MATH1318 Time Series Analysis

Assignment 1

Amal Gopidasan Nair - S3820786

RMIT University

1 Introduction

The ozone layer is the region present in the Earth's stratosphere which filters out most of the sun's ultraviolet radiations. The ozone layer has been depleting for many years and causing increased level of UV radiation on Earth's surface affecting all living organisms.
We analyze the yearly changes in the thickness of ozone layer from 1927 to 2016 in this assignment.
The following task is performed in this assignment:

- Find the best model
- Forecast for the next five years

- Propose a set of possible ARIMA

2 Data Description

- The dataset provided in in csv format with 90 observations.
- The data is from 1927 to 2016.
- The measurement is done in Dobson units.
- Negative value represents a decrease in the thickness and positive value represents increase in thickness.

3 Data Preparation

```
options(warn=-1)
library(TSA)
```

```
##
## Attaching package: 'TSA'
```

```
## The following objects are masked from 'package:stats':
##
##     acf, arima
```

```
## The following object is masked from 'package:utils':
##
##     tar
```

```
library(tseries)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
# Importing the data file in R
ozone<-read.csv("data1.csv", header = FALSE)
rownames(ozone) <- seq(from=1927, to=2016)
class(ozone)
```

```
## [1] "data.frame"
```

- The class of the dataset is found to be data frame.
- The data frame is then converted into time series object.

```
ozonets <- ts(as.vector(ozone), start=1927, end=2016)
class(ozonets)
```

```
## [1] "ts"
```

```
summary(ozonets)
```

```
##          V1
##  Min.    :-11.5794
##  1st Qu.: -4.9281
##  Median : -2.6639
##  Mean   : -3.2023
##  3rd Qu.: -0.8746
##  Max.   :  3.4072
```

4 Model Descriptive Analysis

After conversion of the data into Time series object and then the time series plot and scatter plot is generated to understand the five major points associated with the series and the correlation with subsequent years. The time series plot is generated and evaluated.

```
plot(ozonets,type='o',ylab='Thickness of ozone layer (Dobson Units)', xlab='Year', ma
in ='Time Series Plot-Change in Ozone layer thickness - Figure 1' )
```


Time Series Plot-Change in Ozone layer thickness - Figure 1
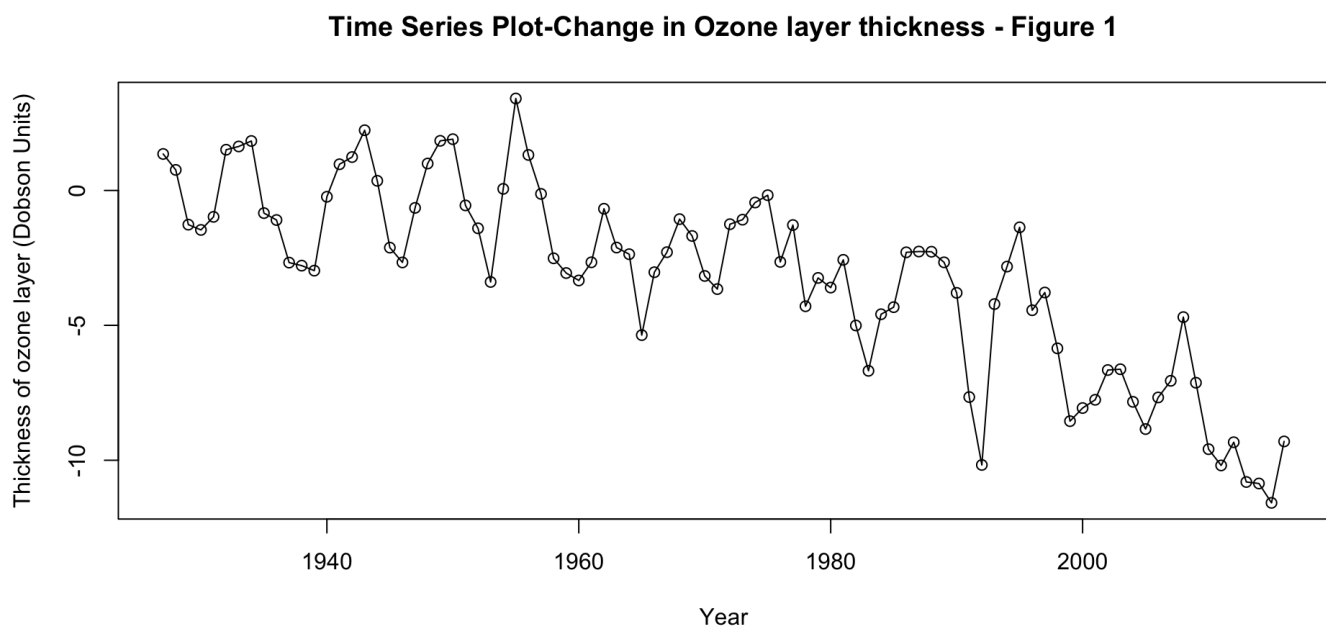
From figure 1 we discuss on the 5 major point:

1.Trend

We can observe that the time series plot follows a decreasing downward trend which denote that the ozone layer's thickness is reducing over the time period.

2.Seasonality

We don't observe any repeating patterns denoting no seasonality in the time series.

3.Changing Variance

We don't observe any significant change in variance overall.

4.Behavior

We can observe both Auto regressive and moving average like behavior with local trends and successive points related to on another.

5.Intervention

We don't observe any sudden change point in the time series.

To understand the correlation between the succeeding data points, which if present enables us to forecast the values for the consecutive year we plot a scatter plot and find the correlation value.

```
plot(y=ozonets,x=zlag(ozonets),ylab='Thickness of ozone layer', xlab='Previous Year T
hickness of ozone layer', main = "Scatter plot of thickness of Ozone layer- Figure 2"
)
```

Scatter plot of thickness of Ozone layer- Figure 2
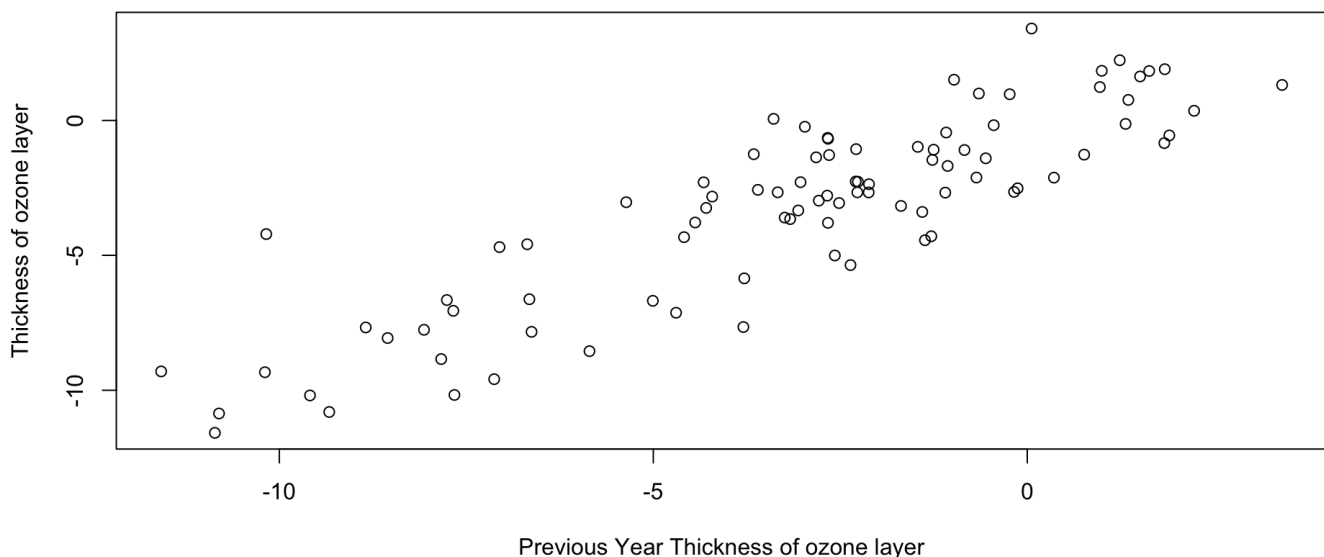

```
y = ozonets
x=zlag(ozonets)
index = 2:length(x)
cor(y[index],x[index])
```

```
## [1] 0.8700381
```

From figure 2 and the correlation factor(~87%) we can observe that there is a high correlation between the succeeding year and the thickness of ozone layer of one year can impact the thickness in future years.

5 Model Fitting

The time series object is analyzed for finding the trend. We consider linear, quadratic, cosine and seasonal trends to find the bet fitting model.

5.1 Linear Regression Model

The deterministic linear trend model is expressed as :

$\mu t = \beta_0 + \beta_1 t$

where $\beta_0 = $ *Intercept* $\beta_1 = $ *Slope of the linear trend*

5.1.1 Summary Statistics and Plot

The slope and intercept of the linear trend is estimated using least-square regression.

```
ozonelm=lm(ozonets~time(ozonets))
summary(ozonelm)
```

```
##
## Call:
## lm(formula = ozonets ~ time(ozonets))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7165 -1.6687  0.0275  1.4726  4.7940
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   213.720155  16.257158   13.15   <2e-16 ***
## time(ozonets)  -0.110029   0.008245  -13.34   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.032 on 88 degrees of freedom
## Multiple R-squared:  0.6693, Adjusted R-squared:  0.6655
## F-statistic: 178.1 on 1 and 88 DF,  p-value: < 2.2e-16
```

β_0= 213.72

β_1= -0.11

```
plot(ozonets,type='o',ylab='Thickness of ozone layer (Dobson Units)', xlab='Year', ma
in ='Time Series Plot-Change in Ozone layer thickness - Figure 3' )
abline(ozonelm, col='red')
legend("bottomleft" ,lty=1, bty = "n", text.width = 10, col=c("black","red"),
       c("Thickness of Ozone Layer", "Linear Slope"))
```

**Time Series Plot-Change in Ozone layer thickness - Figure 3**
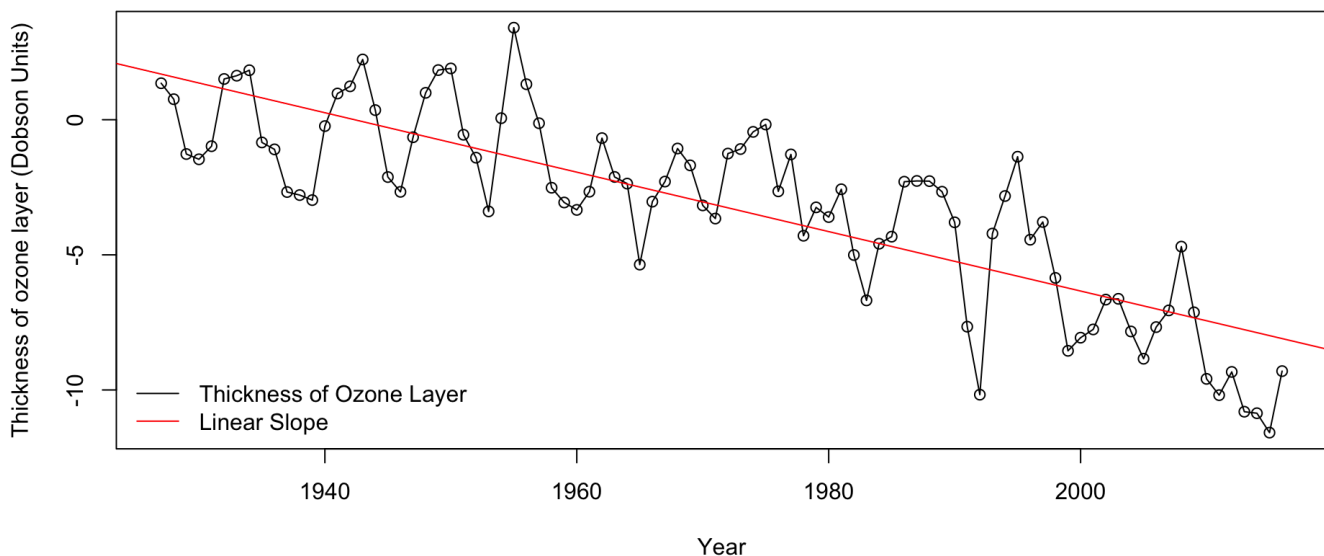
From the summary statistics we can observe that the intercept and slope are significant and the model is also significant as the p-value and Pr value is less than 0.05. With multiple R^2 value we can understand that around 67% of the variation in the series is explained by the linear trend. We prefer between 75%-85% of variation. Also the adjusted R^2 value and the multiple R^2 are approximately the same value. From the time series plot we can observe that overall the data follows a linear trend along the slope.

5.1.2 Residual Analysis

Residual Analysis is performed to validate the model. The estimator and predictor of observed stochastic component, X_t is called the residual corresponding to the tth observation. $X_t = Y_t - \mu_t$

5.1.2.1 Time Series Plot of Standardized Residuals

```
ozonelmresidual <- rstudent(ozonelm)
plot(y=ozonelmresidual,x=as.vector(time(ozonets)), xlab='Time', ylab='Standardized Re
siduals', type = 'o',main = "Time Series Plot of Residuals - Figure 4")
```


Time Series Plot of Residuals - Figure 4
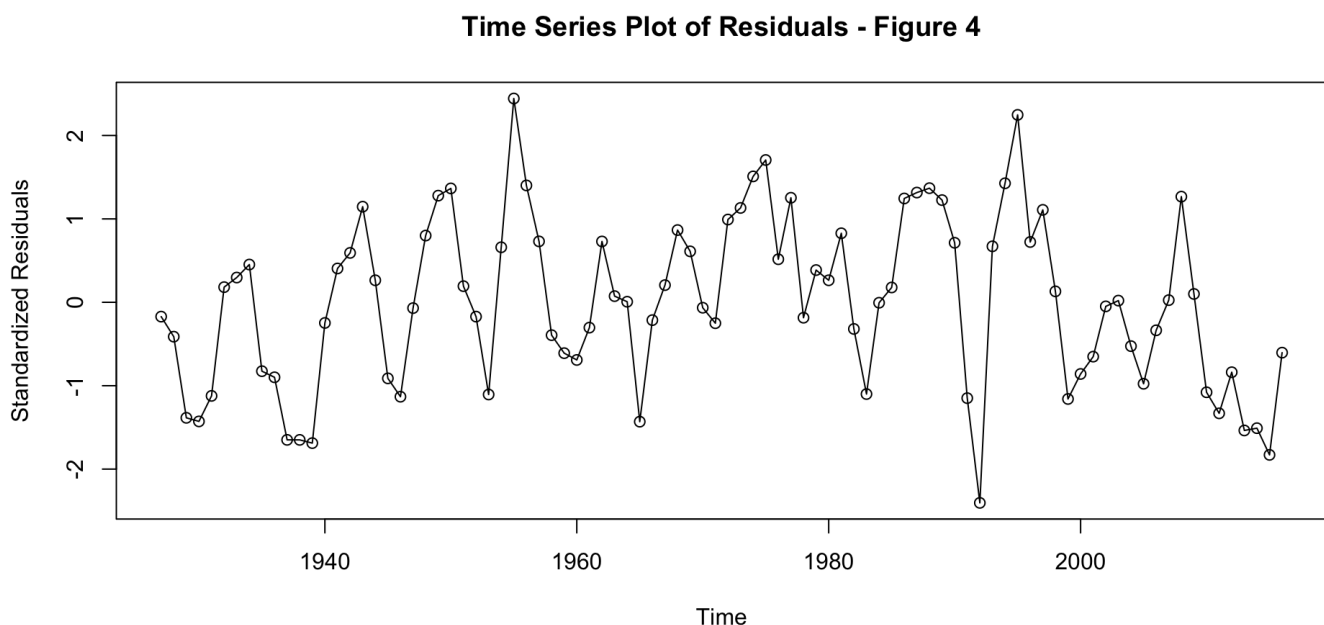
From Figure-4 we can observe that the residuals are around the mean (mean=0). Since most of the values are between 2 and -2 the residuals values are not big values.

