MATH1318 Time Series

Assignment 1 - Semester 1, 2019

Phil Steinke s3725547

Executive Summary

This report examines the decline of the thickness of Ozone layer over 90 years.

To find this, data was collected from 1927 to 2016 in Dobson units (yearly changes). Where a negative value in the dataset represents a decrease in Ozone thickness and a positive value represents an Ozone increase in the thickness.

Goals:

> Your task is to analyse the data by using the analysis methods covered in the first two modules of MATH1318 Time Series Analysis course in this semester.

R code 15%

```
packages <- c('beepr', 'RColorBrewer', 'tidyr', 'TSA', 'TTR', 'smooth')
InstallAndLoadPackages(packages, setLocalWorkingDirectory)
```

```
## [1] "# Check for installed packages, then install them"
```

```
##        beepr RColorBrewer        tidyr          TSA          TTR
##         TRUE         TRUE         TRUE         TRUE         TRUE
##       smooth
##         TRUE
```

```
ozone <- read.csv("raw-data/data1.csv", header = FALSE)
ozone.ts <- ts(as.vector(ozone), start = 1927, end = 2016) # convert to timeseries
colourPallete <- brewer.pal(n = 10, name = "BrBG")
# display.brewer.pal(10, 'BrBG')
cat("\014")
```

```
# ozone.ts %>% tail(n = 5)
ozone.ts %>% dim()
```

```
## [1] 90  1
```

The Data

The ozone thickness series is given by `dataset1.csv` which represents:

- thickness of Ozone layer 1927 to 2016 in Dobson units (yearly changes)
- A negative value in the dataset represents a decrease in Ozone thickness
- A positive value represents an Ozone increase in the thickness
- The dataset is timeseries and includes 90 records

Data was obtained from our lecturer ‘Haydar Demirhan’. Given there is only a single variable `thickness of Ozone layer`, and that the ozone layer varies in thickness around the globe, this statistic would either be from a single location, or an average around the globe.

What are Dobson units?

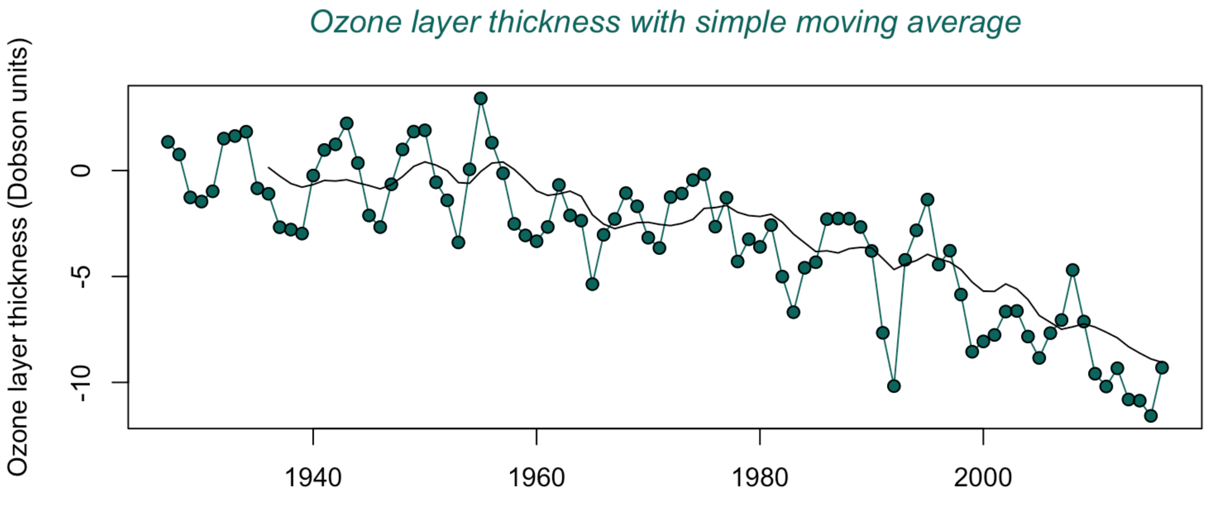
> The Dobson Unit is the most common unit for measuring ozone concentration. Over the Earth’s surface, the ozone layer’s average thickness is about 300 Dobson Units or a layer that is 3 millimeters thick. - https://ozonewatch.gsfc.nasa.gov/facts/dobson_SH.html (https://ozonewatch.gsfc.nasa.gov/facts/dobson_SH.html)

Initial Plot Declare

```
t = time(ozone.ts) # Create time points for model fitting


default_ylab = 'Ozone layer thickness (Dobson units)'
default_subtitle = 'thickness of Ozone layer 1927 to 2016 in Dobson units - yearly
 changes'
default_xlab = 'Year (1927-2016)'


doTimeSeriesPlot <- function(
  x,
  xlab = default_xlab,
  y,
  ylab = default_ylab,
  lines,
  type = 'o',
  title = '',
  subtitle = default_subtitle
  ) {
  # par(bg = colourPallete[5]) # plot background
  plot(
    x,
    y,
    col = colourPallete[9],
    pch = 19,
    type = type,
    xlab = xlab,
    ylab = default_ylab)
  title(
    main = title,
    sub = subtitle,
    col.main = colourPallete[9],
    font.main = 3)
  points(x)
  if(!missing(lines)) { lines }
}
```

