MAPE      MASE
Training set 0.003398959 0.06620819 0.05158229 -7.609886 33.99151 0.2139711
                   ACF1
Training set -0.5555735
```

In [66]:

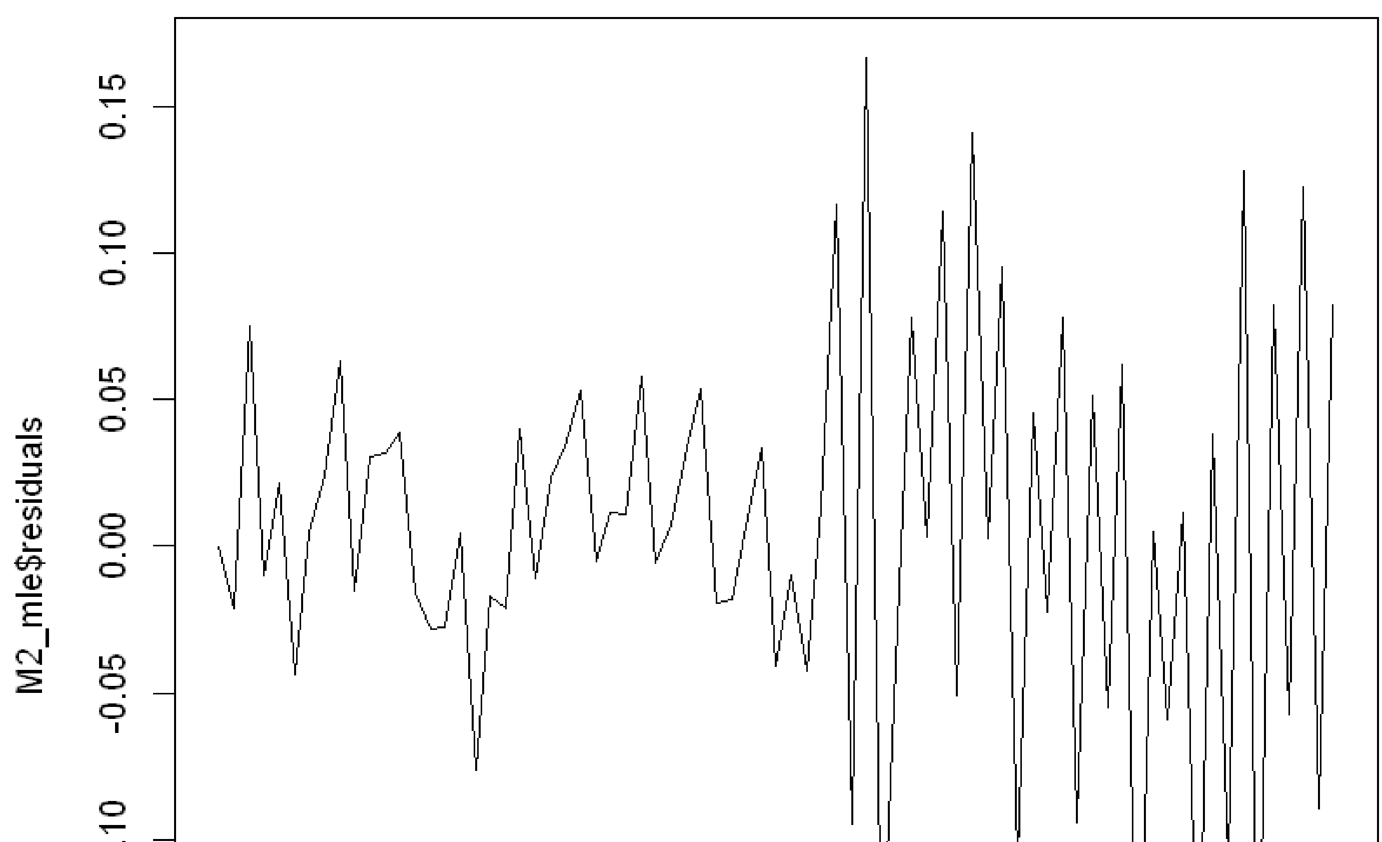
```
summary(M2_mle)
```

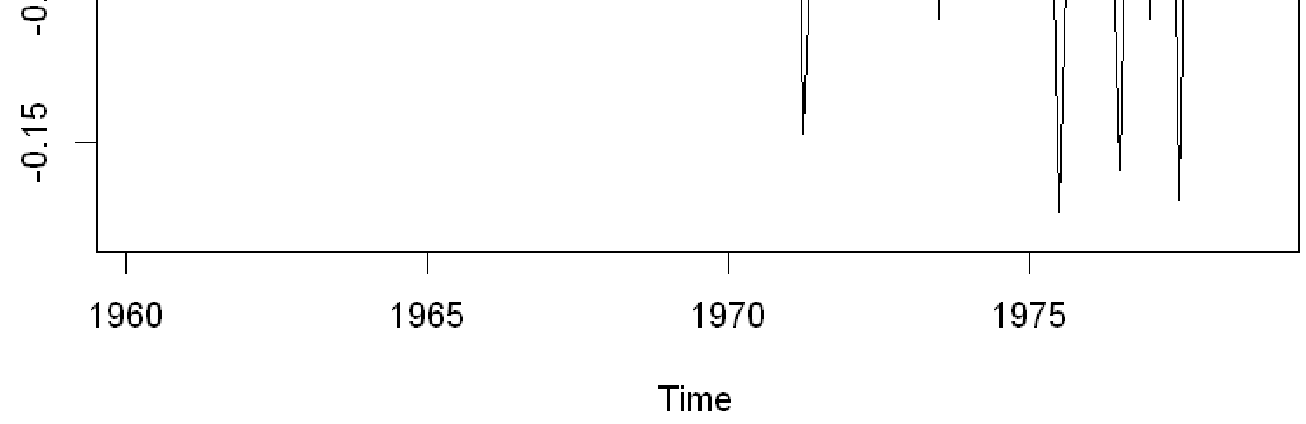
```
Call:
arima(x = log_D_1_UKgas, order = c(2, 1, 2), method = "ML")

Coefficients:
          ar1      ar2      ma1     ma2
      -0.0801  -0.8945  -1.8131  0.8638
s.e.   0.0537   0.0494   0.0650  0.0530

sigma^2 estimated as 0.004913:  log likelihood = 86.06,  aic = -162.12

Training set error measures:
                      ME       RMSE        MAE       MPE    MAPE      MASE
Training set 0.003333348 0.06962407 0.05250501 -5.698118 33.3285 0.2177986
                   ACF1
Training set -0.5467014
```

Residual Checking for model M2

In [67]:

```
plot(M2_mle$residuals)
```

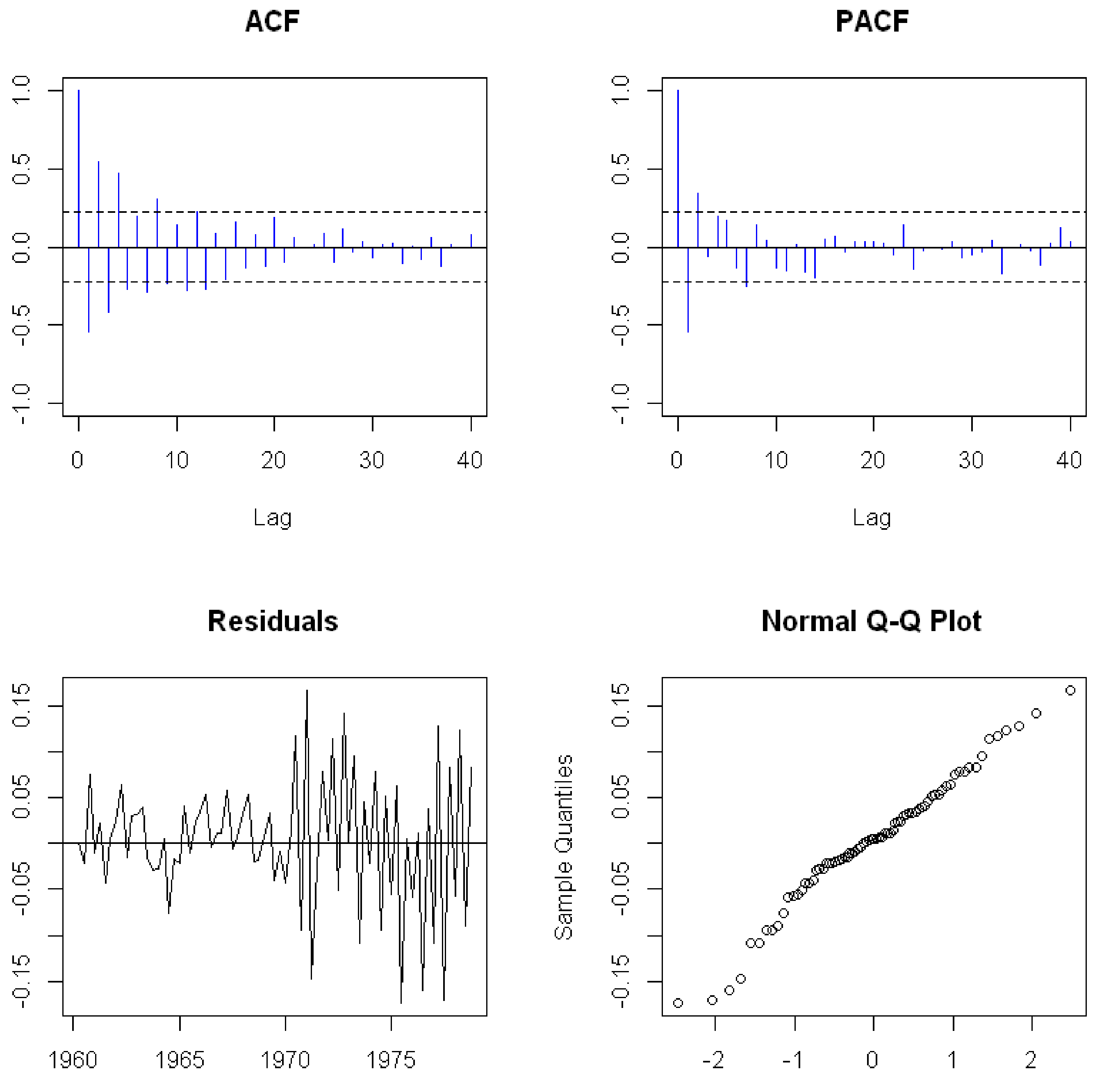




In [68]:

```
test(M2_mle$residuals)
```

```
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                Q ~ chisq(20)    142.69         0 *
McLeod-Li Q                Q ~ chisq(20)     51.52     1e-04 *
Turning points T    (T-48.7)/3.6 ~ N(0,1)       59    0.0042 *
Diff signs S          (S-37)/2.5 ~ N(0,1)       43    0.0171 *
Rank P          (P-1387.5)/109.3 ~ N(0,1)     1355    0.7662
```



Time Theoretical Quantiles

According to Ljung-Box Q Test, McLeod-Li Q Test, Turning points T Test & Diff signs S Test, we found the Residuals may not to be iid noise. Test Distribution Statistic p-value Ljung-Box Q Q ~ chisq(20) 142.69 0 * McLeod-Li Q Q ~ chisq(20) 51.52 1e-04 * Turning points T (T-48.7)/3.6 ~ N(0,1) 59 0.0042 * Diff signs S (S-37)/2.5 ~ N(0,1) 43 0.0171 * According to Rank P test results only we found significant p value & hence do not reject the Null hypothesis (Residuals are iid noise.). Null hypothesis: Residuals are iid noise. Test Distribution Statistic p-value Rank P (P-1387.5)/109.3 ~ N(0,1) 1355 0.7662 According to all the test results it strongly suggests that the data has been completely been stationarized.

Still we decided to apply Box-Cox Transformation.

Answer to the Question No 4

Box-Cox Transformation

In [69]:

```
library(forecast)
```

In [70]:

```
lambda = BoxCox.lambda(train) #Optimal Lambda Selection
```

In [71]:

```
print(lambda)
```

```
[1] -0.4092527
```

In [72]:

```
BoxCox_UKgas = BoxCox(train, lambda)
```

In [73]:

```
print(BoxCox_UKgas)
```

```
         Qtr1     Qtr2     Qtr3     Qtr4
1960 2.137384 2.109835 2.046463 2.099168
1961 2.137384 2.104646 2.046463 2.095341
1962 2.144592 2.120955 2.055486 2.102853
1963 2.156422 2.123906 2.061012 2.099168
1964 2.149087 2.126765 2.055486 2.102853
1965 2.155413 2.133547 2.071299 2.111505
1966 2.164084 2.138627 2.076099 2.116347
1967 2.166782 2.149087 2.089316 2.120955
1968 2.178284 2.161294 2.093372 2.122442
1969 2.186256 2.171918 2.097273 2.132230
1970 2.186256 2.172743 2.157420 2.122442
1971 2.207078 2.162235 2.116347 2.195306
1972 2.212036 2.179797 2.130895 2.217539
1973 2.226561 2.184164 2.136123 2.222616
1974 2.242932 2.202287 2.151248 2.233775
1975 2.250061 2.213455 2.150174 2.235122
1976 2.264476 2.215755 2.149087 2.248758
1977 2.263278 2.232049 2.156422 2.249021
1978 2.273010 2.237409 2.172743 2.252826
```

Augmented Dickey-Fuller Test (For Box-Cox Transformed Data)

In [74]:

```
adf.test(BoxCox_UKgas)
```

```
 Augmented Dickey-Fuller Test

data:  BoxCox_UKgas
Dickey-Fuller = -2.3674, Lag order = 4, p-value = 0.4256
alternative hypothesis: stationary
```

For the Augmented Dickey-Fuller Test

Null Hypothesis: Not Stationary

Alternate Hypothesis: Stationary

As per the significant p value .42 , we do not reject the null hypothesis & Hence, the Series is not Stationary

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test (For Box-Cox Transformed Data)

In [75]:

```
kpss.test(BoxCox_UKgas)
```

```
Warning message in kpss.test(BoxCox_UKgas):
"p-value smaller than printed p-value"
```