5.1.2.2 Histogram of Standardized Residuals

The normality of the residual is checked by plotting histogram.

```
hist(ozonelmresidual,xlab='Standardized Residuals', main = "Histogram of the standard
ized residuals - Figure 5", freq = FALSE, ylim = c(0,0.5))
curve(dnorm(x, mean=mean(ozonelmresidual), sd=sd(ozonelmresidual)), col="red", lwd=2,
add=TRUE, yaxt="n")
```

Histogram of the standardized residuals - Figure 5
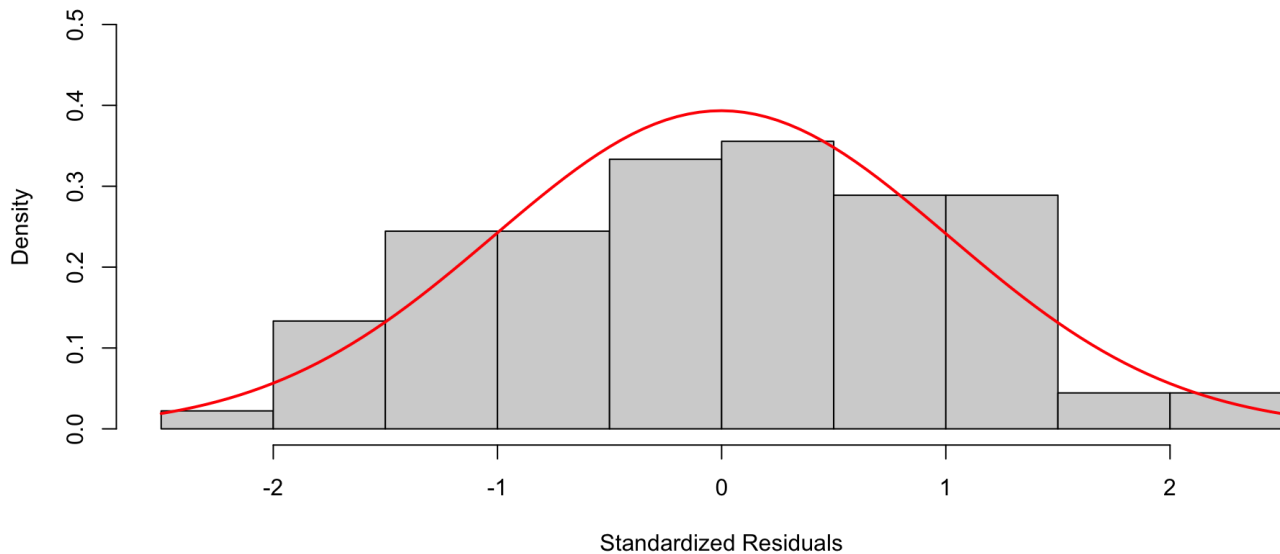
From Figure-5 we can observe that the distribution follows a normal curve approximately.

5.1.2.3 Q-Q Plot of Standardized Residuals

```
qqnorm(ozonelmresidual, main="Q-Q Plot for the Linear Model- Figure 6")
qqline(ozonelmresidual, col="red")
```

Q-Q Plot for the Linear Model- Figure 6
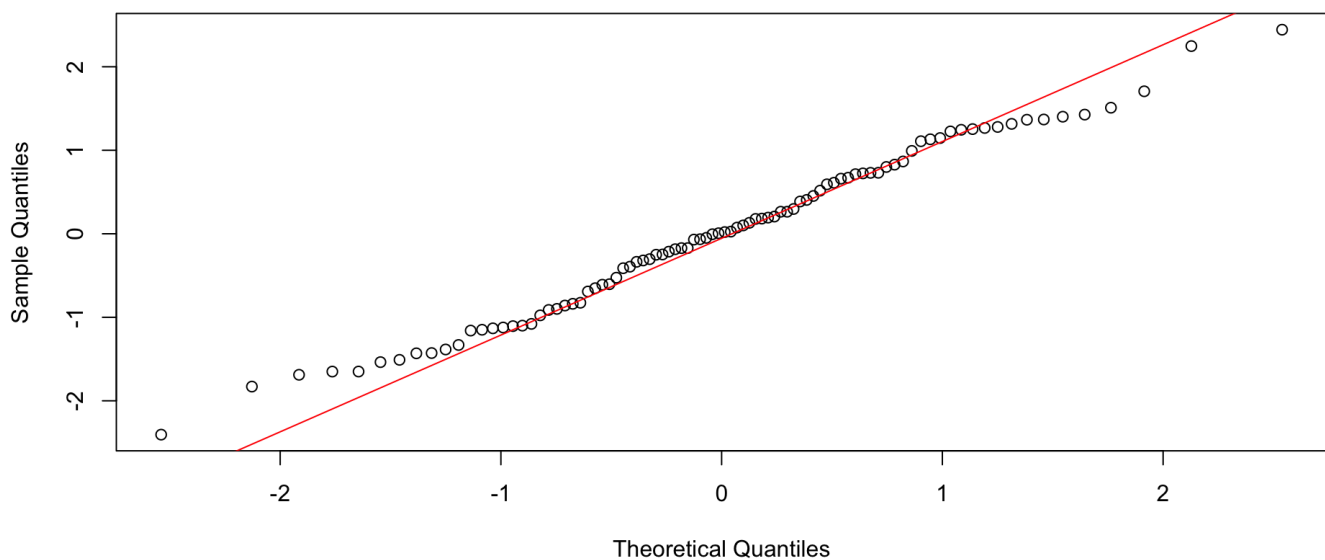
From Figure-6 we can observe that the data doesn't follow a normal distribution in the beginning and end as it isn't following a straight line pattern.

5.1.2.4 Shapiro- Wilk Test

We perform a hypothesis test called Shapiro Wilk test to check the normality assumption. Shapiro-Wilk test calculates the correlation between the residuals and corresponding normal quantiles.

```
shapiro.test(ozonelmresidual)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ozonelmresidual
## W = 0.98733, p-value = 0.5372
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis and conclude that the stochastic component of this model is normally distributed.

5.1.2.5 Auto-correlation Function Plot of Standardized Residuals

ACF plot is used to validate the correlation in the Standardized Residuals.

```
acf(ozonelmresidual,main="ACF of Standardized Residuals - Figure 7")
```

ACF of Standardized Residuals - Figure 7
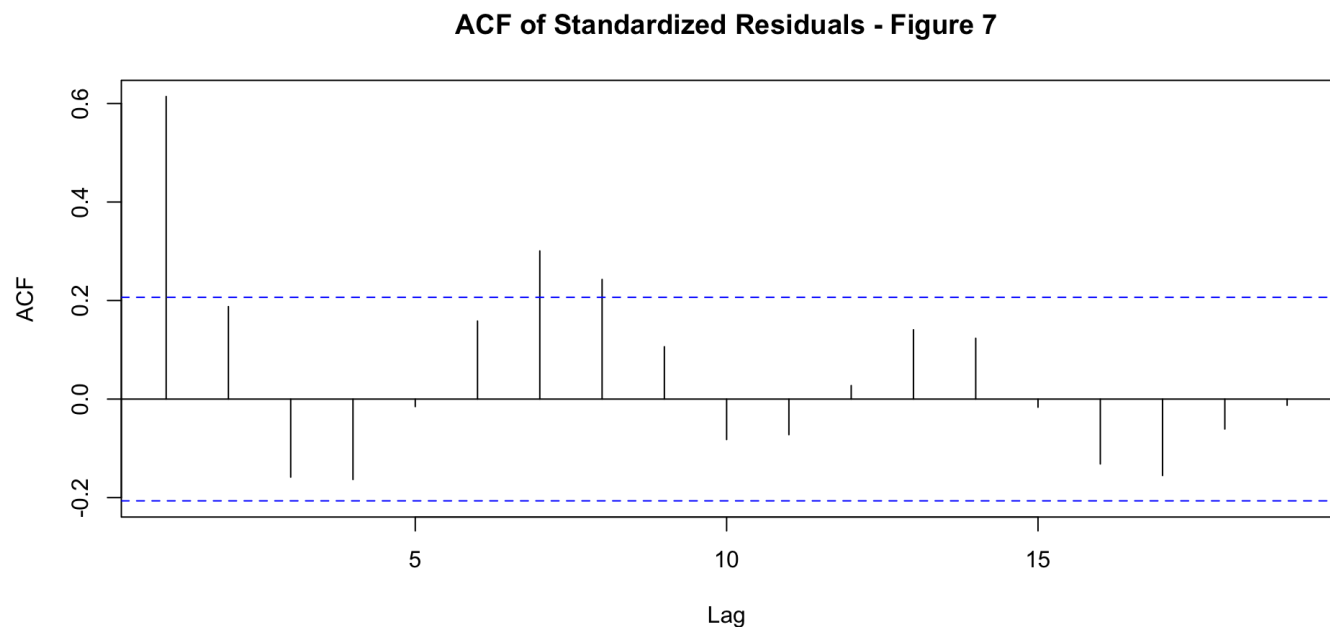
The correlation value is significantly higher than the confidence interval at one instance and sightly significant at another 2 instance considering different lags. This concludes that there is significant auto-correlation left with the residuals.

5.2 Quadratic Trend Model

The deterministic quadratic trend model is expressed as follows:

$\mu t = \beta_0 + \beta_1 t + \beta_2 t^2$

where β_0=*Intercept* β_1=*Slope of the linear trend* β_2=*Quadratic trend in time*

5.2.1 Summary Statistics and Plot

The slope, intercept of the linear trend and quadratic trend is estimated using least-square regression.

```
t = time(ozonets)
t2 = t^2
ozoneqm = lm(ozonets~ t + t2) # label the quadratic trend model as ozoneqm
summary(ozoneqm)
```

```
##
## Call:
## lm(formula = ozonets ~ t + t2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1062 -1.2846 -0.0055  1.3379  4.2325
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.733e+03  1.232e+03  -4.654 1.16e-05 ***
## t            5.924e+00  1.250e+00   4.739 8.30e-06 ***
## t2          -1.530e-03  3.170e-04  -4.827 5.87e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.815 on 87 degrees of freedom
## Multiple R-squared:  0.7391, Adjusted R-squared:  0.7331
## F-statistic: 123.3 on 2 and 87 DF,  p-value: < 2.2e-16
```

$\beta 0$= -5.733e+03

$\beta 1$= 5.924e+00

$\beta 2$= -1.530e-03

```
plot(ts(fitted(ozoneqm)), ylim=c(min(c(fitted(ozoneqm),                     as.v
ector(ozonets))),max(c(fitted(ozoneqm),as.vector(ozonets)))),col=c('red'), ylab='Thic
kness of ozone layer (Dobson Units)',
main = "Time Series plot with fitted Quadratic curve - Figure 8")
lines(as.vector(ozonets),type="o")
legend("bottomleft" ,lty=1, bty = "n", text.width = 10, col=c("black","red"),
       c("Thickness of Ozone Layer", "Quadratic Curve"))
```

**Time Series plot with fitted Quadratic curve - Figure 8**
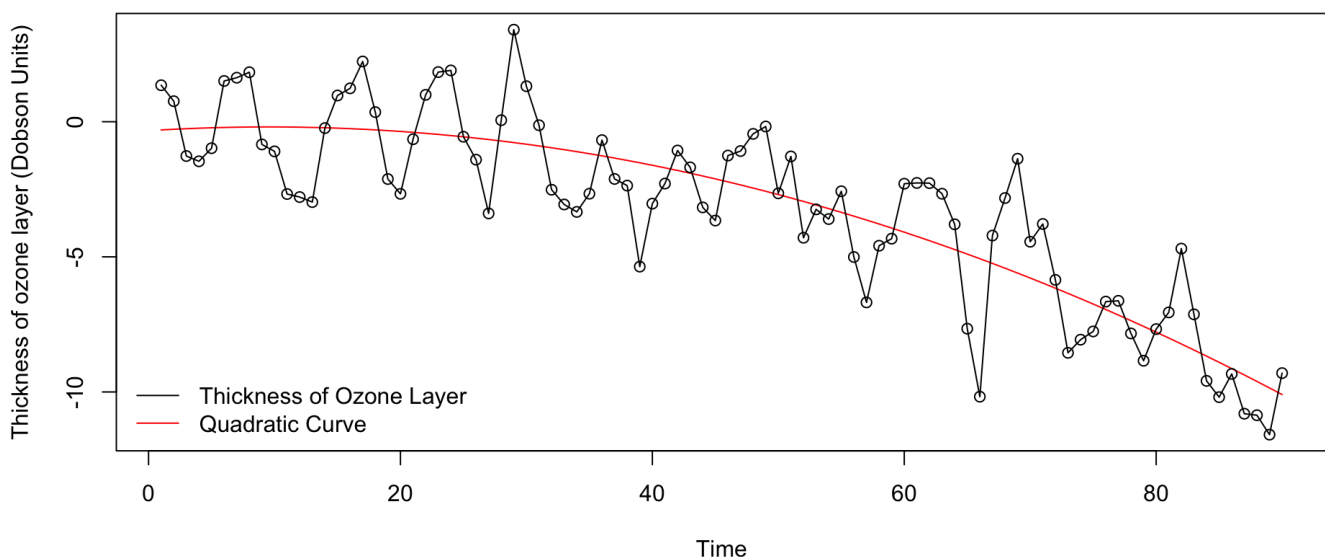
From the summary statistics we can observe that the model is significant as the p-value is less than 0.05 and hence we fail to reject the null hypothesis. The adjusted R-squared value explains 73.31% of the variation in the series and is almost in the preferred percent of variation. From the time series plot we can observe that the

time series data follows the quadratic trend.

5.2.2 Residual Analysis

5.2.2.1 Time Series Plot of Standardized Residuals

```
ozoneqm.res <- rstudent(ozoneqm)
plot(y=ozoneqm.res,x=as.vector(time(ozonets)), xlab='Time', ylab='Standardized Residu
als', type = 'o',main = "Time Series Plot of Residuals - Figure 9")
```

Time Series Plot of Residuals - Figure 9
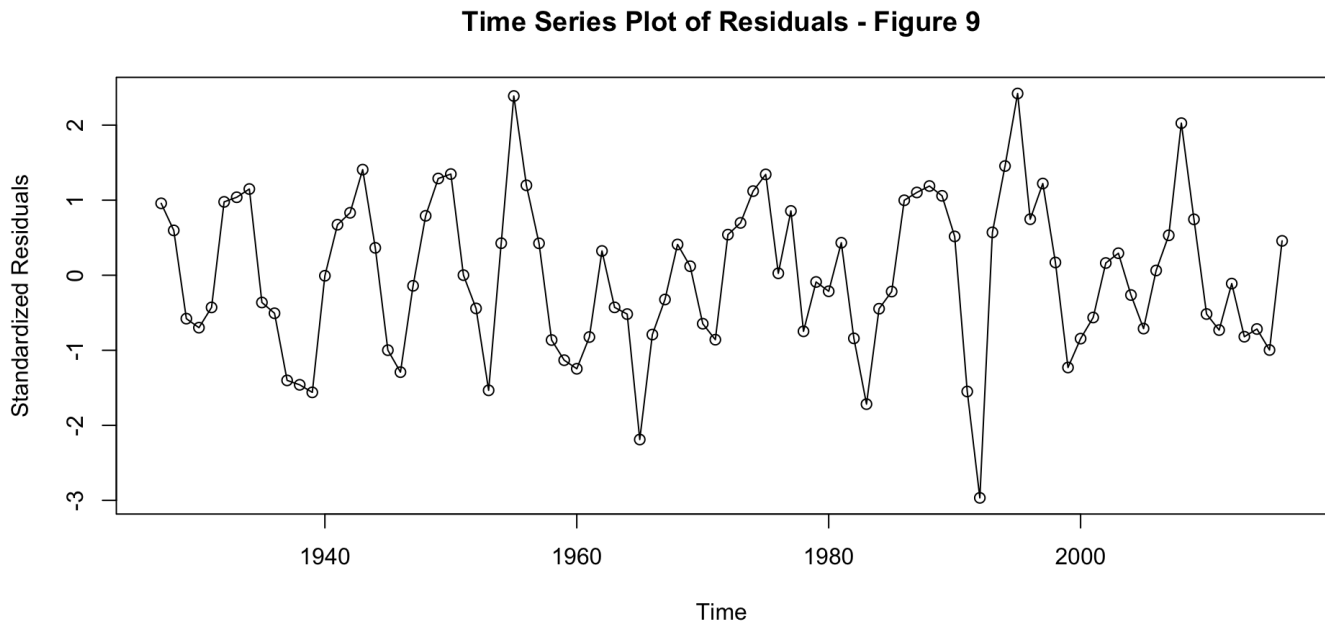
From Figure 9 we observe that the residuals are present around the mean of the standardized residuals.

5.2.2.2 Histogram of Standardized Residuals

The normality of the residuals are plotted.