Ozone Layer Thickness

```
doTimeSeriesPlot(
  x = ozone.ts,
  y = NULL,
  title = 'Ozone layer thickness with simple moving average',
  lines = lines(SMA(ozone.ts)))
```



Year (1927-2016)
thickness of Ozone layer 1927 to 2016 in Dobson units - yearly changes

This plot shows:

- We can see a continuous decreasing trend
- A negative slope
- a decreasing simple moving average in thickness of ozone layer over 90 years

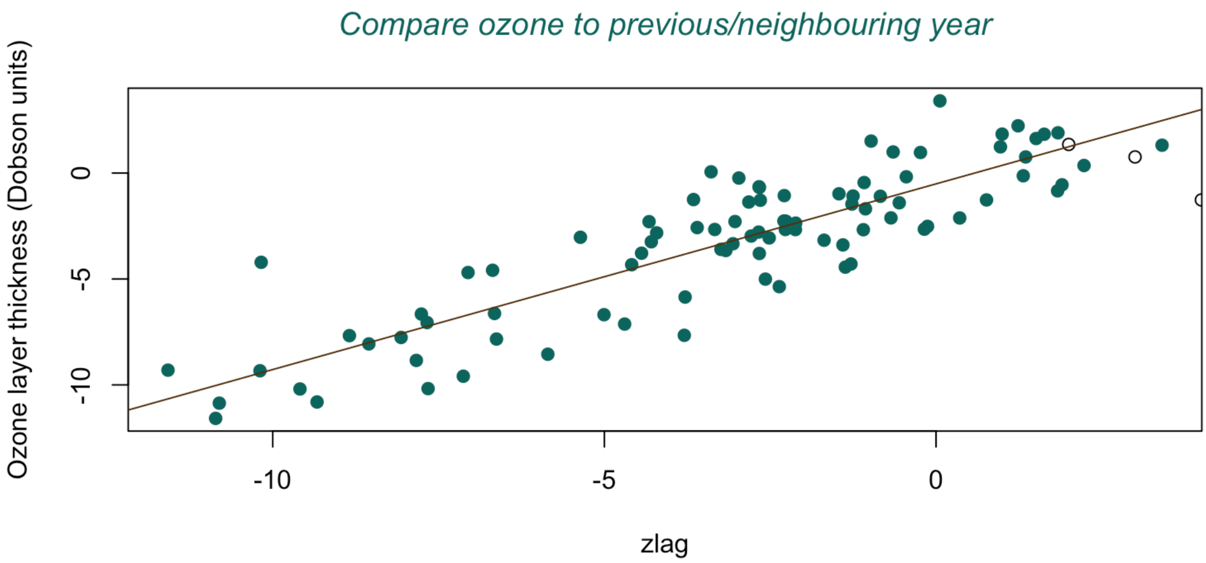
Compare data with it’s previous year

```
y = ozone.ts               # Read data into y
x = zlag(ozone.ts)         # Generate first lag
index = 2:length(x)            # Create an index to get rid of the first NA value and
  the last 5 missing values in x
cor(y[index], x[index])
```

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

- A Correlation of 0.87 implies a strong positive relationship between the Ozone Layer thickness over time

```
doTimeSeriesPlot(
  x = zlag(ozone.ts), # Generate first lag of the series
  xlab = 'zlag',
  y = ozone.ts, # Read the data into y
  ylab = 'ozone.ts',
  title = "Compare ozone to previous/neighbouring year",
  type='p')
linearTrendModel_previousYear =
  lm(ozone.ts~zlag(ozone.ts))
abline(
  linearTrendModel_previousYear,
  col = colourPallete[1])
```



thickness of Ozone layer 1927 to 2016 in Dobson units - yearly changes

There is a strong positive autocorrolation of the data for concurrent years.

