TIME SERIES ANALYSIS – ASSIGNMENT 1

Vishal Patnaik Damodarapatruni - s3811521

13/04/2021

Table of Contents

Introduction

Ozone (O3) is present as a layer within Earth's atmosphere with some thickness which absorbs the UV rays from the Sun. Higher the thickness the greater will be the absorption of the UV rays by the layer. [1]

Here, the thickness of the Ozone layer is analyzed, and modelling is done on the Time Series data. The data is the yearly change in the thickness of the Ozone layer over a period of 90 years i.e., from 1927 to 2016. The date has both negative and positive values, decrease in the thickness is indicated in negative values whereas increase in positive values.

In this analysis, time series techniques are applied to the data and the best model among several models is selected to predict the thickness change in future (i.e., say for example 5 years). This helps both the public and environmental government to understand how the state of the Ozone layer is going to be in future.

Scope

This analysis has two parts: Part 1: Regression Part 2: Time series approach.

- **Part 1:**
    - Finding the best model among Linear, Quadratic, Cosine, Cyclical or Seasonal trend models.
    - Using the best fit model, the data for the next 5 years is predicted.
- **Part 2:**

  Here, possible ARIMA (p, d, q) models are proposed using model specification tools like ACF-PACF, EACF and BIC.

Method

Using the below packages (TSA, tseries, fUnitRoots etc.) the time series data is visualized, and the best model selected is fitted on the data. This model selection is justified by model specification tools like ACF-PACF, EACF and BIC table.

```
library(TSA) # Time Series Analysis.
## Warning: package 'TSA' was built under R version 4.0.5
##
## Attaching package: 'TSA'
## The following objects are masked from 'package:stats':
##
##     acf, arima
## The following object is masked from 'package:utils':
##
##     tar
library(tseries) # Time Series Analysis and Computational Finance. [2]
## Warning: package 'tseries' was built under R version 4.0.5
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
library(fUnitRoots) # To analyze trends and unit roots in financial time seri
es. [3]
## Warning: package 'fUnitRoots' was built under R version 4.0.5
## Loading required package: timeDate
##
## Attaching package: 'timeDate'
## The following objects are masked from 'package:TSA':
##
##     kurtosis, skewness
## Loading required package: timeSeries
## Warning: package 'timeSeries' was built under R version 4.0.5
## Loading required package: fBasics
## Warning: package 'fBasics' was built under R version 4.0.5
library(dplyr)
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:timeSeries':
##
##     filter, lag
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Data

The data is the yearly change in the thickness of the Ozone layer over a period of 90 years i.e., from 1927 to 2016. The date has both negative and positive values, decrease in the thickness is indicated in negative values whereas increase in positive values. The dataset is in csv format and hence it is loaded using "read.csv()" function.

```
v_ozone_thickness <- read.csv("data1.csv", header = FALSE)
head(v_ozone_thickness)
##           V1
## 1  1.3511844
## 2  0.7605324
## 3 -1.2685573
## 4 -1.4636872
## 5 -0.9792030
## 6  1.5085675
```

V1 - Thickness change in ozone layer. Now, let us add years from 1927 to 2016 as row names. So that each value of V1 corresponds to the change in thickness of Ozone layer in that respective year.

```
rownames(v_ozone_thickness) <- seq(from=1927, to=2016)
head(v_ozone_thickness)
##              V1
## 1927  1.3511844
## 1928  0.7605324
## 1929 -1.2685573
## 1930 -1.4636872
## 1931 -0.9792030
## 1932  1.5085675
```

Checking the class of v_ozone_thickness. (It should be data frame.)

```
class(v_ozone_thickness)
## [1] "data.frame"
```

Now convert the data frame into a time series object.

```
v_ozone_thickness_1 <- ts(as.vector(as.matrix(t(v_ozone_thickness))), start =
1927, end = 2016, frequency = 1)
```

Checking the class of v_ozone_thickness_1. (It should be time series.)

```
class(v_ozone_thickness_1)
## [1] "ts"
```

As we got the time series object now let us visualize it.

```
# Function to plot a single data.
v_Plot <- function(v, m){
  plot(v, type = "b", pch = 19, col = "blue", xlab = "years", ylab = "Thickne
ss", main = m)
}
# Function to form a legend corresponding to a single data plot.
v_leg1 <- function(t, l, c, p){
  legend("bottomleft", inset = .03, title = t, legend = l, col = c, horiz = T
RUE, cex = 0.8, lty = 1, box.lty = 2, box.lwd = 2, box.col = "blue", pch = p)
}
v_Plot(v_ozone_thickness_1, "Ozone layer thickness change from 1927 to 2016 i
n Dobson units")
v_leg1("Ozone layer Thickness change over years.", c("Thickness Change"), c("
blue"), c(19))
```

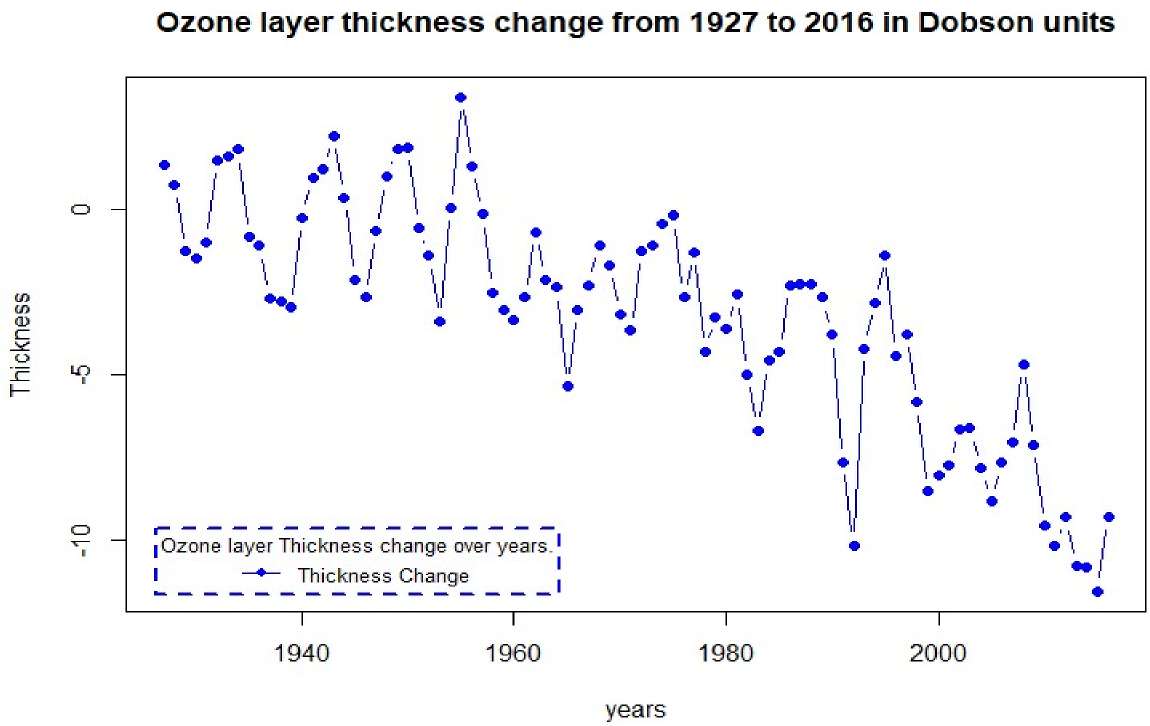

Fig 1: Ozone layer thickness change - Time series plot.

- From the plot we can observe that,
- The data follows a downward trend. This suggests that the ozone layer is depleting over years.
- There is no seasonality in the trend.
- We can also absorb some intervention in the just before 1990. According to research Ozone layer depletion is more at that time. [4]

```
# Function to plot using two different data (Scatter Plot).

v_Plot1 <- function(v, v1, x, y, m){
  plot(x = v, y = v1, pch = 19, col = "blue", xlab = x, ylab = y, main = m)
}
# Function to form a legend corresponding to a two data plot.

v_leg <- function(t, l){
  legend("topleft", inset = .03, title = t, legend = c(l), col = c("blue"), h
oriz = TRUE, cex = 0.8, box.lty = 2, box.lwd = 2, box.col = "blue", pch = c(1
9))
}
```

Now let us analyze the relationship in the data using scatterplot and correlation.

Scatterplot

```
v_data_lag = zlag(v_ozone_thickness_1) # First lag of thickness change.

v_Plot1(v_data_lag, v_ozone_thickness_1, "Thickness change in previous years.
", "Thickness change", "Scatter plot for Ozone Layer Thickness in Dobson unit
s.")
v_leg("Ozone layer Thickness change over years.", "Thickness Change.")
```

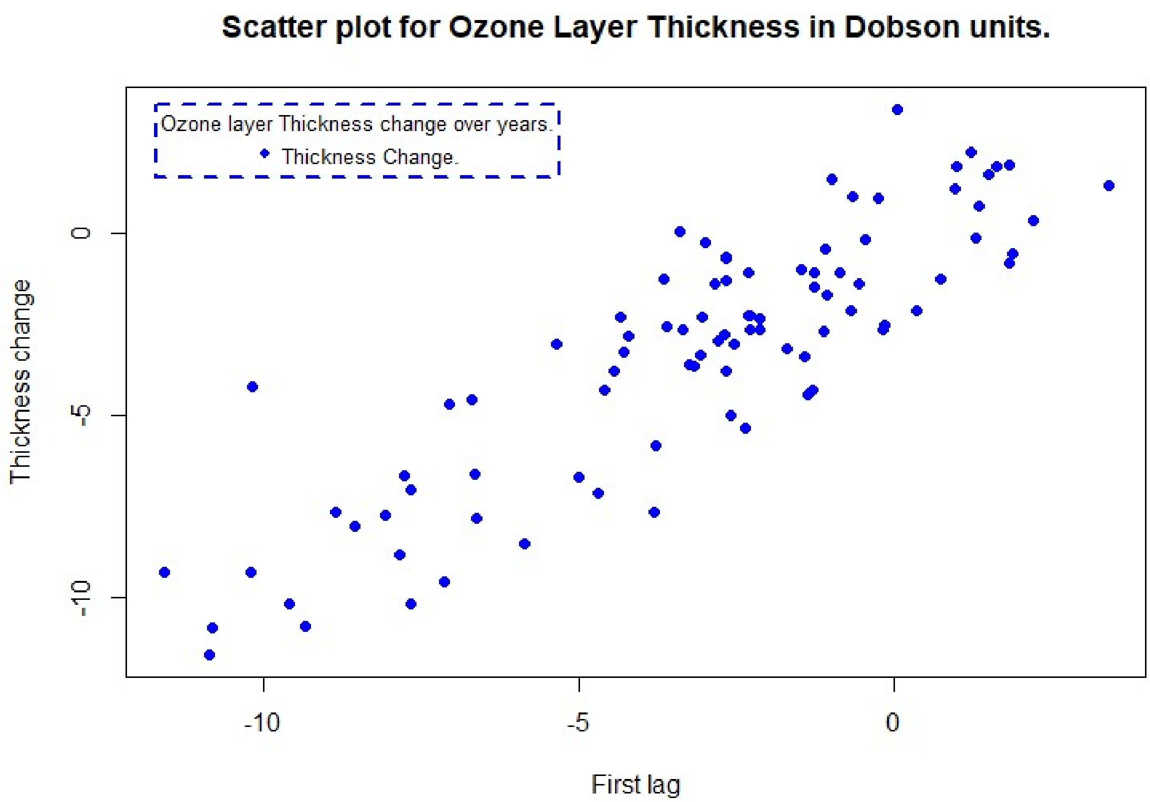

Fig 2: Ozone layer thickness change - Scatterplot.

The relationship seems to be almost linear as at some points the linearity is not seen. This might be due to the skewness in the variables. We can transform the data but here it seemed to be an inefficient process as the results and analysis might mis-lead. Therefore, it is better not to transform data.

Correlation