```
hist(ozoneqm.res,xlab='Standardized Residuals', main = "Histogram of the standardized
residuals - Figure 10", freq = FALSE, ylim = c(0,0.5))
curve(dnorm(x, mean=mean(ozoneqm.res), sd=sd(ozoneqm.res)), col="red", lwd=2, add=TRU
E, yaxt="n")
```

Histogram of the standardized residuals - Figure 10
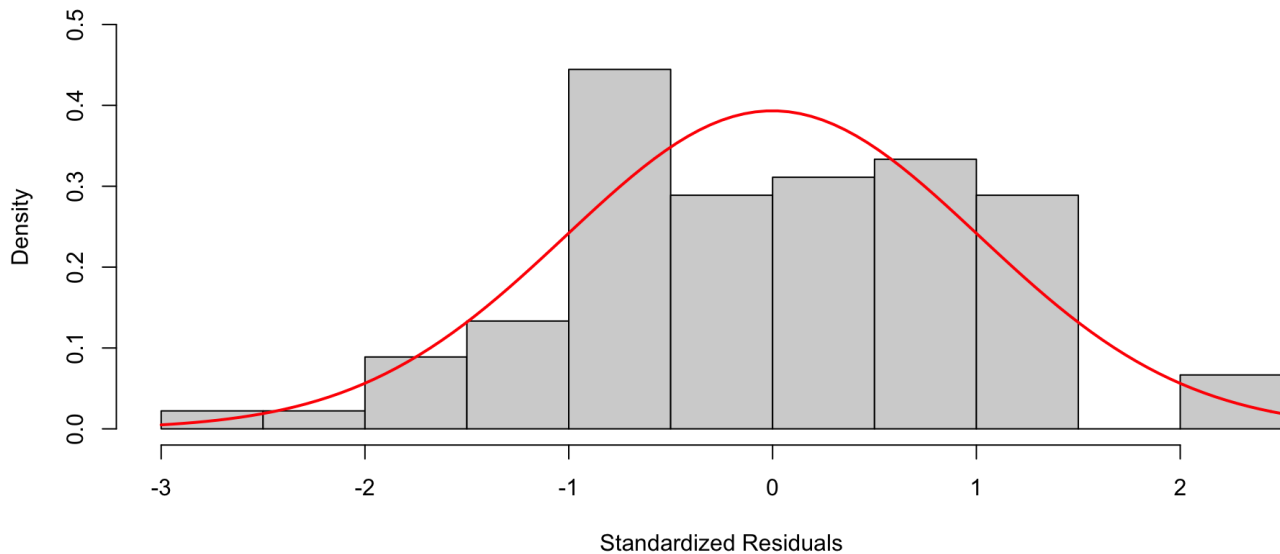
From Figure-10 we can observe that the distribution follows a normal curve approximately with slight left skewness.

5.2.2.3 Q-Q Plot of Standardized Residuals

```
qqnorm(ozoneqm.res, main="Q-Q Plot for the Linear Model- Figure 11")
qqline(ozoneqm.res, col="red")
```

Q-Q Plot for the Linear Model- Figure 11
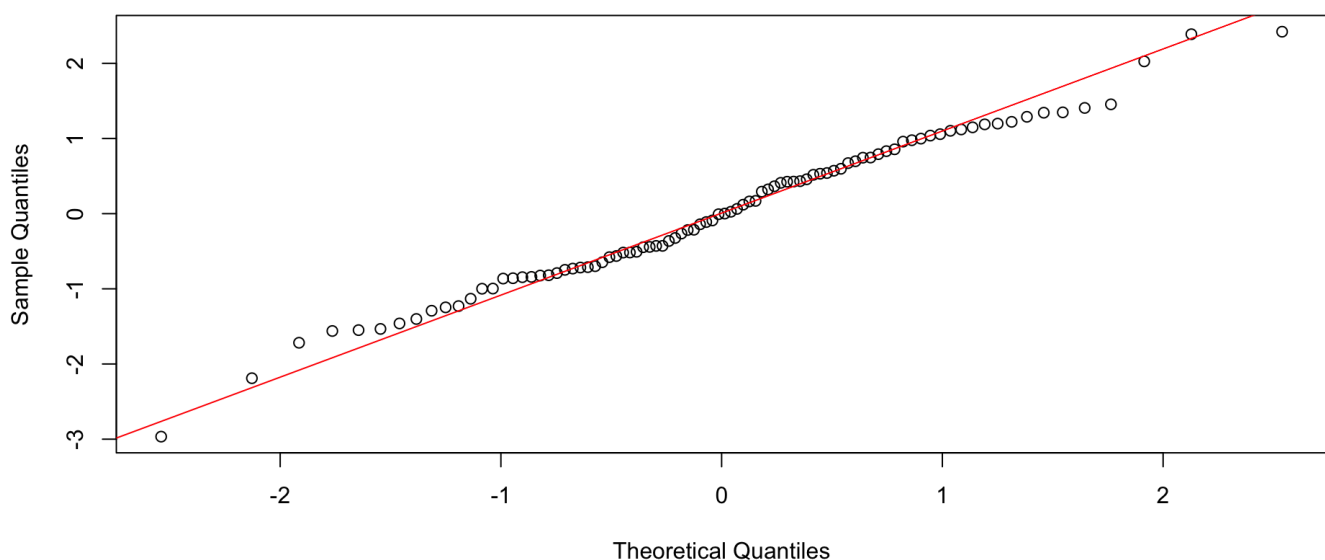
From the QQ plot(Figure 11) we observe that the data follows a normal distribution and follows the straight line with a few observation out of the line with small residual value.

5.2.2.4 Shapiro- Wilk Test

We perform a hypothesis test called Shapiro Wilk test to check the normality assumption. Shapiro-Wilk test calculates the correlation between the residuals and corresponding normal quantiles.

```
shapiro.test(ozoneqm.res)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ozoneqm.res
## W = 0.98889, p-value = 0.6493
```

Since the p-value is greater than 0.05 we fail to reject the null hypothesis and conclude that the stochastic component of this model is normally distributed.

5.2.2.5 Auto-correlation Function Plot of Standardized Residuals

```
acf(ozoneqm.res,main="ACF of Standardized Residuals - Figure 12")
```

ACF of Standardized Residuals - Figure 12
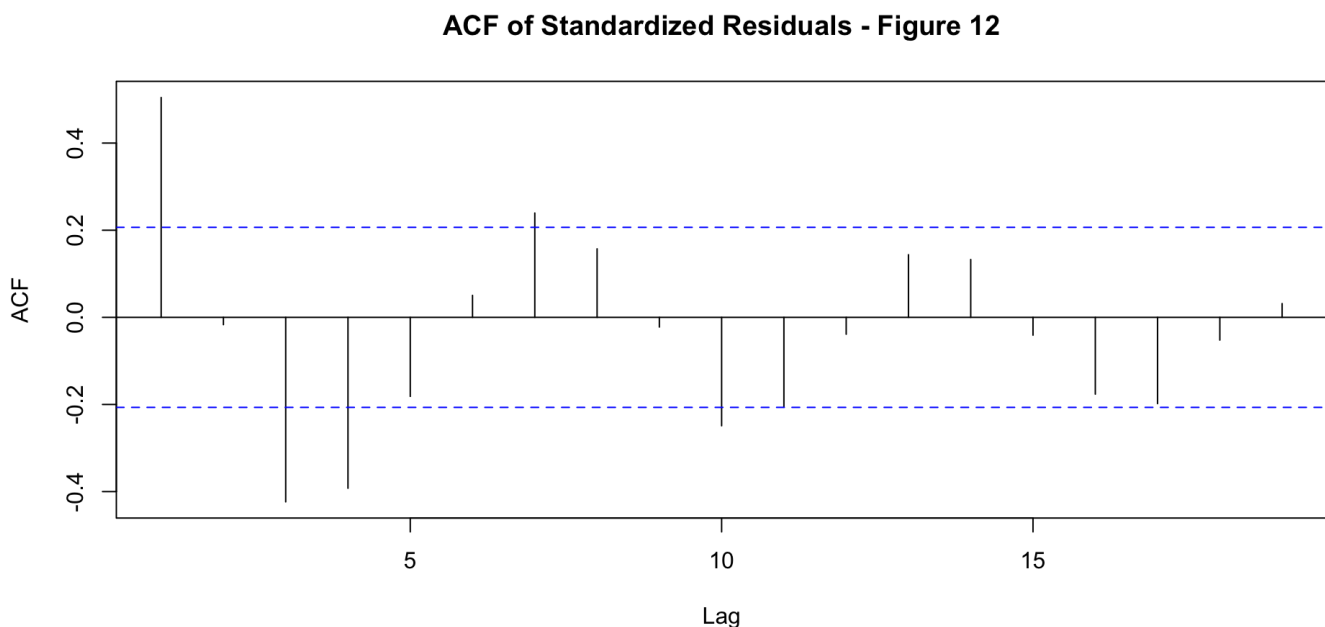
The correlation value is significantly higher than the confidence interval at several lags.

5.3 Harmonic/Cosine Trend Model

The time series data provided is yearly reading and hence we don't find any seasonality in the time series object. The model is tested for fit with the harmonic trend despite no seasonality.

5.3.1 Summary Statistics

For getting the summary statistics we use the harmonic() function, where we need to add the time series object and an m value which is the number of pairs of harmonic functions to be created; 2m must be less than or equal to the frequency of x. Since here our frequency is 1 we consider m to be any value below 0.5.

```
harm.=harmonic(ozonets,m=0.4) # calculate cos(2*pi*t) and sin(2*pi*t)
ozonehm=lm(ozonets~harm.)
summary(ozonehm)
```

```
##
## Call:
## lm(formula = ozonets ~ harm.)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -8.3520  -1.8906   0.4837   2.3643   6.4248
##
## Coefficients: (1 not defined because of singularities)
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.970e+00  4.790e-01  -6.199 1.79e-08 ***
## harm.cos(2*pi*t)        NA         NA      NA       NA
## harm.sin(2*pi*t)  5.462e+11  7.105e+11   0.769    0.444
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.522 on 88 degrees of freedom
## Multiple R-squared:  0.006672,   Adjusted R-squared:  -0.004616
## F-statistic: 0.5911 on 1 and 88 DF,  p-value: 0.4441
```

From the summary statistics we find that the p-value is greater than alpha, 0.05 and hence reject the hypothesis, .i.e. the harmonic model fit is insignificant. Also the R^2 is negative.

5.3.2 Residual Analysis

5.3.2.1 Time Series Plot of Standardized Residuals

```
ozonehm.res<-rstudent(ozonehm)
plot(ozonehm.res,x=as.vector(time(ozonets)),xlab='Time', ylab='Standardized Residual
s',type='o',main="Time series Plot of Residuals - Figure 13")
```
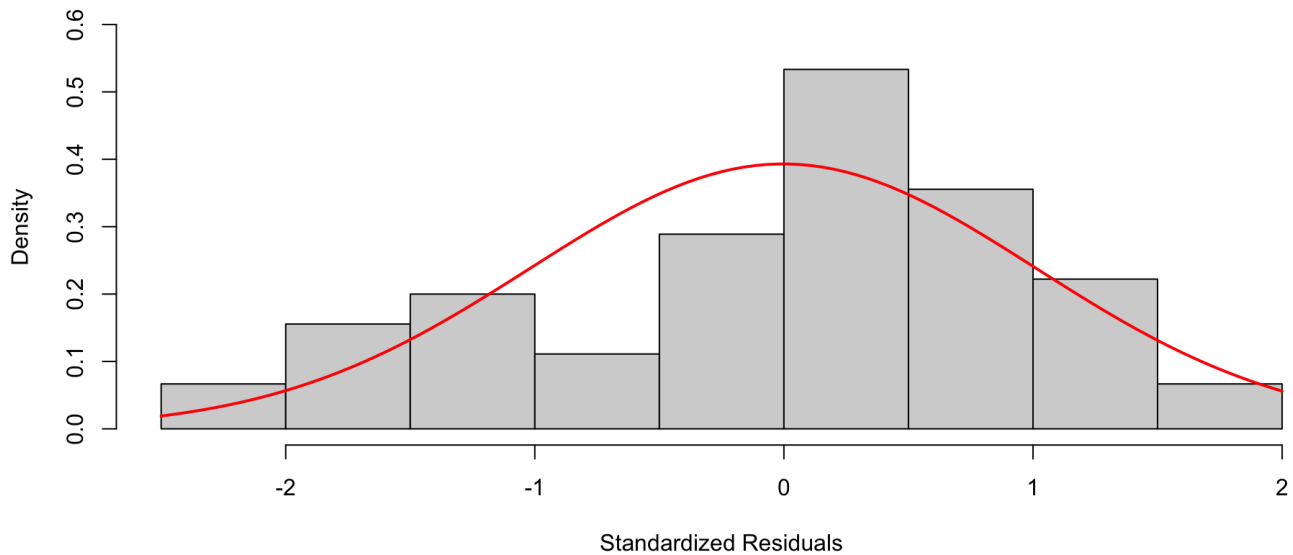

Time series Plot of Residuals - Figure 13
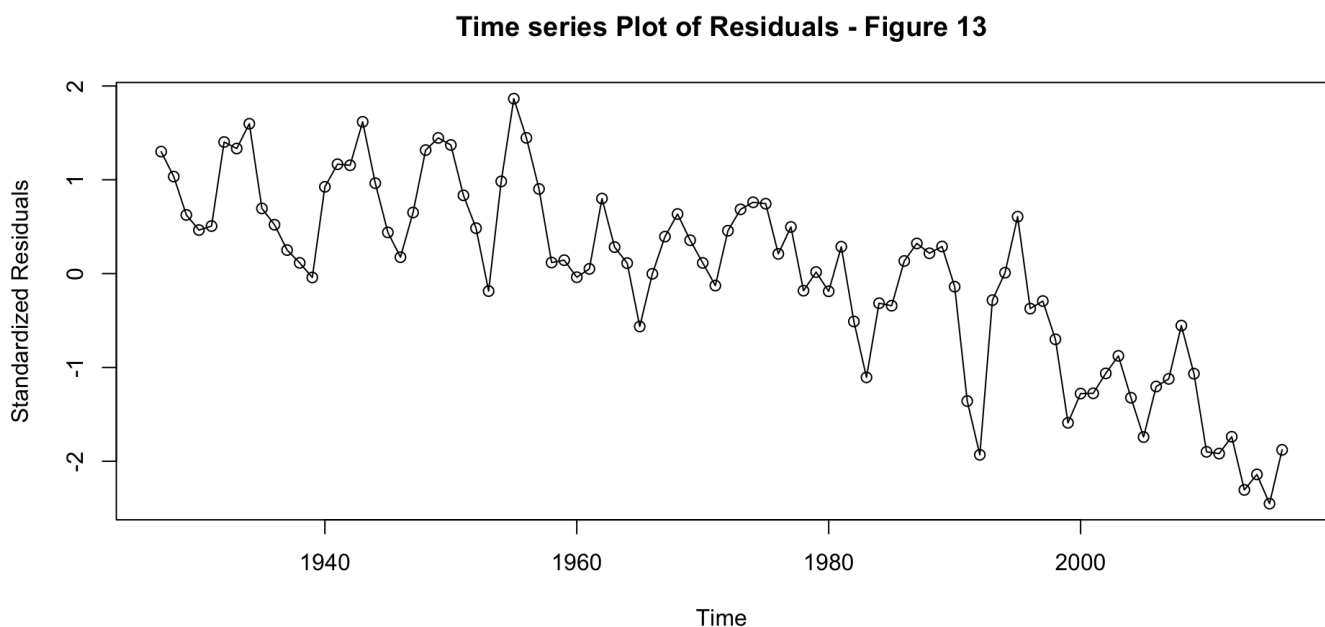
From figure 13 we can observe that the residuals are present around the mean of the standardized residuals and is have a downward trend.

5.3.2.2 Histogram of Standardized Residuals

The normality of the residual are plotted as histogram.