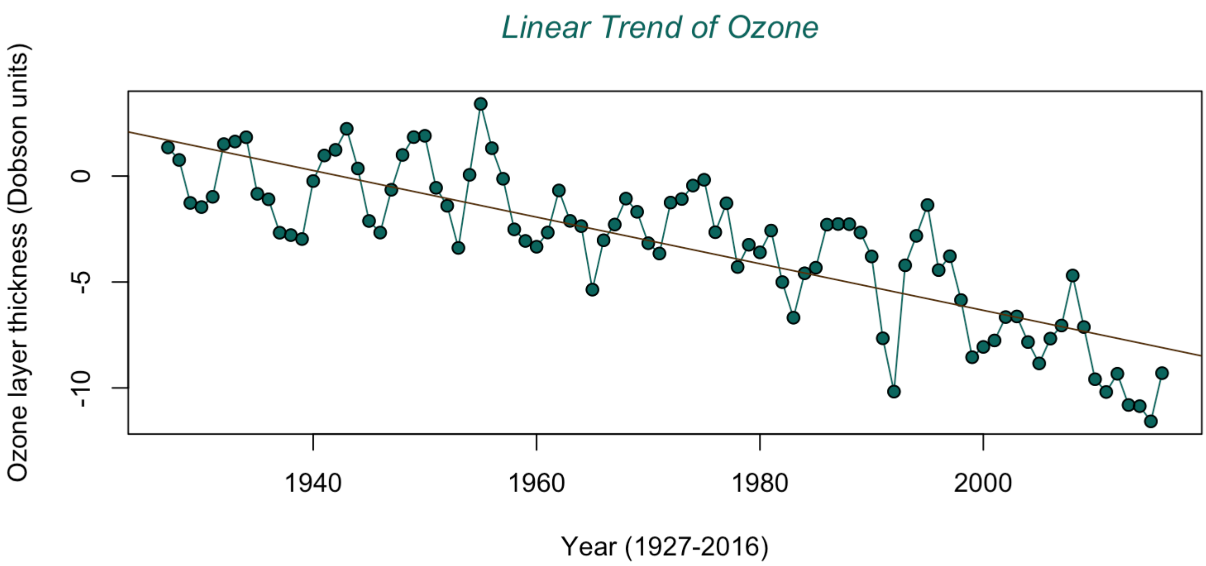
Linear Trend

Model 1

Define Linear Trends Model

```
linearTrendModel = lm(ozone.ts~t)
```

```
doTimeSeriesPlot(
  lines = abline(
            linearTrendModel,
            col = colourPallete[1]),
  title = "Linear Trend of Ozone ",
  x = ozone.ts,
  y = NULL
)
```



thickness of Ozone layer 1927 to 2016 in Dobson units - yearly changes

Initial visual inspection shows a linear trend does not continue to fit the later years of the data

Linear Residuals

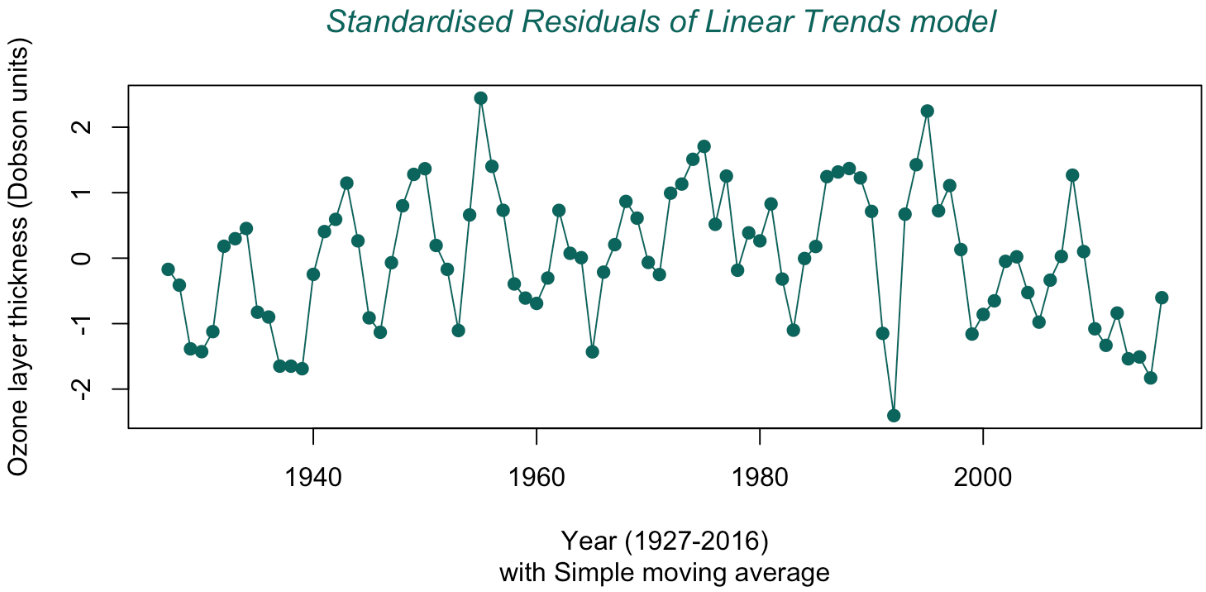
```
t = time(ozone.ts) # Create time points for model fitting
linearTrendModel = lm(ozone.ts~t) # label the model as model1
summary(linearTrendModel)
```

```
##
## Call:
## lm(formula = ozone.ts ~ t)
##
## 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 ***
## t            -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
```

