Code

Time Series Analysis

To find best fitting model for ozone thickness along with 5 year forecast

Student Details

- Saurabh Mallik

Introduction

The purpose of this research is to infer and report certain research questions from the Ozone thickness series provided in the dataset "dataset1.csv", to be able to understand the nature of the series and provide a model fitting along with the forecasts. The dataset provides yearly changes in the thickness of Ozone layer from 1927 to 2016 in Dobson units.

Methodology

To undertake this research, time series analysis and forecasting methods on R Studio are being used to infer from the dataset.

Analysis and Inferences

To check for the nature of the data in the "Ozone Thickness" dataset, we first load relevant packages and the dataset.

Hide

```
ozone <- read_csv("data1.csv", col_names = FALSE)
```

```
Parsed with column specification:
cols(
  X1 = col_double()
)
```

We check for the class of the data.

Hide

```
class(ozone)
```

```
[1] "tbl_df"     "tbl"        "data.frame"
```

We need to now convert the datset into a time series to be able to analyse it properly.

Hide

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

We now check again for the class of the dataset and see that the it has now been converted to a time series.

Hide

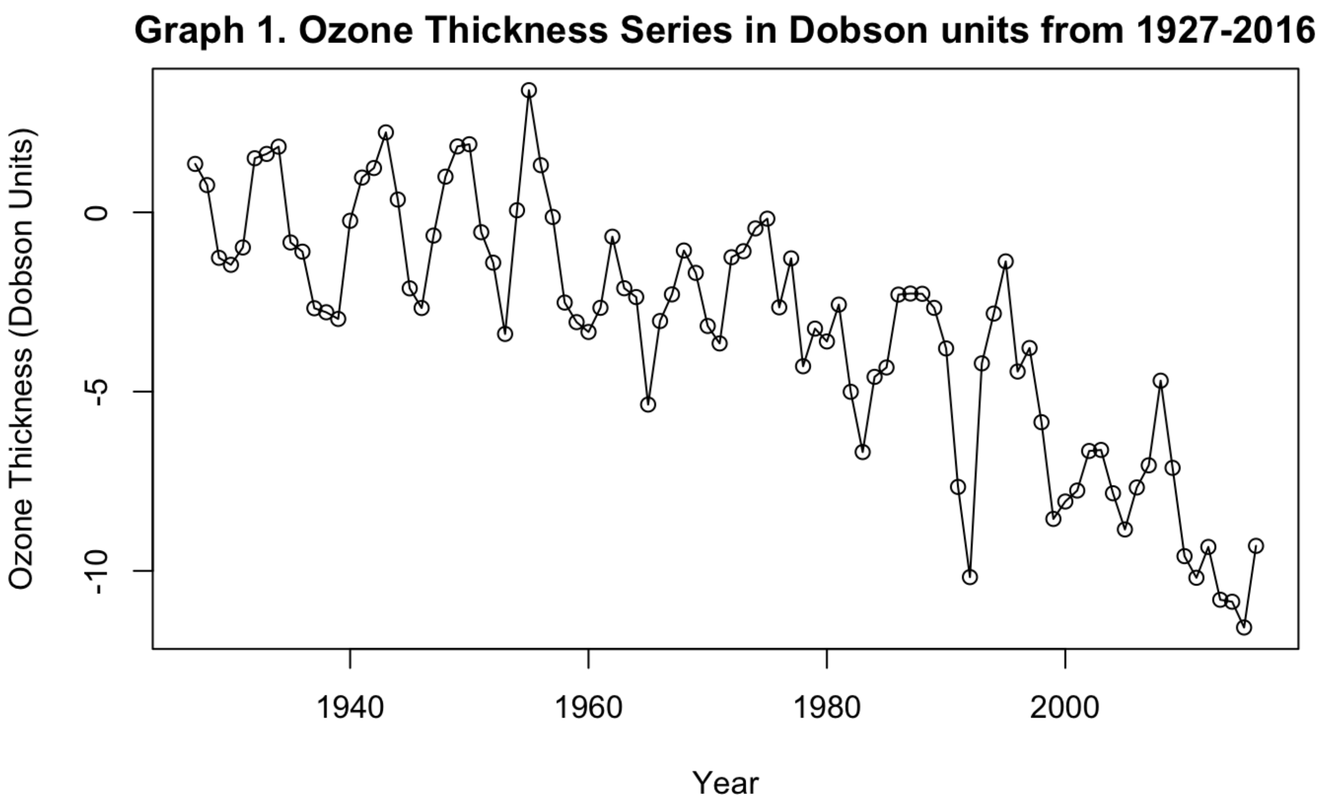
```
class(ozone.ts)
```

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

Next, we plot the series and check for any insights from it.

Hide

```
plot(ozone.ts,type='o', main ='Graph 1. Ozone Thickness Series in Dobson units from 1
927-2016', ylab = "Ozone Thickness (Dobson Units)", xlab = "Year")
```



From the above plot, it is seen that there is no evidence of seasonality. The time series shows auto regressive behaviour, and there are slight hints of changing variance over time. It is also visible that there is a hint of trend and that over time there has been a gradual linear decline in the ozone thickness.

We now start to model fit the data. We start with a simple linear trend model first, and get a summary for it.

