s3668855

Suraj kannan

31 March 2019

Contents

INTRODUCTION

The ozone layer is the common term for the high concentration of ozone that is found in the stratosphere around 15-30km above the earth's surface. It covers the entire planet and protects life on earth by absorbing harmful ultraviolet-B (UV-B) radiation from the sun. Ozone is a naturally occurring molecule. An ozone molecule is made up of three oxygen atoms. It has the chemical formula O3. Depletion of the ozone layer occurs globally, however,the severe depletion of the ozone layer over the Antarctic is often referred to as the 'ozone hole'. Increased depletion has recently started occurring over the Arctic as well.

Load Data

```r
library(TSA)
```

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

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

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

```r
library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

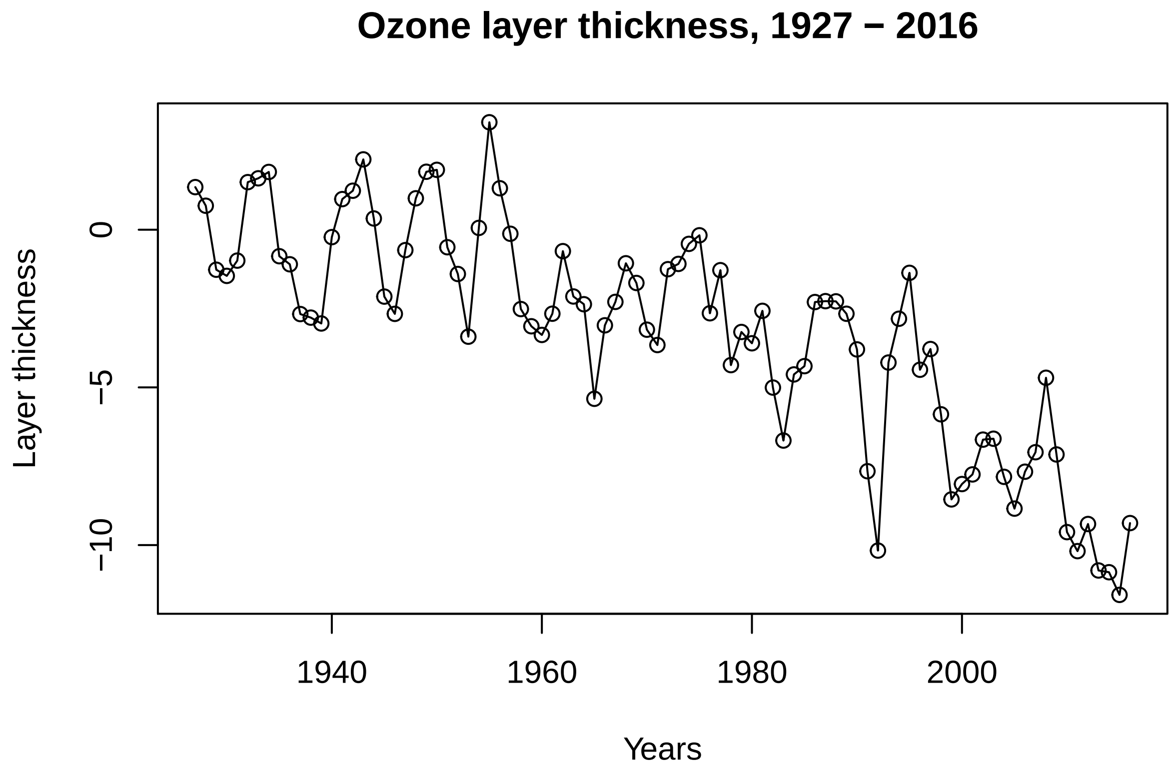
```r
library(ggplot2)
library(forecast)
library(readr)
```

```
##
## Attaching package: 'readr'