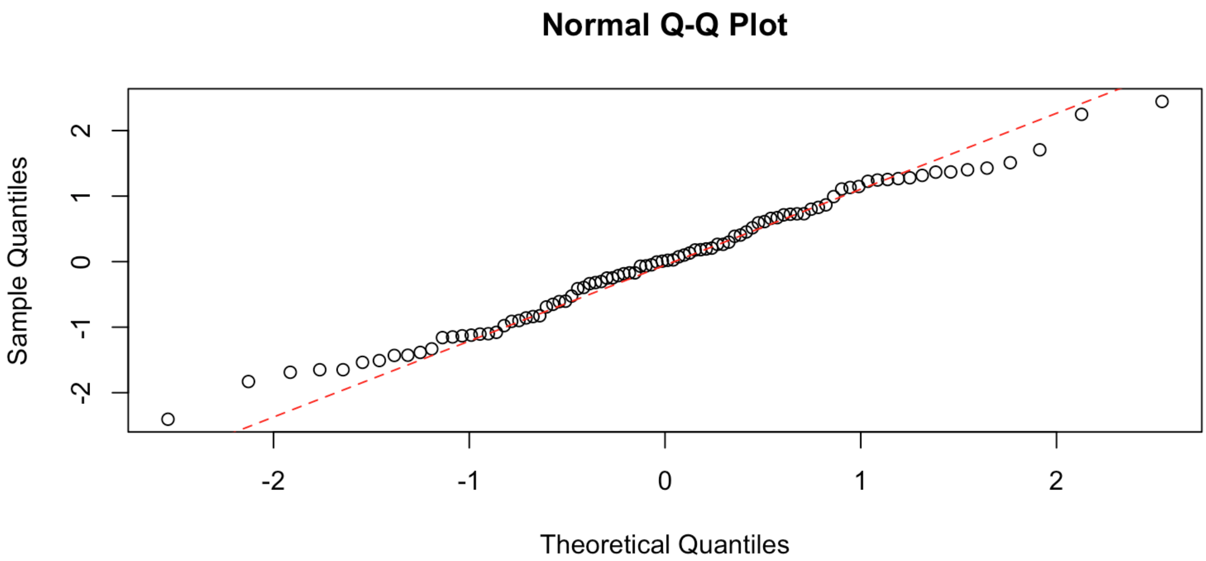
- a p-value < 0.05 shows that the model is significant
- R^2 = 0.6655 shows the model is partially significant

Slope: $\hat{\beta}_1$= -0.110029 Intercept: $\hat{\beta}_{00= 213.72015}$

```
res.linearTrendModel = rstudent(linearTrendModel)

doTimeSeriesPlot(
  x = as.vector(time(ozone.ts)),
  y = res.linearTrendModel,
  title = 'Standardised Residuals of Linear Trends model ',
  subtitle = 'with Simple moving average')
```




```
qqnorm(res.linearTrendModel) # Q-Q Plot for normality
abline(linearTrendModel)
qqline(
  col = 2,
  res.linearTrendModel,
  lty = 2,
  lwd = 1)
```

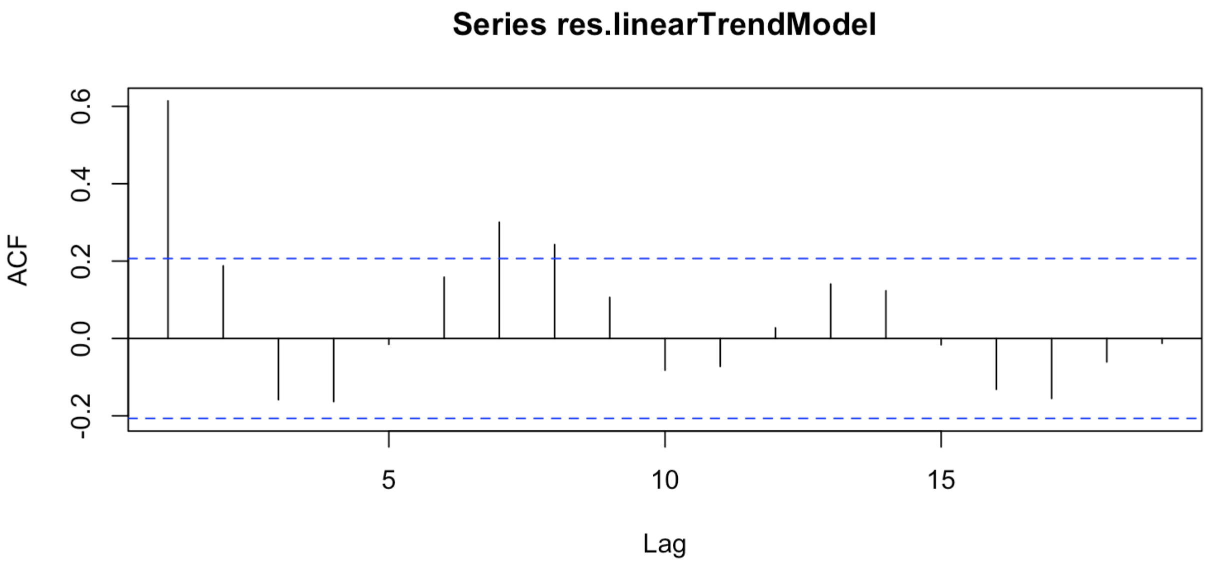


Q-Q plot shows significant skew indicating possible non-normality or multi-modal data

```
shapiro.test(res.linearTrendModel) # Check normality
```

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

```
acf(res.linearTrendModel)
```



ACF shows that this data is not white noise. Shows outliers outside of the 0.2 confidence region.