```
i = 2 : length(v_data_lag) # Creating index by negletting first null values.
cor(v_ozone_thickness_1[i], v_data_lag[i]) # Calculating the correltion coeff
icient.
## [1] 0.8700381
```

The data of the first lag thickness change is correlated to the time series data. The correlation coefficient "0.8700381" proves this. There exists nearly 87 % correlation among the data. This means there exists 87% linear relationship among the data.

Task 1

Introduction

After the best model among Linear, Quadratic, Cosine, Cyclical or Seasonal trend models best fitted model is selected and used to predict the data for the next 5 years.

Modeling

Model building strategy

- Model specification.
- Model fitting.
- Model diagnostics.

Model specification:

From the scatter plot we can observe a linear positive relationship. So, Linear regression might be a better choice. Now let us fit a Simple linear regression model on the data.

Simple linear regression

The deterministic model trend:

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

where: β_0 = Intercept and β_1 = Slope of Linear Trend.

Model fitting:

```
v_time <- time(v_ozone_thickness_1)
V_LM = lm(v_ozone_thickness_1 ~ v_time)
summary(V_LM)
##
## Call:
## lm(formula = v_ozone_thickness_1 ~ v_time)
##
## 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 ***
## v_time       -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
```

Let us interpret the visualization of the data with linear regression line.

```
v_Plot(v_ozone_thickness_1, "Ozone Layer Thickness change in Dobson units - S
imple Linear Model")
abline(V_LM, lty = 1, col = "red")
v_leg1("Ozone layer Thickness change over years.", c("Thickness change", "Lin
ear Model"), c("blue", "Red"), c(19, NA))
```

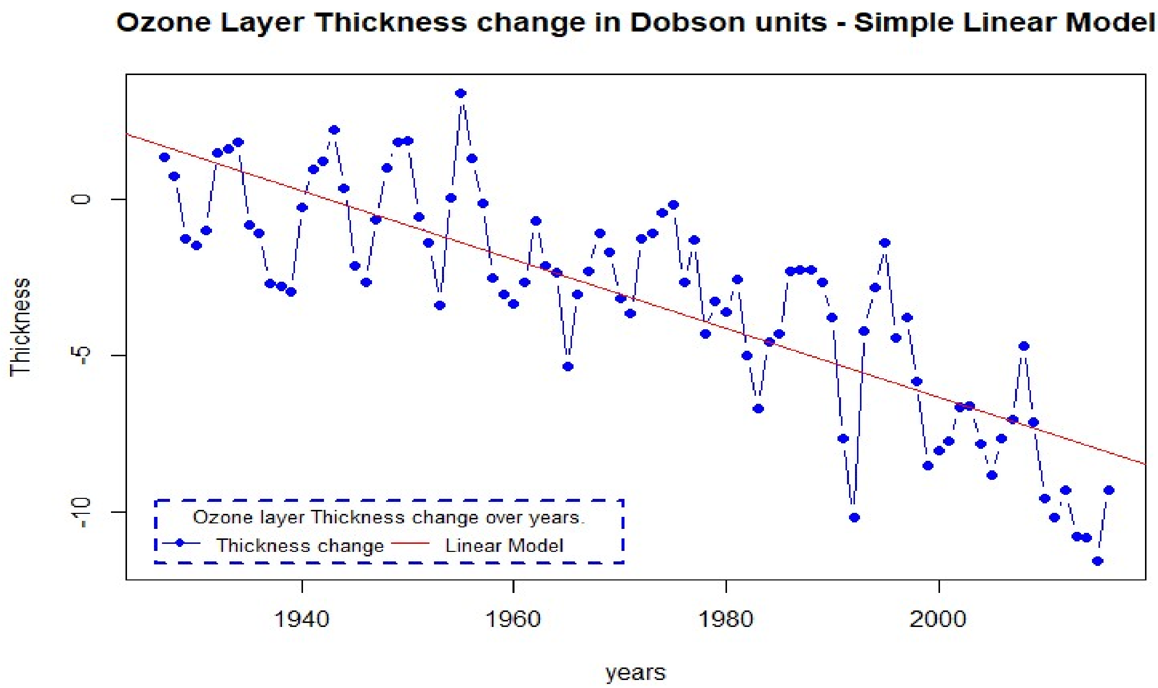

Fig 3: Thickness change - Simple Linear model.

```
summary(V_LM)
##
## Call:
## lm(formula = v_ozone_thickness_1 ~ v_time)
##
## 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 ***
## v_time       -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
```

Hypothesis:

H_0: The data doesn't fit the simple linear regression model.

H_A: The data fits the simple linear regression model.

Interpretations:

- The adjusted F - statistic "F = -13.34".
- R - squared is 0.6693.
- Adjusted R - squared is 0.6655.
- Degrees of freedom - DF are (1, 88)
- p - value is < 0.05 and therefore, it is statistically significant. Therefore, Null hypothesis is rejected. Hence, the model fits the simple linear regression model.
- Slope $\beta 1 = -0.110029$, which suggests the mean is not constant and the data is not stationary. Also, there is a trend in data as the slope is significant.
- As the slope is negative the plot shows a negative or downward trend.
- This model suggests that there is a 66.93% of data change. Suggesting that the model explains only 66.93% of the trend. Which implies that the model shows some trend.

Model diagnostics:

Now let us check whether the model is the best fit on the data or not. This can be checked by analyzing the residuals and performing a Shapiro test.

Note: For a best fit model, a true stochastic nature and normality should be seen in the residuals.

Analyzing the residuals of Simple linear model trend:

```
# Function for residual analysis.