Hide

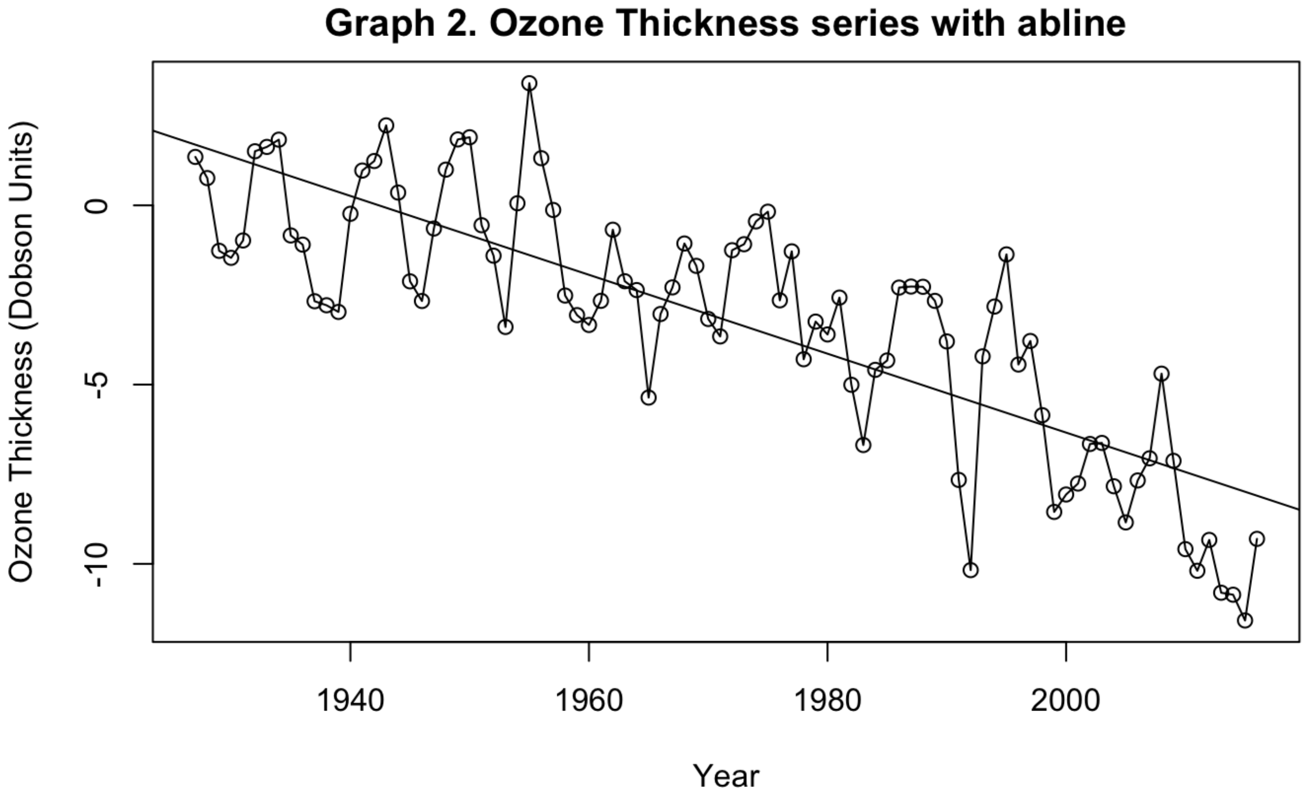
```
model1 <- lm(ozone.ts~time(ozone.ts))
summary(model1)
```

```
Call:
lm(formula = ozone.ts ~ time(ozone.ts))

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.ts)  -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
```

We see that the model is significant, having an R squared value is .6693 which shows that the model is a decent fit but not very good. The coefficients are siginificant with extremely small p values. We again plot the time series and add an abline.

Hide

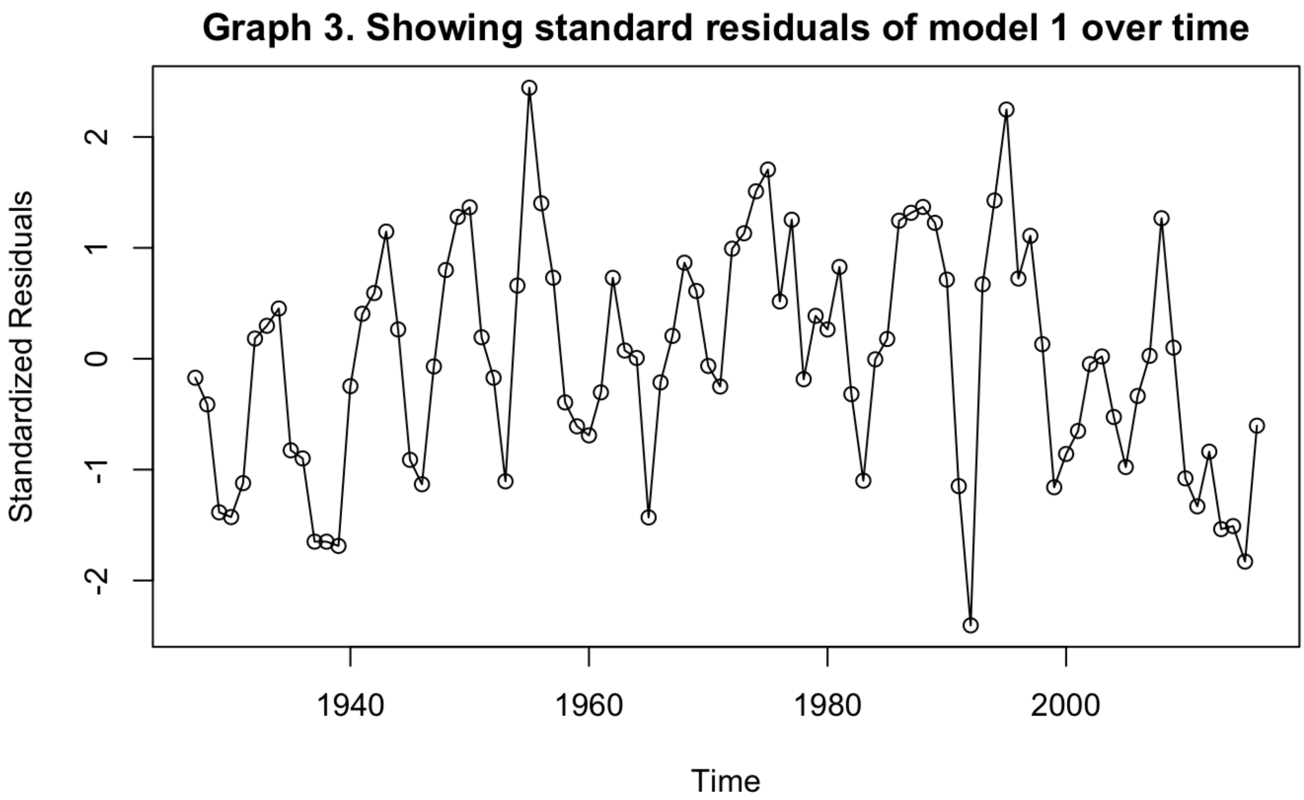
```
plot(ozone.ts,type='o', main ='Graph 2. Ozone Thickness series with abline', ylab =
"Ozone Thickness (Dobson Units)", xlab = "Year")
abline(model1)
```