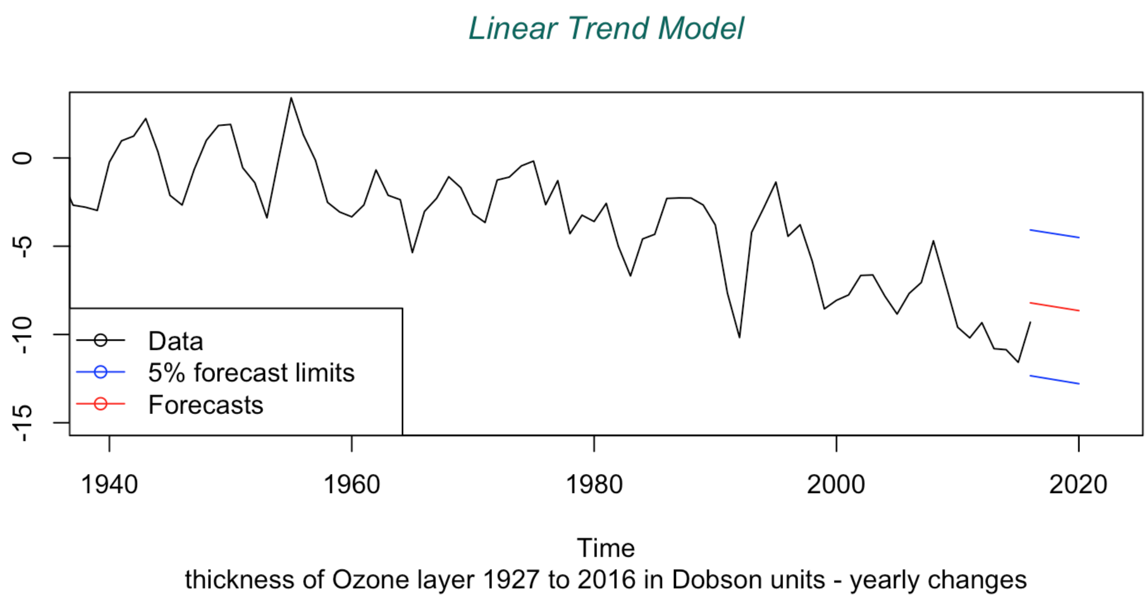
Forecast for Linear

Task: Give predictions of yearly changes for the next 5 years

```
doForecast <- function(ourModel, start, plotTitle, ...) {
  args <- list(...)
  exist <- "t2" %in% names(args)

  t = c(2017:2021)
  t2 = t ^ 2

  if(exist) { new = data.frame(t, t2) }
  else { new = data.frame(t) }

  # Two-step algorithm
  forecasts = predict(ourModel, new, interval = "prediction")
  print(forecasts) # predicted interval

  plot(
    ozone.ts,
    xlim = c(1940,2022),
    ylim = c(-15,3),
    ylab = "")
  lines(ts(as.vector(forecasts[,1]), start), col="red", type="l")
  lines(ts(as.vector(forecasts[,2]), start), col="blue", type="l")
  lines(ts(as.vector(forecasts[,3]), start), col="blue", type="l")
  legend(
    "bottomleft",
    c("Data","5% forecast limits", "Forecasts"),
    col=c("black","blue","red"),
    lty=1,
    pch=1,
    text.width = 20)
    title(
    main = plotTitle,
    sub = 'thickness of Ozone layer 1927 to 2016 in Dobson units - yearly changes',
    col.main = colourPallete[9],
    font.main = 3)
}
```

```
doForecast(linearTrendModel, 2016, 'Linear Trend Model')
```

```
##          fit       lwr       upr
## 1 -8.208590 -12.33732 -4.079864
## 2 -8.318619 -12.45033 -4.186902
## 3 -8.428648 -12.56342 -4.293878
## 4 -8.538677 -12.67656 -4.400792
## 5 -8.648706 -12.78977 -4.507643
```



Hypothesis for Linear

H0:μΔ=0 HA:μΔ≠0 reject hypothesis because data is not normal The stochastic component of model 2 is not normally distributed

H_0: $y_t \neq \beta_0 + \beta_1 x_t + \varepsilon_t$ H_a: $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$

We found that as the p-value was less than 5% and the null mean (i.e. zero) fell within the 95% confidence interval. As such the null hypothesis was rejected.