v_analysis <- function(res_m) {
    # Scatter plot for model residuals
    plot(res_m, type = "b", pch = 19, col = "blue", xlab = "years", ylab = "S
tandardized Residuals", main = "Plot of Residuals over Time")
    abline(h = 0)
    # Standard distribution
    hist(res_m, xlab = 'Standardized Residuals', freq = FALSE, ylim = c(0, 0.
6))
    curve(dnorm(x, mean = mean(res_m), sd = sd(res_m)), col = "red", lwd = 2,
add = TRUE, yaxt = "n")
    # QQplot for model residuals
    qqnorm(res_m, col = c("blue"))
    qqline(res_m)
    # Auto-Correlation Plot
    acf(res_m, main = "ACF of Standardized Residuals",col=c("blue"))
}
# Displaying both Mean and Standard deviation of LM.
sprintf("Mean : %f & Standard Deviation : %f", mean(rstudent(V_LM)), sd(rstud
ent(V_LM)))
## [1] "Mean : -0.002349 & Standard Deviation : 1.014063"
v_analysis(rstudent(V_LM))
```

1. Scatterplot of residuals over time analysis.

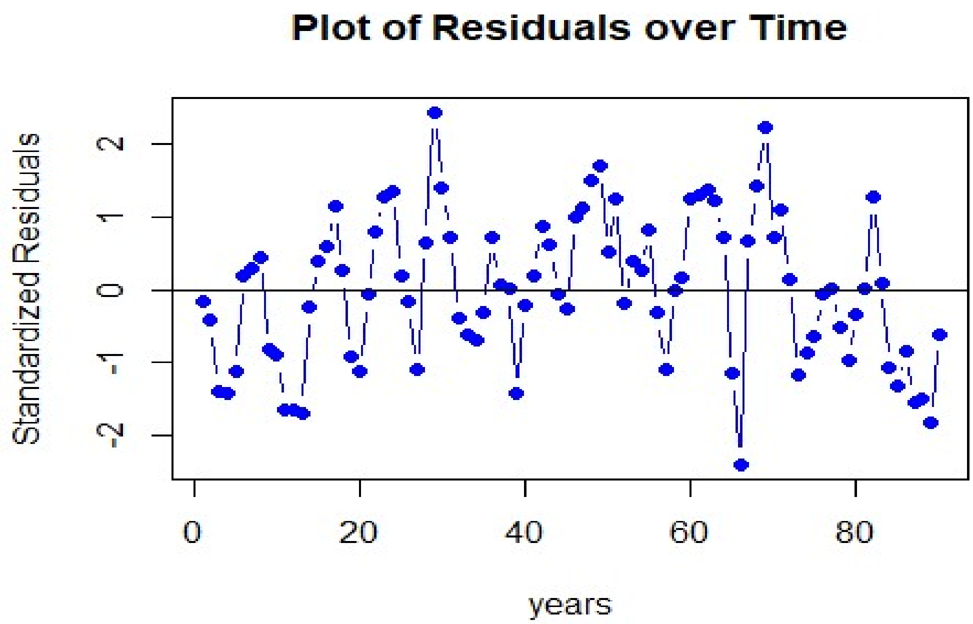

Fig 4: Scatter plot of linear model residual.

The data points are below the line at both the start and end of the trend. Randomness is not seen in the scatterplot due to similarity in the residuals. So, we cannot decide anything at this stage. Further analysis is required.

2. Distribution analysis.

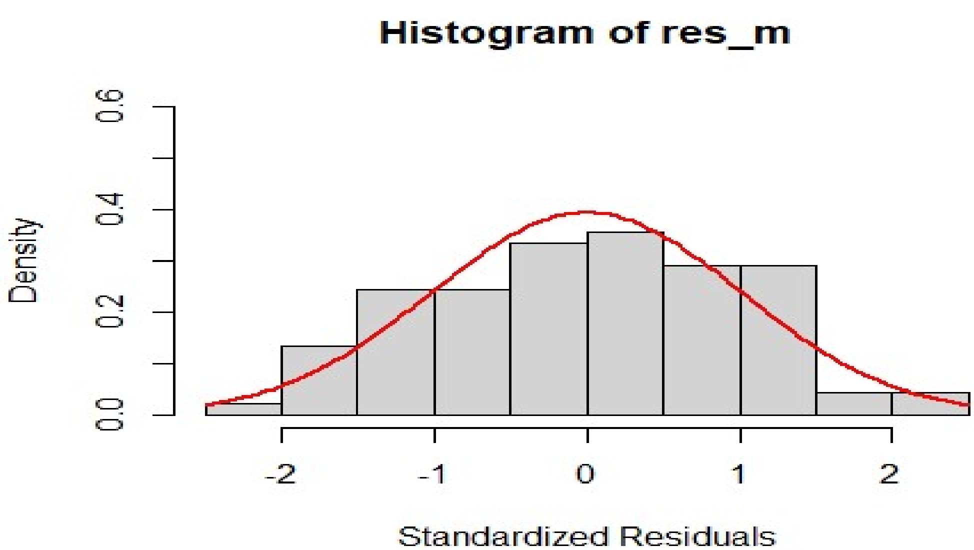

Fig 5: Distribution of standardized residuals from linear model.

Almost symmetric. This suggests a good fit with a very few data falling outside the normal curve indicating Kurtosis.

3. QQ plot analysis.

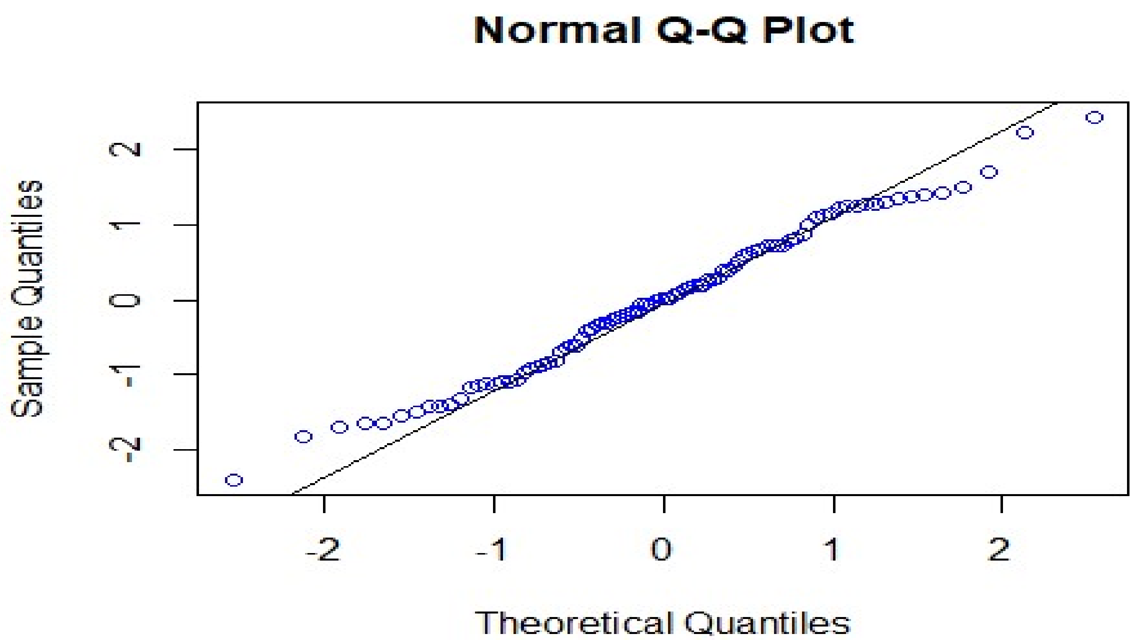

Fig 6: QQ Plot of linear model residuals.

The data is away from the normal at the tails, but it fitted in the middle with a small deviation. The mean is not 0 with identical distribution of data and hence there is no white noise.

4. Auto-Correlation Plot

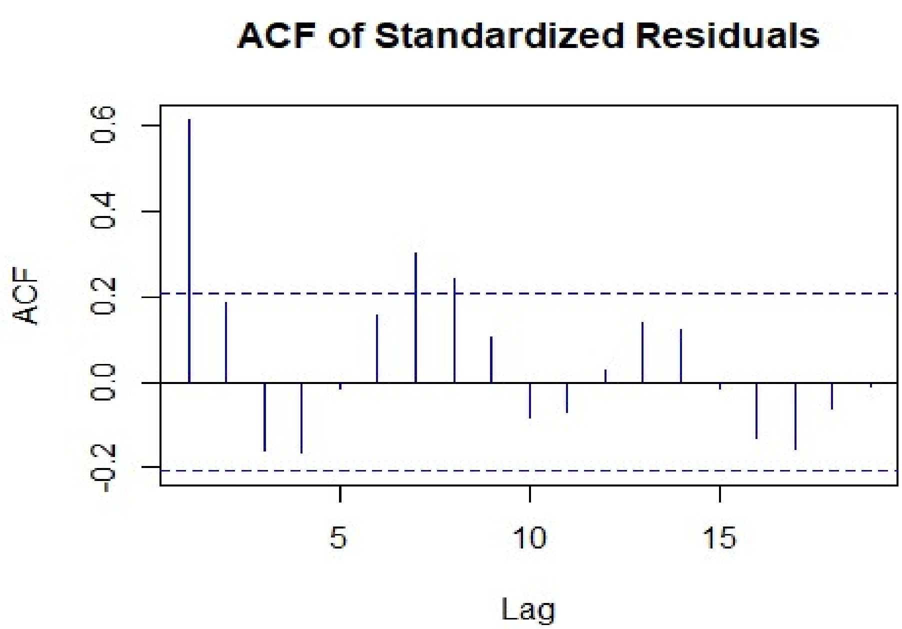

Fig 7: Auto-Correlation Plot of linear model residuals.

The correlation values are higher than the confidence bound at lag 1. Overall, nearly 3 values are above the dotted line suggesting that the stochastic component is not white noise.

Now let us conduct a Shapiro test to check whether the stochastic component is normally distributed or not.

5. Shapiro-Wilk normality test

Hypothesis:

H_0: The stochastic component is normally distributed.

H_A: The stochastic component is not normally distributed.

```
shapiro.test(rstudent(V_LM))
##
##  Shapiro-Wilk normality test
##
## data:  rstudent(V_LM)
## W = 0.98733, p-value = 0.5372
```

Interpretation:

P - value = 0.5372 > 0.05 (not statistically significant). Therefore, we fail to reject Null hypothesis. Hence, we can say that the stochastic component is normally distributed.

Model specification

Now let us again use the Model building strategy. Since, Linear Regression is our choice now let us fit data on Quadratic regression.

Quadratic regression

The deterministic model trend:

$$\mu t = \beta 0 + \beta 1t + \beta 2t2$$

Model fitting

```
v_time2 = v_time ^ 2
V_QM = lm(v_ozone_thickness_1 ~ (v_time + v_time2))
plot(ts(fitted(V_QM)), ylim = c(min(c(fitted(V_QM), as.vector(v_ozone_thickne
ss_1))), max(c(fitted(V_QM),as.vector(v_ozone_thickness_1)))), col = "red", x
lab = "years", ylab = "Thickness change", main = "Quadratic Curve (Fitted) of
Ozone layer Thickness change - Quadratic Model")
lines(as.vector(v_ozone_thickness_1), type = "b", col = "blue", pch = 19)
v_leg1("QuadPlot", c("Time Series Plot", "Quadratic Trend Line"), c("blue", "
red"), c(19, NA))
```

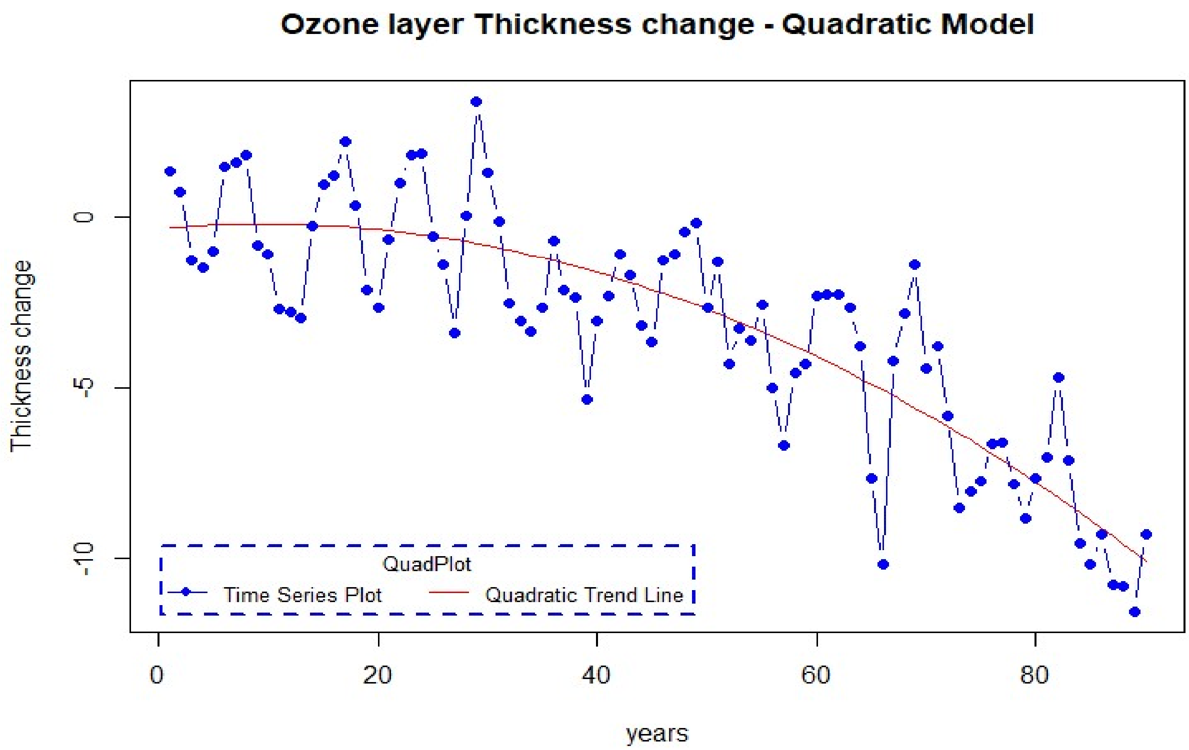

Fig 8: Thickness change - Quadratic model.