From Graph 2, we can see that the fitted least sqaures lines follows the overall trend in the series. We next check the residuals and plot it to gain further inferences.

Hide

```
res.model1 = rstudent(model1)
plot(y = res.model1, x = as.vector(time(ozone.ts)),xlab = 'Time', ylab='Standardized
 Residuals',type='o', main = "Graph 3. Showing standard residuals of model 1 over tim
e")
```

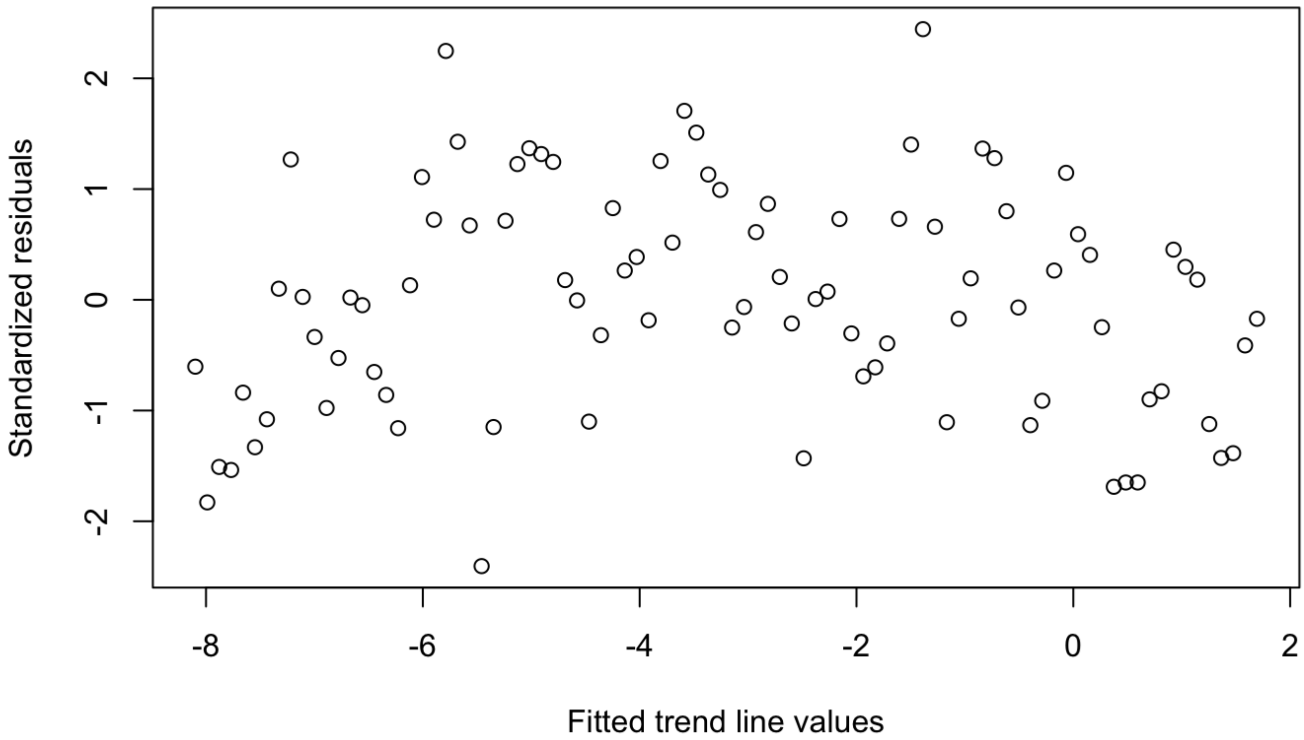


From the residuals plot, we can see that they hang together too much for white noise, the plot is too smooth. Also there seems to be more variation towards the last third of the series. We plot the fitted residuals next.