Quadratic time trends

Model 2

- Use the least squares approach to fit a quadratic time trend to the Ozone data

```
# t = time(ozone.ts)
t2 = t^2 # square it
model.ozone.qa =
  lm(ozone.ts ~ t + t2) # label the quadratic trend model
summary(model.ozone.qa)
```

```
##
## 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
```

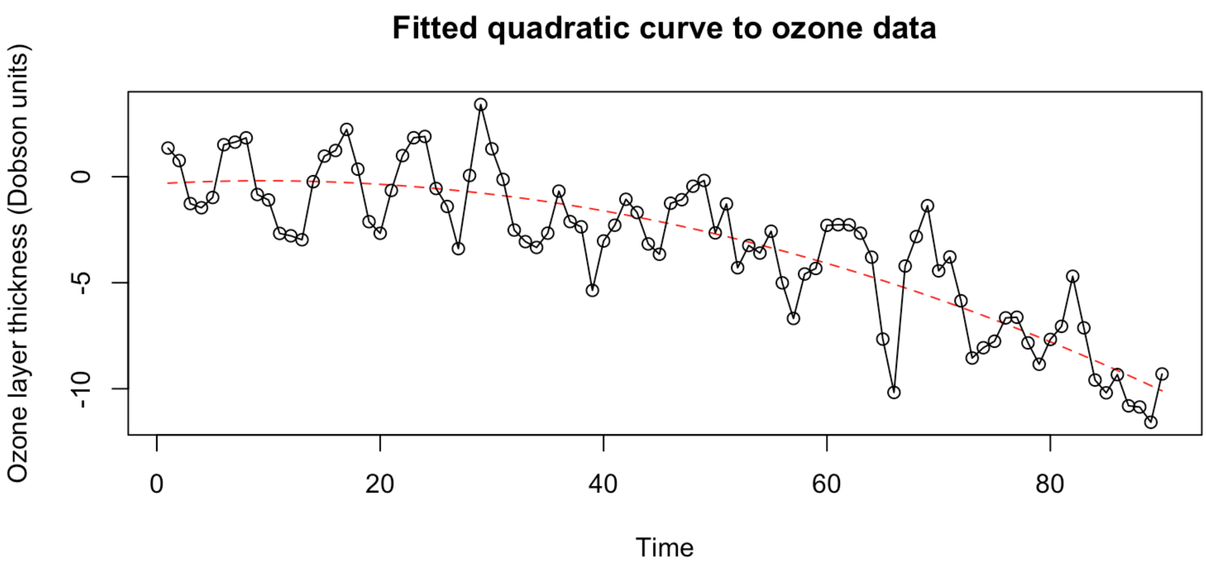
p: 2.2e-16 < 0.05

F: 87 DF

R^2: 0.7331

shows the model is more significant than our linear model

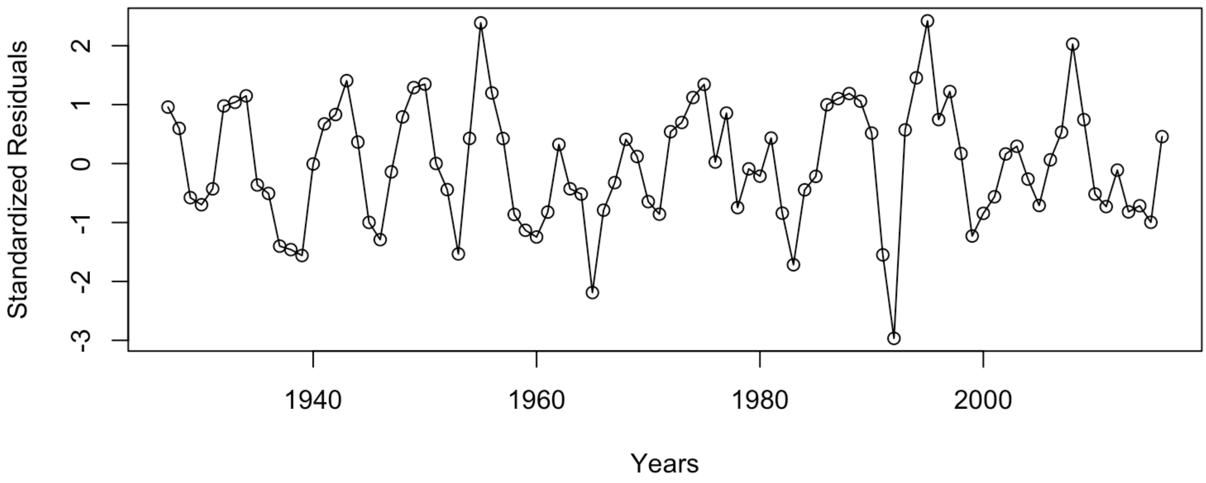
```
plot(
  ts(fitted(model.ozone.qa)),
  ylab = default_ylab,
  ylim = c(
    min(c(
      fitted(model.ozone.qa),
      as.vector(ozone.ts)
    )),
    max(c(
      fitted(model.ozone.qa),
      as.vector(ozone.ts)
    ))
  ),
  col = "red",
  lty = 2,
  main = "Fitted quadratic curve to ozone data",
  type = "l",
  lines = lines(as.vector(ozone.ts), type = "o")
  )
lines(as.vector(ozone.ts), type = "o")
```



