Moving Averages

Lesson 2.1

Introduction

- Smoothing is the process of removing random variations that appear as coarseness in a plot of raw time series data.
- Smoothing is usually done to help us better see patterns, trends for example, in time series.
- Generally, we smooth out the irregular roughness to see a clearer signal.
- Analysts also refer to the smoothing process as filtering the data
- For seasonal data, we might smooth out the seasonality so that we can identify the trend.
- Smoothing doesn't provide us with a model, but it can be a good first step in describing various components of the series.

Moving Averages

- The traditional use of the term moving average is that at each point in time we determine (possibly weighted) averages of observed values that surround a particular time.
- This might be done by looking at a "one-sided" moving average in which you average all values for the previous year's worth of data or a centered moving average in which you use values both before and after the current time.

One-sided moving averages

- One-sided moving averages include the current and previous observations for each average. For example, the formula for a moving average (MA) of X at time t with a length of 7 is the following:

$$\frac{x_{t-6} + x_{t-5} + x_{t-4} + +x_{t-3} + x_{t-2} + x_{t-1} + x_t}{7}$$

Centered moving averages

- For instance, at time t, a "centered moving average of length 3" with equal weights would be the average of values at times $t-1$, t, $t+1$.
- To take away seasonality from a series so we can better see trend, we would use a moving average with a length = seasonal span.
- Thus in the smoothed series, each smoothed value has been averaged across all seasons.
- For quarterly data, for example, we could define a smoothed value for time t as

$$\frac{x_t + x_{t-1} + x_{t-2} + x_{t-3}}{4}$$

the average of this time and the previous 3 quarters. In R code this will be a one-sided filter.

- To smooth away seasonality in quarterly data, in order to identify trend, the usual convention is to use the moving average smoothed at time t is

$$\frac{1}{8}x_{t-2} + \frac{1}{4}x_{t-1} + \frac{1}{4}x_t + \frac{1}{4}x_{t+1} + \frac{1}{8}x_{t+2}$$

- To smooth away seasonality in monthly data, in order to identify trend, the usual convention is to use the moving average smoothed at time t which is given by

$$\frac{1}{24}x_{t-6} + \frac{1}{12}x_{t-5} + \frac{1}{12}x_{t-4} + \cdots + \frac{1}{12}x_{t+4} + \frac{1}{12}x_{t+5} + \frac{1}{24}x_{t+6}$$

- That is, we apply weight $\frac{1}{24}$ to values at times $t-6$ and $t+6$ and weight $\frac{1}{12}$ to all values at all times between $t-5$ and $t+5$.
- A centered moving average creates a bit of a difficulty when we have an even number of time periods in the seasonal span (as we usually do).
- For example, the formula for a centered moving average with a length of 8 is as follows:

$$\frac{1}{2}x_{t-4} + \frac{1}{8}x_{t-3} + \frac{1}{8}x_{t-2} + \frac{1}{8}x_{t-1} + \frac{1}{8}x_t + \frac{1}{8}x_{t+1} + \frac{1}{8}x_{t+2} + \frac{1}{8}x_{t+3} + \frac{1}{2}x_{t+4}$$

Some examples

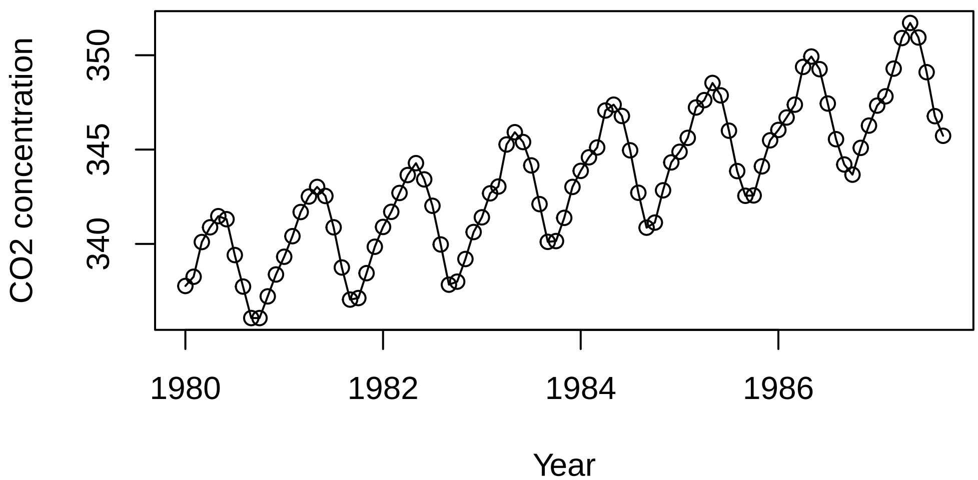
Let us load some packages and the CO_2 data.

```
library(readxl)
library(ggplot2)
library(tidyverse)
library(forecast)

co2 <- read_excel("CO2 1980-1987.xlsx")
```

```
co2_ts<-ts(co2$CO2,start=c(1980,1),end=c(1987,9),frequency=12)

plot(co2_ts,type="o",xlab="Year",
     ylab="CO2 concentration")
```



Again, this time series shows both seasonality and trend.The peaks are observed in the month of May every year.

Let us try computing moving averages as outlined above and observe which period or length best captures the patterns in the data.

```
cma.3<-ma(co2_ts,order=3)
cma.5<-ma(co2_ts,order=5)
cma.7<-ma(co2_ts,order=7)
plot.ts(cbind(co2_ts,cma.3,cma.5,cma.7),
    plot.type = 'single',
```

```r
    col = c("black","red","blue","green"),
    ylab = "CO2 concentration")
```