```
hist(ozonehm.res,xlab='Standardized Residuals', main = "Histogram of the standardized
residuals - Figure 14", freq = FALSE, ylim = c(0,0.6))
curve(dnorm(x, mean=mean(ozonehm.res), sd=sd(ozonehm.res)), col="red", lwd=2, add=TRU
E, yaxt="n")
```
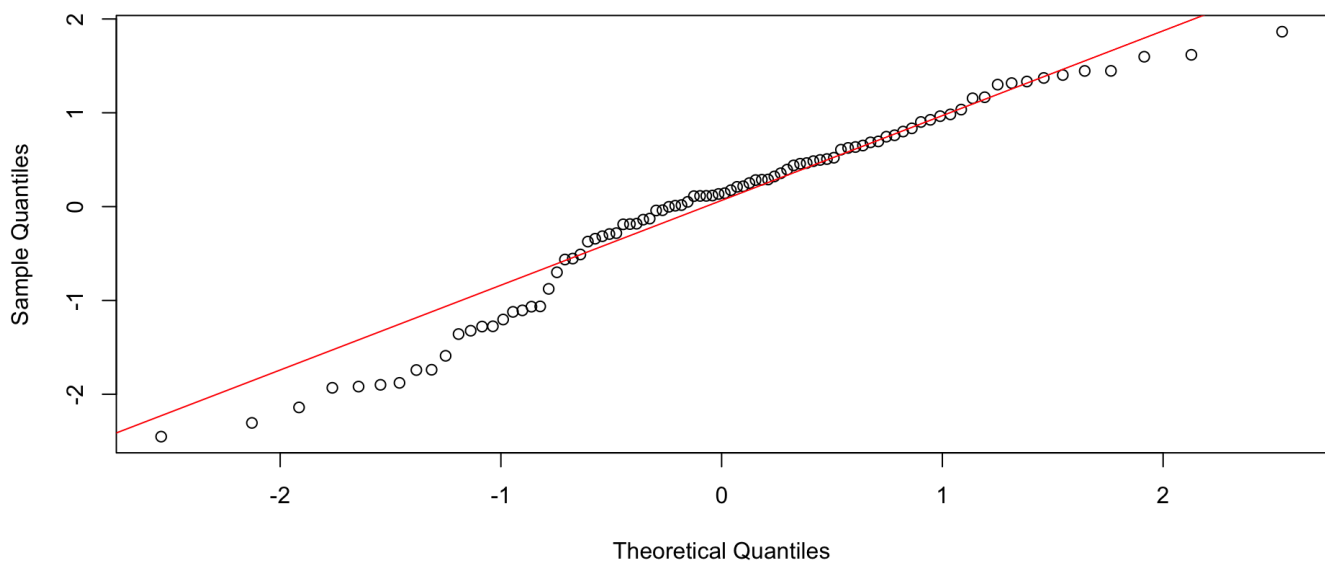
Histogram of the standardized residuals - Figure 14

5.3.2.3 Q-Q Plot of Standardized Residuals

```
qqnorm(ozonehm.res, main="Q-Q Plot for the Linear Model- Figure 15")
qqline(ozonehm.res, col="red")
```

Q-Q Plot for the Linear Model- Figure 15

From the Q-Q plot we can observe that the time series data doesn't follow the straight line and hence not following normality.

5.3.2.4 Shapiro-Wilk Test

```
shapiro.test(rstudent(ozonehm))
```

```
##
##   Shapiro-Wilk normality test
##
## data:  rstudent(ozonehm)
## W = 0.95856, p-value = 0.005875
```

The p-value is less than the alpha significance and hence we reject the null hypothesis and conclude that the time series is not normally distributed.

5.3.2.5 Auto-correlation Function Plot of Standard Residuals

```
acf(ozonehm.res,main="ACF of Standardized Residuals - Figure 16")
```

### ACF of Standardized Residuals - Figure 16
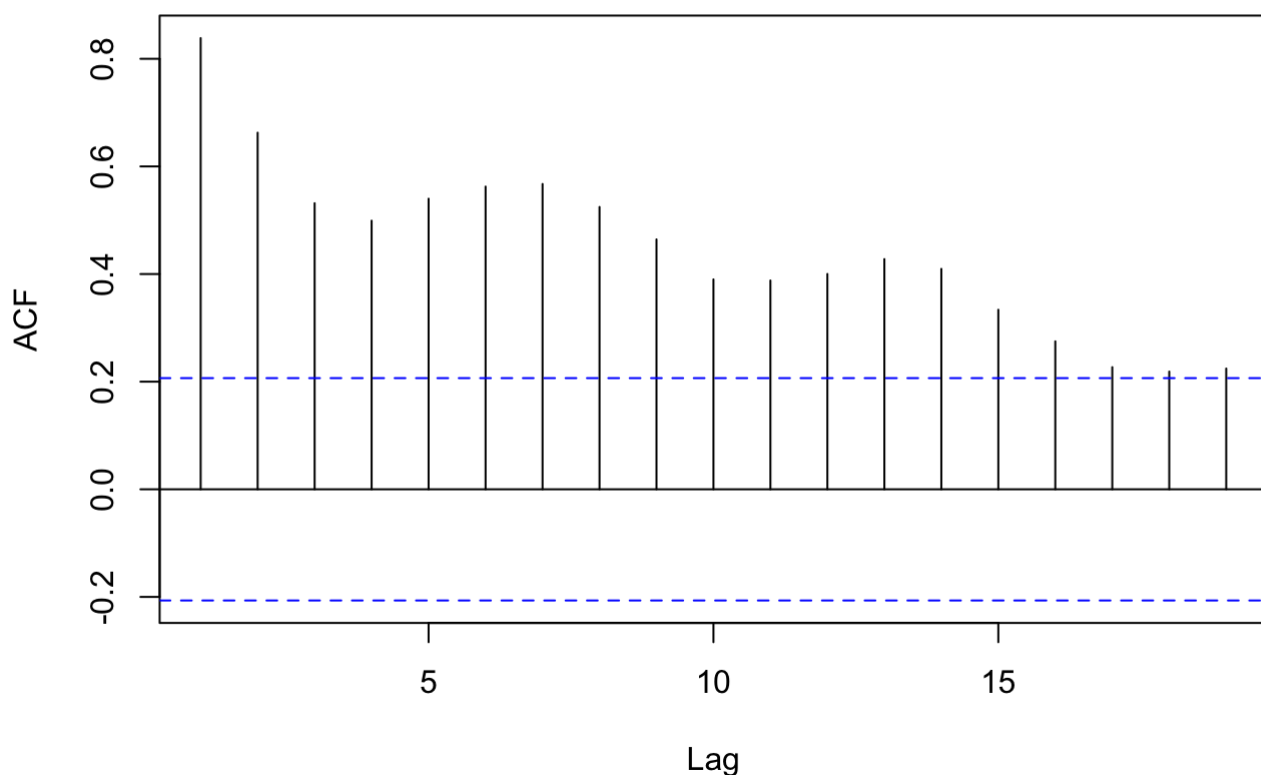
The correlation value is significantly higher at almost all points than the confidence interval and is also having slow decaying pattern showing that the time series is non-stationary in nature which will be further analyzed in other chapters.

5.4 Cyclic or Seasonal Trend

We can't use cyclic or seasonal trend to fit the model as it has to be a time series with frequency greater than 1 and since we are having a yearly time series data we don't use cyclic or seasonal trend model. But by analyzing the ACF of time series dataset we can find a possible frequency for the seasonality and plot the seasonal trend for annual series.

```
acf(ozonets, main="ACF of Time series - Figure 17")
```

ACF of Time series - Figure 17
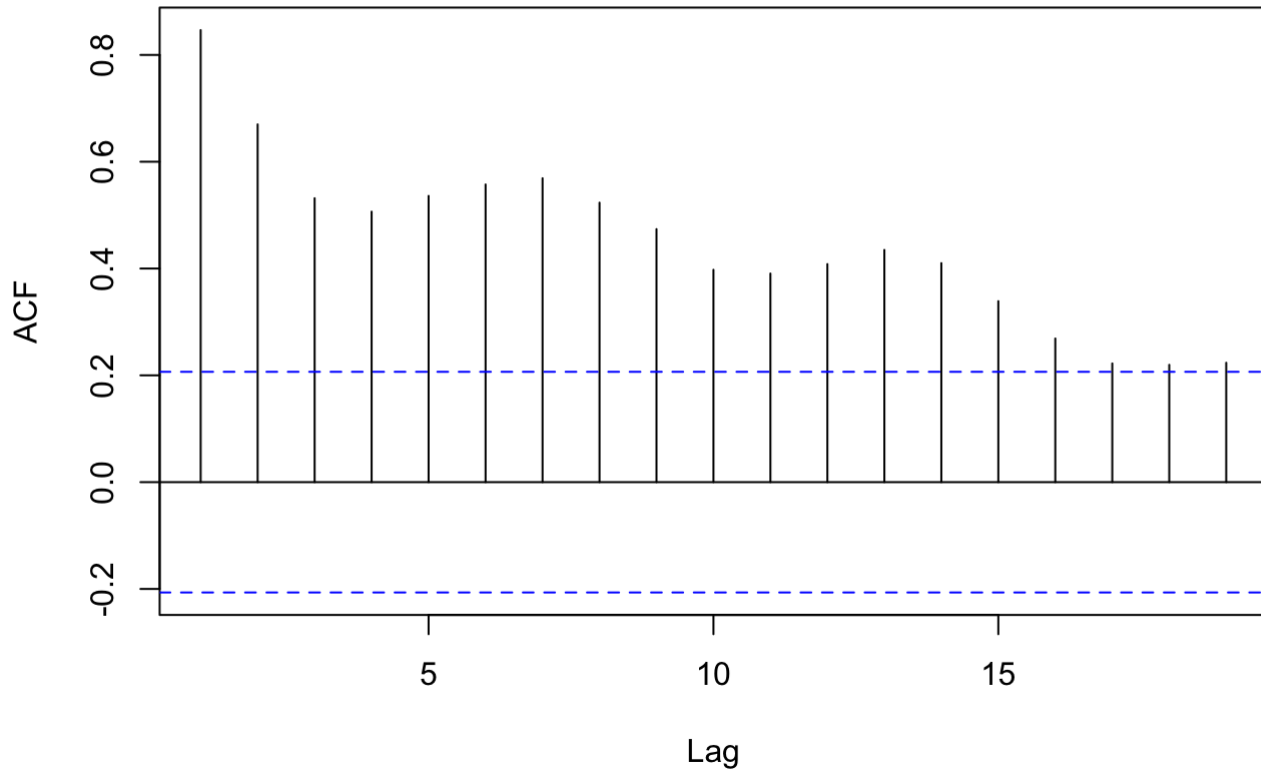
From the ACF plot we can find a possible frequency for the time series dataset. In the above ACF we consider the wave pattern and count the number frequency between two high frequency line. Here the frequency is assumed as 3.

5.4.1 Plot and Summary Statistics

```
ozone1ts<-ts(ozone,start=c(1927,1),end=c(2016,3),frequency = 3)
plot(ozone1ts,lwd=2,ylab="Ozone layer thickness",main="Time Series Plot with Seasonal
ity - Figure 18")
```

Time Series Plot with Seasonality - Figure 18
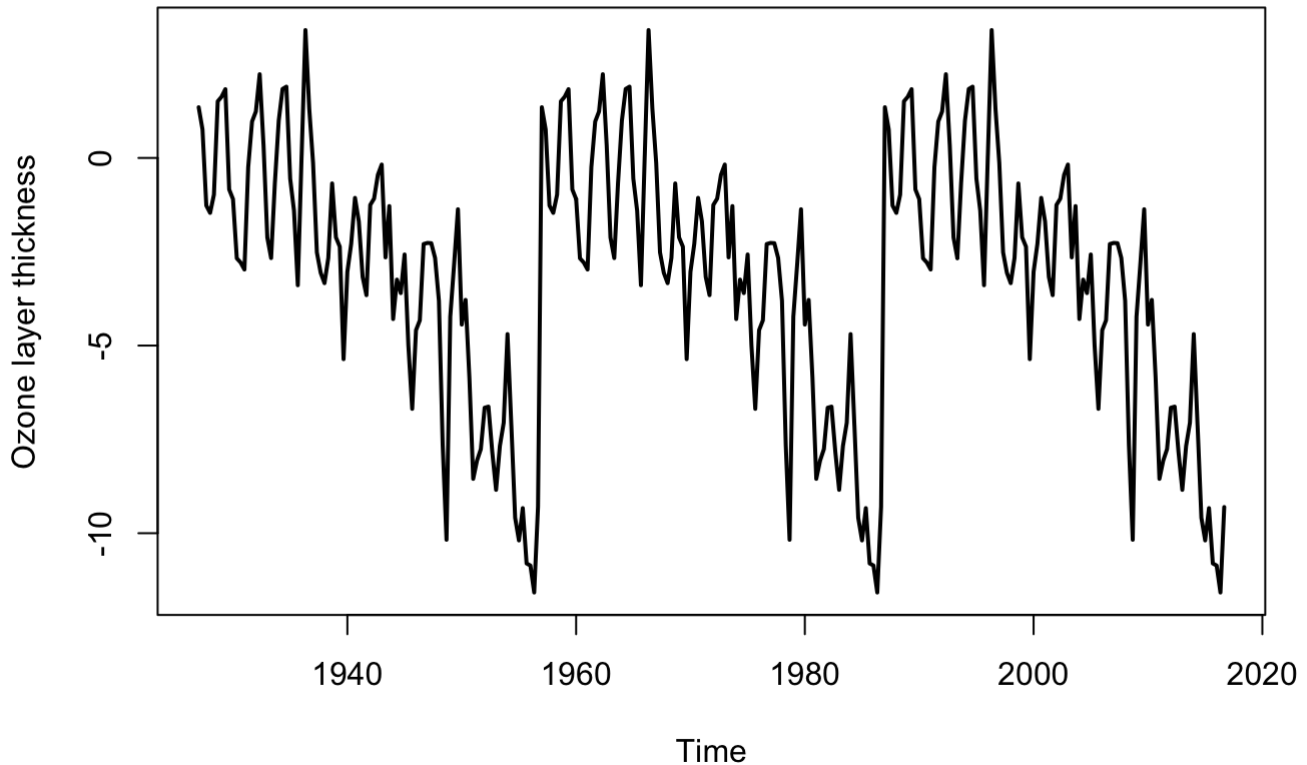
We can observe the seasonal trend from the plot.