```
 KPSS Test for Level Stationarity

data:  BoxCox_UKgas
KPSS Level = 1.9376, Truncation lag parameter = 3, p-value = 0.01
```

For the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

Null Hypothesis: Stationary

Alternate Hypothesis: Not Stationary

As per the small p value .01 , we reject the null hypothesis & Hence, the Series is not Stationary

Now we will try take differences to Make The Box-Cox Transformed Data Stationary

For Difference = 1

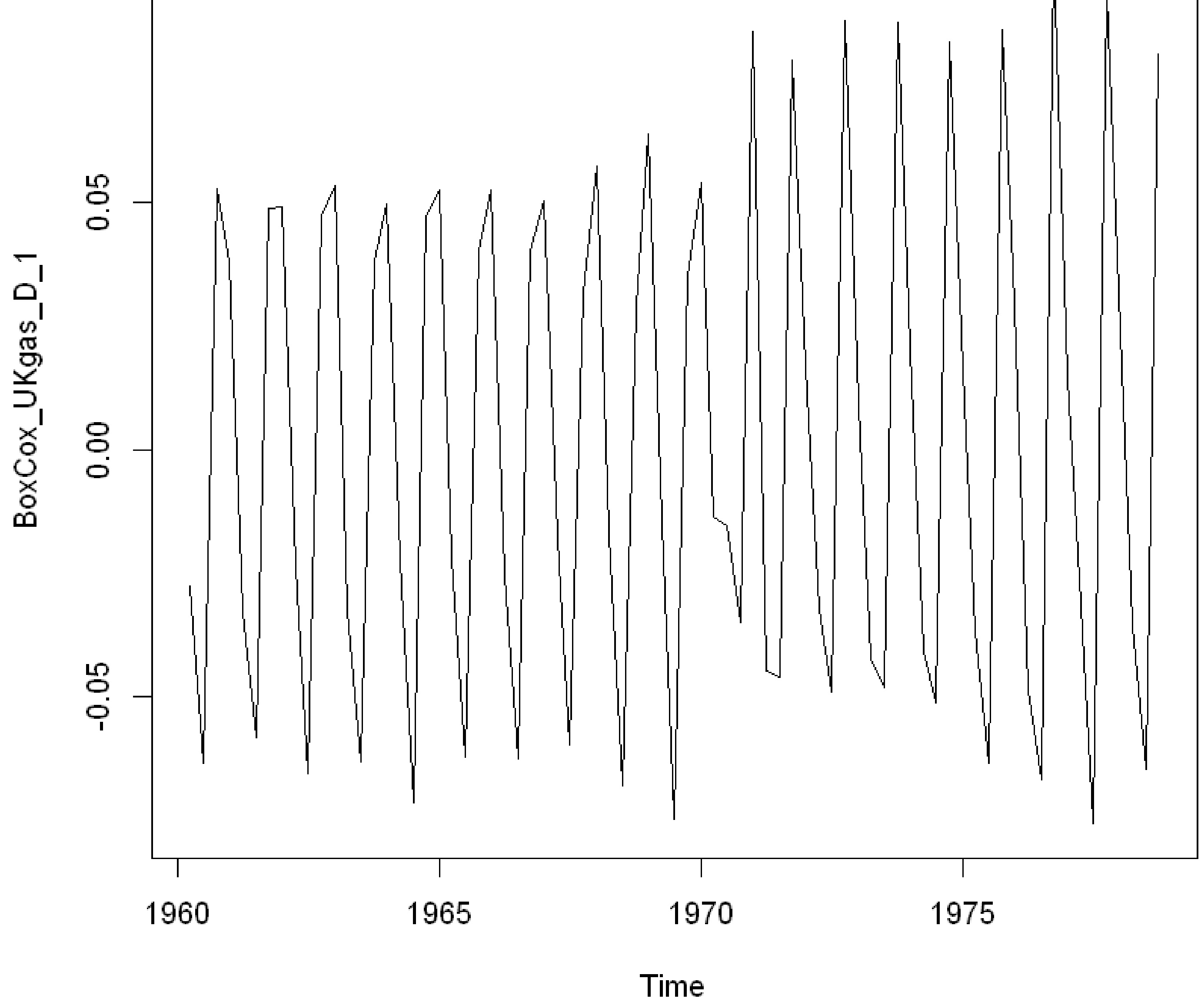
In [76]:

```
BoxCox_UKgas_D_1 = diff(BoxCox_UKgas, differences =1)
```

Time Plot

In [77]:

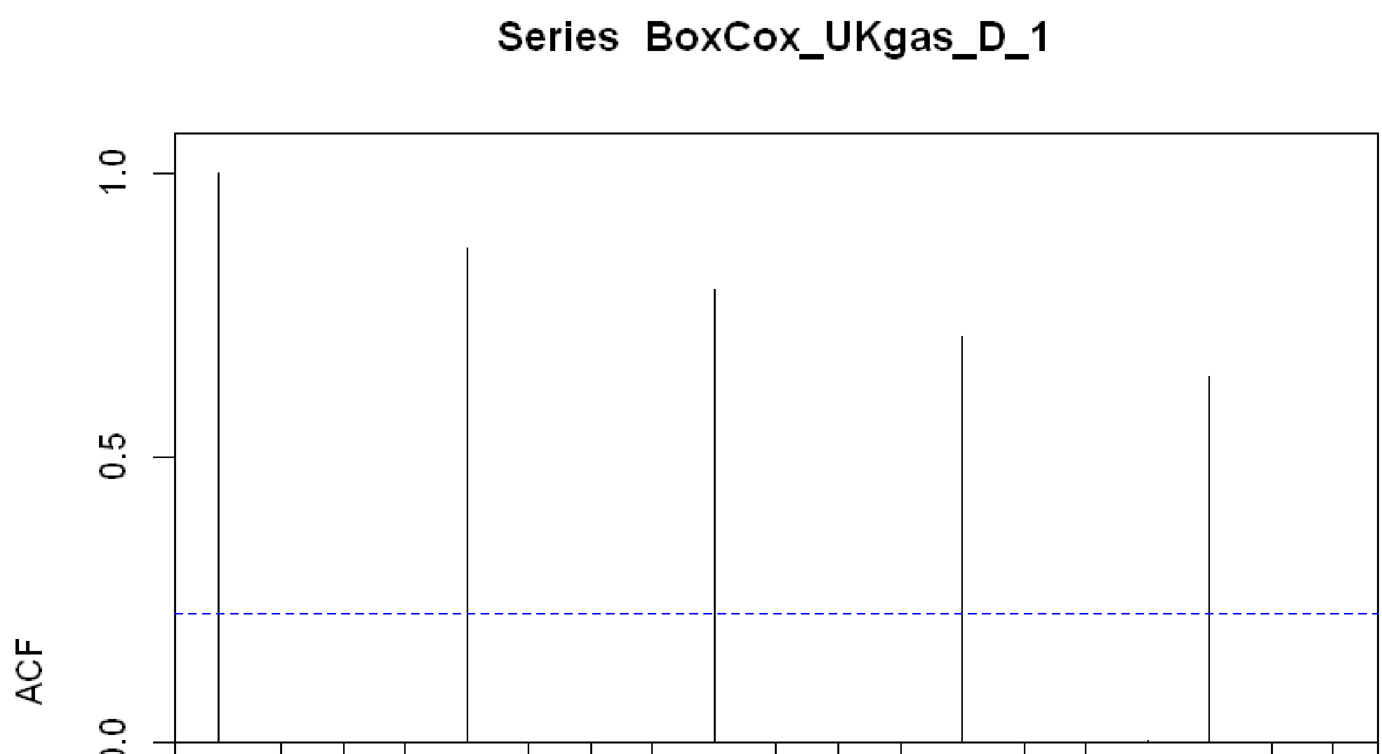
```
plot(BoxCox_UKgas_D_1)
```

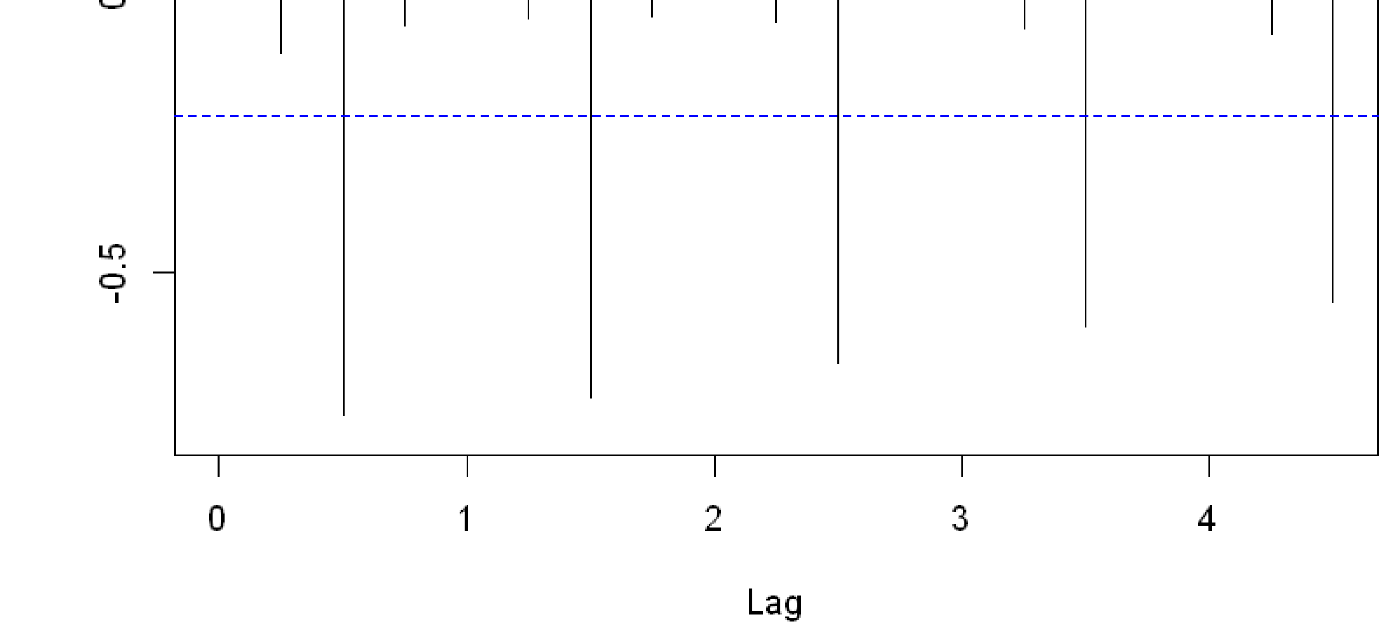


ACF Plot

In [78]:

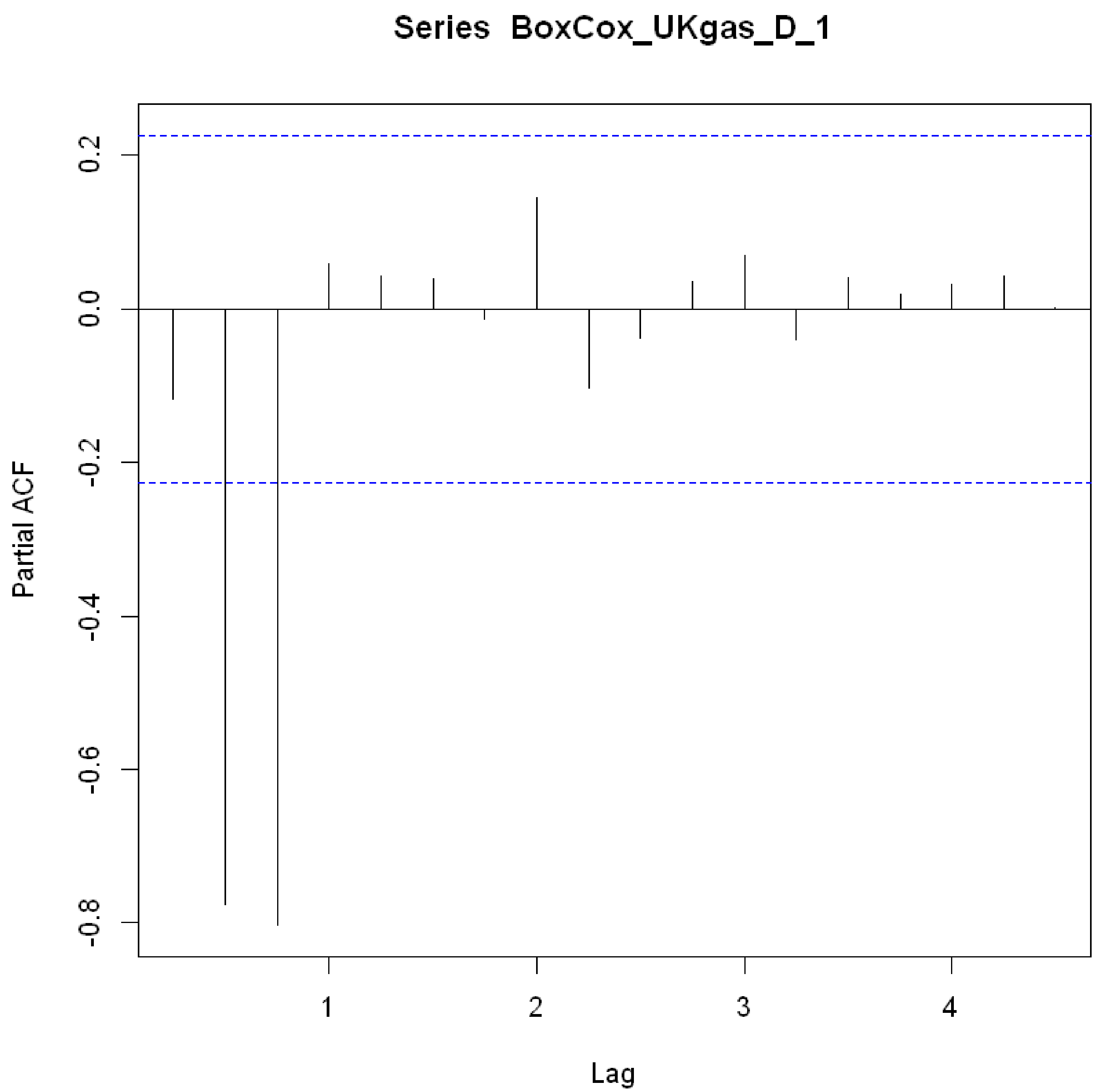
```
acf(BoxCox_UKgas_D_1)
```





In [79]:

```
pacf(BoxCox_UKgas_D_1)
```



Augmented Dickey-Fuller Test for Box-Cox Transformed Data (Difference = 1)

In [80]:

```
adf.test(BoxCox_UKgas_D_1)
```

```
Warning message in adf.test(BoxCox_UKgas_D_1):
"p-value smaller than printed p-value"
```

```
	Augmented Dickey-Fuller Test

data:  BoxCox_UKgas_D_1
Dickey-Fuller = -6.5594, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

For the Augmented Dickey-Fuller Test for Box-Cox Transformed Data with Difference = 1

Null Hypothesis: Not Stationary

Alternate Hypothesis: Stationary

As per the small p value .01 , we reject the null hypothesis & Hence, the Series is Stationary

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test for the Box-Cox Transformed Data (Difference = 1)

In [81]:

```
kpss.test(BoxCox_UKgas_D_1)
```

```
	KPSS Test for Level Stationarity

data:  BoxCox_UKgas_D_1
KPSS Level = 0.37361, Truncation lag parameter = 3, p-value = 0.08853
```

For the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test for the Box-Cox Transformed Data with Difference = 1

Null Hypothesis: Stationary

Alternate Hypothesis: Not Stationary

As per the significant p value .08 , we do not reject the null hypothesis & Hence, the Series is Stationary

As per the ADF & KPSS Test Result the Box-Cox Transformed data might be Stationary. Hence we decide to fit a Model.

Model Fitting (M3)

Using ACF & PACF plots of the Box-Cox Transformed stationary series.