Visual inspection shows a closer fit to the data than linear

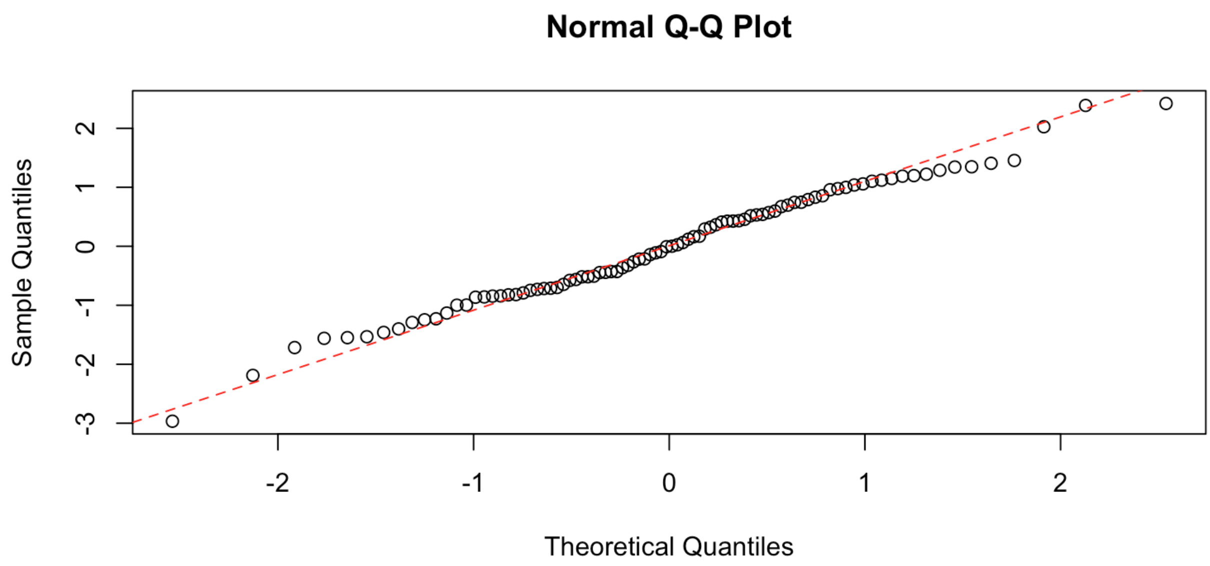
Quadratic residuals

```
res.model.ozone.qa = rstudent(model.ozone.qa)
plot(
  x = as.vector(time(ozone.ts)),
  xlab = 'Years',
  ylab = 'Standardized Residuals',
  y = res.model.ozone.qa,
  type = 'o')
```



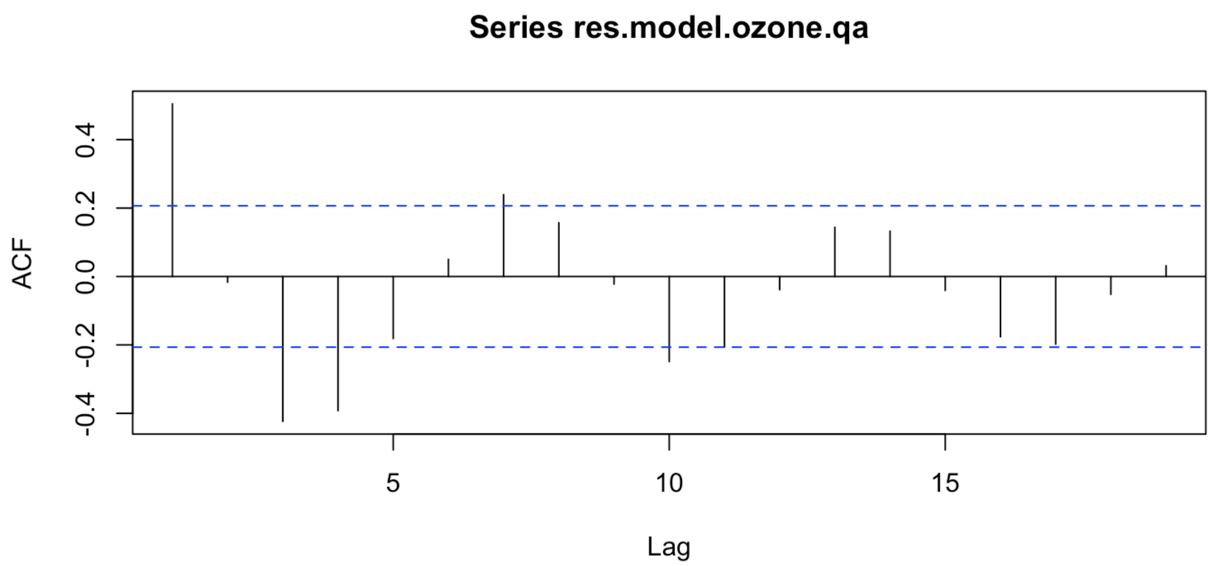
Standardised residual shows a moving average

```
qqnorm(res.model.ozone.qa) # Q-Q Plot for normality
qqline(
  res.model.ozone.qa,
  col = 2,
  lty = 2,
  lwd = 1)
```