```
freq.=season(ozone1ts)
ozonesm=lm(ozone1ts~freq.)
summary(ozonesm)
```

```
##
## Call:
## lm(formula = ozone1ts ~ freq.)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.4413 -1.8678  0.4816  2.4662  6.5452
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -3.0211     0.3699  -8.168 1.26e-14 ***
## freq.Season-2  -0.1170     0.5231  -0.224    0.823
## freq.Season-3  -0.4266     0.5231  -0.815    0.416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.509 on 267 degrees of freedom
## Multiple R-squared:  0.002653,   Adjusted R-squared:  -0.004818
## F-statistic: 0.3551 on 2 and 267 DF,  p-value: 0.7014
```

From the summary we can observe that the p-value of the overall model and the seasons is greater than the alpha value of significance, 0.05 and both the multiple R^2 and adjusted R^2 are very low and insignificant.

5.4.2 Residual Analysis

5.4.2.1 Time Series Plot of Standardized Residuals

```
ozonesm.res<-rstudent(ozonesm)
plot(ozonesm.res,x=as.vector(time(ozone1ts)),xlab='Time', ylab='Standardized Residual
s',type='o',main="Time series Plot of Residuals - 19")
```

Time series Plot of Residuals - 19
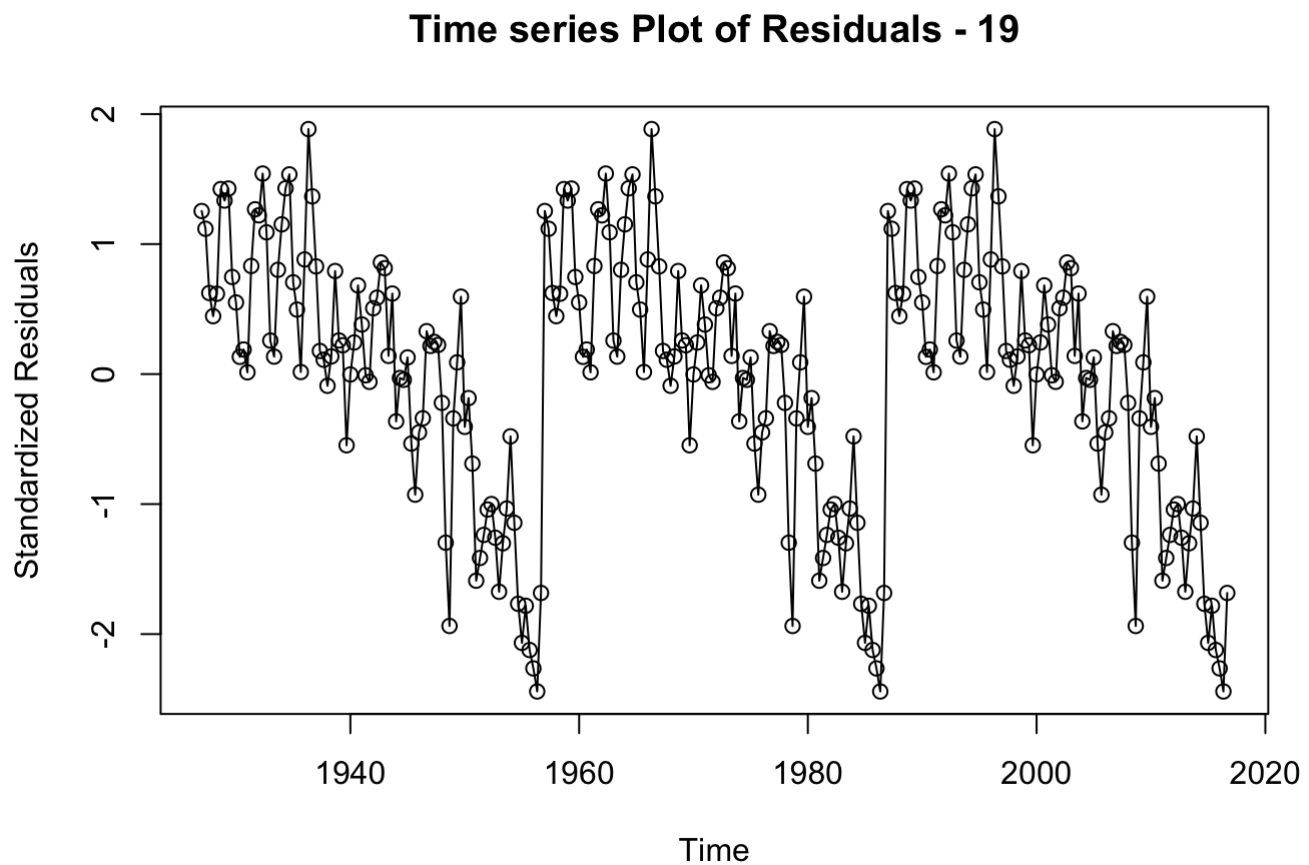
From Figure 19 the residuals follows a seasonality pattern and is found to be around 0.

5.4.2.2 Histogram of Standardized Residuals

```
hist(ozonesm.res,xlab='Standardized Residuals', main = "Histogram of the standardized
residuals - Figure 20", freq = FALSE, ylim = c(0,0.5))
curve(dnorm(x, mean=mean(ozonehm.res), sd=sd(ozonehm.res)), col="red", lwd=2, add=TRU
E, yaxt="n")
```

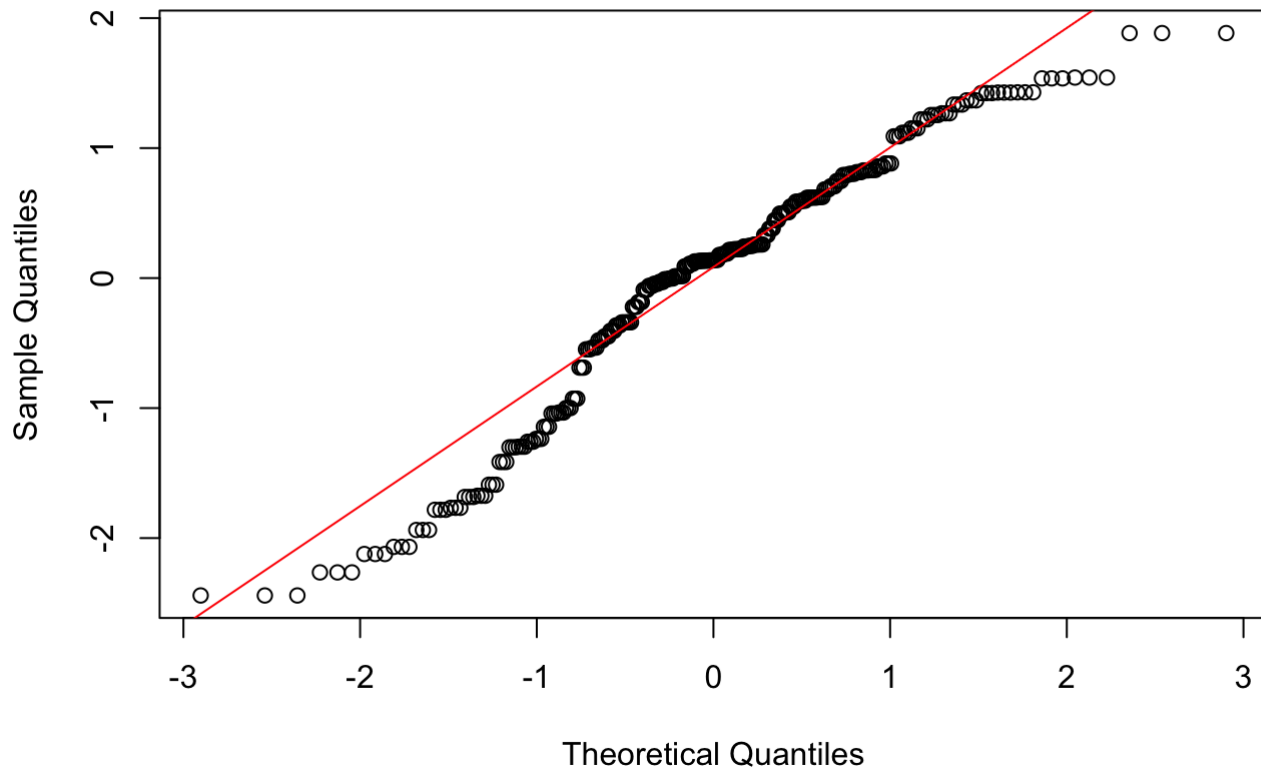## Histogram of the standardized residuals - Figure 20
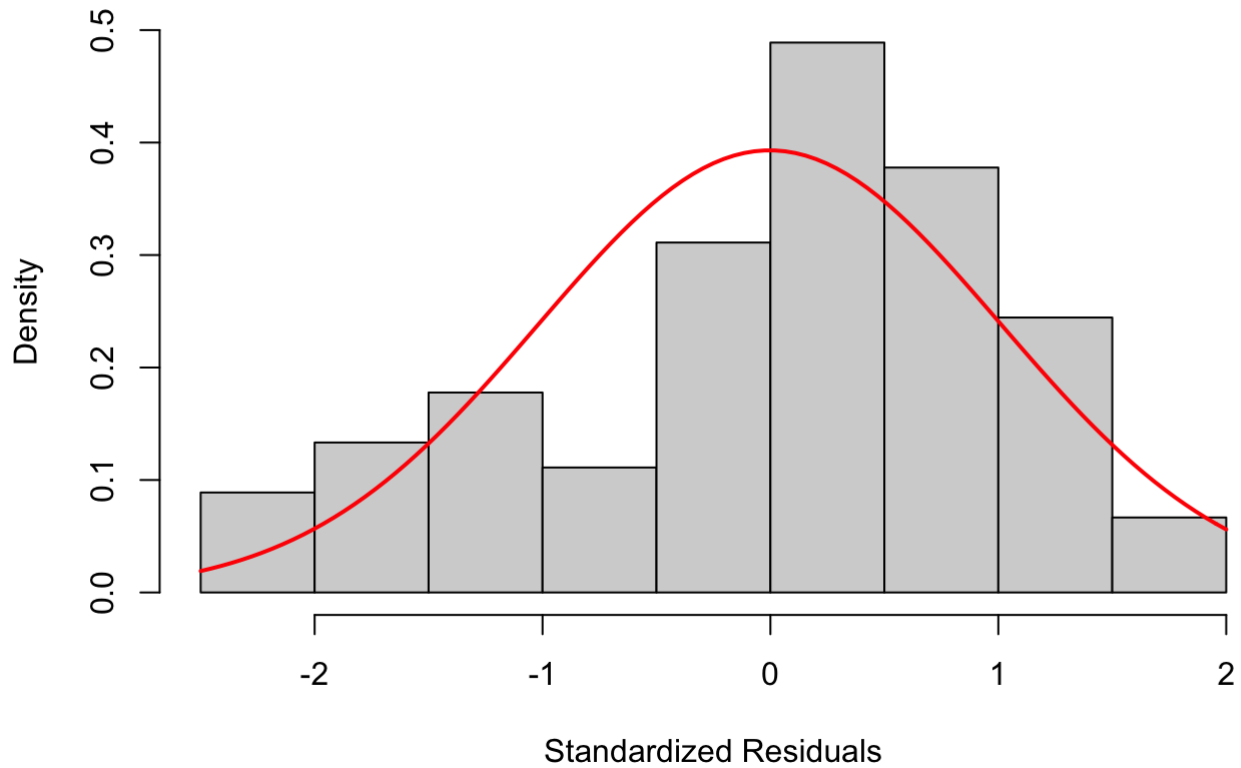
From the above histogram we can observe that the plot is not following normality.

5.4.2.3 Q-Q Plot of Standardized Residuals

```
qqnorm(ozonesm.res, main="Q-Q Plot for the Linear Model- Figure 21")
qqline(ozonesm.res, col="red")
```

Q-Q Plot for the Linear Model- Figure 21

From the figure 21 we observe that the data doesn't follow the trend and hence is not normally distributed.

5.4.2.4 Shapiro-Wilk Test

```
shapiro.test(ozonesm.res)
```

```
##
##   Shapiro-Wilk normality test
##
## data:  ozonesm.res
## W = 0.95719, p-value = 3.814e-07
```

The p-value is less than the significant alpha, 0.05 and hence we reject null hypothesis and conclude that the data is not normally distributed.

5.4.2.5 Auto-correlation Function of Standardized Residuals

