**TIME SERIES ANALYSIS**

**REPORT**

Shamini Puthooppallil Baby

[Shaminibaby000@gmail.com](mailto:Shaminibaby000@gmail.com)

Masters In Data Science

Contents

[Introduction 3](#_Toc4940984)

[Data Preparation 4](#_Toc4940985)

[Time Series Analysis 5](#_Toc4940986)

[Time Series Plot Analysis and Findings 5](#_Toc4940987)

[Model Building -Regression Approach 7](#_Toc4940988)

[Linear Trend 7](#_Toc4940989)

[Quadratic Trend 8](#_Toc4940990)

[Cyclical or Seasonal trend. 9](#_Toc4940991)

[Data Splitting 11](#_Toc4940992)

[Basic Diagnostic Checking 14](#_Toc4940993)

[Linear model 14](#_Toc4940994)

[Quadratic model 14](#_Toc4940995)

[Split Model 15](#_Toc4940996)

[Residual Analysis 16](#_Toc4940997)

[Linear model 16](#_Toc4940998)

[Quadratic model 19](#_Toc4940999)

[Prediction 22](#_Toc4941000)

[Conclusion 23](#_Toc4941001)

[References 24](#_Toc4941002)

[Appendix 25](#_Toc4941003)

[1. Required Libraries 25](#_Toc4941004)

[2. Pre validations 25](#_Toc4941005)

[3. Time series plot of R time series object 25](#_Toc4941006)

[4. Plot - relationship between pairs of consecutive ozone thickness: 26](#_Toc4941007)

[5. Linear Trend 26](#_Toc4941008)

[6. Quadratic Trend 26](#_Toc4941009)

[7. Cyclic or Seasonal Trend 27](#_Toc4941010)

[8. Splitting Data 28](#_Toc4941011)

[9. Residual Analysis- Linear model 29](#_Toc4941012)

[10. Residual Analysis - Quadratic model 30](#_Toc4941013)

[11. Prediction 30](#_Toc4941014)

## Introduction

This report describes time series analysis of yearly changes in the thickness of Ozone layer from 1927 to 2016 in Dobson units. A negative value in the dataset represents a decrease in the thickness and a positive value represents an increase in the thickness.

The main aim is to analyse the data by using the Time series analysis methods. The analysis trying to find the best fitting trend model to the ozone thickness dataset and give predictions of yearly changes for the next 5 years.

The analysis starts with a data cleansing part followed by model specification by descriptive analysis. Various models will be testing using statistical methods and out of the fitted model, a suitable model will be selected for future prediction values.

## Data Preparation

Before starting the data analysis and modelling, the main duty of the data scientist is to make sure that the data provided is in correct format. If the dataset is not in proper format, the entire work needs to be repeated.

The required libraries for the time series analysis is in [Appendix 1.](#_Required_Libraries)

Dataset is loaded to the R software using read.csv function and perform required pre validations for the loaded dataset. The analysis is performing using the R Markdown and the common statistical tools in the upcoming sessions.

Structure of the dataset, column values, null values and impossible values are checked using some basic R data preprocessing packages. Sample data is viewed and make sure that the data loaded correctly to the R software. This code snippet is available in [Appendix 2](#_Pre_validations).

## Time Series Analysis

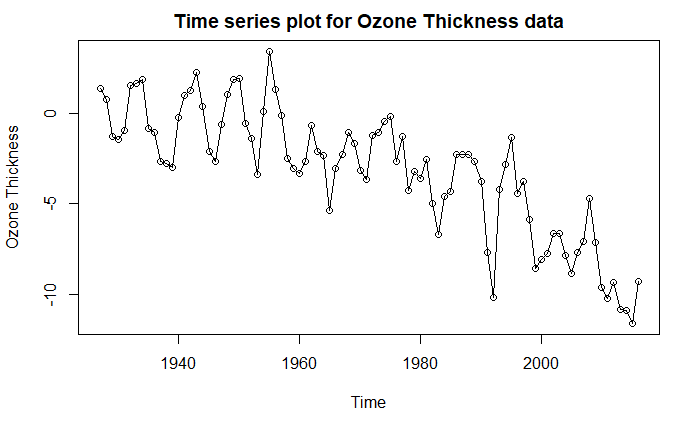
In this stage, the dataset is analysed using more statistical and descriptive tools to get more insights of the time series characteristics.

The data analysis stage considered the time series characteristics of the dataset and figured out the elements of suitable and successful data analysis of time series data. The dataset needs to be a R time series object to perform the time series analysis. So, the dataset converted to timeseries object using the TSA package function.

The ts() function will convert the dataframe into an R time series object. Next step is to plot the time series dataset and analyses the common characteristics of time series plot.