In [82]:

```
M3_css = arima(BoxCox_UKgas_D_1, order=c(2,1,2))
```

In [83]:

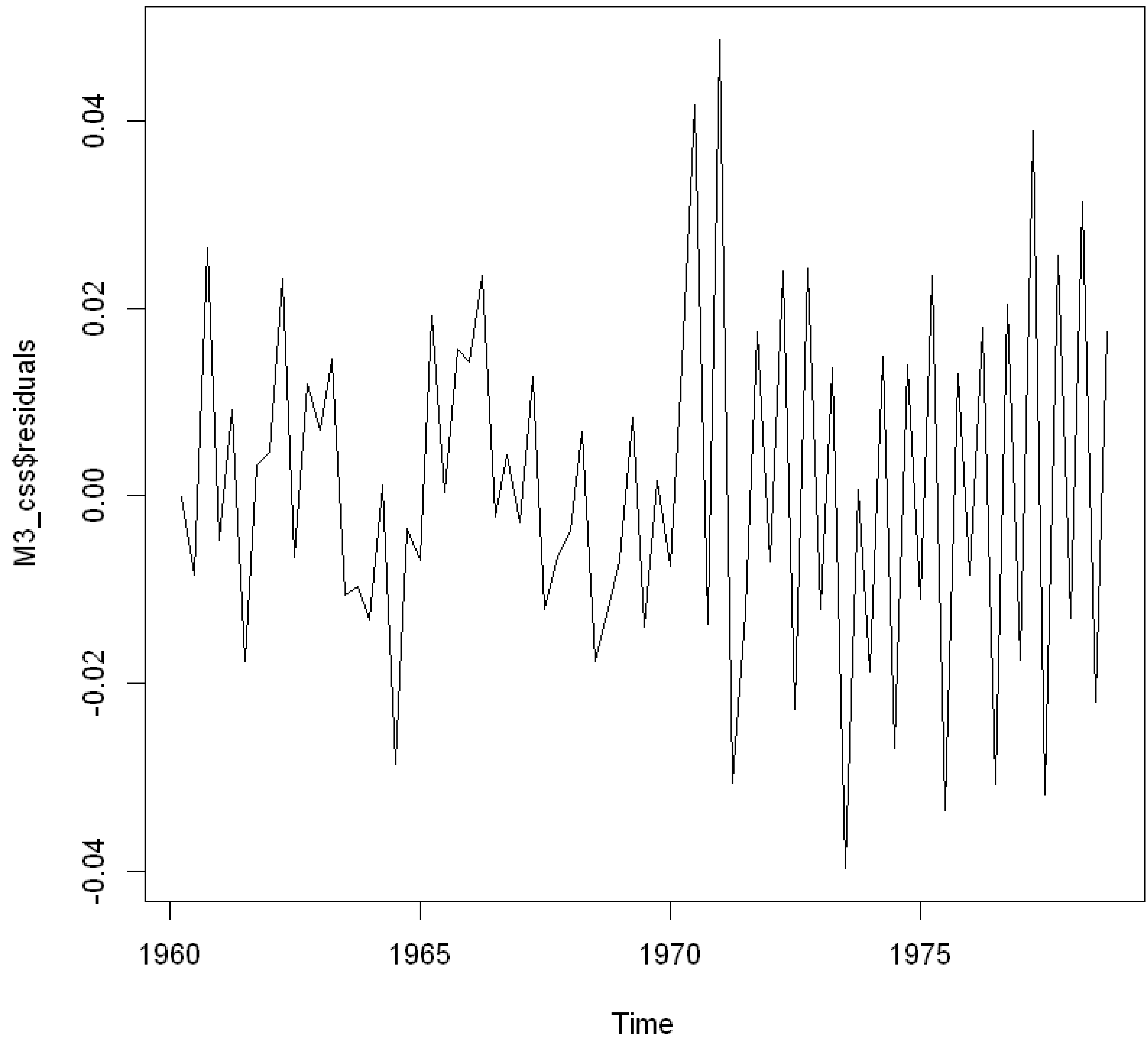
```
summary(M3_css)
```

```
Call:
arima(x = BoxCox_UKgas_D_1, order = c(2, 1, 2))

Coefficients:
          ar1      ar2      ma1     ma2
      -0.0866  -0.8754  -1.8842  1.0000
s.e.   0.0575   0.0533   0.0748  0.0779

sigma^2 estimated as 0.0003555:  log likelihood = 181.46,  aic = -352.92

Training set error measures:
                       ME      RMSE        MAE       MPE     MAPE      MASE
Training set 0.0008397286 0.0187278 0.01543708 -13.54607 42.02302 0.2373626
                   ACF1
Training set -0.5235836
```

Residual Checking for model M3

In [84]:

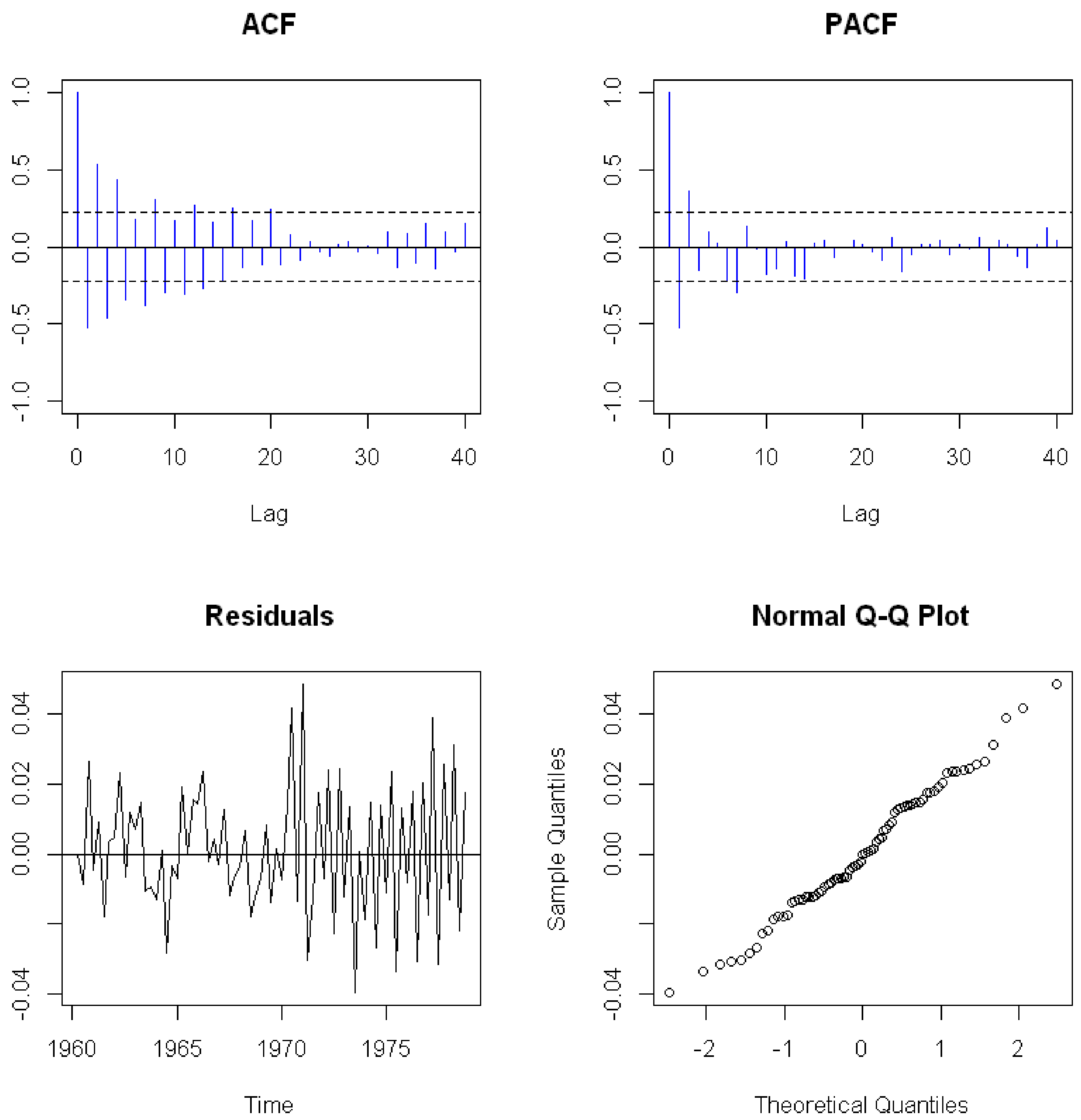
```
plot(M3_css$residuals)
```



In [85]:

```
test(M3_css$residuals)
```

```
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                Q ~ chisq(20)    166.42         0 *
McLeod-Li Q                Q ~ chisq(20)     19.31    0.5017
Turning points T    (T-48.7)/3.6 ~ N(0,1)       65         0 *
Diff signs S          (S-37)/2.5 ~ N(0,1)       41     0.112
Rank P          (P-1387.5)/109.3 ~ N(0,1)     1335     0.631
```