```
acf(ozonesm.res,main="ACF of Standardized Residuals – Figure 22")
```

ACF of Standardized Residuals - Figure 22
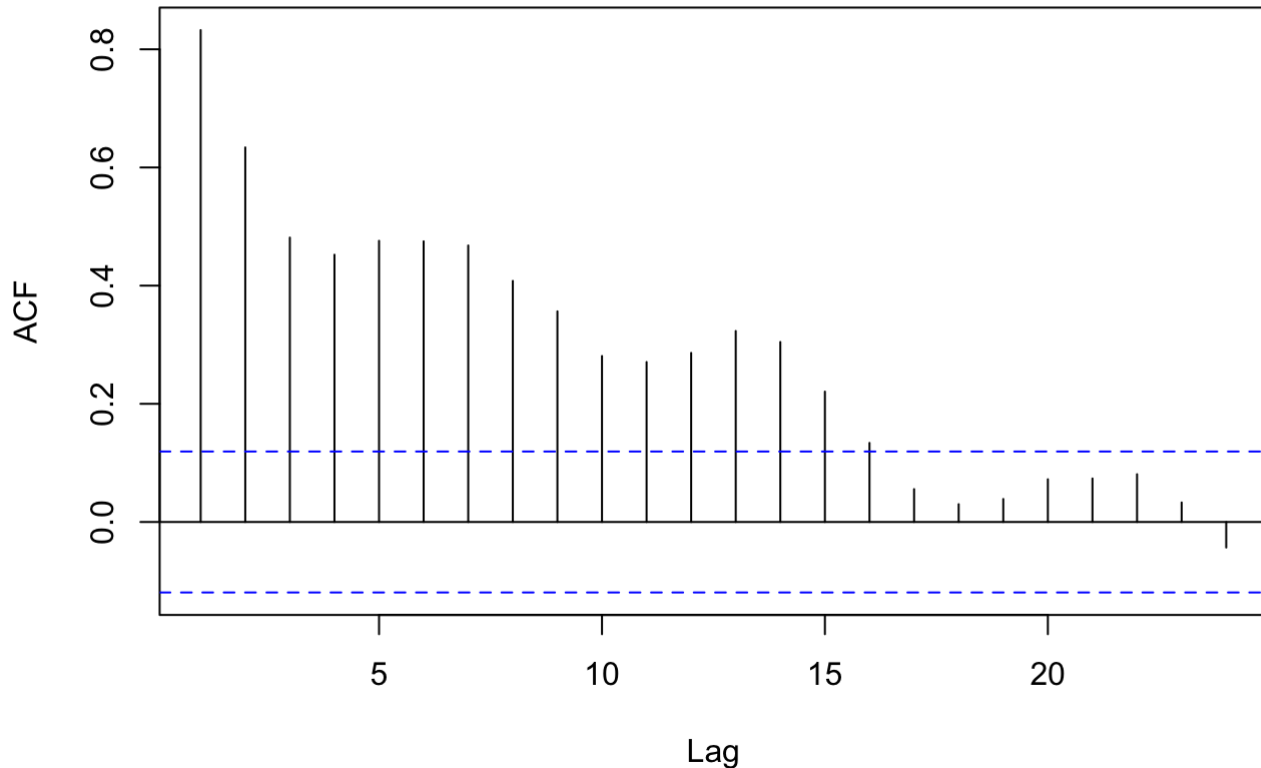
The ACF plot has slow decaying pattern and correlation value is significantly high in most of the lags.

6 Conclusion on Model Fitting

- By observing our analysis done on various model fitting we can narrow down the model to be either linear trend or quadratic trend model.
- With the residual analysis results, we can conclude that quadratic trend model had good fit over linear trend model.
- The coefficient of determination namely adjusted R^2 is larger in case of quadratic trend model which describes the variation in the time series.
- Although we conclude that quadratic trend is good fit, it is not the best fit. To find best fit we could further analyze the time series data with ARIMA models as the time series has both AR and MA behavior (Analysis done in coming chapters).

7 Forecast - 5 years Prediction

- Post finding the fit for model we predict the value of ozone thickness for the next 5 years.
- We use the quadratic trend model fit for forecasting.
- The forecasting is done with forecast limit of 5% and is expected to be with the range.

```
t=time(ozonets)
t2=t^2
ozonef = lm(ozonets~t+t2)
Year = c(2017, 2018, 2019, 2020, 2021)
ozone1 = data.frame(t = Year, t2 = Year^2, 1)
forecast = predict(ozonef,ozone1, interval = "prediction")
forecast
```

```
##           fit        lwr        upr
## 1 -10.34387 -14.13556 -6.552180
## 2 -10.59469 -14.40282 -6.786548
## 3 -10.84856 -14.67434 -7.022786
## 4 -11.10550 -14.95015 -7.260851
## 5 -11.36550 -15.23030 -7.500701
```

```
plot(ozonets,type='o',xlim = c(1927,2021),ylim=c(-15,5),xlab="Year",ylab='Thickness o
f Ozone Layer(Dobson Units)',main="Time series Plot - Change in Ozone Layer Thickness
- Figure 23" )
#We convert forecast to time series object
lines(ts(as.vector(forecast[,1]),start = 2017),col="red",type="l")
lines(ts(as.vector(forecast[,2]),start = 2017),col="blue",type="l")
lines(ts(as.vector(forecast[,3]),start = 2017),col="blue",type="l")
legend("bottomleft", lty=1, bty = "n",pch=1, col=c("black","blue","red"), text.width
=12,
       c("Data","5% Forecast limits", "Forecasts"))
```

Time series Plot - Change in Ozone Layer Thickness - Figure 23
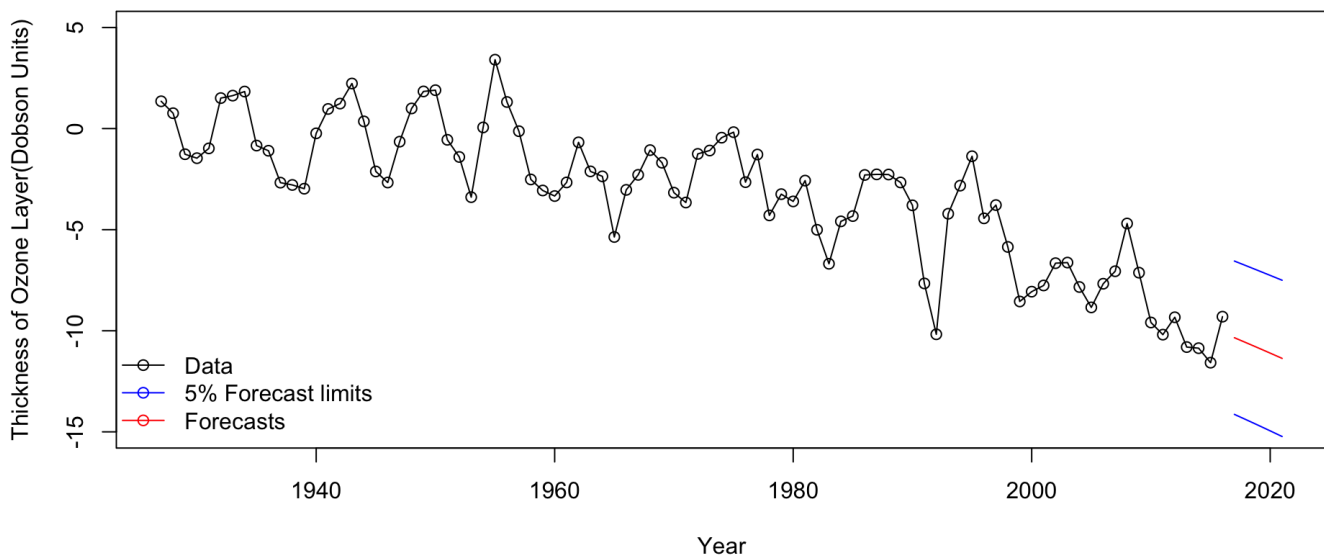
The prediction from the data and plot is that the thickness of Ozone layer has a downward decreasing trend in future which is matter of concern for our environment.

8 ARIMA Model

8.1 ADF Test

In ADF we test the null hypothesis that the series is non-stationary and alternate hypothesis that the series is stationary.

```
adf.test(ozonets,alternative = c("stationary"))
```

```
##
##   Augmented Dickey-Fuller Test
##
## data:   ozonets
## Dickey-Fuller = -3.2376, Lag order = 4, p-value = 0.0867
## alternative hypothesis: stationary
```

From ADF Test we can observe that the p-value is closer to the significance level of 0.05. Hence we change the k value in the ADF Test from default value. The k value is taken 1 less and 1 greater than the default value, 4.

```
adf.test(ozonets,k=3,alternative = c("stationary"))
```

```
##
##   Augmented Dickey-Fuller Test
##
## data:   ozonets
## Dickey-Fuller = -3.9376, Lag order = 3, p-value = 0.01596
## alternative hypothesis: stationary
```

```
adf.test(ozonets,k=5,alternative = c("stationary"))
```

```
##
##   Augmented Dickey-Fuller Test
##
## data:   ozonets
## Dickey-Fuller = -2.5704, Lag order = 5, p-value = 0.3413
## alternative hypothesis: stationary
```
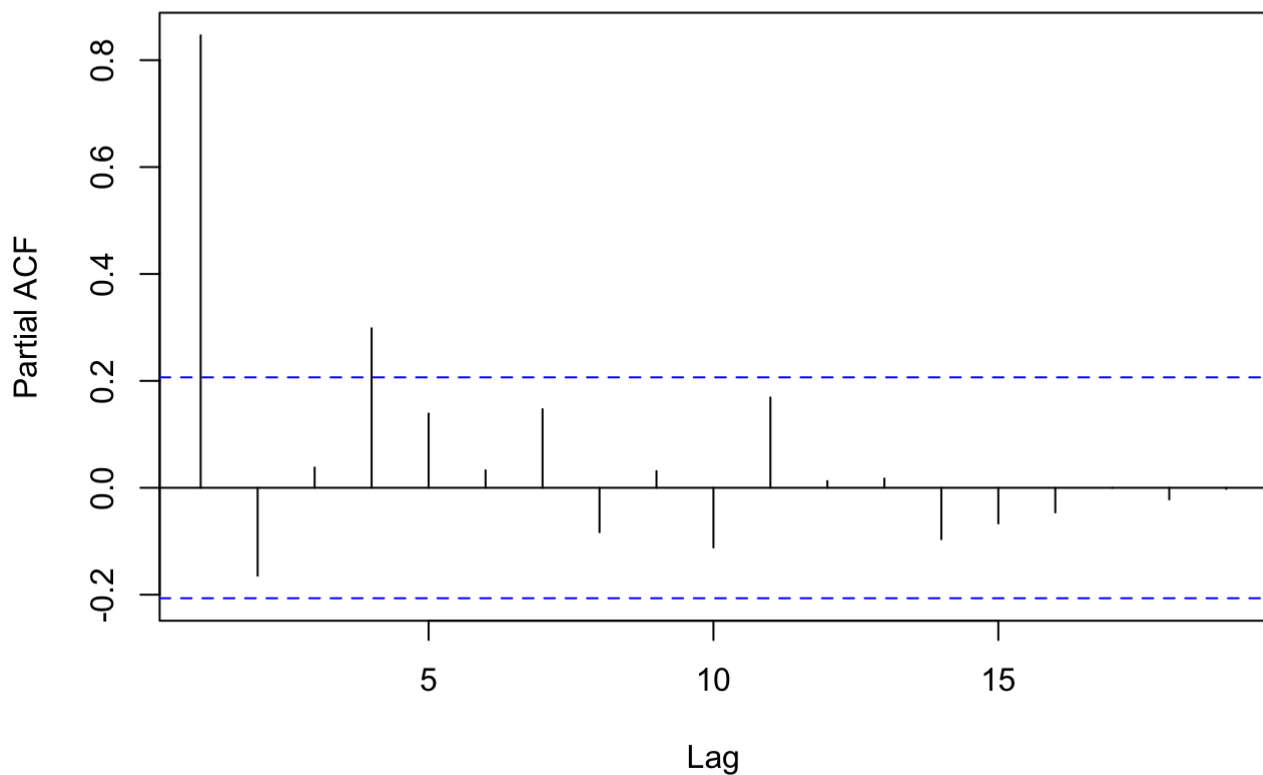
From the test we can observe that for k=5 the p-value is significantly greater than alpha value of 0.05.

8.2 ACF & PACF Plots

```
acf(ozonets,main="ACF of Time Series Data - Figure 24")
```

ACF of Time Series Data - Figure 24
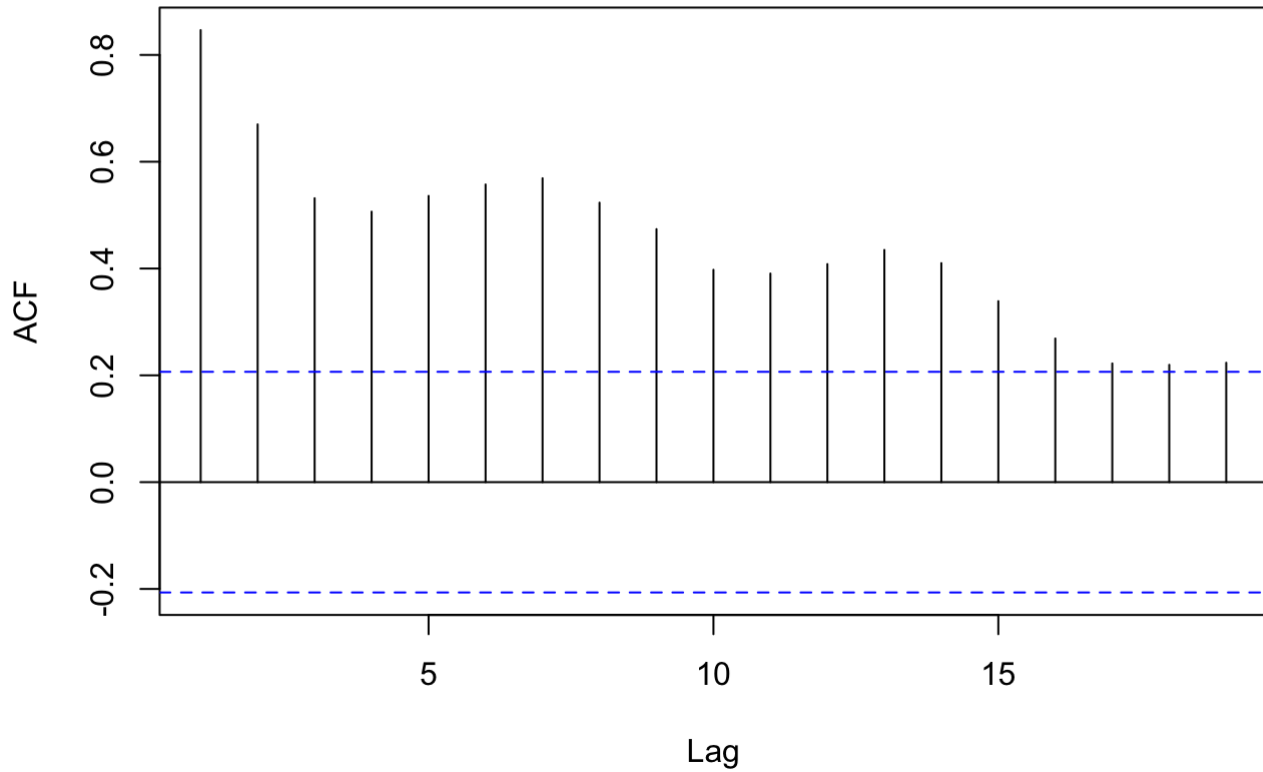

```
pacf(ozonets,main="PACF of Time Series Data – Figure 25")
```

PACF of Time Series Data - Figure 25

We can observe that there is slowly decaying pattern in the ACF plot and high correlation and significance in the first lag of PACF giving strong evidence on non-stationary property of the series.

In order to remove the non-stationary property we apply a few transformation techniques.

8.3 Transformation

8.3.1 Log Transformation

Using the log transformation method we transform the non-stationary series into stationary. Since there is negative values in the series, the series is shifted by adding a value which the absolute value of the minimum value in the ozone layer thickness time series and add with a constant to avoid zero value if any.