You can find the code snippet in [Appendix 3](#_Time_series_plot)

### Time Series Plot Analysis and Findings

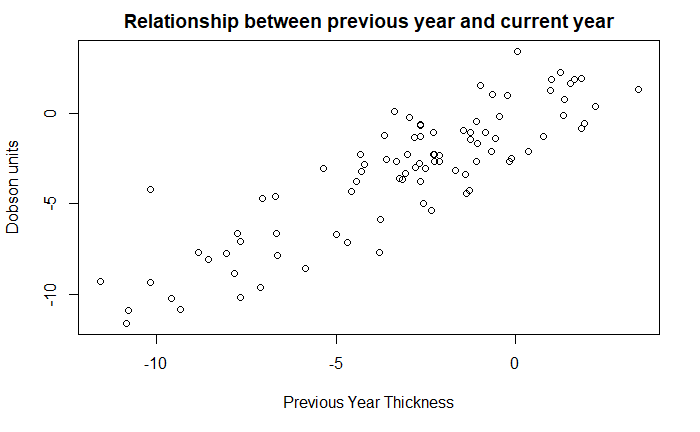


We observe considerable variation in thickness of ozone value over the years from this time series plot. The year 1955 reported the highest ozone thickness, while 2015 was the lowest. Also, there is no clear intervention points in the plot which shows a sudden upward or downward change. We cannot spot any obvious changing in variance from the above plot. The data set is not showing full seasonality during this year period but if we observe clearly, the first2,3 decades shows aseasonality with a frequency of 7 years. And we can say the overall plot depicts a downward trend from 1927 to 2016.

The above are the assumptions we made from the first glance of the time series plot. We will analyse the above points further in the subsequent sessions.

To investigate more on the relationship between pairs of consecutive ozone thickness a scatter plot is drawn with current year and previous year data in the ozone thickness dataset.

The code chunk for this relationship is available in [Appendix 4.](#_Plot_-_relationship)



This plot shows a strong upward trend. We observe a high correlation between thickness of succeeding years However, it is impossible to observe seasonality for this scatter plot. Here, neighboring values are very closely related. Large changes in thickness do not occur from one year to the next. Correlation between the previous year observation and the current year observation has the value 0.8700381 which is a strong one.

## Model Building -Regression Approach

In this approach we are trying to fit our dataset to the best suitable model. The main objective of this process is to find a suitable model which is appropriate and satisfy most of the statistically significant tests and assumptions.

### Linear Trend

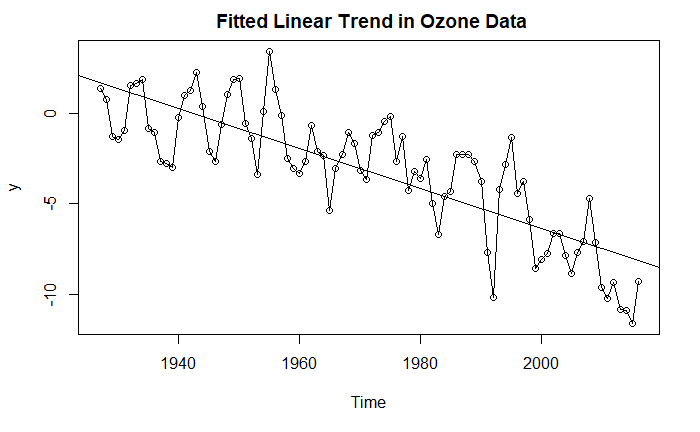
Initially the data is fitted to a linear trend model and analyzed the characteristics and compared these characteristics with the rest of the models.

## Call:  
## lm(formula = ozone\_thickness ~ time(ozone\_thickness))  
##   
## 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(ozone\_thickness) -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

Estimates of slope and intercept are -0.110029 and 213.720155 respectively. Here slope and intercept are statistically significant at 5% significance level.

After fitting the model, a trend line is drawn along with the original dataset to analyze the behavior of the linear model.

The R codes used for linear model is available in [Appendix 5](#_Linear_Trend)

 The overall performance of the trend line is good. But we can observe that at the end, the mean level is not captured by the trend line. So, we can try some better models instad of linear.

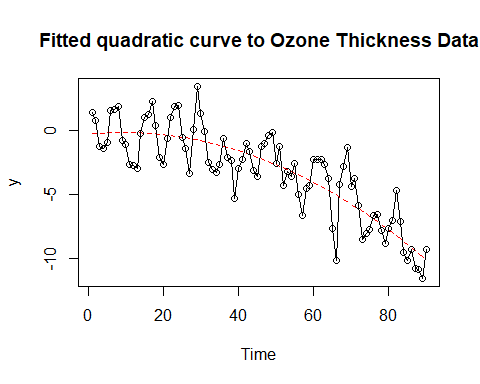
### Quadratic Trend

Next the data is fitted to a quadratic trend model and analyzed the characteristics and compared these characteristics with the rest of the models.

## lm(formula = ozone\_thickness ~ 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

Estimates of intercept and t^2 are -5.733e+03 ,5.924e+00 and -1.530e-03 respectively. These 3 values are statistically significant at 5% significance level.Quadratic trend term is found significant.The value of multiple R-squared 0.7391 which is slightly higher than the linear trend model.

Fitted quadratic trend is shown below:



In this quadratic model, it’s clear that the fitted curve captures the mean level throughout the time period. So, this model may be fit in this dataset more suitable than linear model.

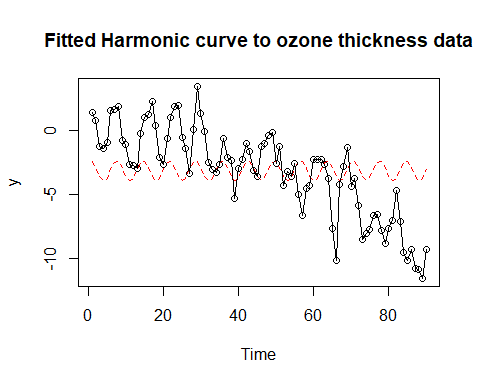
The R codes used for linear model is available in [Appendix 6.](#_Quadratic_Trend)

### Cyclical or Seasonal trend.

The frequency of the time series object created with TSA package. Many models are tried using different frequencies. One of the examples with frequency 7 is shown below.

## lm(formula = ozone\_thickness ~ har.)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.8294 -1.8422 0.7481 2.4701 5.8635   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.1949 0.3700 -8.636 2.5e-13 \*\*\*  
## har.cos(2\*pi\*t) 0.7386 0.5226 1.413 0.161   
## har.sin(2\*pi\*t) -0.2544 0.5239 -0.486 0.628   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.509 on 87 degrees of freedom  
## Multiple R-squared: 0.02487, Adjusted R-squared: 0.002453   
## F-statistic: 1.109 on 2 and 87 DF, p-value: 0.3344

The R codes used for linear model is available in [Appendix 7](#_Cyclic_or_Seasonal)



None of the values fitted with significant values. So, there is no point of modelling the ozone data with harmonic or cosine trends. Since, [the time series plot of ozone data](#linear-trend) doesn’t shows any notable overall seasonal or cyclic trends, these model fitting can be ignored in this stage.

### Data Splitting

The data shows almost sesonal trend during the first 3,4 decades and then lost the seasonality.

Here the data is divided into two parts according to the trend. The first part is fitted with harmonic trend and the second part is fitted with quadratic trend.

##### 

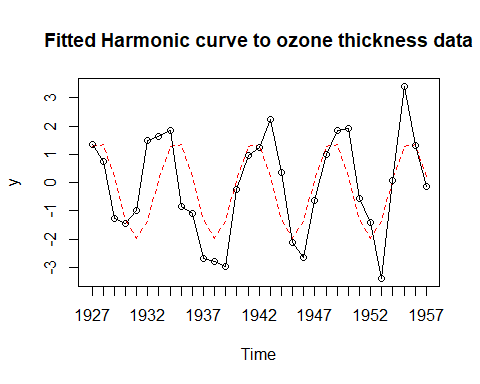
#### Split Data-Harmonic Trend

The first part is fitted with harmonic trend and the result of the harmonic trend shown below.

## lm(formula = ozone\_thickness1 ~ har.)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.2054 -0.7580 -0.1387 0.6408 2.9017   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.2403 0.2366 -1.016 0.3185   
## har.cos(2\*pi\*t) 1.5316 0.3343 4.582 8.7e-05 \*\*\*  
## har.sin(2\*pi\*t) 0.8329 0.3333 2.499 0.0186 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.311 on 28 degrees of freedom  
## Multiple R-squared: 0.4957, Adjusted R-squared: 0.4597   
## F-statistic: 13.76 on 2 and 28 DF, p-value: 6.883e-05

From the first part, it’s clear that the data from 1927 to 1957 follows harmonic trend with seasonality. We can observe a cyclic trend with frequency of 7. Since, from 1953 the time series plot shows different trend.So, we must analyze the rest of the dataset with other modelling techniques.

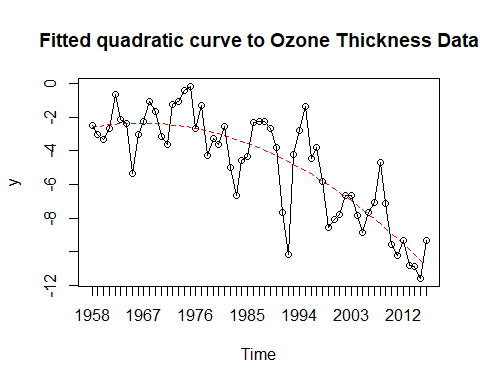
The R codes used for harmonic model is available in [Appendix 8](#_Splitting__Data)



#### Split Data-Quadratic Trend

The second part is fitted with quadratic trend and the result of this trend shown below.

## lm(formula = ozone\_thickness2 ~ t + t2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.5591 -0.9004 0.0660 1.3471 3.8072   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.333e+04 3.395e+03 -3.927 0.000238 \*\*\*  
## t 1.356e+01 3.417e+00 3.967 0.000209 \*\*\*  
## t2 -3.447e-03 8.599e-04 -4.008 0.000183 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.713 on 56 degrees of freedom  
## Multiple R-squared: 0.7042, Adjusted R-squared: 0.6937   
## F-statistic: 66.67 on 2 and 56 DF, p-value: 1.535e-15



From the s second part, it’s clear that the data from 1958 to 2016 the graph follows almost a perfect quadratic trend. We can observe almost prefect quadratic trend here.

The R codes used for harmonic model is available in [Appendix 8](#_Splitting__Data)

## 

## Basic Diagnostic Checking

**(Standard Deviation, Coefficient of determination, Probabilistic value)**

This session includes the diagnostic checks to validate the accuracy of each nodel by some statistical parameters.

Usually for the Residual standard deviation(s),the smaller the value of s, the better the fit for the model. Also, for coefficient of determination(R^2), high value but not close to 1 values of R^2 implies a satisfactory fit for the selected model. Likewise, small probability - value (p-value) (typically <=0.05) indicates strong evidence against the null hypothesis, so you reject the null hypothesis.

### Linear model

**Residual standard deviation: 2.032 on 88 degrees of freedom**

**coefficient of determination: Multiple R-squared: 0.6693, Adjusted R-squared: 0.6655**

**probability-value: < 2.2e-16**

According to multiple R^2, about 66.9% of the variation in the linear model series is explained by the linear time trend. The adjusted version of multiple R^2 provides an approximately unbiased estimate of true R^2(66.55%). In this linear model, slope and intercept is statistically signicant at 5% signicance level(< 2.2e-16).

### Quadratic model

**Residual standard deviation: 1.815 on 87 degrees of freedom**

**coefficient of determination: Multiple R-squared: 0.7391, Adjusted R-squared: 0.7331**

**probability-value: < 2.2e-16**

According to multiple R^2, about 73.91% of the variation in the linear model series is explained by the quadratic time trend. The adjusted version of multiple R^2 provides an approximately unbiased estimate of true R^2 (73.31%) In this quadratic model, Intercept ,t and t^2 is statistically signicant at 5% signicance level(1.16e-05,8.30e-06 and 5.87e-06 respectively).

From the above basic diagnostic checking, its clear that the quadrartic trend is more suitable for ozone thickness dataset. Because the coificient of determination and and the probability value says that quadratic model is more significant and best fit for this dataset.

### Split Model

#### Harmonic Trend

**Residual standard error: 1.311 on 28 degrees of freedom**

**Multiple R-squared: 0.4957, Adjusted R-squared: 0.4597**

**F-statistic: 13.76 on 2 and 28 DF,**

**p-value: 6.883e-05**

According to multiple R^2, about 49.57% of the variation in the linear model series is explained by the harmonic time trend. The adjusted version of multiple R^2 provides an approximately unbiased estimate of true R^2 (45.97%) In this harmonic model, har.cos(2pit) and har.sin(2pit) are statistically signicant at 5% signicance level. But here the intercept value is not statistically significant. I have tried to split the data with maximum possibilities. All the cases show insignificant intercept values for harmonic plot. So, it’s not suitable to fit this model. Also, the accuracy of fit is less than 50 %. Even though the plot shows the harmonic trend, the statistical analysis shows insignificant values.

#### Quadratic Trend

**Residual standard deviation: 1.713 on 56 degrees of freedom**

**coefficient of determination: 0.7042, Adjusted R-squared: 0.6937**

**probability-value: 1.535e-15**

According to multiple R^2, about 70.42% of the variation in the linear model series is explained by the quadratic time trend. The adjusted version of multiple R^2 provides an approximately unbiased estimate of true R^2 (69.37%) In this quadrtatic model, intercept,t1 and t^2 are statistically signicant at 5% signicance level (0.000238,0.000209,0.000183 respectively).

Even if the second splitted part fit nicely, the harmonic trend doesnt shows a significant fit.

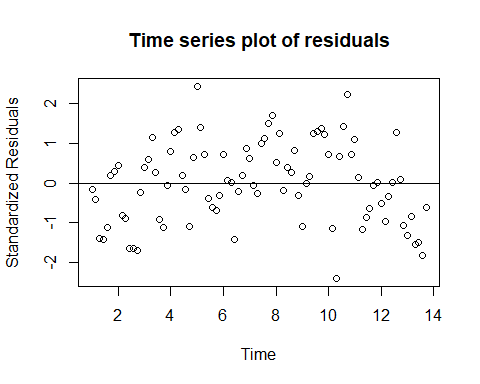
**The above basic diagnostics from the all possible outcomes shows that the quadratic modelling trend is more suitable for ozone thickness dataset.**

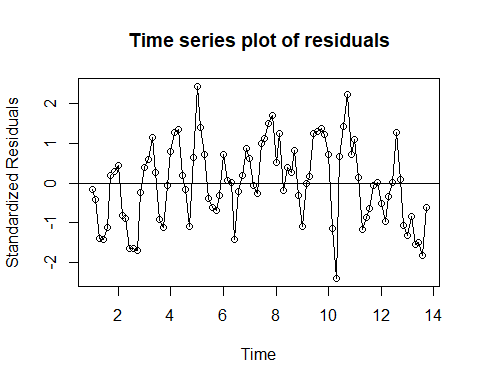
## Residual Analysis

Error on the output need to be interpreted carefully. If the trend model is reasonably correct, then the residuals should behave roughly like the true stochastic component, and various assumptions about the stochastic component can be assessed by looking at the residuals

### Linear model

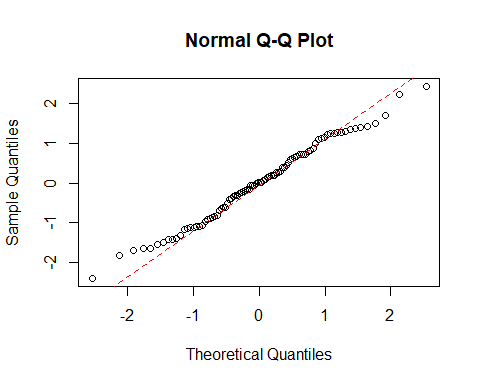
The following figure shows standardized residuals from linear model of the ozone thickness data fitted by means of the data (1 with trend line and other without the trend line):





we would expect a plot to suggest a rectangular scatter with no discernible trends whatsoever. There is a striking departures from randomness in this plot in the last 6 years.So we cannot assume that the residuals follow randomness.

The QQ plot for the normality check is shown below.

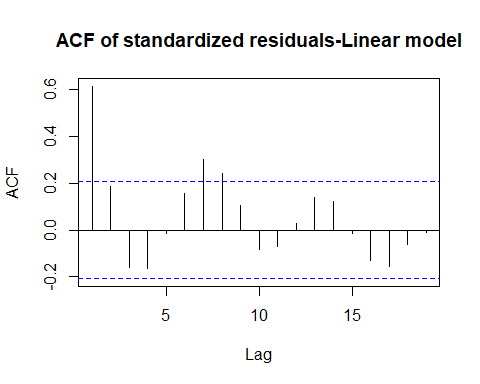


There is a fluctuation from the noraml line from -2 to -1 and from 1 to 3. From -1 to 1 the residuals follow the normal line This points are far away from the normal line.This means that the residuals does not follow the assumption of normality.

## Shapiro-Wilk normality test  
##   
## data: res.model.ozone\_thickness.ln  
## W = 0.98733, p-value = 0.5372

Shapiro test shows that the residuals follow normality and the P value is .5372 means the test is significant

Autoregressive function very useful and important tool in the analysis of time series data.

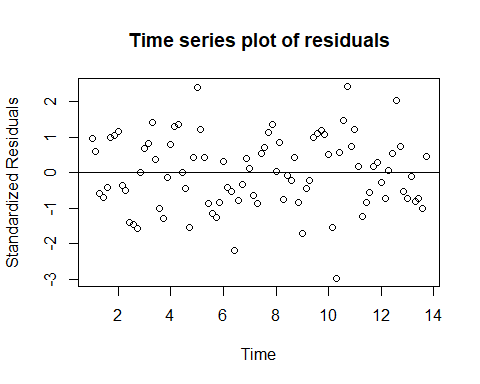


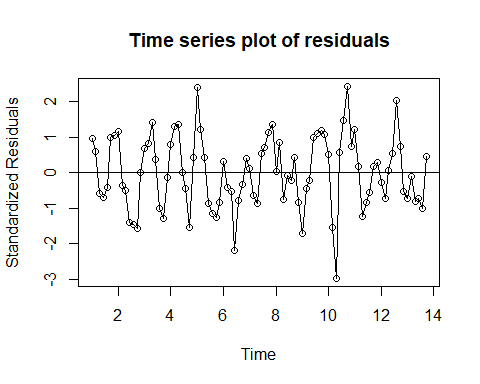
There is a highly insignificant value in the beginnig in the acf plot. Also, we can see that two more insignificant residual values at 7 and 8 th position. Most of the significant values are near -2and 2 except 3,4 low values.

The R codes used for residual analysis of linear model is available in [Appendix 9](#_Residual_Analysis-_Linear)

### Quadratic model

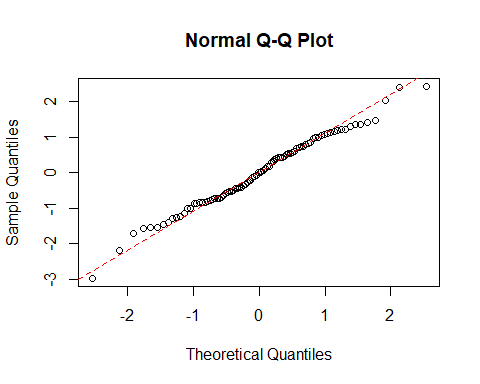
The following chunk generates a time series plot for the standardized residuals from quadratic model of the ozone thickness data fitted by means of the data:





There is a point out of -3 between the years 1980 and 2000. Apart from that value, we can clearly say that the residual fit randomly around the rectange in the above This plot fits the residual more randomly than the linear trend.

The normality check for the residual is done by plotting the QQ plot below.

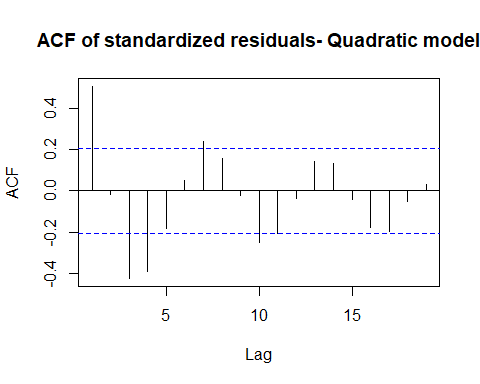


There is a small fluctuation from the normal line from -0 to -1 and from 1 to 3. From 1 to 1 the residuals follow the normal line. And these fluctuations are less as compared to linear model. Even though, there is a fluctuation from the normality, the value of the fluctiatin is less as compared to linear trend.

##   
## Shapiro-Wilk normality test  
##   
## data: res.model.ozone\_thickness.qa  
## W = 0.98889, p-value = 0.6493

Shapiro test shows that the residuals follow normality and the P value is 0.6493 means the test is significant. Also, W value for this test is greater than linear model.

The R codes used for residual analysis of quadratic model is available in [Appendix 10](#_Residual_Analysis_-)



As compared with the linear model, we can see that there is 4 to 5 values which are not significant.But the values are <=.04 (It was .06 in the case of linear)Also we can see that the most of the significant residual value is closer to 0.So,from this findings, I am considering this quadratic model is more suitable for the ozone data.

**Note:** Since this Analysis is covering only the basic models such as linear, quadratic and harmonic etc. Scope of this report is limited to these models. The acf plot shows a cyclic wave like structure. So we could have analyze more on other models (AR,MA,ARMA).

## Prediction

After ensuring that the fitted model is suitable for prediction purposes, we use the model to find forecasts. For time series regression models, this task is simply based on the straightforward use of the fitted regression model.

Predicted values from the quadratic trend model is showing below.

fit lwr upr

1 -10.34387 -14.13556 -6.552180

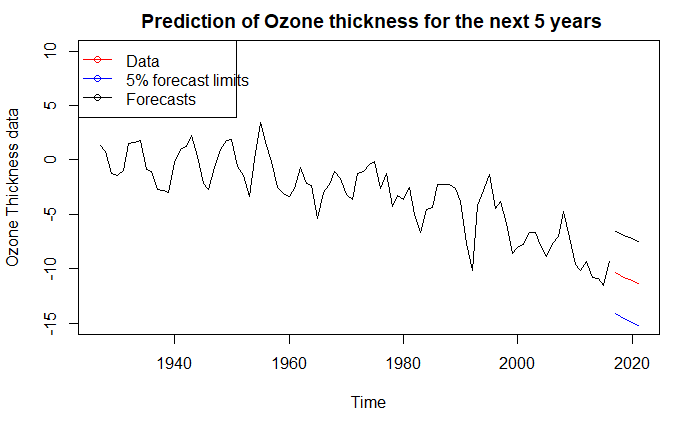
2 -10.59469 -14.40282 -6.786548

3 -10.84856 -14.67434 -7.022786

4 -11.10550 -14.95015 -7.260851

5 -11.36550 -15.23030 -7.500701

We can plot these forecasts next to the time series of interest as shown below.



Forecasts from the quadratic model successfully follow the repeating pattern in the original series. Here the overall trend of the series going down.The forecasts and the 5% forecast limits shows the same downward trend from 2017 to 2021.

So, we can conclude that quadratic model fits the data almost perfectly as compared to another models.

## Conclusion

The goal of the time series analysis is to find the best fitting trend model to this dataset and give predictions of yearly changes for the next 5 years. The dataset is provided with ozone thickness with 90 observations. During the initial analysis, the time series plot shows an upward trend with no seasonal objects and no interventions.

Model specification and model building performed on different trends such as linear, quadratic and harmonic. For better understanding of the dataset, the data is divided into two and perform separate model fitting for each part. The split model and harmonic model were rejected due to poor model performance.

Residual analysis is performed for the remaining linear and quadratic model. The statistical tests and significance tests show that the quadratic model is more suitable for ozone dataset. So, quadratic model fitted for the dataset.