Hide

```
plot(y = rstudent(model1), x = fitted(model1),
     ylab = "Standardized residuals",
     xlab = "Fitted trend line values",
     main = "Graph 4. Time series plot of the standardized residuals")
```

Graph 4. Time series plot of the standardized residuals



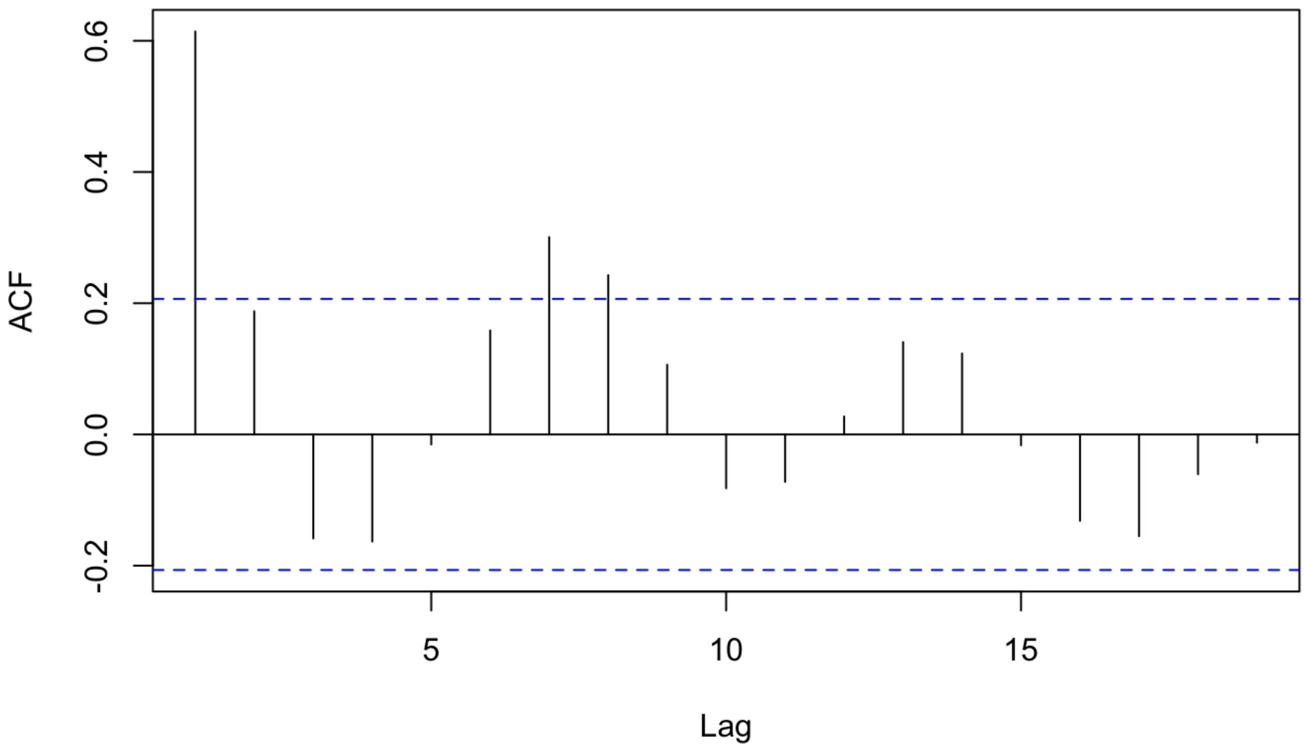
From graph 4, we can see that the points tend to hang together as well.

The sample acf of the standardized residuals is below.

Hide

```
acf(rstudent(model1), main = "Graph 5. ACF of the standardized residuals")
```

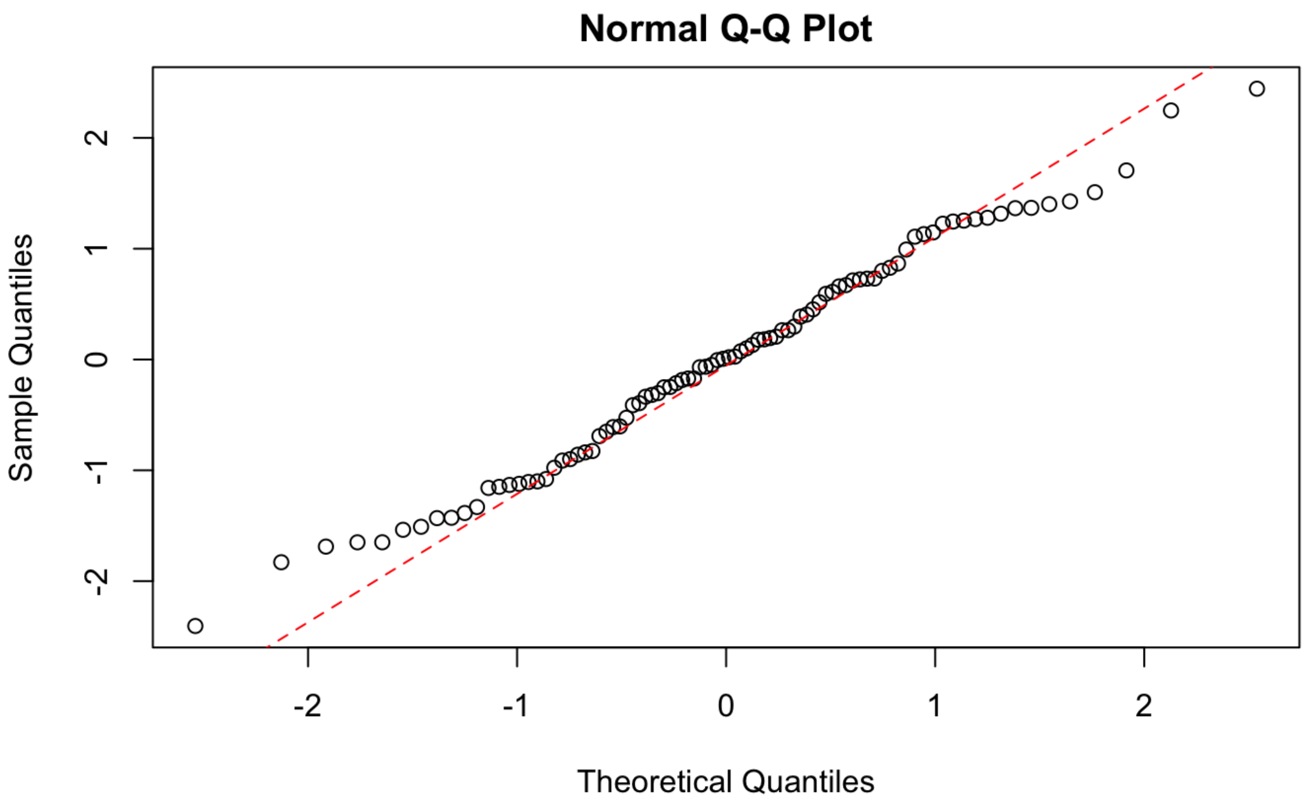
Graph 5. ACF of the standardized residuals