According to Ljung-Box Q Test & Turning points T Test, we found the Residuals may not to be iid noise. Test Distribution Statistic p-value Ljung-Box Q Q ~ chisq(20) 166.42 0 * Turning points T (T-48.7)/3.6 ~ N(0,1) 65 0 * According to Rank P test results only we found significant p value & hence do not reject the Null hypothesis (Residuals are iid noise.) Test Distribution Statistic p-value McLeod-Li Q Q ~ chisq(20) 19.31 0.5017 Diff signs S (S-37)/2.5 ~ N(0,1) 41 0.112 Rank P (P-1387.5)/109.3 ~ N(0,1) 1335 0.631

Answer to the Question No 5

In [86]:

```
AIC(M1_mle)
```

737.3443929293

In [87]:

```
BIC(M1_mle)
```

748.86471839532

In [88]:

```
AIC(M2_mle)
```

-162.120463675724

In [89]:

```
BIC(M2_mle)
```

-150.600138209703

In [90]:

```
BIC(M3_css)
```

-341.399637469183

In [91]:

```
BIC(M3_css)
```

-341.399637469183

Selection of Model among M1, M2 & M3 using AIC,AIC_c & BIC

Model	AIC	AICc	BIC
M1	737.3443929293		748.86471839532
M2	-162.120463675724		-150.600138209703
M3	-341.399637469183		-341.399637469183

According to the AIC & BIC score of the above models the M3 model with 1 Difference seems to be more accurate among Model M1, M2 & M3.

On the other hand, in terms of the test result of the residuals of the above models the Model M2 seems to be more Stationary.

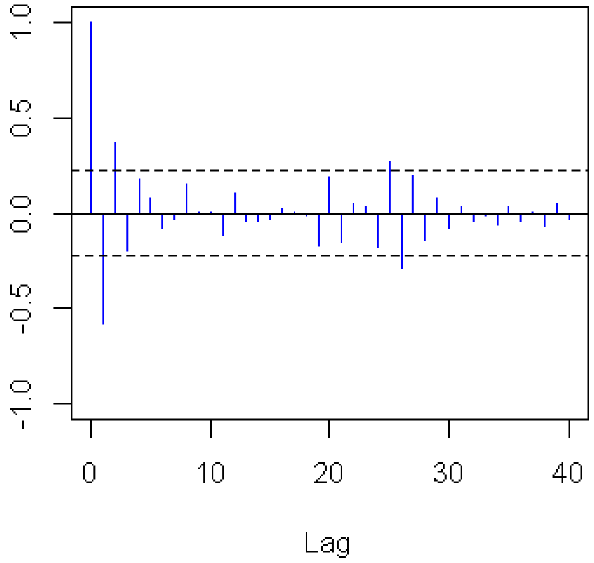
M1 Residuals Test

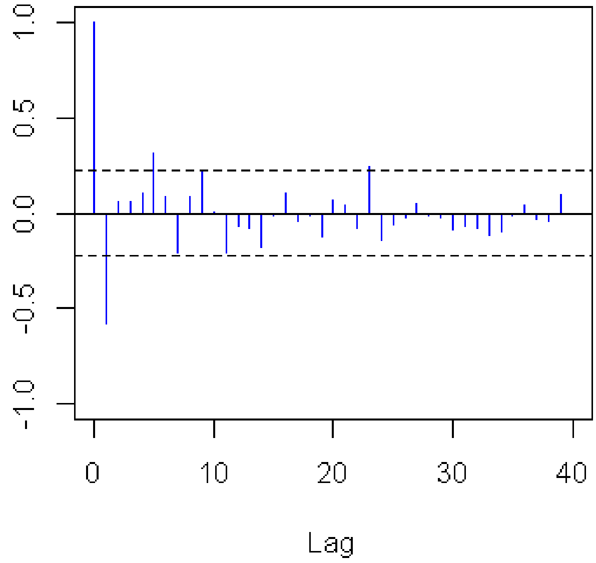
In [92]:

```
test(M1_mle$residuals)
```

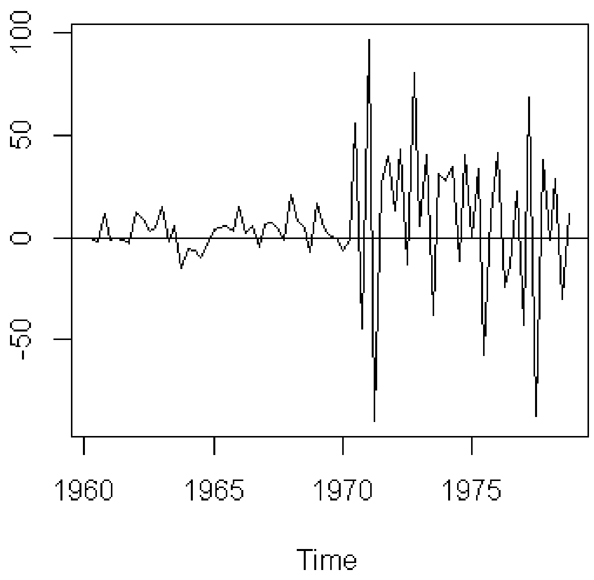
```
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                Q ~ chisq(20)     54.88         0 *
McLeod-Li Q                Q ~ chisq(20)     31.11    0.0537
Turning points T    (T-48.7)/3.6 ~ N(0,1)       55    0.0791
Diff signs S          (S-37)/2.5 ~ N(0,1)       37         1
Rank P          (P-1387.5)/109.3 ~ N(0,1)     1531    0.1892
```

ACF **PACF**

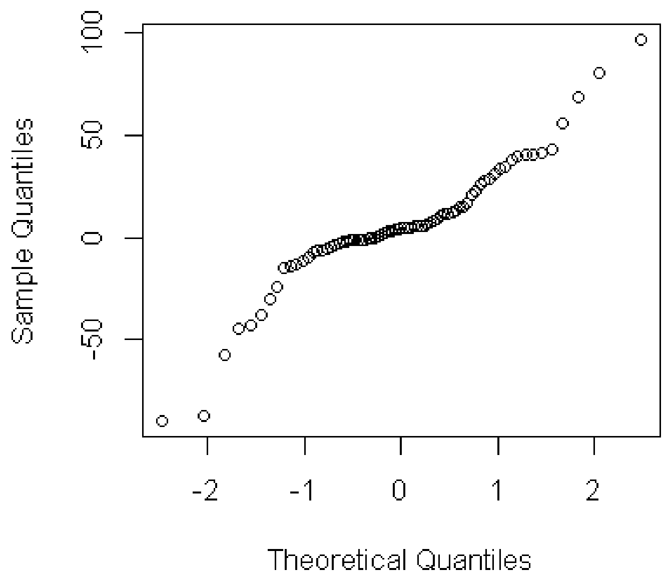




Residuals



Normal Q-Q Plot



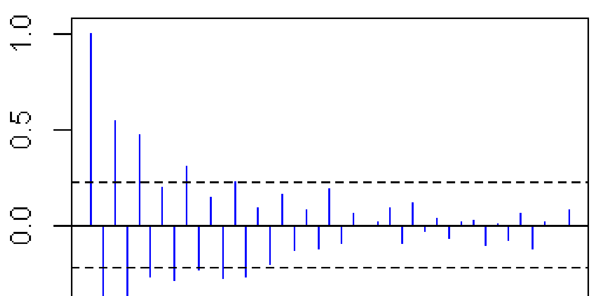
M2 Residuals Test

In [93]:

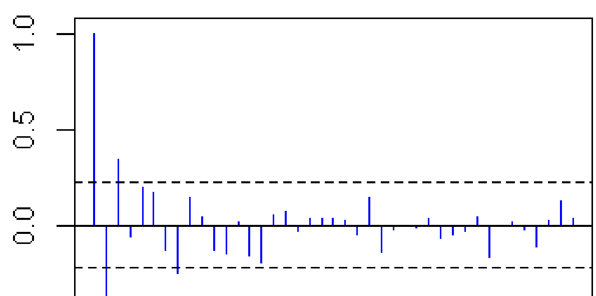
```
test(M2_mle$residuals)
```

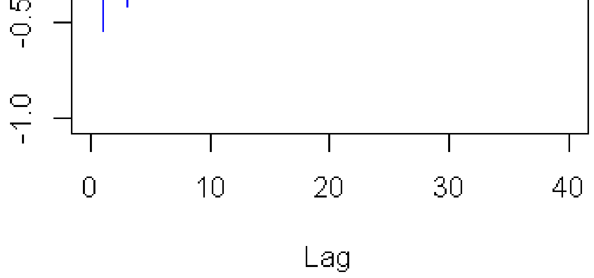
```
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                  Q ~ chisq(20)    142.69         0 *
McLeod-Li Q                  Q ~ chisq(20)     51.52     1e-04 *
Turning points T    (T-48.7)/3.6 ~ N(0,1)        59    0.0042 *
Diff signs S          (S-37)/2.5 ~ N(0,1)        43    0.0171 *
Rank P          (P-1387.5)/109.3 ~ N(0,1)      1355    0.7662
```