```
summary(V_QM)
##
## Call:
## lm(formula = v_ozone_thickness_1 ~ (v_time + v_time2))
##
## 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 ***
## v_time       5.924e+00  1.250e+00   4.739 8.30e-06 ***
## v_time2     -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
```

Hypothesis:

H_0: The data doesn′t fit the Quadratic model.

H_A: The data fits the Quadratic model.

Interpretations:

- The adjusted F - statistic "F = 13.34".
- R - squared is 0.7391.
- DF - (2, 87)
- p - value is < 0.05 and therefore, it is statistically significant. Therefore, Null hypothesis is rejected. Hence, the model fits the Quadratic model.
- This model suggests that there is a 73.91% of change in data. This implies model explains 73.91% of the trend, which is greater than Simple Linear model.

Model diagnostics

Now let us check whether the model is the best fit on the data or not. This can be checked by analyzing the residuals and performing a Shapiro test.

Note: For a best fit model, a true stochastic nature and normality should be seen in the residuals.

Analyzing the residuals of Quadratic model trend:

```
# Displaying both Mean and Standard deviation of QM.
sprintf("Mean : %f & Standard Deviation : %f", mean(rstudent(V_QM)), sd(rstud
ent(V_QM)))
## [1] "Mean : -0.000732 & Standard Deviation : 1.014339"
v_analysis(rstudent(V_QM))
```

1. Scatterplot of residuals over time analysis.

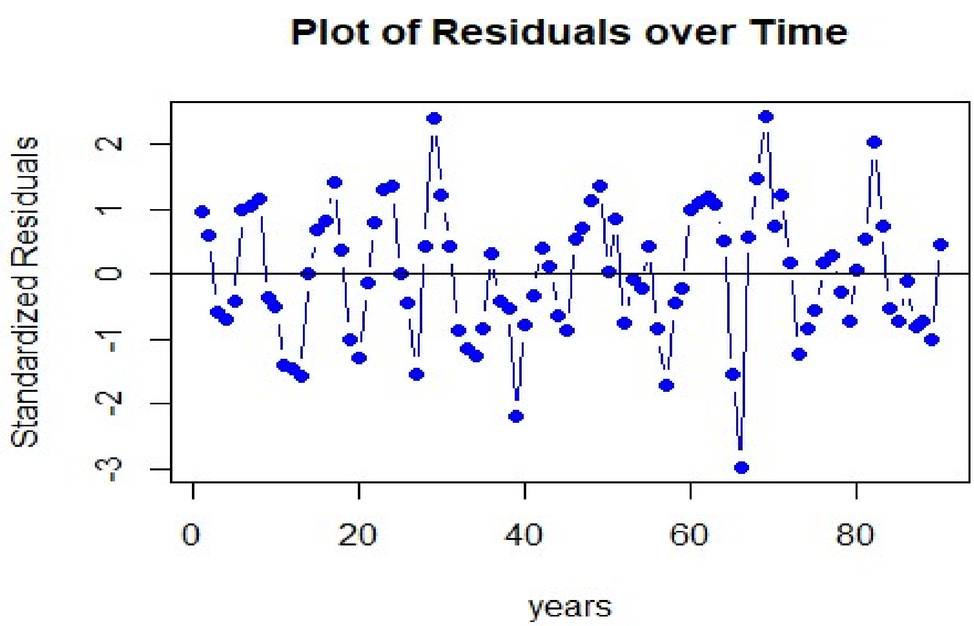

Fig 9: Scatter plot of quadratic model residual.

The data points are above the line at both the start and the end of the trend. The variance looks constant. Here also randomness is not seen. So, we cannot decide anything at this stage. Further analysis is required.

2. Distribution analysis.

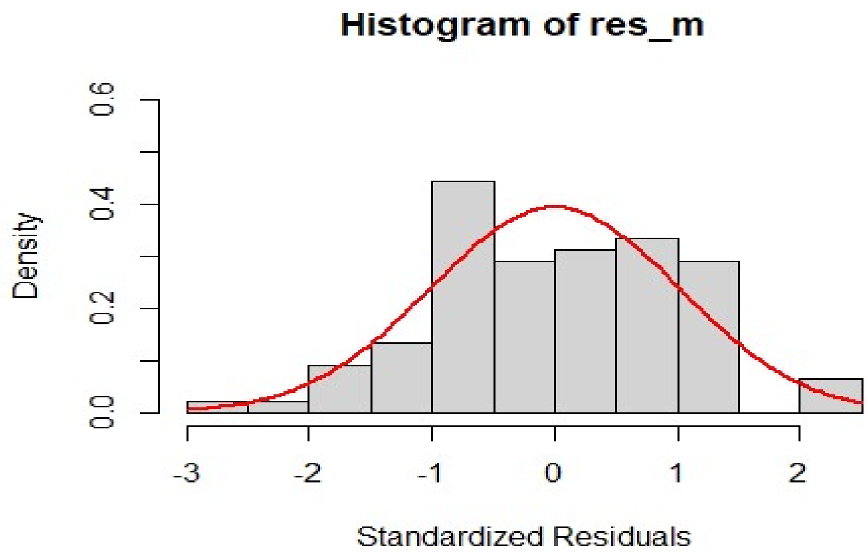

Fig 10: Distribution of standardized residuals from quadratic model.

Almost symmetric, but comparatively less symmetric than simple linear regression. This suggests a good fit with a very few data falling outside the normal curve indicating Kurtosis.

3. QQ plot analysis.

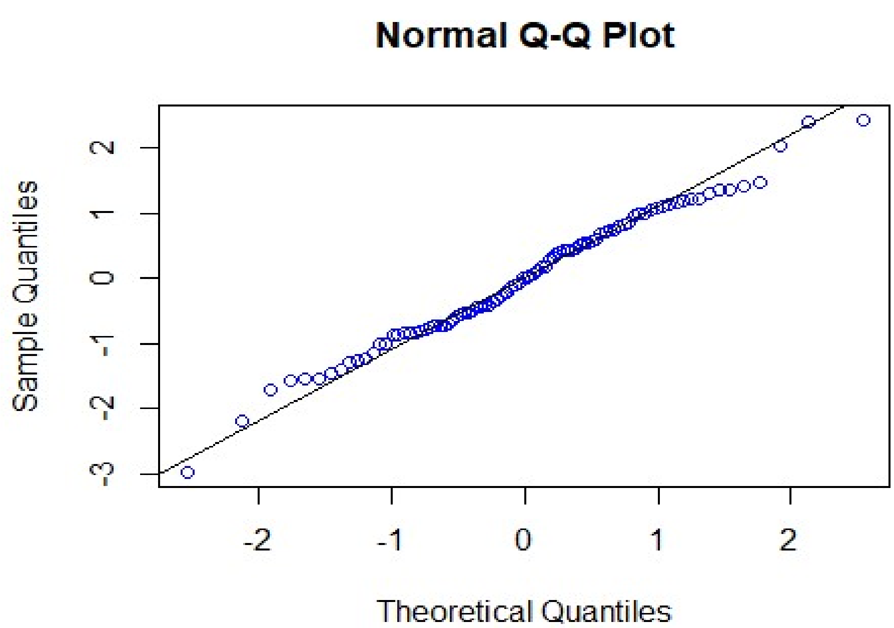

Fig 11: QQ Plot of quadratic model residuals.

The data at the tails is away from the normal with a small deviation in the middle which is better than simple linear regression. The mean is not 0 with identical distribution of data and hence there is no white noise.

4. Auto-Correlation Plot

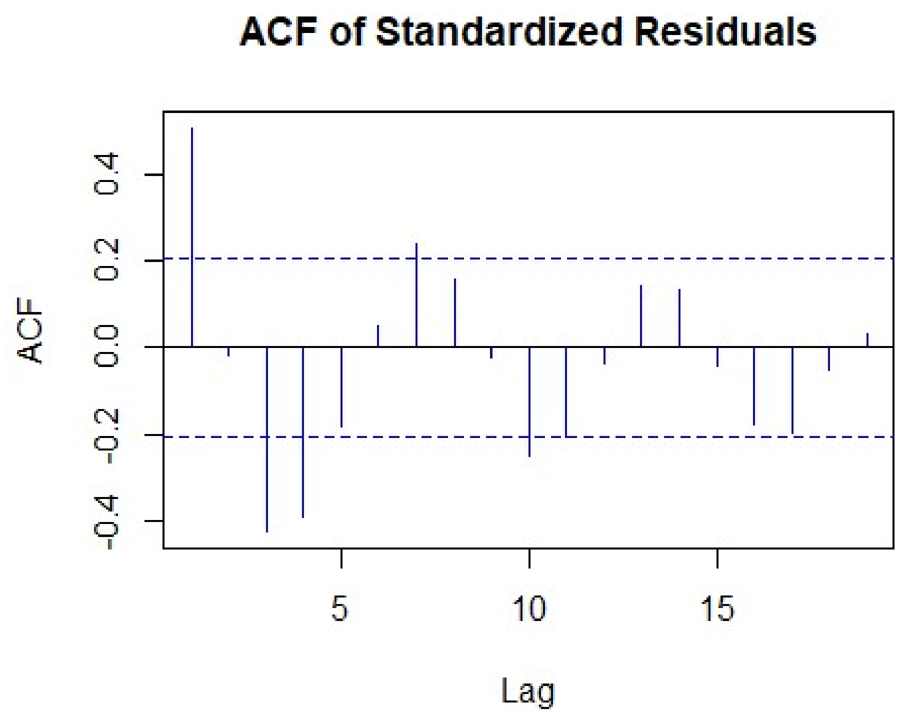

Fig 12: Auto-Correlation Plot of quadratic model residuals.

With significant correlation values the models shows some trend at 1, 3 and 4 lags.

Now let us conduct a Shapiro test to check whether the stochastic component is normally distributed or not.

5. Shapiro-Wilk normality test

Hypothesis:

H_0: The stochastic component is normally distributed.

H_A: The stochastic component is not normally distributed.