The ACF plot confirms the smootheness of the series, as we have correlation values higher than the confidence bound at several lags.

Hide

```
qqnorm(res.model1)
qqline(res.model1, col = 2, lwd = 1, lty = 2)
```



Because we see a considerable amount of movement away from the reference line, we conclude that the normality assumption does not hold for the ozone thickness dataset.

Hide

```
shapiro.test(res.model1)
```

```

	Shapiro-Wilk normality test

data:  res.model1
W = 0.98733, p-value = 0.5372
```

The Shapiro-Wilk normality test shows that the model isnt significant.

We next try another model adding a time period, in the form of a quadratic trend model.

Hide

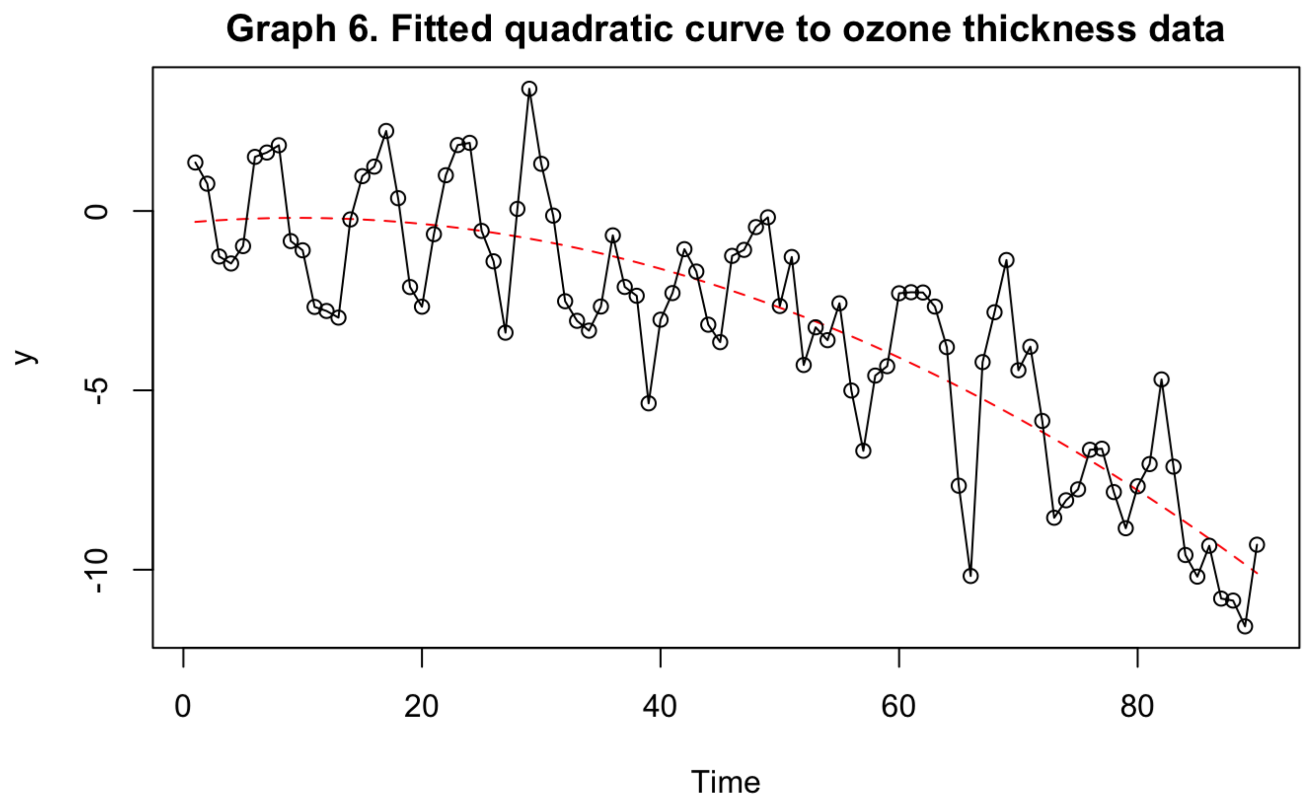
```
t = time(ozone.ts)
t2 = t^2
model1.1 = lm(ozone.ts~ t + t2)
summary(model1.1)
```

```
Call:
lm(formula = ozone.ts ~ 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
```

We see that the model is significant having a p-value < 2.2e-16, and a better R-squared value of .73 as compared to .66 of the previous model. We see that the trend coefficients are all significant as well. We proceed to diagnostic checking for the model.