Data is not normal because Q-Q plot has some skew towards the final data. However, it is less skewed than the linear model

```
acf(res.model.ozone.qa)
```



ACF shows not white noise. Shows 3 clear and 5 total outliers outside of the 0.2 confidence region.

```
shapiro.test(res.model.ozone.qa) # Check normality
```

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

Hypothesis for Quadratic

H0:μΔ=0 HA:μΔ≠0 reject hypothesis because data is not normal The stochastic component of model 2 is not normally distributed

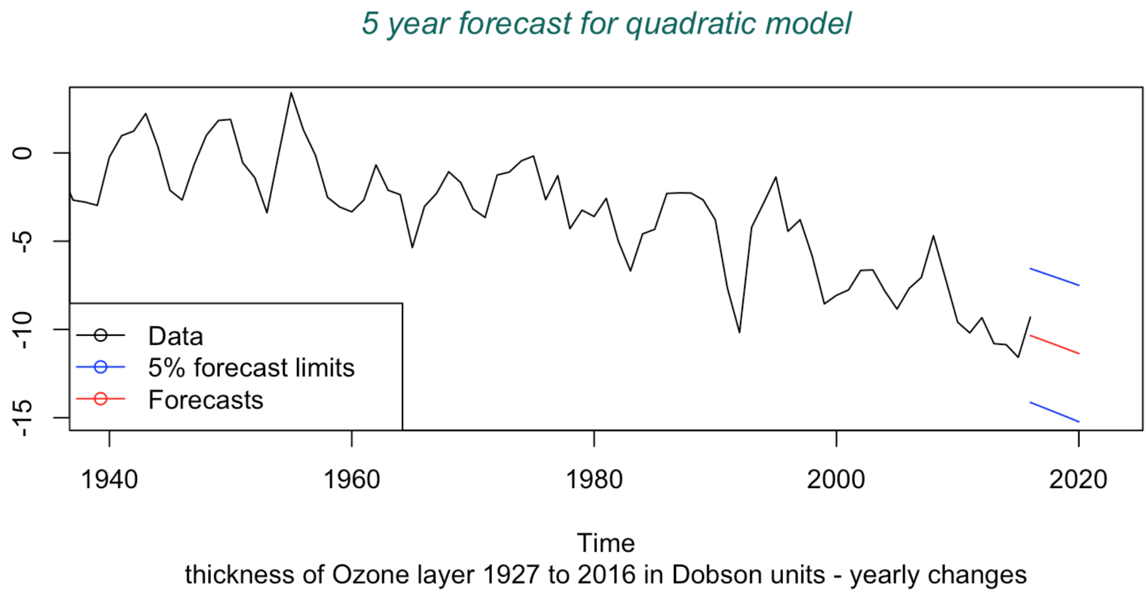
H_0: $\mu_t \neq \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t$ H_a: $\mu_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t$

We found that as the p-value was less than 5% and the null mean (i.e. zero) fell within the 95% confidence interval. As such the null hypothesis was rejected.

Forecast for Quadratic model

```
doForecast(model.ozone.qa, 2016, '5 year forecast for quadratic model', t2=T)
```

```
##         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
```



Harmonic Model

```
har=harmonic(ozone.ts,0.45)
model.ozone.har=lm(ozone.ts~har)
summary(model.ozone.har)
```

```
##
## 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 ***
## harcos(2*pi*t)         NA         NA      NA       NA
## harsin(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
```

- p-value = 0.4441 > 0.05, shows the model is insignificant
- Cos = NA
- No seasonality
- Therefore warrants no further investigation

Conclusion

Task: Find the best fitting trend model to this dataset

Of our three models, the quadratic trend model was the most successful fit because it better correlates with the decline towards the end of the series.

Task: Give predictions of yearly changes for the next 5 years

Utilising the linear and quadratic models, graphs and numeric outputs with a 5% variance have been provided.

The data shows an increased rate of decay over time. This does not appear to be a stationary series, with seasonal pattern and moving average. Further investigation to find a better fitting model is warranted.