```
shapiro.test(rstudent(V_QM))
##
##  Shapiro-Wilk normality test
##
## data:  rstudent(V_QM)
## W = 0.98889, p-value = 0.6493
```

Interpretation:

P - value = 0.6493 > 0.05 (not statistically significant). Therefore, we fail to reject Null hypothesis. Hence, the stochastic component is normally distributed.

Model specification

Firstly, the data is linear. This is found from the scatter plot which showed a positive linear relationship or linear positivity. Hence, the data doesn't fit Seasonal / Cyclic or Cosine models.
More precisely,
Since the data is in years there will be no seasonal trend.
The data is checked for a 7 - year period for any cycles or patterns by setting the frequency to 7.

```
v_seven = ts(v_ozone_thickness_1, frequency = 7)

ts_plot1 = plot(v_seven, type = 'l',
                main ="Time series plot of the yearly changes in thickness of Ozone layer",
                ylab = "Change in Ozone thickness",
                xlab = "7 years cycle", col = "blue")
v_leg1("Ozone layer thickness change - 7 year cycle.", c("Thickness change - 7 years"), c("blue"), c(19))

v_cycle = factor(rep(1 : 7, length.out = length(v_seven)), ordered = TRUE)
points(y = v_seven, x = time(v_seven), pch = as.character(v_cycle), col = 2, cex = 1.15)
```

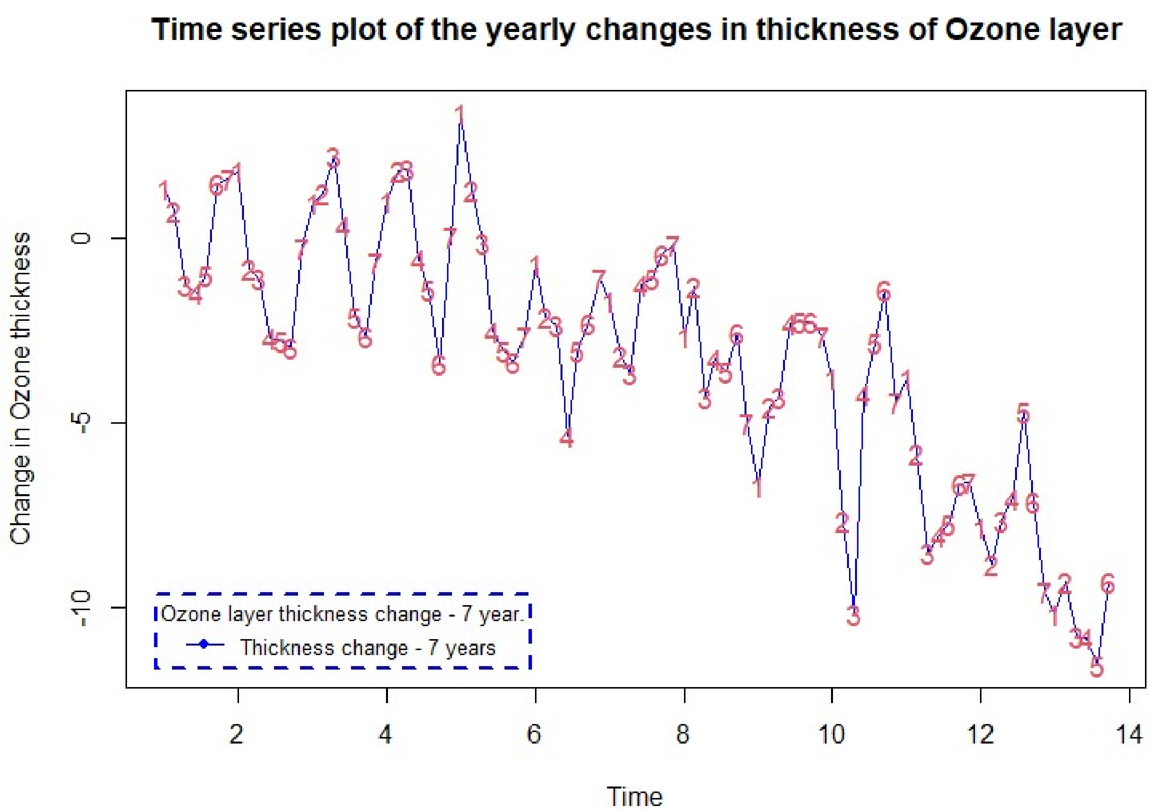

Fig 13: Pattern check for model specification.

From the plot we can observe that there are no patterns in the data.
Therefore, the data doesn't fit the Cosine model either.

Model analysis Summary

- The residuals of both simple linear and quadratic are normally distributed and have significant autocorrelation.
- Among all the models, Quadratic model which fits almost 73.91% can be considered as a best fit.
- The degrees of freedom (s) are smaller for Quadratic model. This adds evidence that quadratic model is best.
- Even though Quadratic model is performed best, it did not show the complete trend in the series data.
- Also, the model failed to perform well in the residual analysis.
- Therefore, further analysis is to be made in the part 2 using ARIMA model and model specific tools.
- But at this point Quadratic model is the best model and the future data can be predicted using it.

Future predictions

Thickness change for the next five years (2017 - 2021):

```
h = 5
new = data.frame(v_time = seq((max(v_time) + 1), (max(v_time) + h), 1))
new$v_time2 = new$v_time^2

# Predicting from new data.
v_pred = predict(V_QM, new, interval = "prediction")

# Based on the prediction forming new data table for the next five years.
v_pred_tab = data.frame(Year = seq(2017, 2021, 1), v_pred)
colnames(v_pred_tab) = c("Year", "Prediction", "LowerCI", "UpperCI")
head(v_pred_tab)
##   Year Prediction   LowerCI   UpperCI
## 1 2017  -10.34387 -14.13556 -6.552180
## 2 2018  -10.59469 -14.40282 -6.786548
## 3 2019  -10.84856 -14.67434 -7.022786
## 4 2020  -11.10550 -14.95015 -7.260851
## 5 2021  -11.36550 -15.23030 -7.500701
```

Plotting the predicted data.

```
plot(v_ozone_thickness_1, type = 'o',
     main ="Next 5 year thickness change prediction using Quadratic model.",
     ylab = "Thickness change",
     xlab = "Years",
     xlim = c(1927, 2021),
     ylim = c(-15, 4),
     col = c("blue"), pch = c(19))
```

```
lines(ts(as.vector(v_pred[, 1]), start = 2017), col = "red", type = "l")
lines(ts(as.vector(v_pred[, 2]), start = 2017), col = "blue", type = "l")
lines(ts(as.vector(v_pred[, 3]), start = 2017), col = "blue", type = "l")

v_leg1("Forecast", c("Data", "Forecast limits", "Forecasts"), c("blue", "blue
", "red"), c(NA, NA, NA))
```

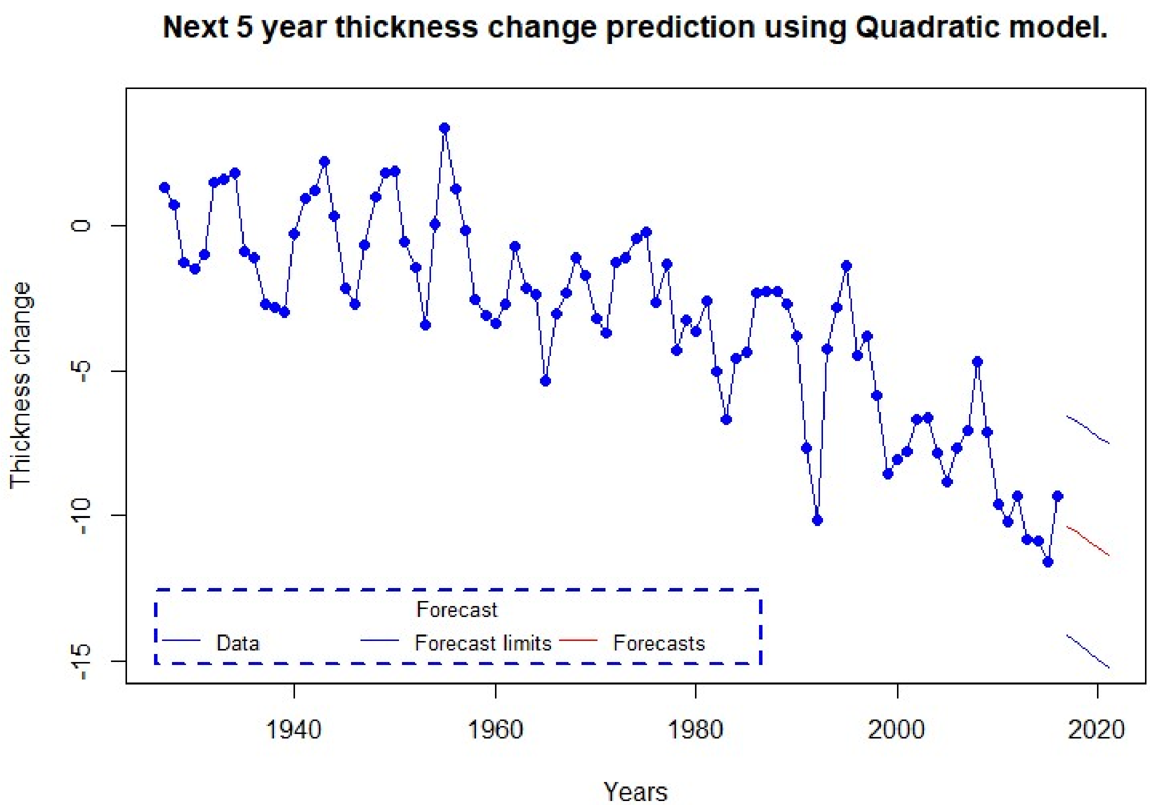

Fig 14: Next 5 - year forecast on the change in thickness in ozone layer.

From the five - year forecast results we can predict that there will be depletion in the ozone layer in the future.

Task 1 – Summary

- From the future prediction we can say that the ozone layer is going to deplete more in the next few years. Which acts as a sign of warning for both people and government organizations.
- Seasonality: There is no seasonality. Therefore, no cyclic / seasonal trend.
- Pattern: No pattern in the data. Hence no Cosine wave patterns.
- Behavior: AR/MA – Auto Regressive and Moving Average.
- Trend: Downward
- Measurements are not independent.
- The mean is not 0 and the variance is constant. Also, there is correlation between variables. Therefore, time series is not white noise.

- There is no need of applying any transformation to the data. But we will check this in Part 2 for checking variance change, which makes the time series stationary.
- There is change in mean. Which implies data is non – stationary.
- Since, non – stationary respective tests are to be conducted to confirm their stationary nature.
- Then differencing is to be applied to make the time series stationary. So that there will be no change in mean and variance of the series.
- Then the respective model specific tools are applied on the differenced data.
- The ARIMA orders are then obtained.

Task 2:

Introduction

In this section, using suitable model specifications a set of possible **Auto Regressive Integrated Moving Average - ARIMA** (p, d, q) models are proposed.

Testing for stationary

In the previous task, the Quadratic model interpretations suggested that the data is non - stationary. Since, mean is not constant.

Analyzing trends by plotting ACF and PACF.