Hide

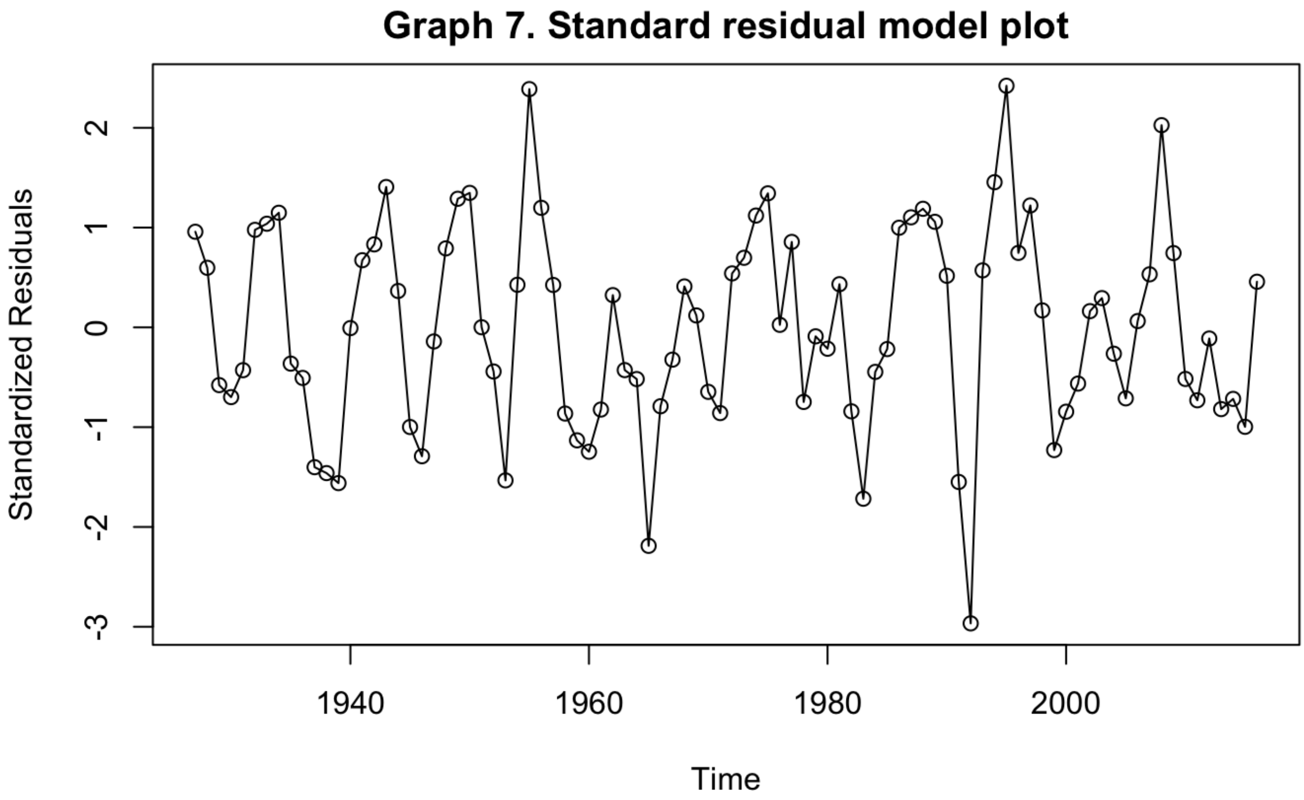
```
plot(ts(fitted(model1.1)), ylim = c(min(c(fitted(model1.1),
    as.vector(ozone.ts))), max(c(fitted(model1.1),as.vector(ozone.ts)))),
    ylab='y' , main = "Graph 6. Fitted quadratic curve to ozone thickness data", type
="l",lty=2,col="red")
lines(as.vector(ozone.ts),type="o")
```



We can see the trend line is fitting the data well. We next undergo diagnostic testing for this model.

Hide

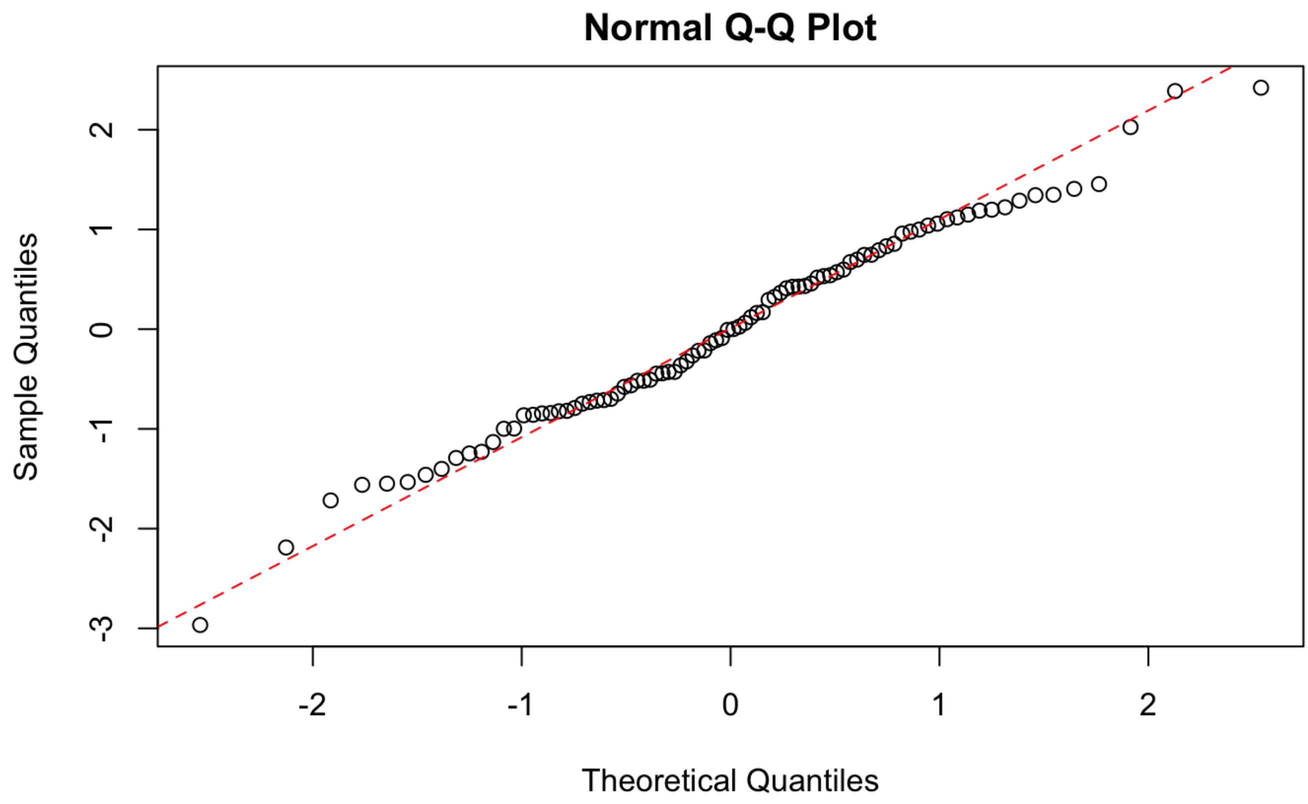
```
res.model1.1 = rstudent(model1.1)
plot(y = res.model1.1, x = as.vector(time(ozone.ts)),xlab = 'Time', ylab='Standardize
d Residuals',type='o', main = "Graph 7. Standard residual model plot")
```