From the selected quadratic model, the ozone thickness for the next 5 years is predicted, and the results are plotted in the dataset. The predicted values almost fit the selected model and concluded that quadratic model is the more suitable fit in this time series analysis of ozone thickness dataset.

## References

1. Time Series Analysis With Applications in R(Jonathan D. Cryer, Kung-Sik Chan)
2. <https://cran.r-project.org/web/packages/HarmonicRegression/HarmonicRegression.pdf>
3. <https://handbook.unimelb.edu.au/subjects/ecom30004>

# Appendix

## Required Libraries

#Required libraries  
library(readr)  
library(TSA)  
library(dplyr)  
library(ks)

## Pre validations

data1 <- read.csv("C:/Users/SIA/Desktop/SEM 3/Time series Analysis/data1.csv",header=FALSE)  
names(data1)[1]<-paste("Thickness")  
#View the sample  
head(data1)

## Thickness  
## 1 1.3511844  
## 2 0.7605324  
## 3 -1.2685573  
## 4 -1.4636872  
## 5 -0.9792030  
## 6 1.5085675

#check the structure  
str(data1)

## 'data.frame': 90 obs. of 1 variable:  
## $ Thickness: num 1.351 0.761 -1.269 -1.464 -0.979 ...

#check for impossible and null values  
data1[is.na(data1$Thickness),]

## numeric(0)

data1[is.nan(data1$Thickness),]

## numeric(0)

## Time series plot of R time series object

The time series plot for this data is generated with the following code chunk

#convert dataset to R time series object

ozone\_thickness <- ts(as.vector(data1), start=1927, end=2016)

#generate time series plot for the imported data

plot(ozone\_thickness,type='o',ylab='Ozone Thickness')

## Plot - relationship between pairs of consecutive ozone thickness:

The following code chunk generates a scatter plot to investigate the relationship between pairs of consecutive ozone thickness:

#Plot the relationship diagram

plot(y=ozone\_thickness,x=zlag(ozone\_thickness),ylab='Inches', xlab='Previous Year Thickness')

#Correlation between previous year and current year

x=ozone\_thickness   
y=zlag(ozone\_thickness)  
index=2:length(x)  
cor(y[index],x[index])

## [1] 0.8700381

## Linear Trend

The below R chunk will fit the data into a linear model.

model.ozone\_thickness.ln = lm(ozone\_thickness~time(ozone\_thickness))   
summary(model.ozone\_thickness.ln)

The trend line is plotted Using the below R code:

plot(ozone\_thickness,type='o',ylab='y')  
plot.new

abline(model.ozone\_thickness.ln)

## Quadratic Trend

The below R chunk will fit the data into a quadratic model.

# create the time variables for quadratic trend

t = time(ozone\_thickness)  
t2 = t^2

# label the quadratic trend  
model.ozone\_thickness.qa = lm(ozone\_thickness~ t + t2) summary(model.ozone\_thickness.qa)

The trend curve is plotted Using the below R code:

# Fitted quadratic trend

plot(ts(fitted(model.ozone\_thickness.qa)), ylim = c(min(c(fitted(model.ozone\_thickness.qa),  
 as.vector(ozone\_thickness))), max(c(fitted(model.ozone\_thickness.qa),as.vector(ozone\_thickness)))),  
 ylab='y' , main = "Fitted quadratic curve to Ozone Thickness Data", type="l",lty=2,col="red")  
plot.new

lines(as.vector(ozone\_thickness),type="o")

## Cyclic or Seasonal Trend

The below R chunk will fit the data into a harmonic model.

# set frequency 7

ozone\_thickness <- ts(as.vector(data1), freq = 7)

# Fitted harmonic trend

har.=harmonic(ozone\_thickness,1)  
model.ozone\_thickness.har=lm(ozone\_thickness~har.)

# view the summary

summary(model.ozone\_thickness.har)

The fitted values are plotted against the original value using the below code.

plot(ts(fitted(model.ozone\_thickness.har)), ylim = c(min(c(fitted(model.ozone\_thickness.har),  
 as.vector(ozone\_thickness))), max(c(fitted(model.ozone\_thickness.har),as.vector(ozone\_thickness)))),  
 ylab='y' , main = "Fitted Harmonic curve to ozone thickness data", type="l",lty=2,col="red")  
plot.new

lines(as.vector(ozone\_thickness),type="o")

## Splitting Data

**Harmonic trend**

# data split

ozone\_thickness1<-data1[1:31,

# fit harmonic model  
ozone\_thickness1 <- ts(as.vector(ozone\_thickness1),frequency=7)   
har.=harmonic(ozone\_thickness1,1)  
model.ozone\_thickness.har1=lm(ozone\_thickness1~har.)  
summary(model.ozone\_thickness.har1)

# plot harmonic trend in original data

plot(ts(fitted(model.ozone\_thickness.har1)), ylim = c(min(c(fitted(model.ozone\_thickness.har1),  
 as.vector(ozone\_thickness1))), max(c(fitted(model.ozone\_thickness.har1),as.vector(ozone\_thickness1)))),  
 ylab='y', main = "Fitted Harmonic curve to ozone thickness data", type="l",lty=2,col="red",xaxt="n")  
plot.new

axis(1, at=1:31, labels=c(1927:1957))

lines(as.vector(ozone\_thickness1),type="o")

**Quadratic Trend**

# data split

ozone\_thickness2<-data1[32:90,]

# fit quadratic model  
ozone\_thickness2 <-ts(as.vector(ozone\_thickness2), start=1958, end=2016)  
t = time(ozone\_thickness2)  
t2 = t^2  
model.ozone\_thickness.qa2 = lm(ozone\_thickness2~ t + t2) # label the quadratic trend  
summary(model.ozone\_thickness.qa2)

# plot quadratic trend in original data

plot(ts(fitted(model.ozone\_thickness.qa2)), ylim = c(min(c(fitted(model.ozone\_thickness.qa2),  
 as.vector(ozone\_thickness2))), max(c(fitted(model.ozone\_thickness.qa2),as.vector(ozone\_thickness2)))),  
 ylab='y' , main = "Fitted quadratic curve to Ozone Thickness Data", type="l",lty=2,col="red",xaxt="n")  
plot.new

axis(1, at=1:59, labels=c(1958:2016))  
plot.new

lines(as.vector(ozone\_thickness2),type="o")

## Residual Analysis- Linear model

The following R chunk shows standardized residuals from linear model of the ozone thickness data fitted by means of the data:

#plot without trend line

res.model.ozone\_thickness.ln = rstudent(model.ozone\_thickness.ln)  
plot(y = res.model.ozone\_thickness.ln, x = as.vector(time(ozone\_thickness)),xlab = 'Time', ylab='Standardized Residuals',main = "Time series plot of residuals")  
plot.new

abline(h=0)

#plot with trend line

res.model.ozone\_thickness.ln = rstudent(model.ozone\_thickness.ln)  
plot(y = res.model.ozone\_thickness.ln, x = as.vector(time(ozone\_thickness)),xlab = 'Time', ylab='Standardized Residuals',main = "Time series plot of residuals",type="o")  
plot.new

abline(h=0)

#qq plot to check normality

qqnorm(res.model.ozone\_thickness.ln)  
  
 qqline(res.model.ozone\_thickness.ln, col = 2, lwd = 1, lty = 2)

# **shapiro test**

shapiro.test(res.model.ozone\_thickness.ln)

# **auto correlation function**

acf(res.model.ozone\_thickness.ln,main="ACF of standardized residuals-Linear model")

## Residual Analysis - Quadratic model

The following chunk generates a time series plot for the standardized residuals from quadratic model of the ozone thickness data fitted by means of the data:

res.model.ozone\_thickness.qa = rstudent(model.ozone\_thickness.qa)  
#plot without trend line

plot(y = res.model.ozone\_thickness.qa, x = as.vector(time(ozone\_thickness)),xlab = 'Time', ylab='Standardized Residuals',main = "Time series plot of residuals")  
plot.new

abline(h=0)

#plot with trend line

plot(y = res.model.ozone\_thickness.qa, x = as.vector(time(ozone\_thickness)),xlab = 'Time', ylab='Standardized Residuals',type='o',main = "Time series plot of residuals")  
plot.new

abline(h=0)

#qq plot to check normality

qqnorm(res.model.ozone\_thickness.qa)  
qqline(res.model.ozone\_thickness.qa, col = 2, lwd = 1, lty = 2)

# **shapiro test**

shapiro.test(res.model.ozone\_thickness.qa)

# **auto correlation function**

acf(res.model.ozone\_thickness.qa,main = "ACF of standardized residuals- Quadratic model")

## Prediction

# Create time points for model fitting  
k = c(2017:20121)  
k2 = k^2

# forecast the model  
forecastdata=data.frame(k,k2)

#plot the new data into the existing data

forecasts=predict(model.ozone\_thickness.qa,forecastdata,interval="prediction")

plot these forecasts next to the time series of interest by the following code chunk:

# plot the dataset

plot(ozone\_thickness, xlim = c(1927,2021), ylim = c(-15, 10), ylab = "Ozone Thickness data", main =" Prediction of Ozone thickness for the next 5 years")  
# We do this for all columns of forecasts  
lines(ts(as.vector(forecasts[,1]), start = 2017), col="red", type="l")  
lines(ts(as.vector(forecasts[,2]), start = 2017), col="blue", type="l")  
lines(ts(as.vector(forecasts[,3]), start = 2017), col="black", type="l")  
legend("topleft", lty=1, pch=1, col=c("red","blue","black"), text.width = 18,  
 c("Data","5% forecast limits", "Forecasts"))