```
acf(v_ozone_thickness_1, ci.type = 'ma', main = "Thickness change - ACF")
```

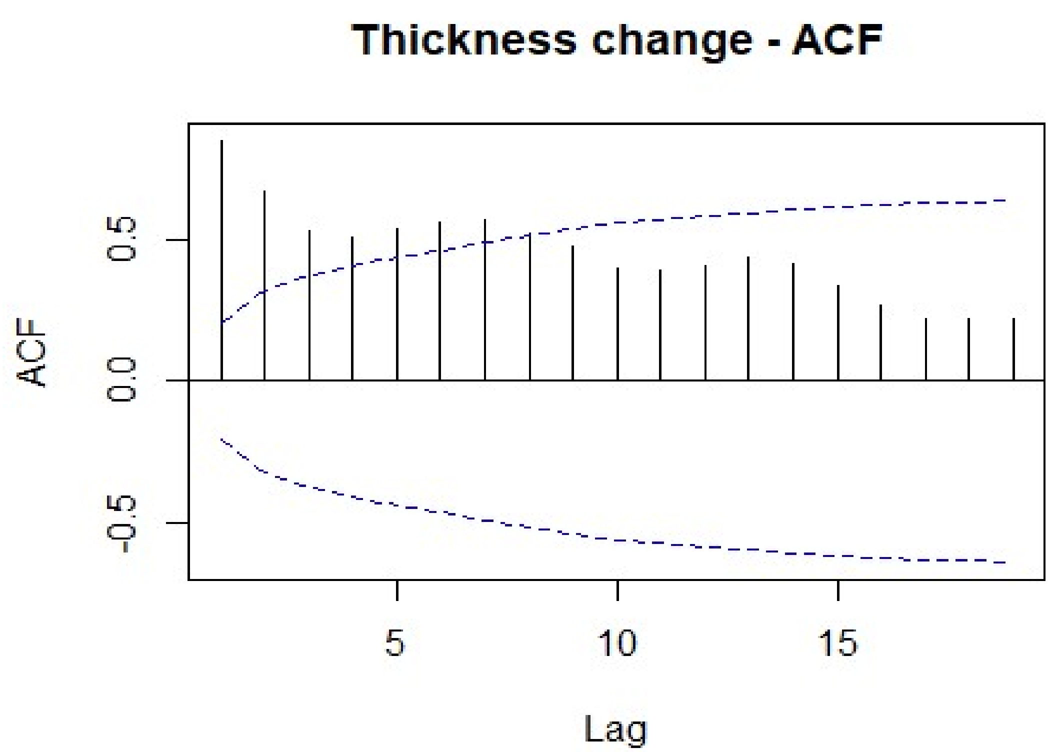

Fig 15: Thickness change with sophisticated error bonds – ACF.

```
pacf(v_ozone_thickness_1, main = "Thickness change - PACF")
```

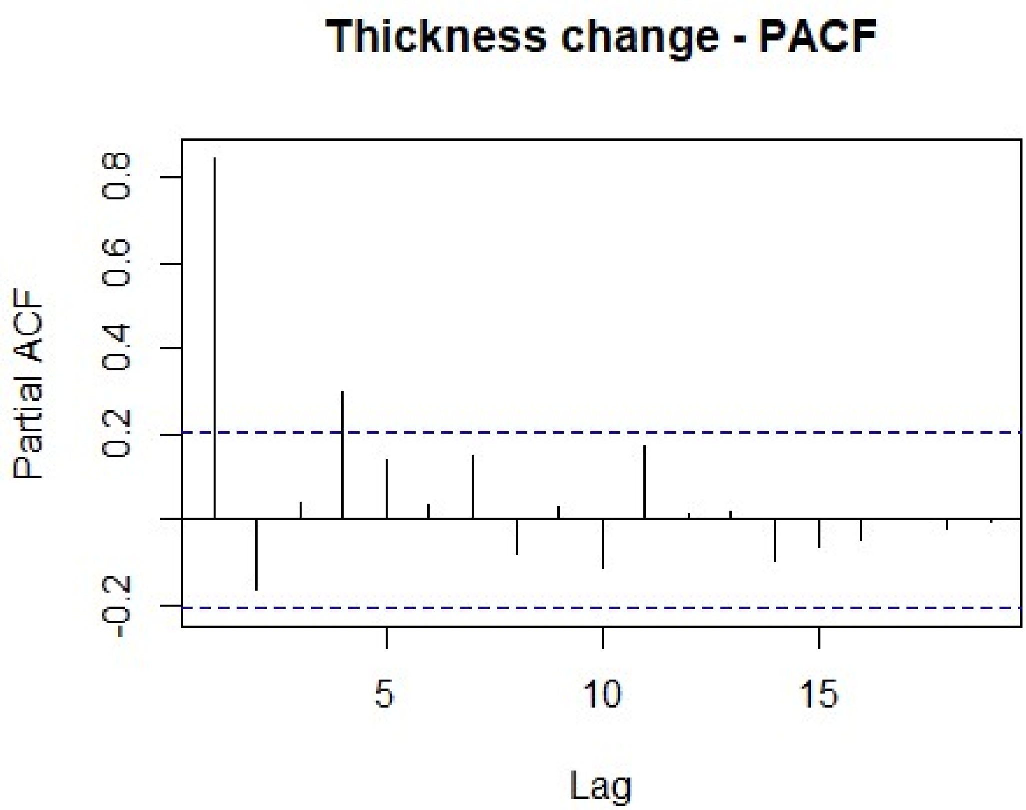

Fig 16: Thickness change – PACF.

The gradually decreasing ACF and PACF with a high peak at the start indicated the presence of randomness, which makes the plots significant. As the data is independent and identically distributed, we cannot say there is white noise based on randomness.

Now let us prove this non – stationary nature using Augmented Dickey-Fuller test. For this we need lags. These lags can be calculated using “ar()” function.

```
# Lag for ADF test
ar(v_ozone_thickness_1)
##
## Call:
## ar(x = v_ozone_thickness_1)
##
## Coefficients:
##       1        2        3        4
##  0.9807  -0.1417  -0.2581   0.2985
##
## Order selected 4  sigma^2 estimated as  3.242
```

Conducting Augmented Dickey-Fuller tests with 4 lags.

```
adfTest(v_ozone_thickness_1, lags = 4, type = "nc", title = 'No constant nor
Time Trend')
##
## Title:
##  No constant nor Time Trend
##
## Test Results:
##   PARAMETER:
##     Lag Order: 4
##   STATISTIC:
##     Dickey-Fuller: 0.5699
##   P VALUE:
##     0.7942
##
## Description:
##  Sun Apr 18 00:36:04 2021 by user: HP
adfTest(v_ozone_thickness_1, lags = 4, type = "c", title = 'With Constant but
no Time Trend')
##
## Title:
##  With Constant but no Time Trend
##
## Test Results:
##   PARAMETER:
##     Lag Order: 4
##   STATISTIC:
##     Dickey-Fuller: -0.4361
##   P VALUE:
##     0.8924
##
## Description:
##  Sun Apr 18 00:36:04 2021 by user: HP
adfTest(v_ozone_thickness_1, lags = 4, type = "ct", title = 'With constant an
d Time Trend')
##
## Title:
##  With constant and Time Trend
##
## Test Results:
##   PARAMETER:
##     Lag Order: 4
##   STATISTIC:
##     Dickey-Fuller: -3.2376
##   P VALUE:
##     0.0867
##
## Description:
##  Sun Apr 18 00:36:04 2021 by user: HP
```

Hypothesis:

H0: The data is not stationary.
HA: The data is stationary.

Interpretations:

No constant and no Time Trend:
p - value - 0.7942 > 0.5
With constant but no Time Trend:
p - value - 0.8924 > 0.5
With constant and Time Trend:
p - value - 0.0867 > 0.5
In all the 3 cases p - value is greater than 0.5 and hence the test is not statistically significant. Therefore, we fail to reject Null hypothesis i.e.,
The data is not stationary.

Changing variance or applying transformation.

In the previous analysis we already observed that the data has no change in variance, let us check whether applying transformation on the data changes its variance or not.