```
ozone1 = ozonets+ abs(min(ozonets))+1
ozonelog<-log(ozone1)
plot(ozonelog,ylab="Log(Ozone Layer thickness", main=" Log transformed time series pl
ot of Ozone layer thickness - Figure 26",type="o")
```

Log transformed time series plot of Ozone layer thickness - Figure 26
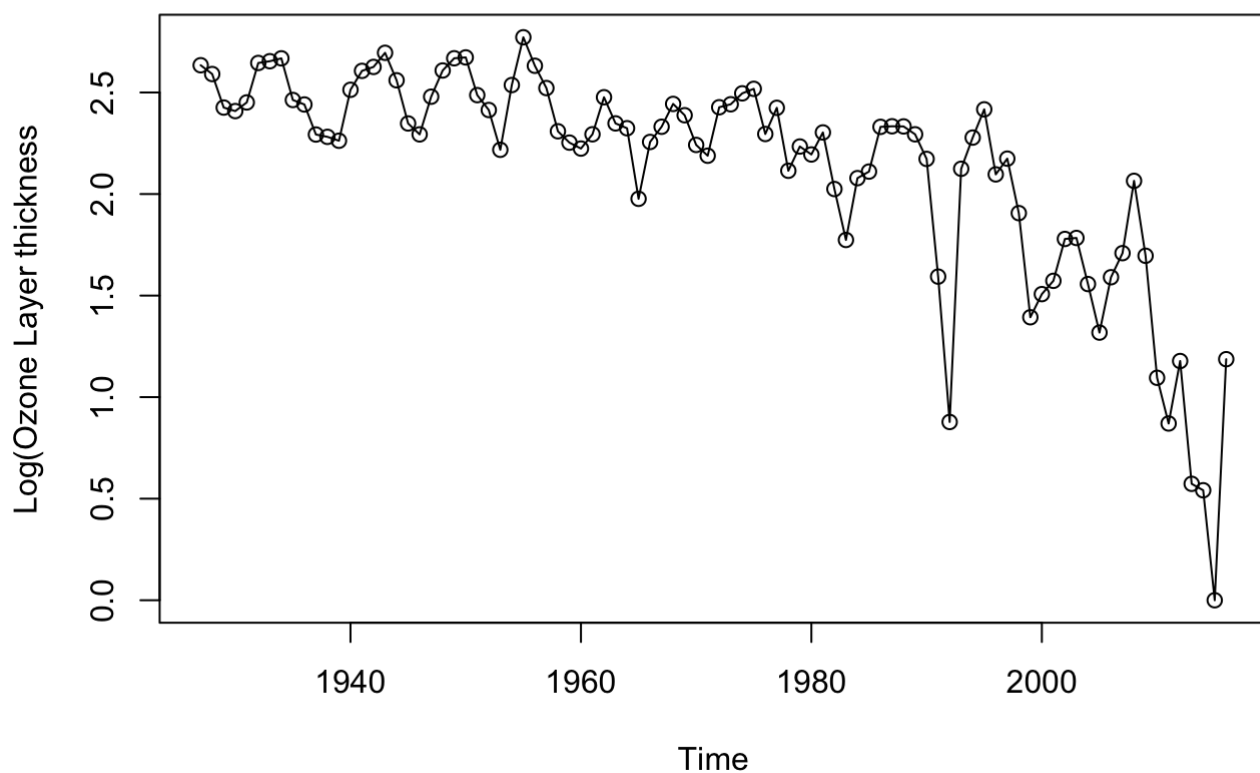
There is very slight difference in the time series plot while comparing the original one but not enough details. Hence we find the first difference of the log transformed series.

```
ozonelgdiff<-diff(ozonelog,differences = 1)
plot(ozonelgdiff,type="o",ylab="First differenced log transformation of ozone layer t
hickness - Figure 27",xlab="Year")
```
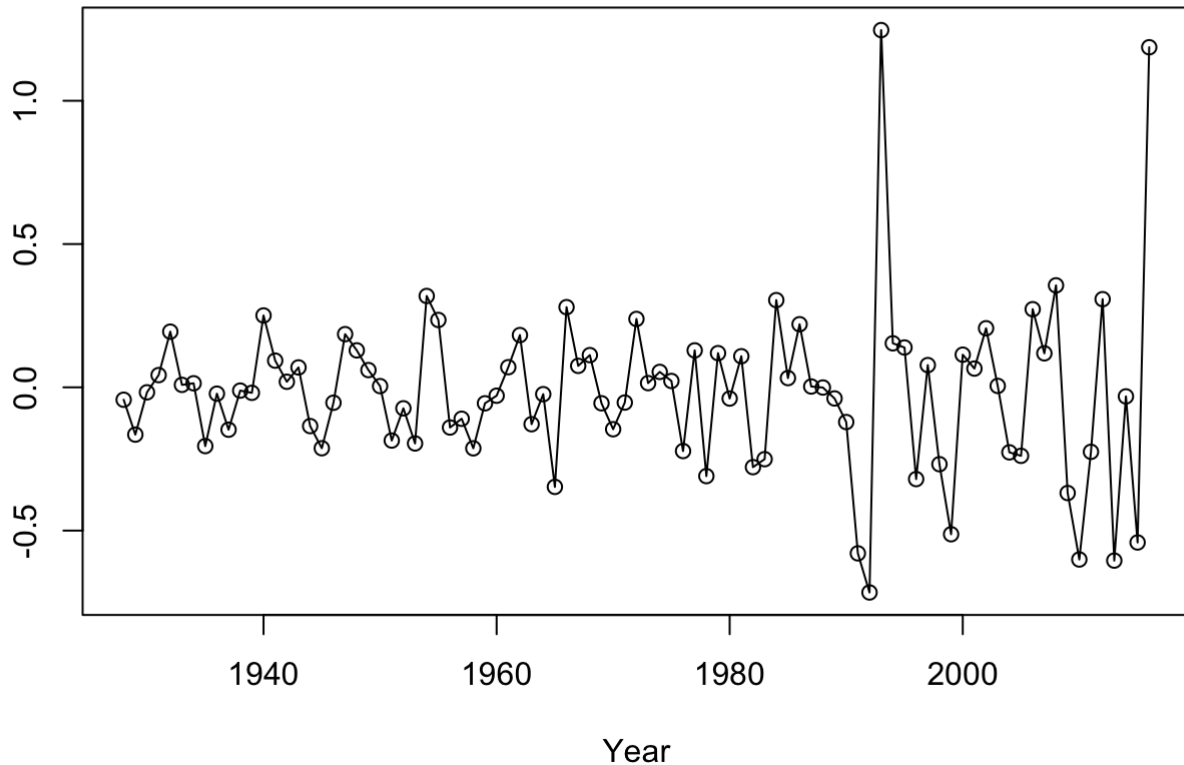
The series looks more as a stationary series. ADF Test is done to understand the null hypothesis .i.e.the series is non-stationary and alternate hypothesis, the series is stationary.

```
adf.test(ozonelgdiff)
```

```
##
##   Augmented Dickey-Fuller Test
##
## data:  ozonelgdiff
## Dickey-Fuller = -6.3119, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary
```

From ADF Test we can observe that the p-value is less than the significance alpha value of 0.05 and hence we reject the null hypothesis, .i.e the series is non-stationary.

8.3.2 Box-Cox Transformation

Using BoxCox function we transform the time series to stationary series. The series is found to have negative values or zero values and hence we add a value to the series. The series gets shifted and it doesn't effect the correlation. We add the absolutes value of the minimum value from series and add a constant to avoid any zero value(1 in this case).

"ozonebc<-BoxCox.ar(ozonets)" creates an error - "Error in BoxCox.ar(ozonets) : Data values must be positive"

```
options(warn=-1)
ozonebc<-BoxCox.ar(ozonets+ abs(min(ozonets))+1)
```
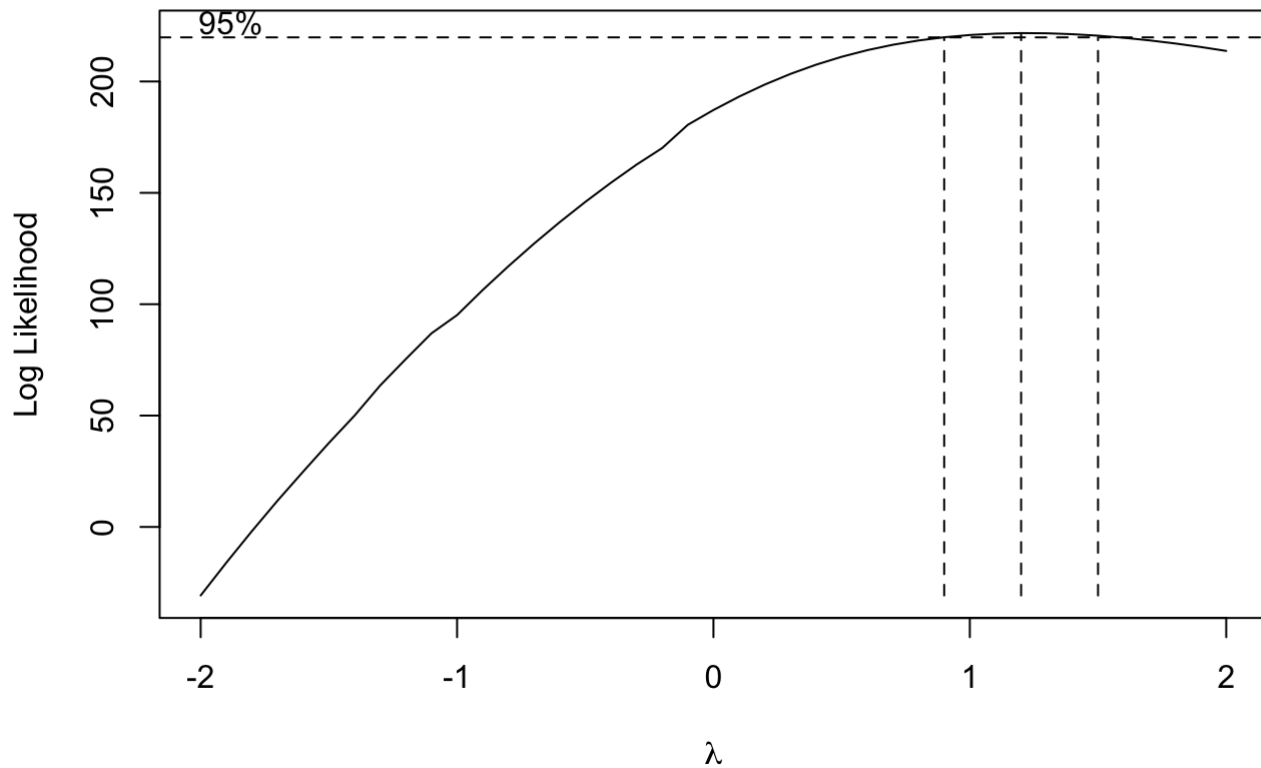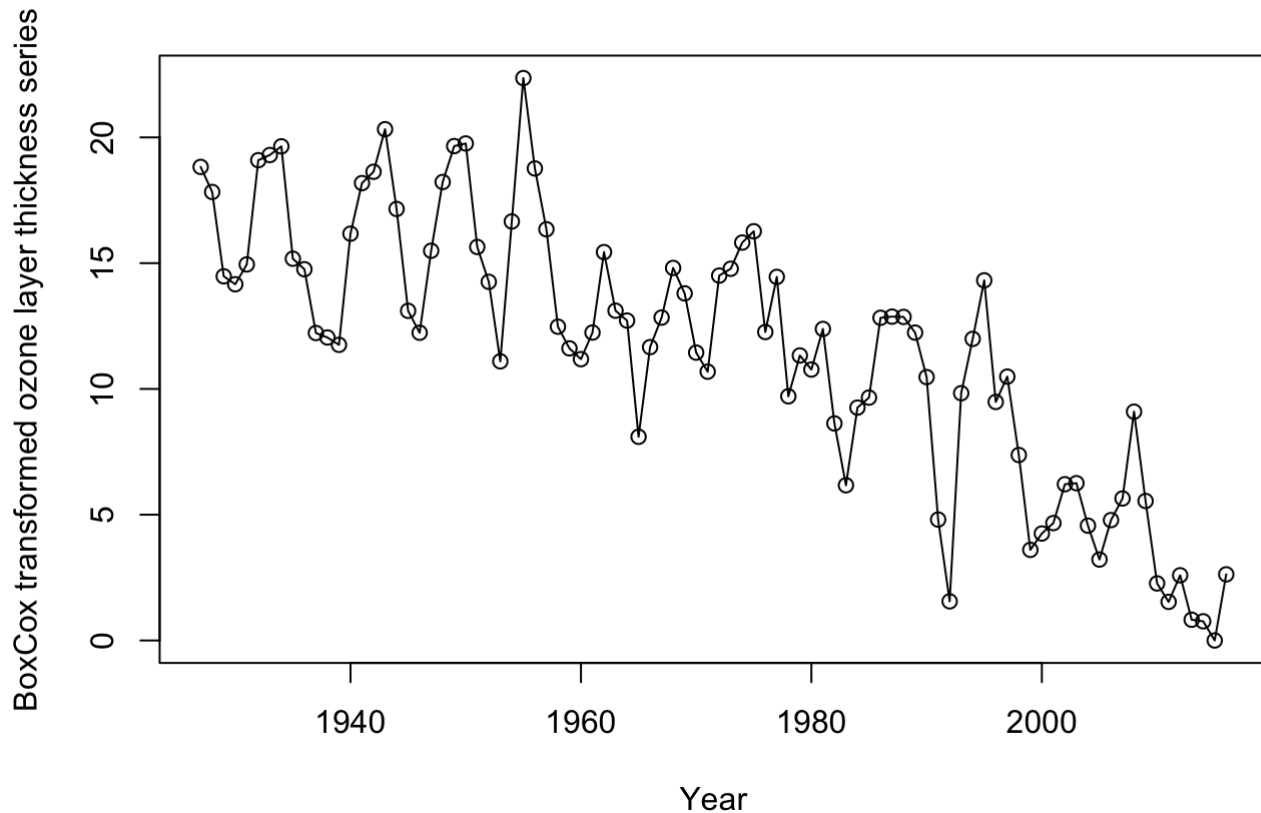
```
ozonebc$ci
```

```
## [1] 0.9 1.5
```

```
mean(ozonebc$ci)
```

```
## [1] 1.2
```

```
lambda <- mean(ozonebc$ci)
ozone1 = ozonets+ abs(min(ozonets))+1
ozonebcts = (ozone1^lambda-1)/lambda
plot(ozonebcts,ylab="BoxCox transformed ozone layer thickness series", xlab="Year",ty
pe="o",main="BoxCox Transformed Time Series - Figure 28")
```

BoxCox Transformed Time Series - Figure 28

There is only a very slight change in plot when compared to the actual time series plot and hence we try to achieve stationarity with differencing.