We see that the values are still bunching together, but not as much as the previous model. We continue to do a normality test.

Hide

```
qqnorm(res.model1.1)
qqline(res.model1.1, col = 2, lwd = 1, lty = 2)
```



The values are still moving slightly away from the reference line.

Hide

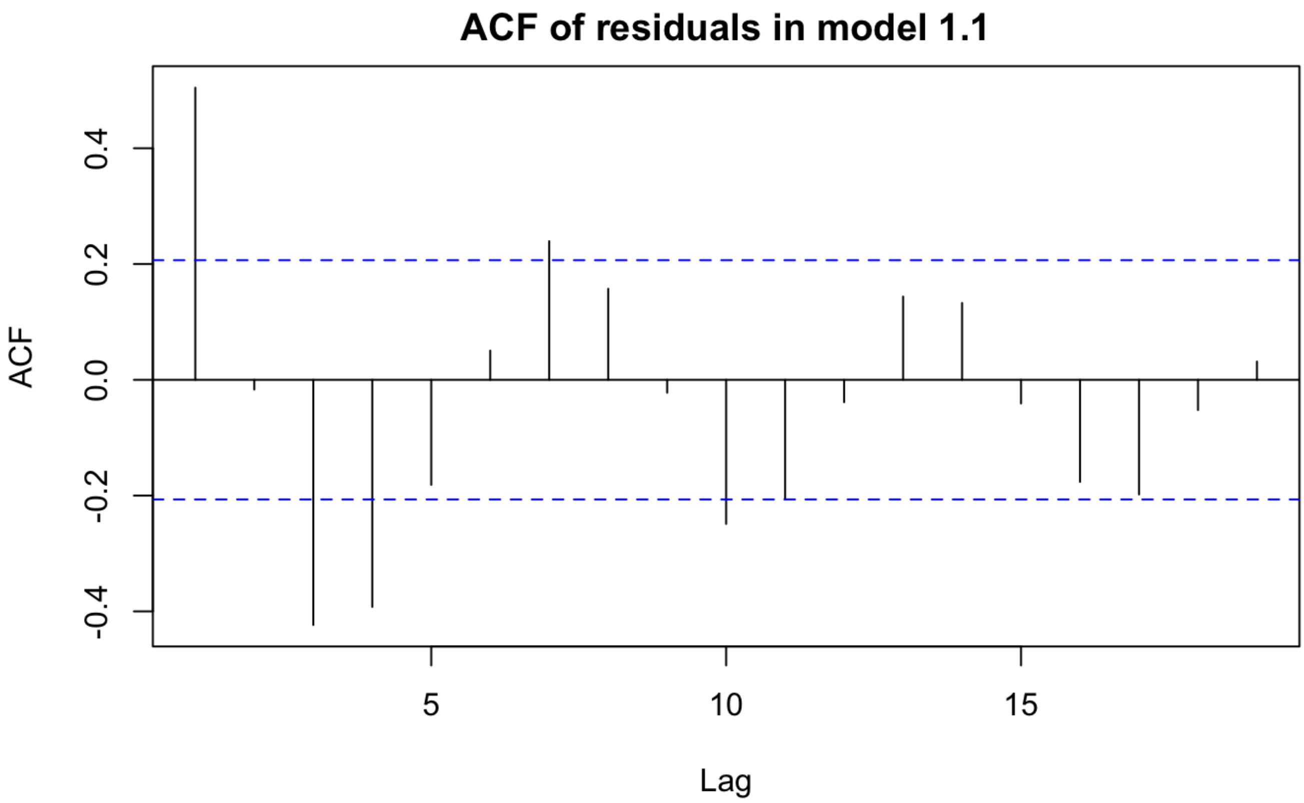
```
shapiro.test(res.model1.1)
```

```


	Shapiro-Wilk normality test

data:  res.model1.1
W = 0.98889, p-value = 0.6493
```

This model also has an insignificant p-value for the Shapiro-Wilk normality test.

Hide

```
acf(res.model1.1, main = "ACF of residuals in model 1.1")
```



The ACF plot confirms the smootheness of the series, as we have correlation values higher than the confidence bound at several lags.