```
# Checking for Box_Cox transformation with best lambda
# Hear i used warning = FALSE. Hence, no warnings.
v_BoxCox <-  BoxCox.ar(v_ozone_thickness_1 + 13)
title(main = "Log-likelihood vs Lambda values.")
```

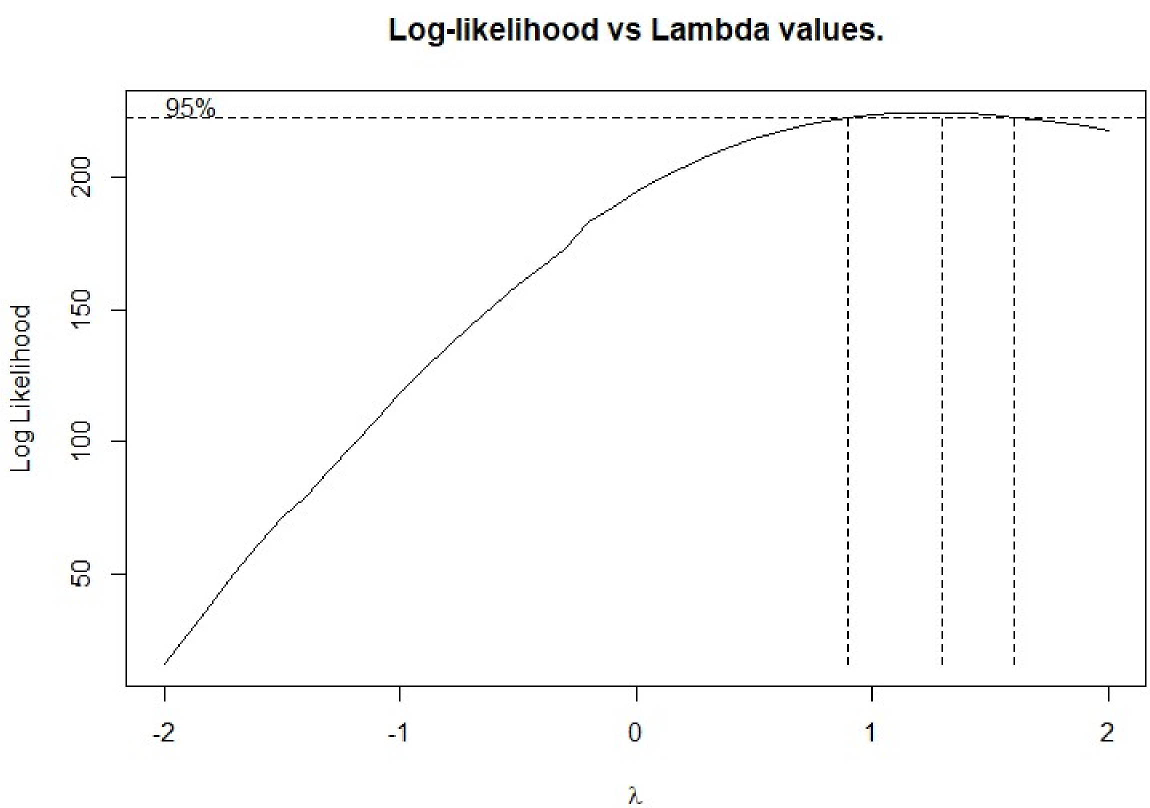

Fig 17: Finding the best lambda.

As the lambda confidence interval includes 1. There is no need to apply transformation on the data. This is already stated in visual analysis in the first part.

Differencing the data to make data stationary.

As no transformation is required, let us take the first difference of the data to make the data stationary.

```
v_diff = diff(v_ozone_thickness_1)

# Plot - First difference.
v_Plot(v_diff, "Differenced data of initial thickness change in Ozone layer."
)
v_leg("Differenced thickness change over years.", c("First ever differenced t
hickness Change"))
```

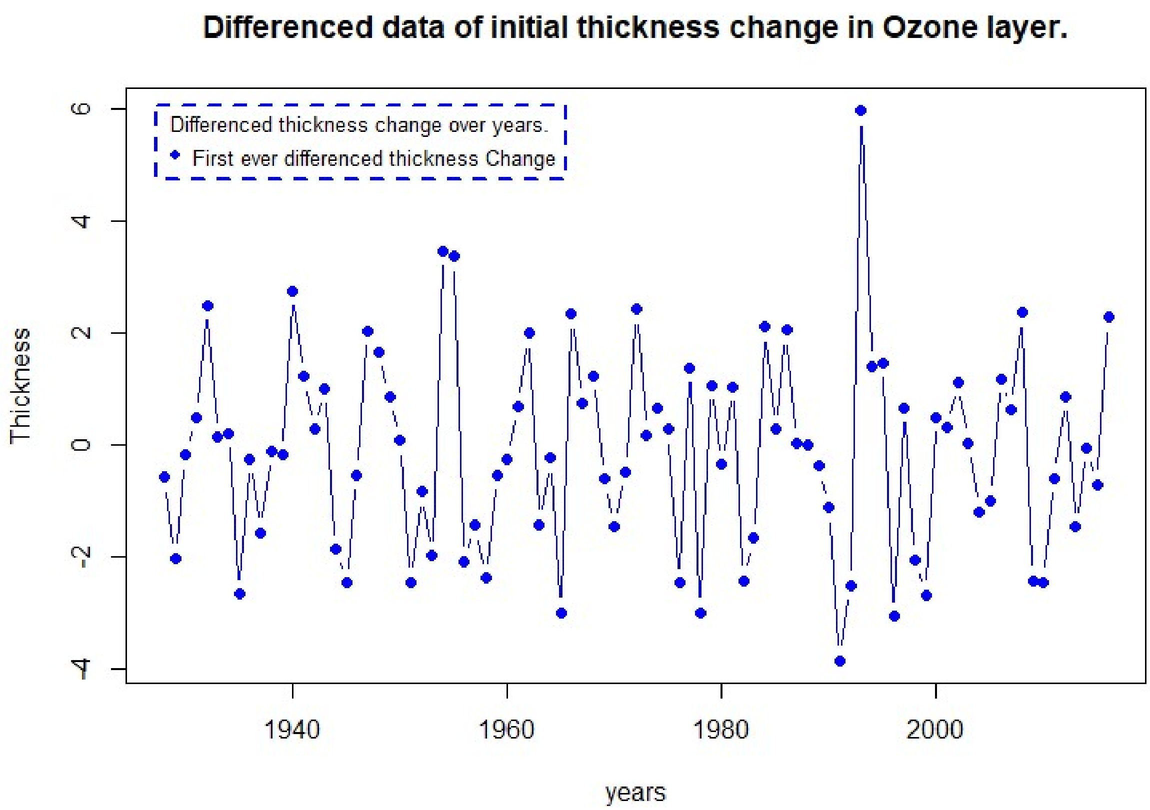

Fig 18: Change in thickness of ozone layer - First difference.

The data seems to be a stationary. To check whether the data is stationary or not apply ADF test. For this it is required to identify the lag on the first differenced data.

```
ar(v_diff)
##
## Call:
## ar(x = v_diff)
##
## Coefficients:
```

```
##       1       2       3       4       5       6
## -0.1976 -0.2628 -0.6019 -0.3064 -0.2253 -0.3045
##
## Order selected 6  sigma^2 estimated as  2.232
```

As, order/lags = 6. Conduct Augmented Dickey-Fuller tests on differenced data with 6 lags.

```
#conducting the ADF tests
adfTest(v_diff, lags = 6, type = "nc", title = 'No constant nor Time Trend')
## Warning in adfTest(v_diff, lags = 6, type = "nc", title = "No constant nor
Time
## Trend"): p-value smaller than printed p-value
##
## Title:
##  No constant nor Time Trend
##
## Test Results:
##   PARAMETER:
##     Lag Order: 6
##   STATISTIC:
##     Dickey-Fuller: -5.0757
##   P VALUE:
##     0.01
##
## Description:
##  Sun Apr 18 00:36:05 2021 by user: HP
adfTest(v_diff, lags = 6, type = "c", title = 'With Constant but no Time Tren
d')
## Warning in adfTest(v_diff, lags = 6, type = "c", title = "With Constant bu
t no
## Time Trend"): p-value smaller than printed p-value
##
## Title:
##  With Constant but no Time Trend
##
## Test Results:
##   PARAMETER:
##     Lag Order: 6
##   STATISTIC:
##     Dickey-Fuller: -5.5901
##   P VALUE:
##     0.01
##
## Description:
##  Sun Apr 18 00:36:05 2021 by user: HP
adfTest(v_diff, lags = 6, type = "ct", title = 'With constant and Time Trend'
)
## Warning in adfTest(v_diff, lags = 6, type = "ct", title = "With constant a
nd
## Time Trend"): p-value smaller than printed p-value
```

```
##
## Title:
##  With constant and Time Trend
##
## Test Results:
##   PARAMETER:
##     Lag Order: 6
##   STATISTIC:
##     Dickey-Fuller: -5.7327
##   P VALUE:
##     0.01
##
## Description:
##  Sun Apr 18 00:36:05 2021 by user: HP
```

Hypothesis:

H_0: The data is not stationary.

H_A: The data is stationary.

Interpretations:

No constant and no Time Trend:
p - value ~ 0.01 < 0.5
With constant but no Time Trend:
p - value ~ 0.01 < 0.5
With constant and Time Trend:
p - value ~ 0.01 < 0.5

P - value rounded to 0.01 as it is very small (exponential value.)
In all the 3 cases p - value is less than 0.5 and hence the test is statistically significant. The Null hypothesis i.e., "The data is not stationary" can be rejected. Therefore, the data is stationary.

Using model specific tools on differenced ozone data for model determination (ARIMA orders).

Using ACF and PACF

```
acf(v_diff, main = "Ozone layer thickness first difference data - ACF")
```

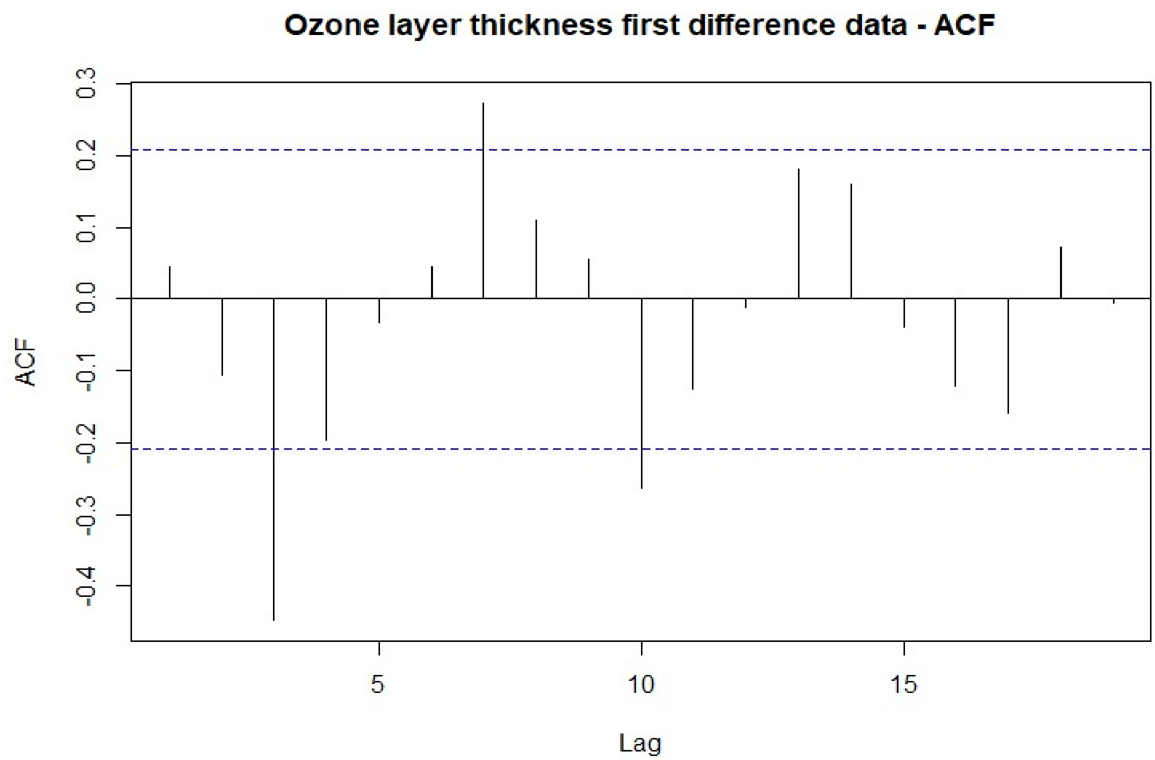

Fig 19: Differenced data – ACF.