```
ozonebctsdiff<-diff(ozonebcts,differences = 1)
plot(ozonebctsdiff,ylab="First differenced BoxCox transformation of ozone layer thick
ness", xlab="Year",type="o",main="BoxCox Transformed and Differenced Time Series - Fi
gure 29")
```

BoxCox Transformed and Differenced Time Series - Figure 29
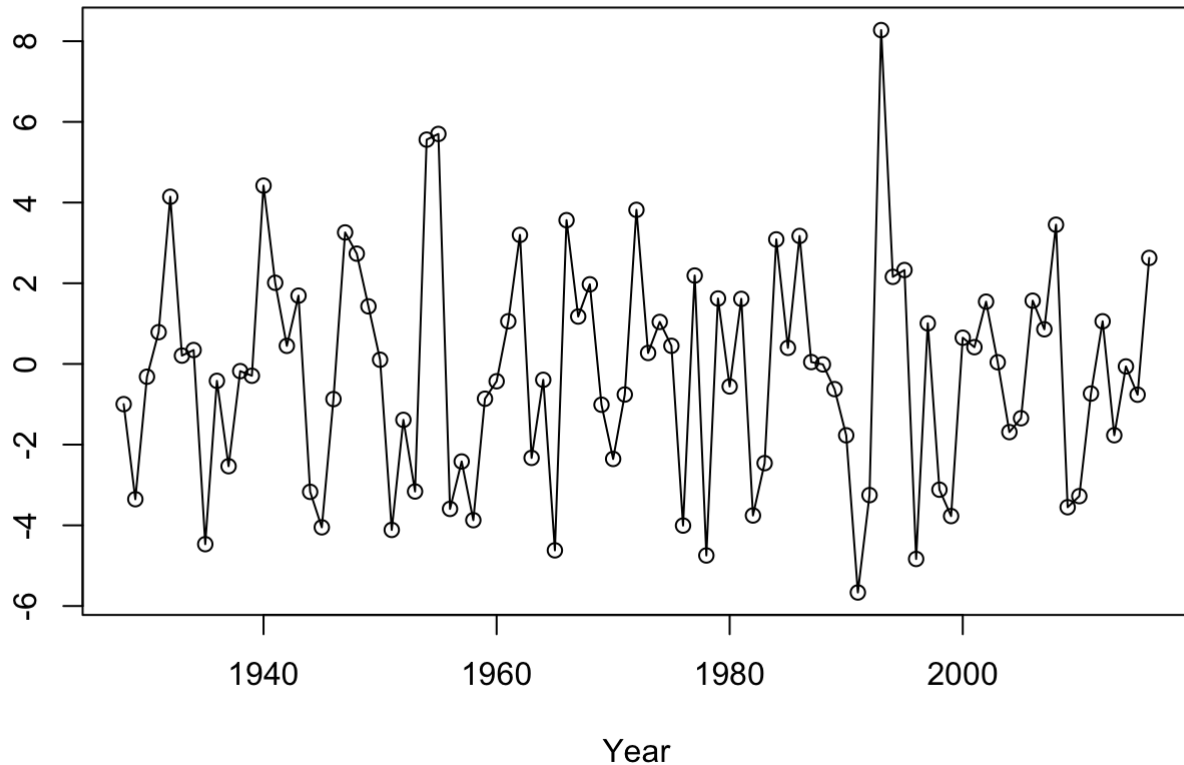
The difference series of the BoxCox transformed time series looks more stationary than the transformed and original series. We conduct a ADF Test to find the hypothesis.

```
adf.test(ozonebctsdiff)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  ozonebctsdiff
## Dickey-Fuller = -7.2426, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary
```

From ADF Test we observe that the p-value is less than the alpha value, 0.05 and hence we reject the null hypothesis and confirm stationarity of the series.

8.3.3 Selection from Transformed series

Observing the time series plot of the first difference BoxCox transformation and log transformation, we select the BoxCox transformation series as it looks more stationary with less variance and high points.

8.4 Ploting ACF and PACF post Transformation

We plot the ACF and PACF plots of the difference-BoxCox transformed data.

```
acf(ozonebctsdiff,main="ACF of first difference BoxCox transformed Time Series Data -
Figure 30")
```

ACF of first difference BoxCox transformed Time Series Data - Figure 3
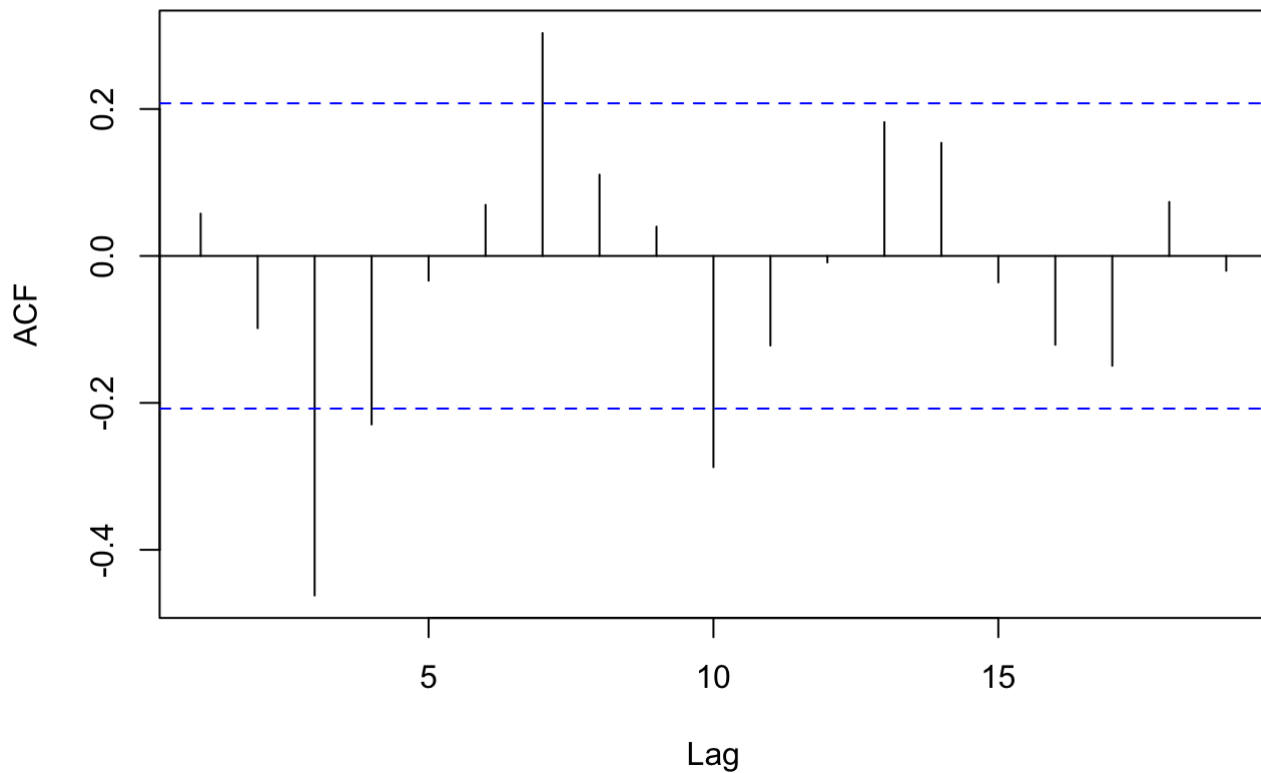

```
pacf(ozonebctsdiff,main="PACF of first difference BoxCox transformed Time Series Data
- Figure 31")
```

PACF of first difference BoxCox transformed Time Series Data - Figure 3
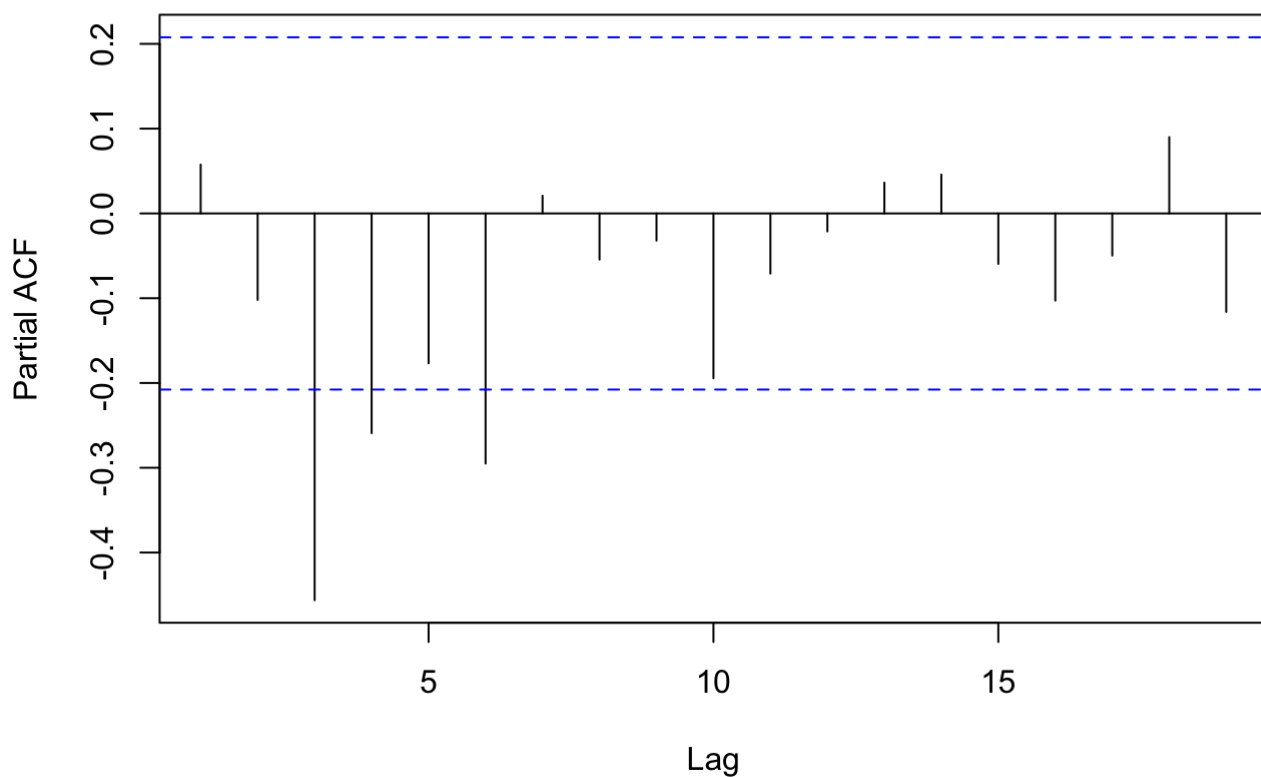
From the ACF plot we observe the q values as 1,2,3 and from PACF the p values are 2 and 3. In ACF plot we don't consider the at 10. In the PACF plot we take 3 scenarios wherein 1 is considered when the lags are considered upto 5 and the second lag is considered to be less significant and in the second scenario then we consider the 2nd one and in the 3rd we consider all the 3.

8.5 EACF Method

From the EACF table we consider the top most left "o" and the neighboring "o"'s while considering p and q value.

p values : 0,1
q values : 0,1

```
eacf(ozonebctsdiff)
```

```
## AR/MA
##    0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 o o o x x o o x o o x o  o  o  o
## 1 x o x o o o x o o x o  o  o  o
## 2 o o x o o o x o o x o  o  o  o
## 3 x o x o o x o o o o o  o  o  o
## 4 x o o x o x o o o o o  o  o  o
## 5 x x x x o x o o o o o  o  o  o
## 6 o o o x o o o o o o o  o  o  o
## 7 x o o x o o o o o o o  o  o  o
```
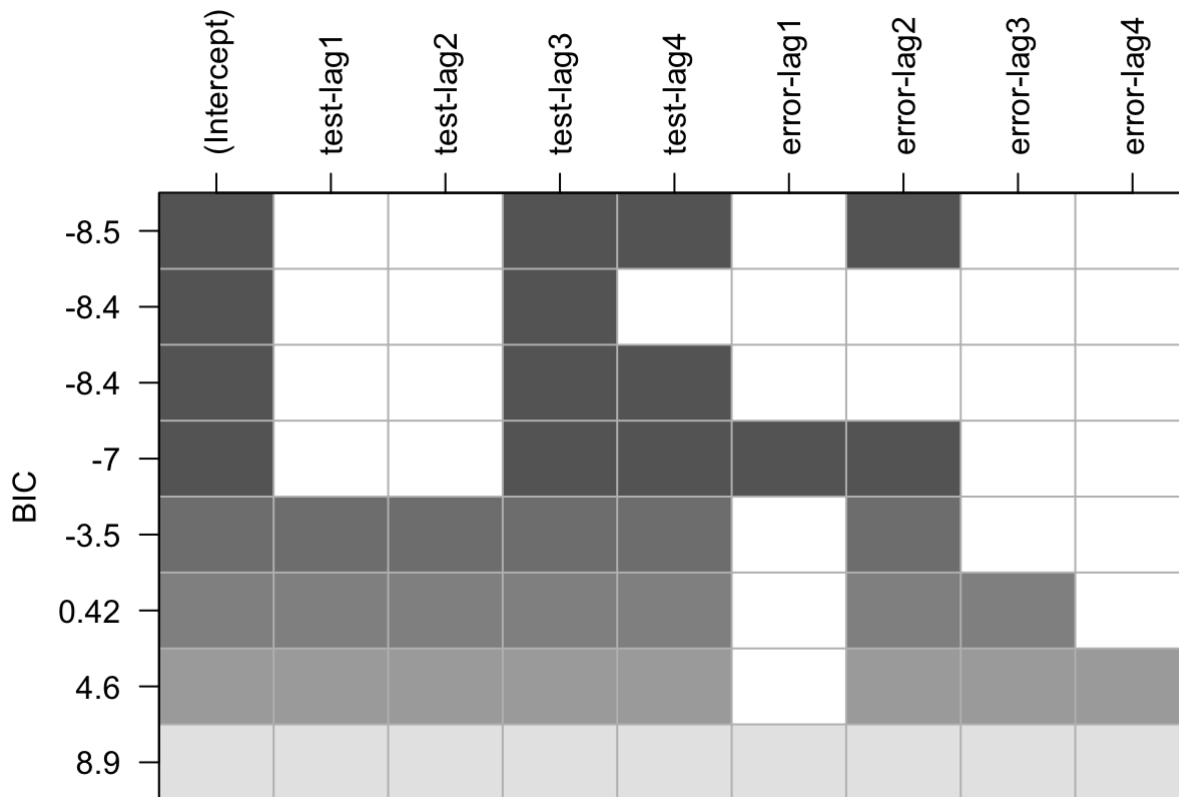
8.6 BIC Table

In BIC table we consider the nar and nma values to be 4 as at higher values the significant value changes drastically. The 'test-lag' represents the p value and 'error-lag' represents q value.

p values : 3,4
q values : 2

```
ozonebic=armasubsets(y=ozonebctsdiff,nar=4,nma=4,y.name='test',ar.method = 'ols')
plot(ozonebic)
```

9 Models

- ARIMA model is represented as ARIMA(p,d,q)
- d value is 1 since the differencing is done once.

From ACF, PACF, EACF and BIC table the probable models are:
* ARIMA (0,1,0)
* ARIMA (0,1,1)
* ARIMA (1,1,1)
* ARIMA (2,1,1)
* ARIMA (2,1,2)
* ARIMA (2,1,3)
* ARIMA (3,1,1)
* ARIMA (3,1,2)
* ARIMA (3,1,3)
* ARIMA (4,1,2)

10 Summary and Conclusion

The Ozone layer thickness data were analyzed. Initially, the data description was done to understand the dimensions of the data and converted it into time series data. The data description was done based on the five points associated with the series. The model fitting is done by considering linear, quadratic, harmonic, and seasonal trends, and the best fit among the four was found. Within the limits of the task in this report, we understand that quadratic trend is not the best fit and we move to fit the best fit using ARIMA Modeling. The time-series data was transformed and differencing was done before modeling. The models were distinguished using ACF-PACF plots, EACF, and BIC methods. The probable set of models were discussed at the end.

11 Reference

1. Home - RDocumentation. (2021). Retrieved 11 April 2021, from https://www.rdocumentation.org/ (https://www.rdocumentation.org/)
2. Cryer, Jonathan D., and Kung-Sik. Chan. Time Series Analysis With Applications in R . 2nd ed. 2008. New York, NY: Springer New York, 2008. Web.
3. Demirhan, H. (2021). Math1318 Time Series Analysis[Lecture Notes].Canvas@RMIT University. https://rmit.instructure.com/courses/79716/modules (https://rmit.instructure.com/courses/79716/modules)
4. Yihui Xie, E. (2021). R Markdown Cookbook. Retrieved 11 April 2021, from https://bookdown.org/yihui/rmarkdown-cookbook/ (https://bookdown.org/yihui/rmarkdown-cookbook/)