## The following object is masked from 'package:TSA':
##
##     spec
```

Representation of time series

This time series dataset is designed in such a way that the mean function is known prior and the rest of the functions form of trend can be determined. The time series trend mentioned here is a deterministic trend. A linear function y =ax+b can be set or in some cases y=a+bx+cx^2. This will produce a straight line with small fluctuations in the middle.

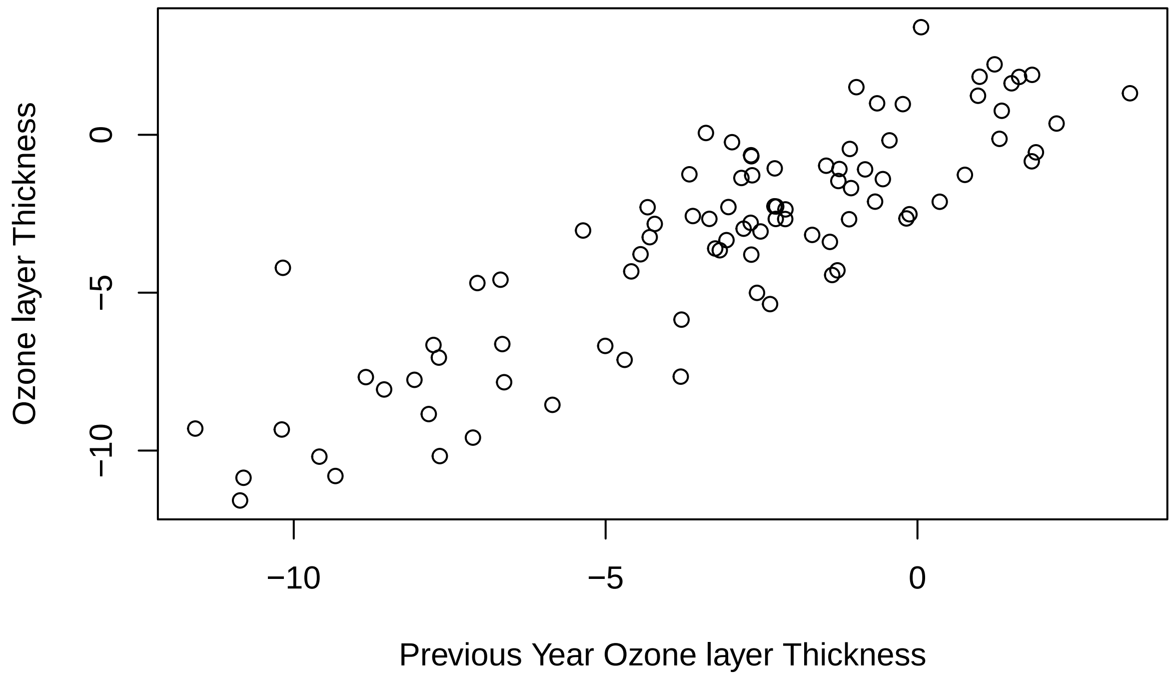
```r
suraj = read.csv('data1.csv',col.names='Layer thickness', header = FALSE)
suraj = ts(as.vector(suraj), start = 1927,end=2016)

plot(suraj, type='o',
     ylab='Layer thickness', xlab='Years',
     main=' Ozone layer thickness, 1927 - 2016')
```



##ZLAG PLOTTING The changes in the ozone layer depletion is plotted. The difference with the previous year s value is defined in the zlag. The variation in thickness of the ozone layer is plotted. The year 1952 had the highest thickness whereas 1992 had very very less ozone layer thickness. The next year s data can be predicted with the previous year s data.

Generation of scatter plot - Pairs of consequetive ozone layer relationship

```
plot(y=suraj,x=zlag(suraj),ylab='Ozone layer Thickness',
     xlab='Previous Year Ozone layer Thickness')
```



Calculation of Correlation of Ozone layer thickness

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

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

The corelation between the previous year and the present year. The correlation is 0.87 High co relation has been observed. Seasonality is not possible to observe. ## Linear and Quadratic MODEL

Linear Model Calculation

a linear trend model with the equation y=ax+b, where a is the intercept and b is the slope. This graph shows the linear trend model. The thickness of the ozone layer decreases as time progresses. In the linear trend model shown above, the thickness of teh ozone layer starts at the top left after which it falls below the mean line.

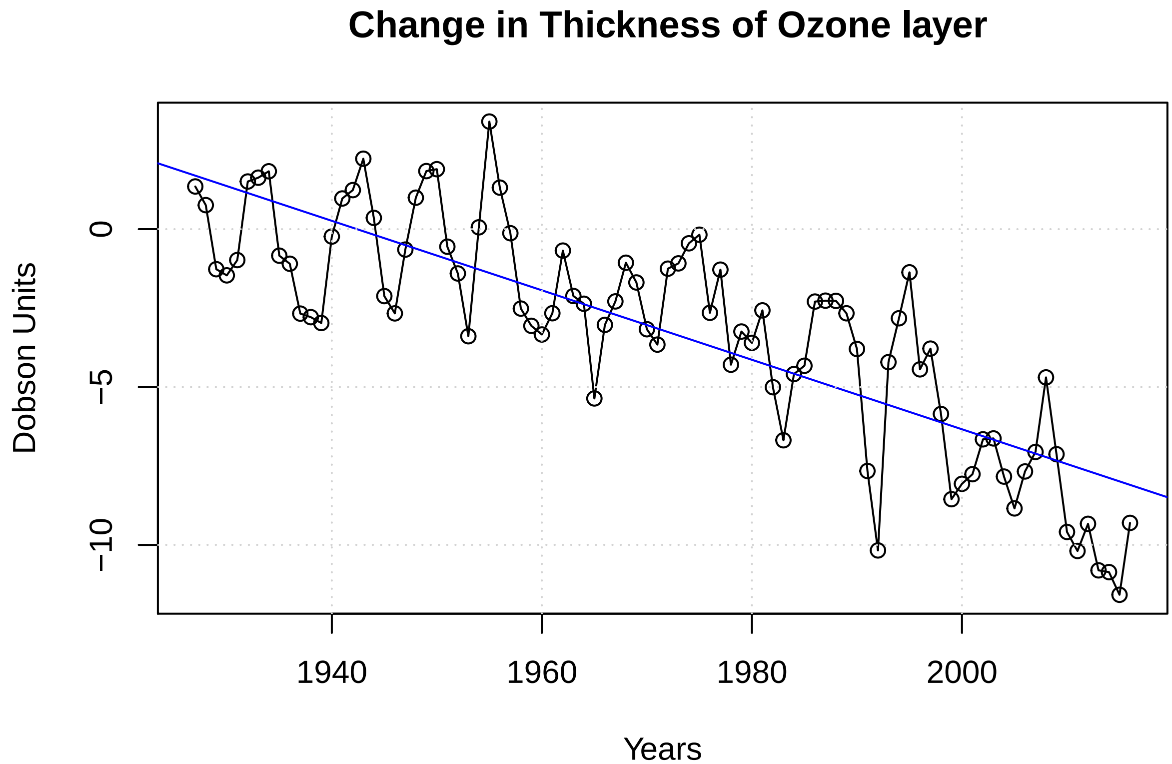
The high variance is attributed with the increase in years. Apparently in the year 1954 the ozone layer is at the highest thickness level. Following this, the ozone layer thickness is continously decreasing towards the negative value from the year 2014 to 2016.

The calculated values of slope and intercept are b=-0.1100 and a=213.72, the significance level of the slope is at 5%.

Intercept and slope estimation using least square regression method

```
model1 = lm(y~time(y))
plot(y, type='o',
     ylab='Dobson Units', xlab='Years',
```

```r
     main=' Change in Thickness of Ozone layer')
abline(model1, col ="blue")
grid()
```

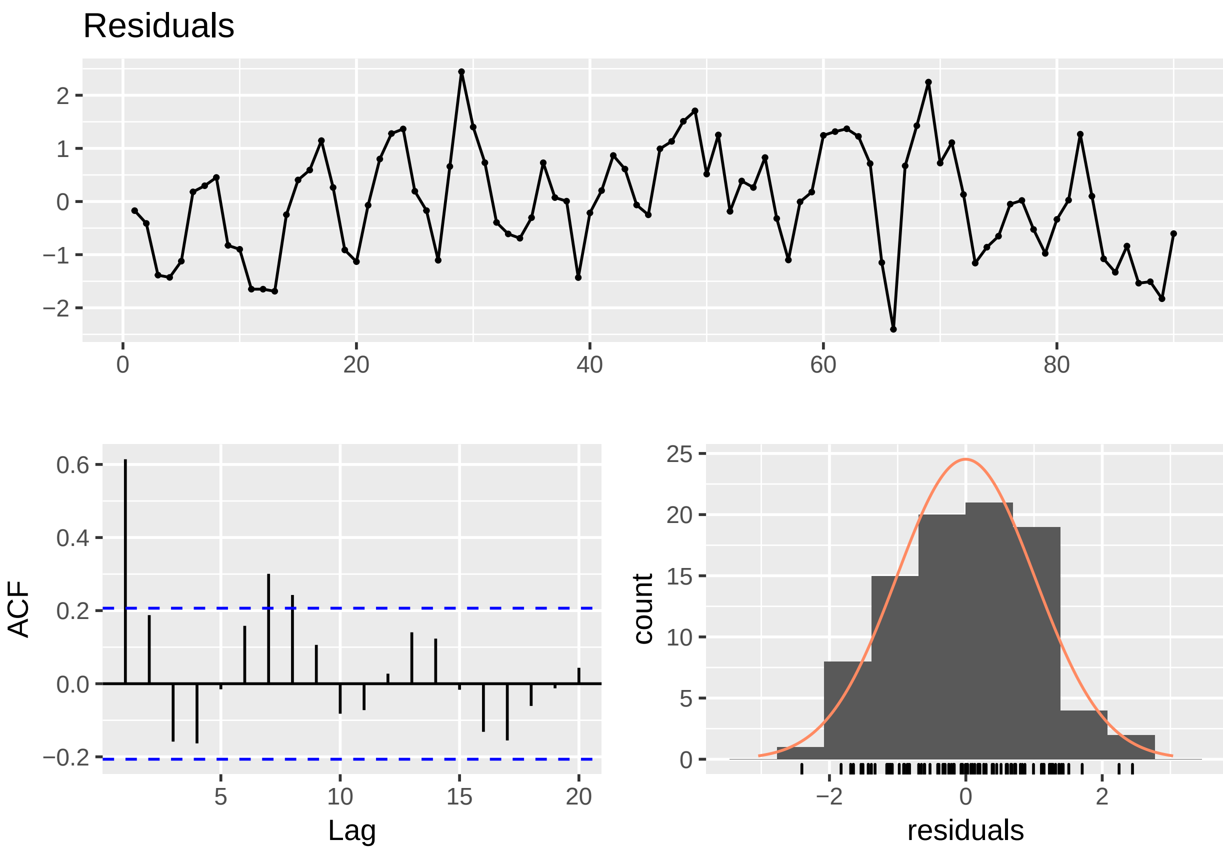



```r
summary(model1)
```

```
##
## Call:
## lm(formula = y ~ time(y))
##
## 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(y)      -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
```

Residual Analysis - Linear trend

```r
checkresiduals(rstudent(model1))
```

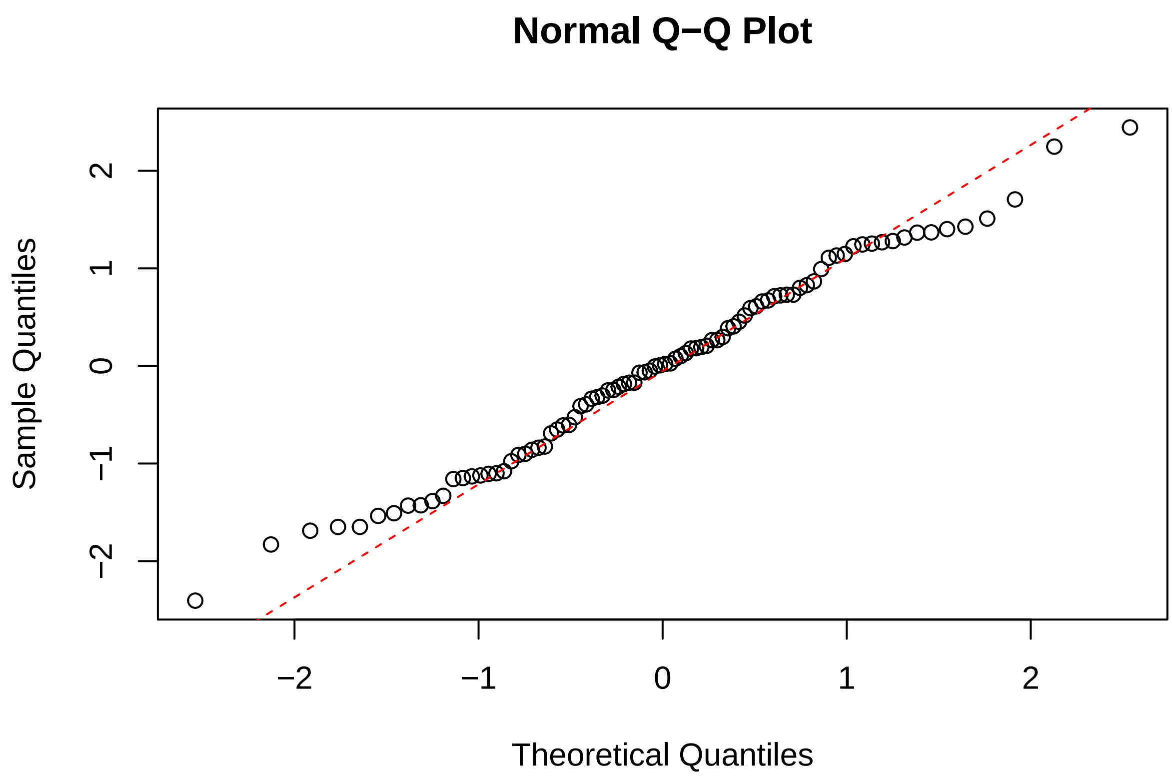


ACF - corelation function for quadratic trends which indicate that some lines are crossing the significance level. It can be concluded that the process is not an white noise process.

HISTOGRAM - used to check the normality of residuals. The frequency histogram of standardised residuals is depicted here. In the above Histogram, the plot is not symmetric at Middle. It does not tail off at both the high and low ends and we cannot say accurately that the histogram of residual values is normally distributed.

Checking The Normality of residuals

```r
y1 = rstudent(model1)
qqnorm(y1)
qqline(y1, col = 2, lwd = 1, lty = 2)
```



Standardised residuals for the dataset. The staright line pattern represents a normally distributed stochastic component in the above model. Comparing the linear and quadratic model, there is staright line with a few a few outliers. Out of the both models the quadratic trend model seems to fit the dataset more accurately than the linear trend model, the reason being there are less outliers in the quadratic model and there are more outliers in the linear trend model. The less number of outliers is attributed with more number of points lying on the line. The normality assumption can be tested with various hypothesis tests

```
shapiro.test(y1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  y1
## W = 0.98733, p-value = 0.5372
```

The values derived from the shapiro test indicate that the distribution of data are not different from a normal distribution as the pvalue > 0.05 for both the trend models. We do not reject the hypothesis.

Quadratic model Calculation

```
t = time(y)
t2 = t^2
model2 = lm(y ~ t+t2) # label the model as model1
#t = time(ozone.ts)
#t2 = t^2
#model.ozone.quad = lm(ozone.ts~ t + t2)
#summary(model.ozone.quad)
summary(model2)
```

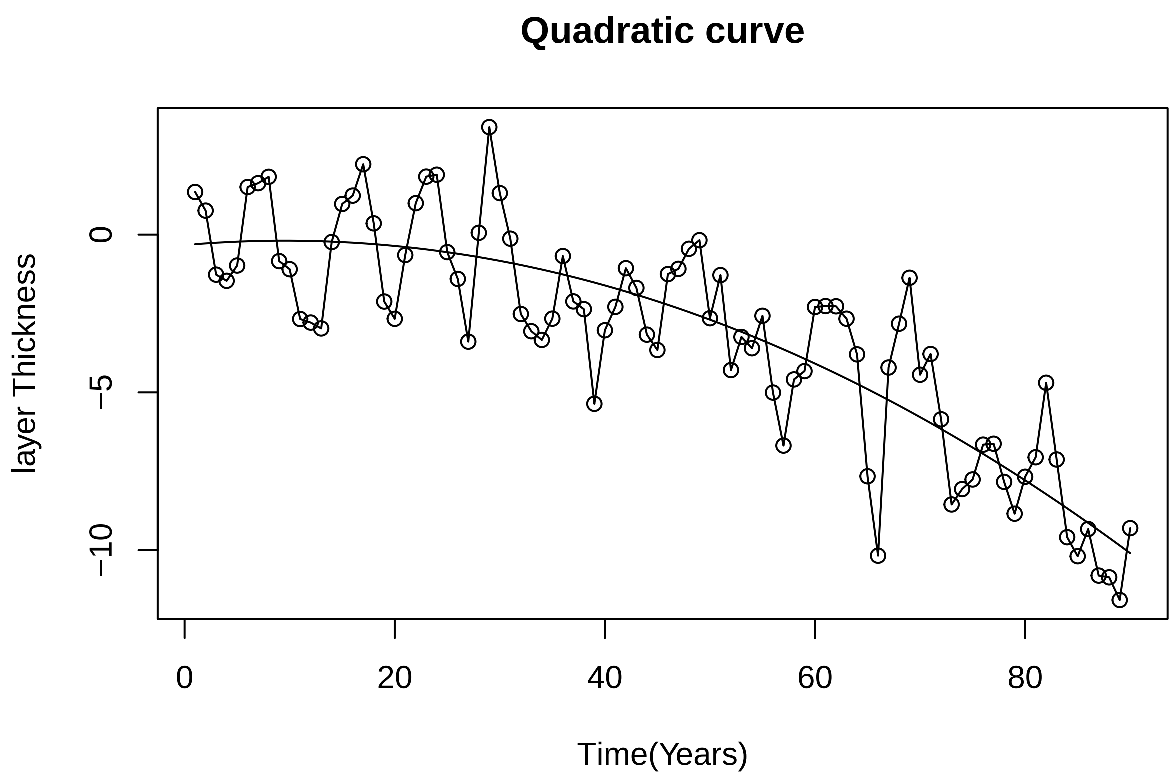
```
##
```

```
## Call:
## lm(formula = y ~ 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
```

mentioned above as y=a+bx+cxˆ2 where "a" is the intercept, "b" is the linear trend, and "c" is the quadratic trend in time a=-5.733e+03,b=5.924e+00 and C=-1.530e-03 .

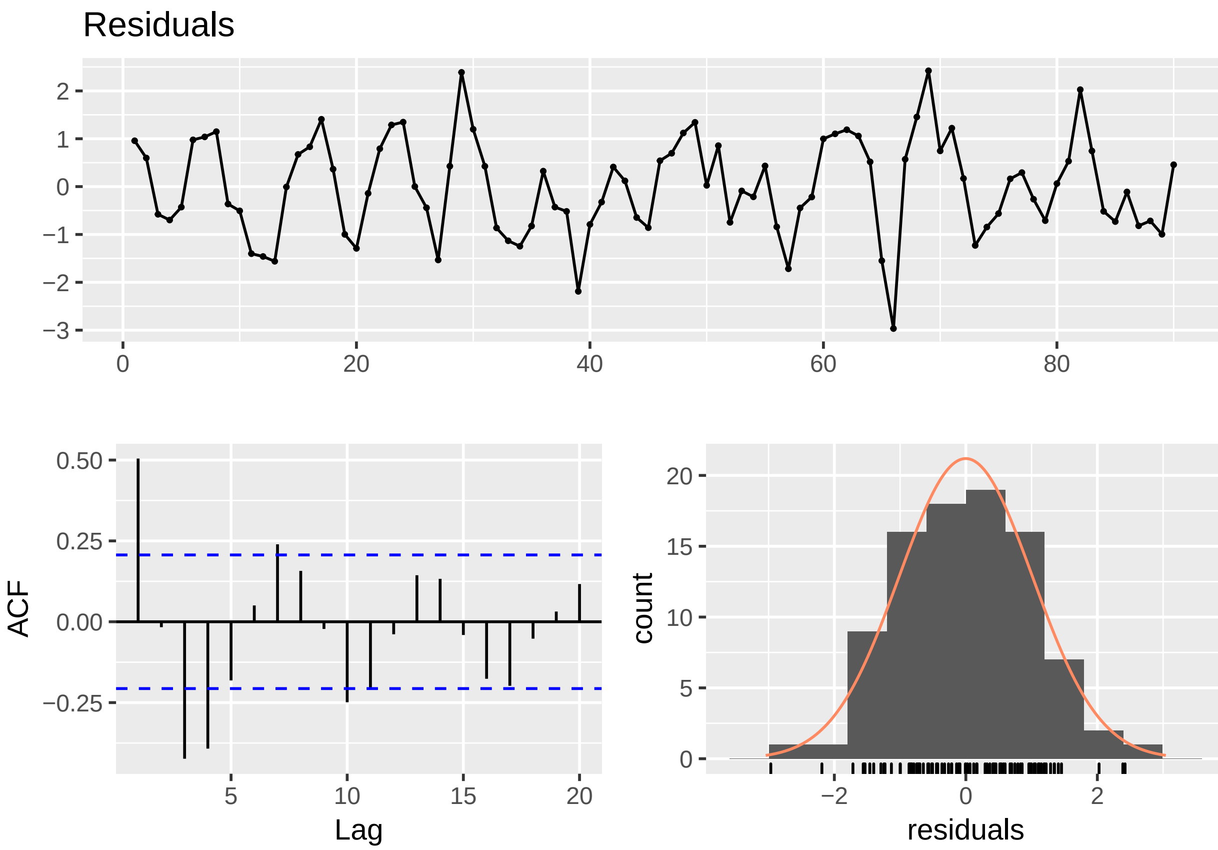
It is evident that on seeing the p -values, the quadratic trend term is significant. It can be concluded that after analysing both linear and quadratic trend, the Quadratic trend suits well for the "Ozone layer thickness of the earth" dataset. The reason being it fits more points in the line than to that of the linear trend model. There are Symmetric points on both sides of the Quadratic curve than the linear trend model curve. If the R-squared is increased it will reduce the error std deviation by approx 10 percent . Since Quadratic trend model has a higher std deviation it is the perfect fit. According to the p-values, quadratic trend is found significant and the value of multiple R-squared is 0.73 which is better than the linear model #Fitted quadratic trend curve claculation

```
plot(ts(fitted(model2)), ylim = c(min(c(fitted(model2),
     as.vector(y))), max(c(fitted(model2),as.vector(y)))),
     ylab='layer Thickness' ,
     main = "Quadratic curve",xlab='Time(Years)' )
lines(as.vector(y),type="o")
```



Residual Analysis of Quadratic trend

```
checkresiduals(rstudent(model2))
```

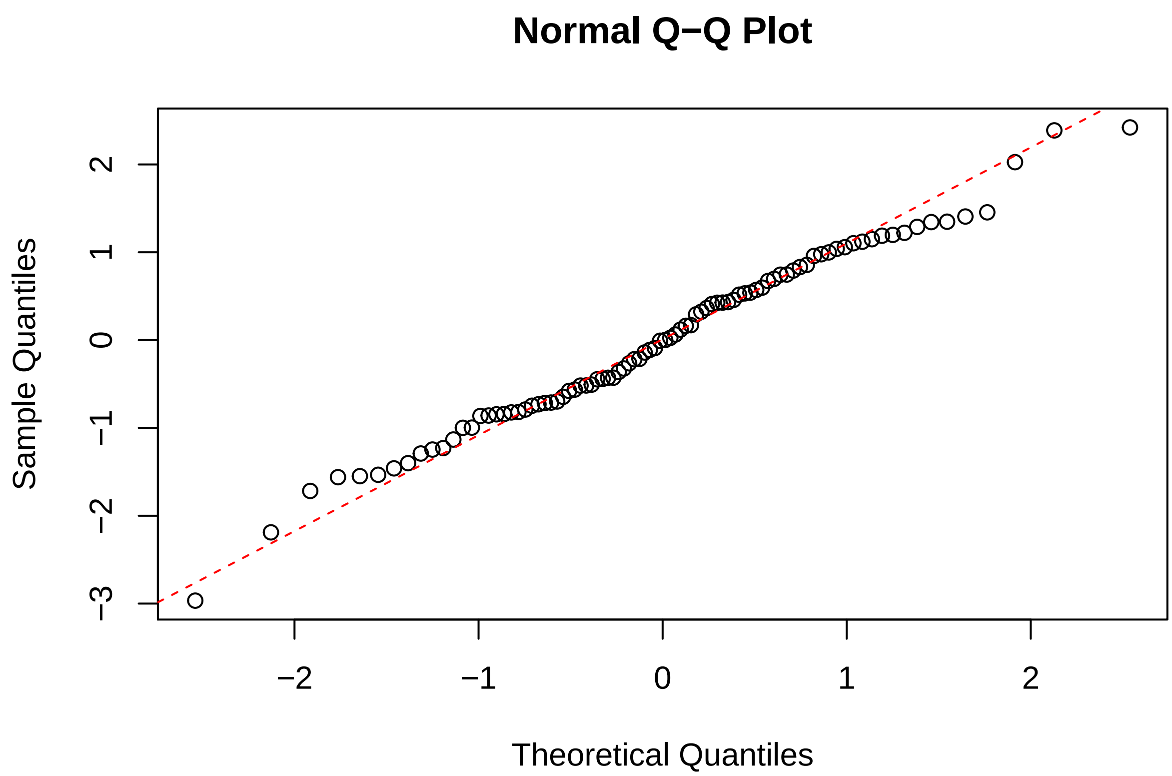


ACF - corelation function for quadratic trends which indicate that some lines are crossing the significance level. It can be concluded that the process is not an white noise process.

HISTOGRAM - used to check the normality of residuals. The frequency histogram of standardised residuals is depicted here. In the above Histogram, the plot is not symmetric at Middle. It does not tail off at both the high and low ends and we cannot say accurately that the histogram of residual values is normally distributed.

Checking The Normality of residuals

```
y1 = rstudent(model2)
qqnorm(y1)
qqline(y1, col = 2, lwd = 1, lty = 2)
```



straight-line pattern here supports the assumption of a normally distributed stochastic component in this model.