Next we move to harmonic model fitting.

Hide

```
har.=harmonic(ozone.ts,.49)
model1.3=lm(ozone.ts~har.)
summary(model1.3)
```

```

Call:
lm(formula = ozone.ts ~ har.)

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 ***
har.cos(2*pi*t)         NA         NA      NA       NA    
har.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
```

As we saw there is no seasonality in the ozone thickness series, we also get no cosine estimate from the model. The model is insignificant and fitting will not be good.

Since we checked diagnostic testing for linear and quadratic model, we got the best r squared value for the quadratic model, and hence we will go ahead with fit and forecast using that model, for the next 5 years.

The forecasted values for the next 5 years are below.

Hide

```
t = c(2017, 2018, 2019, 2020, 2021)
t2 = t^2
new = data.frame(t,t2)
forecasts = predict(model1.1, new, interval = "prediction")
print(forecasts)
```

```
        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
```

Hide

```
plot(ozone.ts, main = "Graph 7. Forecasts of next 5 years of Ozone Thickness (in Dobs
on Units)", type = "o", xlim = c(1920, 2025), ylim = c(-15.5,5))
lines(ts(as.vector(forecasts[,1]), start = 2017), col="red", type="l")
```

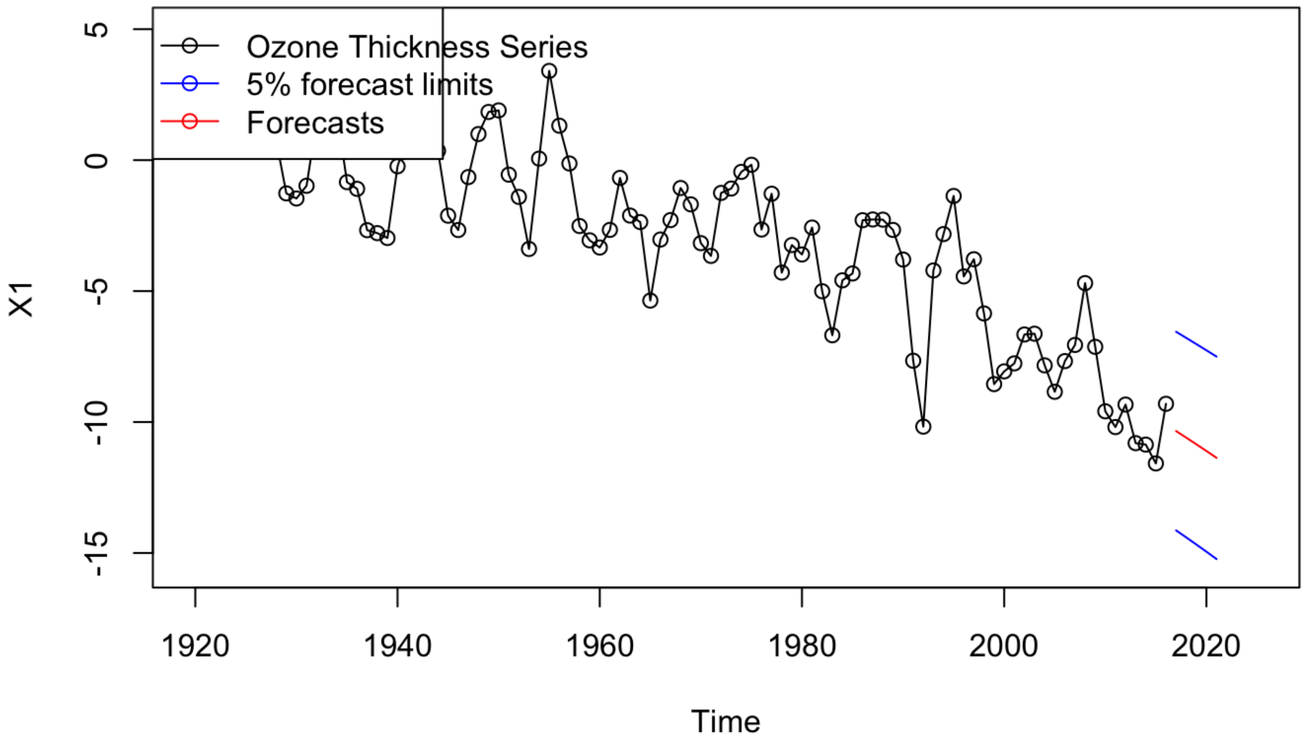
Hide

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

Hide

```
legend("topleft", lty=1, pch=1, col=c("black","blue","red"), text.width = 18,
       c("Ozone Thickness Series","5% forecast limits", "Forecasts"))
```

Graph 7. Forecasts of next 5 years of Ozone Thickness (in Dobson Units)



Conclusion

Based on the different model fitted for forecasting it was decided that the quadratic linear model had the best fit as it had significant trend components and the model itself was significant with a low p-value. The r squared value was also better than that of the linear and harmonic models with an r squared value of 73.91%. Based on these findings, we fitted the model and predict and plot the following forecasts for the next 5 years.

Hide

```
print(forecasts)
```

```
        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
```

Hide

```
plot(ozone.ts, main = "Graph 7. Forecasts of next 5 years of Ozone Thickness (in Dobs
on Units)", type = "o", xlim = c(1920, 2025), ylim = c(-15.5,5))
lines(ts(as.vector(forecasts[,1]), start = 2017), col="red", type="l")
```

Hide

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

Hide

```
legend("topleft", lty=1, pch=1, col=c("black","blue","red"), text.width = 18,
       c("Ozone Thickness Series","5% forecast limits", "Forecasts"))
```