```
pacf(v_diff, main = "Ozone layer thickness first difference data - PACF")
```

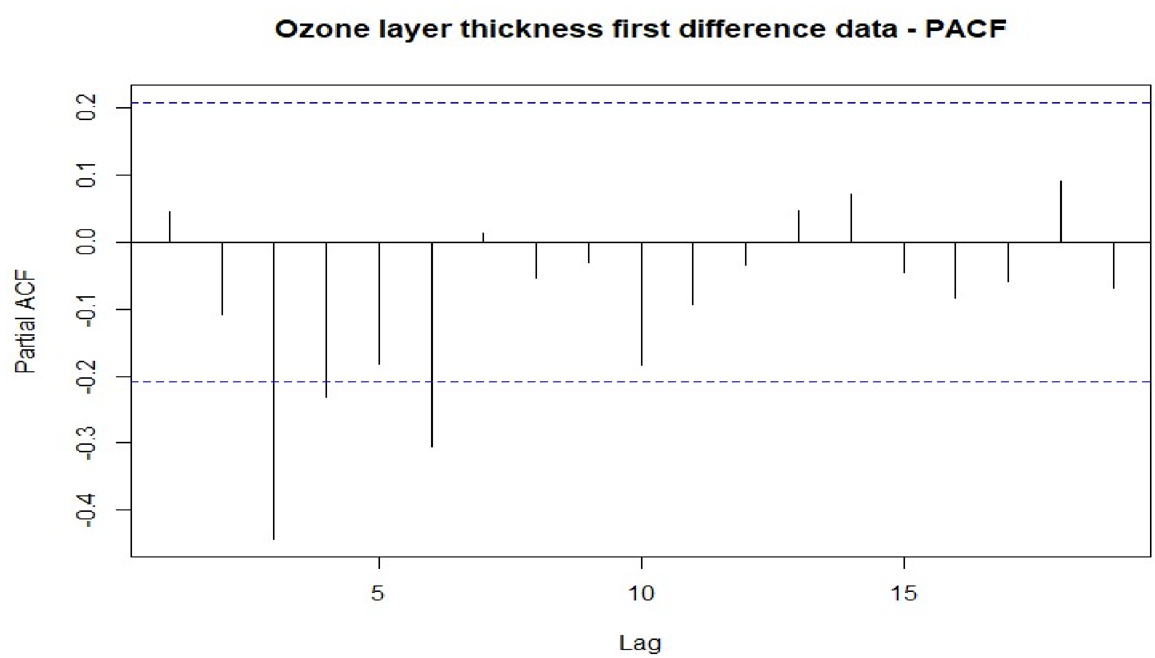

Fig 20: Differenced data – PACF.

From the plots we can find that ACF has an indefinite pattern. For ACF we have 3 significant vales are obtained at lags – 3, 7, 10 and for PACF lags are at 4, 5, 6.
Since, the plots have a damped sine wave, it is Auto Regressive (AR). We can have the model orders from PACF as ARI (3, 1) or ARIMA (3, 1, 3) when the significant values are considered. As, ARIMA (3, 1, 3) is the specified model all the lower orders can be specified.
Therefore, ARIMA (1, 1, 1), ARIMA (1, 1, 2), ARIMA (1, 1, 3), ARIMA (2, 1, 1), ARIMA (2, 1, 2), ARIMA (2, 1, 3), ARIMA (3, 1, 1), ARIMA (3, 1, 2), ARIMA (3, 1, 3) or they can be written as IMA (1, 3) and ARI (3, 1) can all be considered as ARIMA orders.

Using EACF

```
#calculating the eacf
eacf(v_diff, ar.max = 6, ma.max = 9)
## AR/MA
##   0 1 2 3 4 5 6 7 8 9
## 0 o o x o o o x o o x
## 1 x o x o o o o o o x
## 2 x o x o o o x o o x
## 3 x o x o o x o o o o
## 4 x o o x o x o o o o
## 5 x x x x o x o o o o
## 6 o o o x x o o o o o
```

- Since, p = 0 and q = 3 from the top left vertex the model can be MA (3). If we observe, we can have the adjacent model MA (4). Therefore, IMA (1, 3) and IMA (1, 4).
- Finally, the possible ARMA model could be - ARIMA (1, 1, 3) and ARIMA (1, 1, 4).

Using BIC

```
#creating the BIC plot
res = armasubsets(v_diff, nar = 9, nma = 9, y.name = 'ar', ar.method = 'ols')
## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.
in =
## force.in, : 3 linear dependencies found
## Reordering variables and trying again:
plot(res)
title(main = "Thickness difference - BIC", line = 6)
```

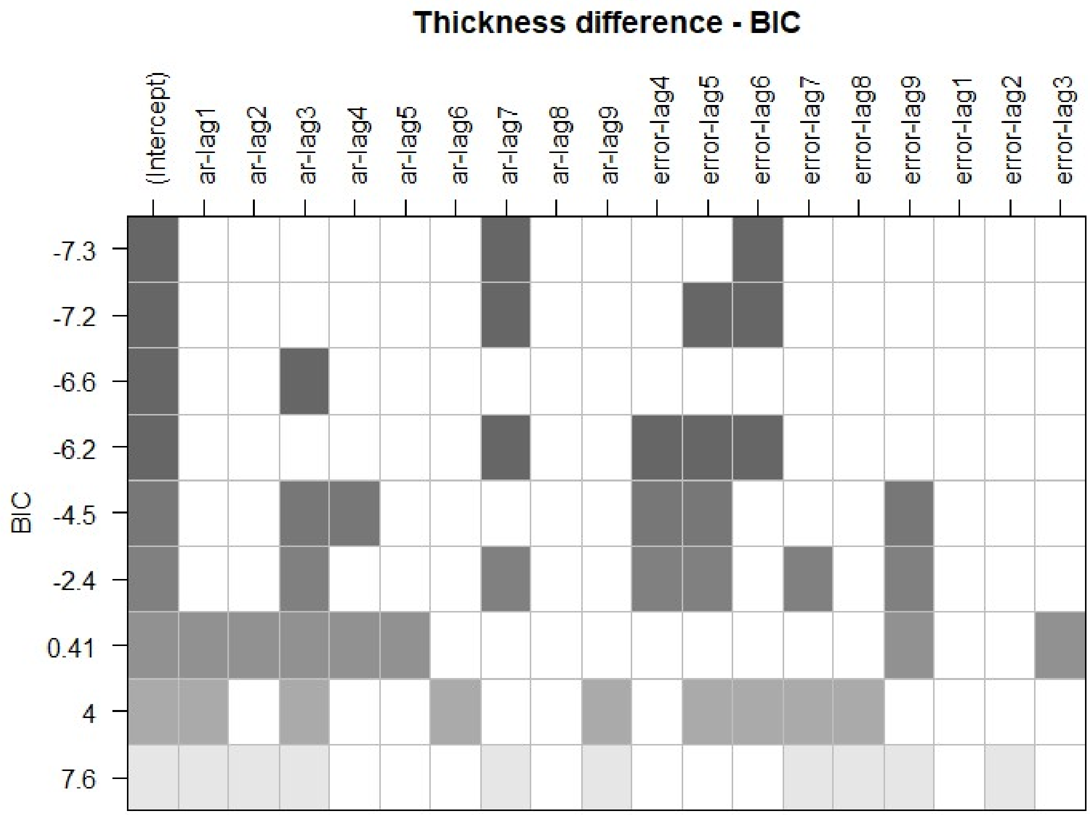

Fig 21: Differenced data – BIC Table.

From the BIC table model specified – ARIMA (7, 1, 6). Since, from the shaded blocks we get AR (7) and MA (6)

Hence to model the ozone data, we can include ARIMA (7,1,6) as a potential model.

Task 2 - Summary

- The data is non – stationary and is converted to stationary by differencing.
- Then the respective tools are used to specify the orders and determine the models.
- Possible models,
  - **ACF – PACF**
    - ARIMA (1, 1, 1), ARIMA (1, 1, 2), ARIMA (1, 1, 3), ARIMA (2, 1, 1), ARIMA (2, 1, 2), ARIMA (2, 1, 3), ARIMA (3, 1, 1), ARIMA (3, 1, 2), ARIMA (3, 1, 3) or they can be written as IMA (1, 3) and ARI (3, 1)
  - **EACF**
    - ARIMA (1, 1, 3), ARIMA (1, 1, 4), IMA (1, 3), and IMA (1, 4)
  - **BIC table**
    - ARIMA (7,1,6)

Both the tools ACF -PACF and EACF proposed ARIMA (1, 1, 3) and IMA (1, 3). Therefore, we can conclude that the ARMA models suitable for the analysis could be ARIMA (1, 1, 3) and IMA (1, 3)

Conclusion

- Quadratic model is the best model on this series data.
- Future predictions suggests that the ozone layer is going to be depleted.
- The best models specified from the tools are, ARIMA (1, 1, 3) and IMA (1, 3).
- The model specification tools are not so efficient and further analysis is required.

References

[1] H. R. a. M. Roser, "Ozone Layer," 2018. [Online]. Available: https://ourworldindata.org/ozone-layer. [Accessed 2021].

[2] K. H. [. c. B. L. [. (. t. c. Adrian Trapletti [aut], "CRAN," 2020. [Online]. Available: https://cran.r-project.org/web/packages/tseries/index.html. [Accessed 2021].

[3] T. S. [. Y. C. [. Diethelm Wuertz [aut], "CRAN," 2017. [Online]. Available: https://cran.r-project.org/web/packages/fUnitRoots/index.html. [Accessed 2021].

[4] "Ozone layer depletion," 2021. [Online]. [Accessed 2021].