```
shapiro.test(y1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  y1
## W = 0.98889, p-value = 0.6493
```

Forecasting and Prediction

Forecasting using Quadratic Model

```
t=time(suraj)

h = 5
z = seq((as.vector(tail(t,1))+1), (as.vector(tail(t,1))+h))
pred = data.frame(t = z, t2 = z^2)
forecasts = predict(model2, pred, interval = "prediction")
knitr::kable(forecasts, caption="Prediction of 5 Year Values")
```

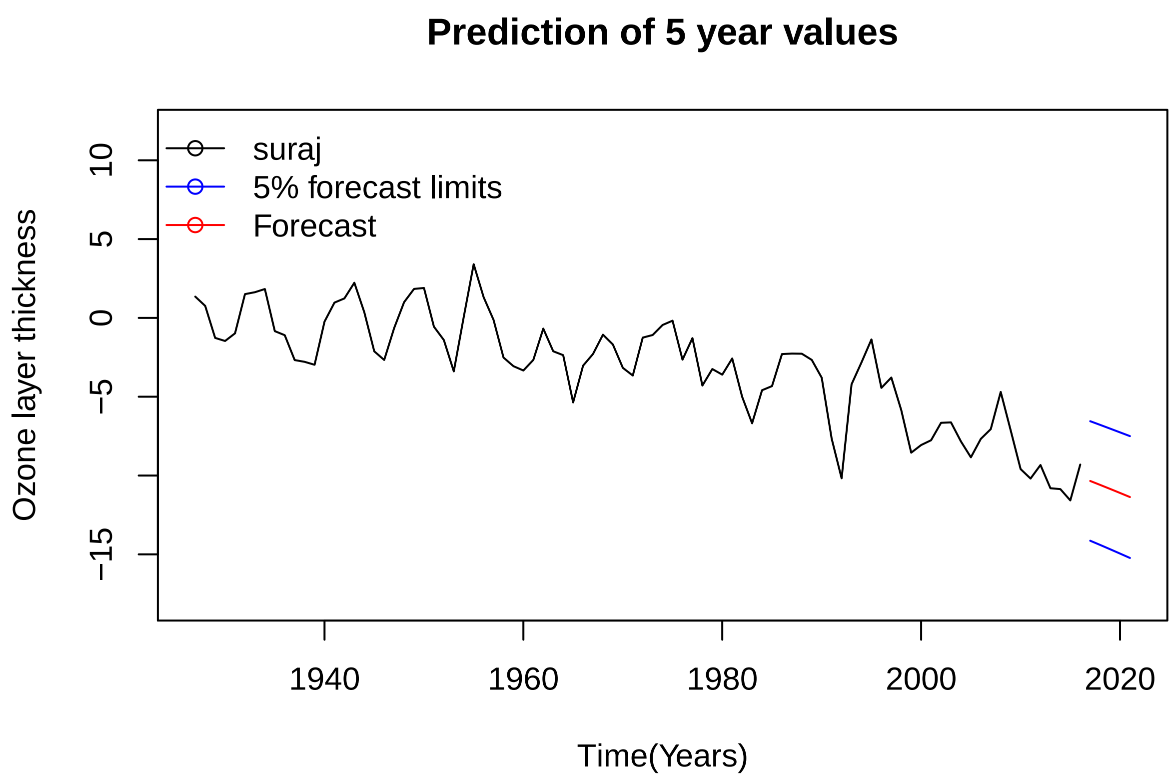
Table 1: Prediction of 5 Year Values

fit	lwr	upr
-10.34387	-14.13556	-6.552180
-10.59469	-14.40282	-6.786548

fit	lwr	upr
-10.84856	-14.67434	-7.022786
-11.10550	-14.95015	-7.260851
-11.36550	-15.23030	-7.500701

Plotting the forecasted Time series values

```
plot(suraj, xlim = c(1927,2021), ylim = c(-18, 12),
     ylab = "Ozone layer thickness", xlab='Time(Years)',
     main='Prediction of 5 year values')
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="blue", type="l")
legend(x="topleft", lty=1, pch=1, col=c("black","blue","red"),text.width = 10,
       c("suraj","5% forecast limits", "Forecast"),
       bty = 'n')
```



CONCLUSION

We conclude that, based on the giveen dataset, the operations of modelling and estimating deterministic trends in time series were performed. It can be found that, the ozone layer thickness can be estimated using a Quadratic trend model and using a hypothesis tests. The hypothesis tests are QQ plot and Shiapro test. For the given dataset the Quadratic trend model is the better at fitting the data than the linear model as the r Squared values are higher. As the time passes by there is a gradual decrease in the ozone layer thickness.