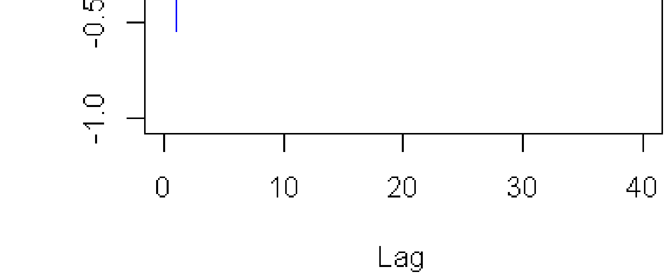
ACF

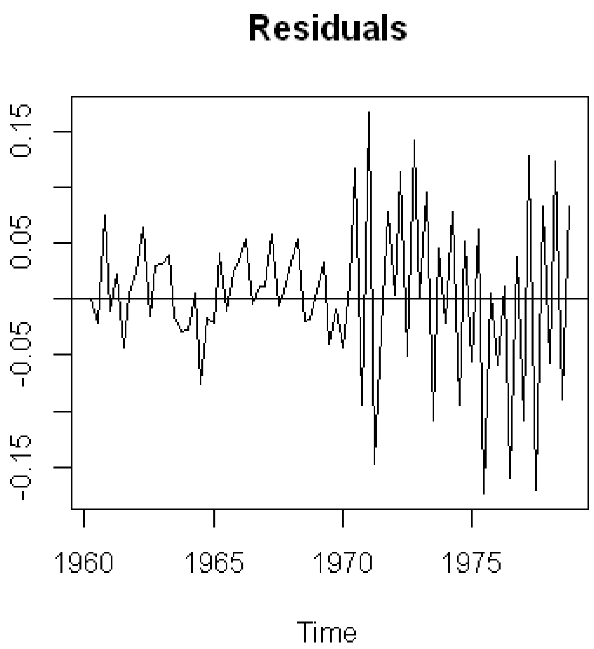


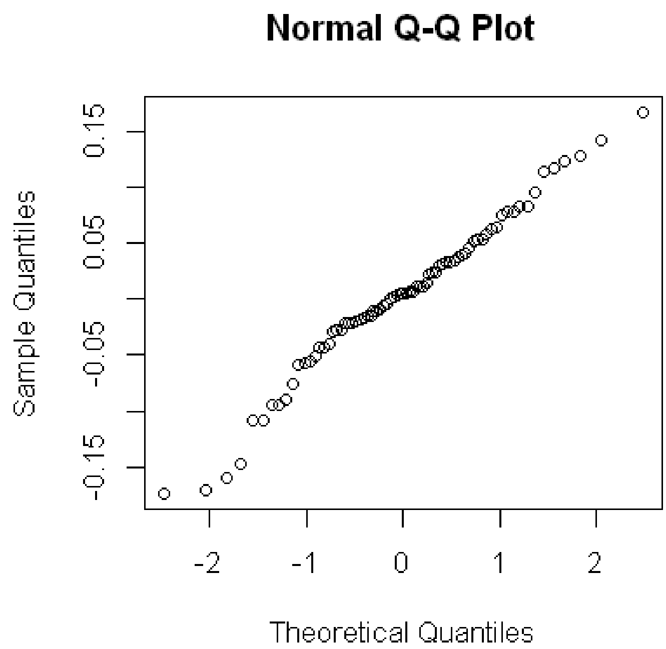
PACF









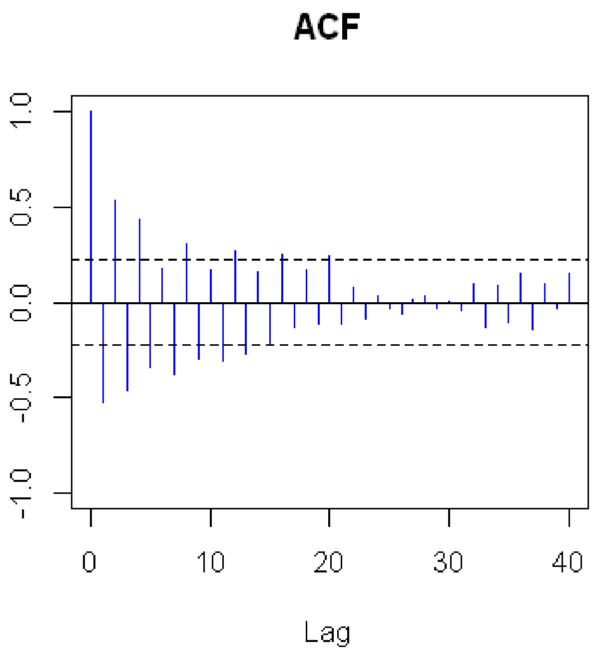


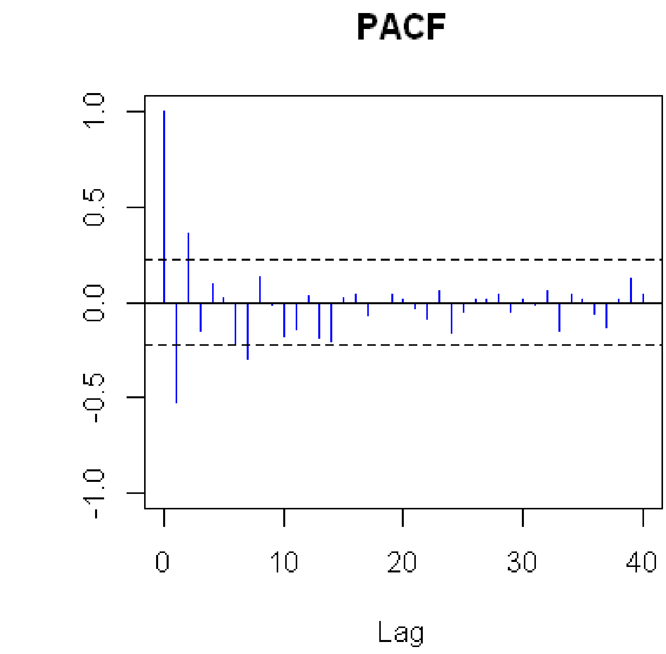
M3 Residuals Test

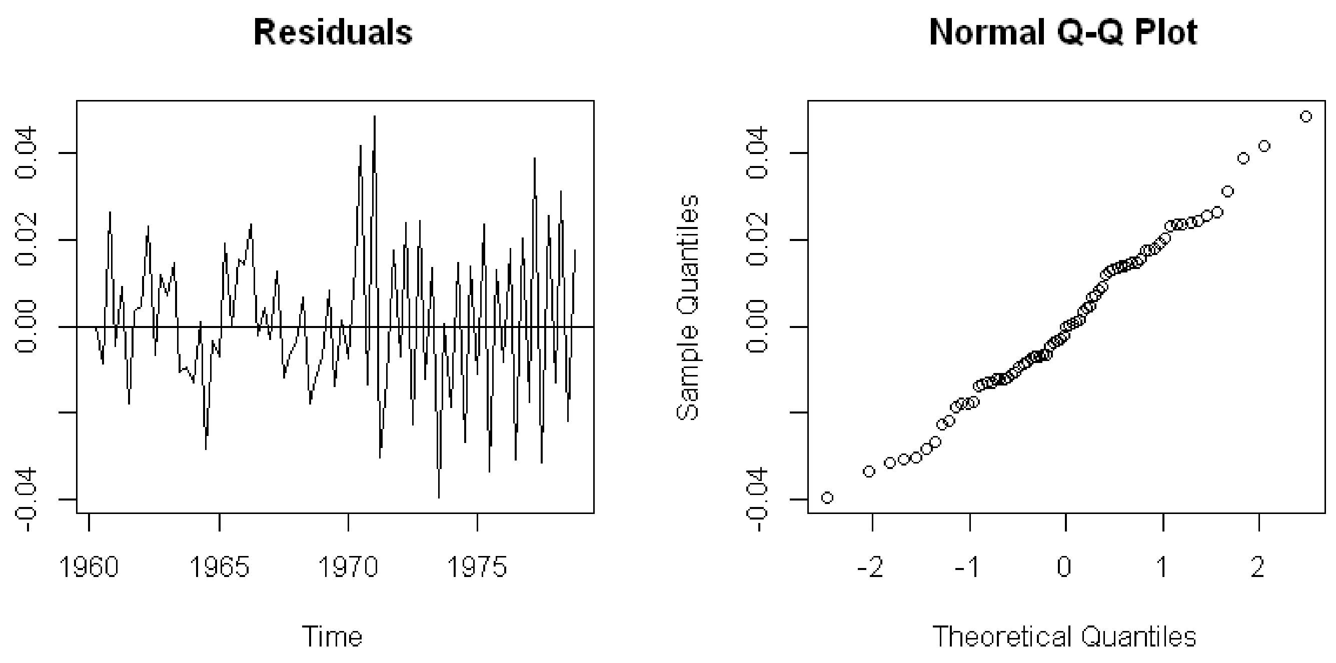
In [94]:

```
test(M3_css$residuals)
```

```
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                Q ~ chisq(20)    166.42         0 *
McLeod-Li Q                Q ~ chisq(20)     19.31    0.5017
Turning points T    (T-48.7)/3.6 ~ N(0,1)       65         0 *
Diff signs S          (S-37)/2.5 ~ N(0,1)       41     0.112
Rank P          (P-1387.5)/109.3 ~ N(0,1)     1335     0.631
```







Answer to the Question No 6

Need More Exploratiom

Answer to the Question No 7

Auto Arima Model

In [97]:

```
arima_model_forecast <- auto.arima(test)

summary(arima_model_forecast)
```

```
Series: test
ARIMA(0,0,0)(0,1,0)[4] with drift

Coefficients:
       drift
      7.6750
s.e.  1.9808

sigma^2 estimated as 1823:  log likelihood=-144.34
AIC=292.67   AICc=293.15   BIC=295.34

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE
Training set 0.06158903 39.21849 31.16784 -1.550546 6.255807 0.7798923
                   ACF1
Training set 0.09422983
```

Holt's Trend Method

In [98]:

```
holt_model <- holt(test, h = 10)

summary(holt_model)
```

```
Forecast method: Holt's method

Model Information:
Holt's method
```

```
Call:
 holt(y = test, h = 10)

  Smoothing parameters:
    alpha = 1e-04
    beta  = 1e-04

  Initial states:
    l = 547.7591
    b = 3.4776

  sigma:  286.964

     AIC     AICc      BIC
478.8294 481.1371 486.1581

Error measures:
                    ME     RMSE      MAE       MPE     MAPE     MASE
Training set -9.879313 268.4303 230.9539 -32.97518 57.48237 5.779006
                      ACF1
Training set -0.0003346537

Forecasts:
        Point Forecast    Lo 80    Hi 80     Lo 95    Hi 95
1987 Q1       661.3948 293.6356 1029.154  98.95567 1223.834
1987 Q2       664.8408 297.0816 1032.600 102.40164 1227.280
1987 Q3       668.2868 300.5276 1036.046 105.84760 1230.726
1987 Q4       671.7328 303.9735 1039.492 109.29354 1234.172
1988 Q1       675.1787 307.4195 1042.938 112.73945 1237.618
1988 Q2       678.6247 310.8654 1046.384 116.18533 1241.064
1988 Q3       682.0707 314.3113 1049.830 119.63117 1244.510
1988 Q4       685.5167 317.7571 1053.276 123.07698 1247.956
1989 Q1       688.9627 321.2030 1056.722 126.52273 1251.403
1989 Q2       692.4087 324.6488 1060.169 129.96843 1254.849
```

Simple Exponential Smoothing

In [100]:

```
se_model <- ses(test, h = 10)

summary(se_model)
```

```
Forecast method: Simple exponential smoothing

Model Information:
Simple exponential smoothing

Call:
 ses(y = test, h = 10)

  Smoothing parameters:
    alpha = 1e-04

  Initial states:
    l = 594.842

  sigma:  281.2945

     AIC     AICc      BIC
475.7601 476.6172 480.1573

Error measures:
                     ME     RMSE      MAE       MPE     MAPE     MASE
Training set 0.03634224 272.3622 234.1532 -31.43043 57.43875 5.859061
                   ACF1
Training set 0.03213703

Forecasts:
        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
1987 Q1       594.8421 234.3487 955.3355 43.51507 1146.169
```

```
1987 Q1         594.8421 234.3487 955.3355 43.51507 1146.169
1987 Q2         594.8421 234.3487 955.3355 43.51507 1146.169
1987 Q3         594.8421 234.3487 955.3355 43.51507 1146.169
1987 Q4         594.8421 234.3487 955.3355 43.51507 1146.169
1988 Q1         594.8421 234.3487 955.3355 43.51506 1146.169
1988 Q2         594.8421 234.3487 955.3355 43.51506 1146.169
1988 Q3         594.8421 234.3487 955.3355 43.51506 1146.169
1988 Q4         594.8421 234.3487 955.3355 43.51506 1146.169
1989 Q1         594.8421 234.3487 955.3355 43.51505 1146.169
1989 Q2         594.8421 234.3487 955.3355 43.51505 1146.169
```

In [ ]: