Lesson 2.1

Moving Averages

Introduction

The classical method of time series decomposition originated in the 1920s and was widely used until the 1950s. It still forms the basis of many time series decomposition methods, so it is important to understand how it works. The first step in a classical decomposition is to use a moving average method to estimate the trend-cycle, so we begin by discussing moving averages.

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

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

Meanwhile, 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:

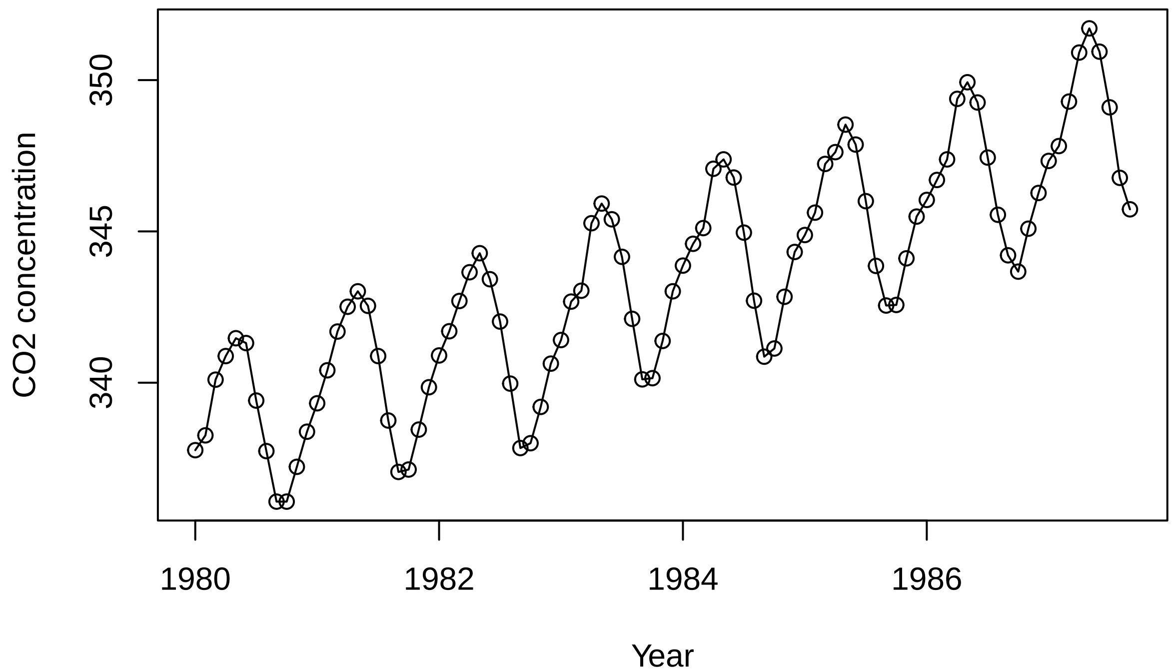
$$\frac{1}{2}x_{t-4} + \frac{1}{8}x_{t-3} + \frac{1}{8}x_{t-2} + \frac{1}{8}x_{t-1} + \frac{1}{8}x_t + \frac{1}{8}x_{t+1} + \frac{1}{8}x_{t+2} + \frac{1}{8}x_{t+3} + \frac{1}{2}x_{t+4}$$

Some examples

Let us load some packages and the CO_2 data.

```
library(readxl)
library(ggplot2)
library(tidyverse)
library(forecast)
library(fpp2)

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

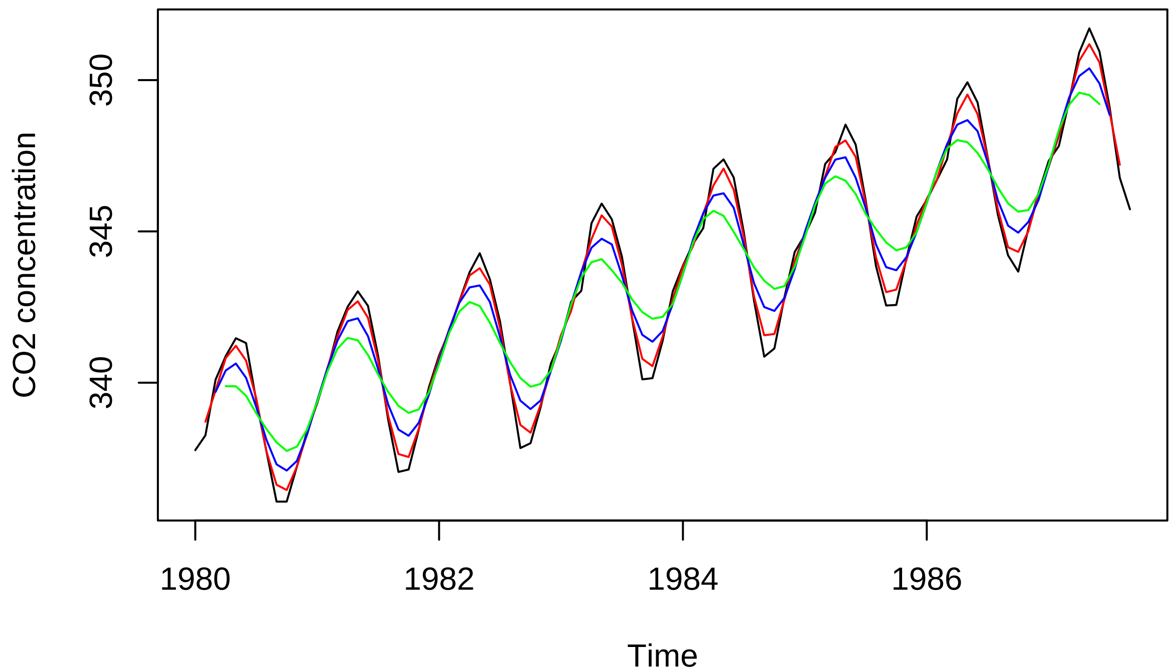
```
co2_ts<-ts(co2$CO2,
           start=c(1980,1),
           end=c(1987,9),
           frequency=12)

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



Moving averages of moving averages

It is possible to apply a moving average to a moving average. One reason for doing this is to make an even-order moving average symmetric.

For example, we might take a moving average of order 4, and then apply another moving average of order 2 to the results. In the following table, this has been done for the first few years of the Australian quarterly beer production data.

```
beer <- window(ausbeer,start=1992)
ma4 <- ma(beer, order=4, centre=FALSE)
ma2x4 <- ma(beer, order=4, centre=TRUE)

as.data.frame(cbind(beer, ma4, ma2x4))
```

```
##    beer    ma4   ma2x4
## 1   443     NA      NA
## 2   410 451.25      NA
## 3   420 448.75 450.000
## 4   532 451.50 450.125
## 5   433 449.00 450.250
```

```
## 6   421 444.00 446.500
## 7   410 448.00 446.000
## 8   512 438.00 443.000
## 9   449 441.25 439.625
## 10  381 446.00 443.625
## 11  423 440.25 443.125
## 12  531 447.00 443.625
## 13  426 445.25 446.125
## 14  408 442.50 443.875
## 15  416 438.25 440.375
## 16  520 435.75 437.000
## 17  409 431.25 433.500
## 18  398 428.00 429.625
## 19  398 433.75 430.875
## 20  507 433.75 433.750
## 21  432 435.75 434.750
## 22  398 440.50 438.125
## 23  406 439.50 440.000
## 24  526 439.25 439.375
## 25  428 438.50 438.875
## 26  397 436.25 437.375
## 27  403 438.00 437.125
## 28  517 434.50 436.250
## 29  435 439.75 437.125
## 30  383 440.75 440.250
## 31  424 437.25 439.000
## 32  521 442.00 439.625
## 33  421 439.50 440.750
## 34  402 434.25 436.875
## 35  414 441.75 438.000
## 36  500 436.25 439.000
## 37  451 436.75 436.500
## 38  380 434.75 435.750
## 39  416 429.00 431.875
## 40  492 436.00 432.500
## 41  428 433.50 434.750
## 42  408 437.00 435.250
## 43  406 438.75 437.875
## 44  506 431.75 435.250
## 45  435 435.50 433.625
## 46  380 431.50 433.500
## 47  421 431.50 431.500
## 48  490 434.00 432.750
## 49  435 431.75 432.875
## 50  390 422.75 427.250
```

```
## 51  412 418.00 420.375
## 52  454 421.25 419.625
## 53  416 420.25 420.750
## 54  403 427.25 423.750
## 55  408 432.75 430.000
## 56  482 428.50 430.625
## 57  438 427.75 428.125
## 58  386 430.00 428.875
## 59  405 427.25 428.625
## 60  491 426.50 426.875
## 61  427 423.75 425.125
## 62  383 419.25 421.500
## 63  394 417.50 418.375
## 64  473 419.25 418.375
## 65  420 423.25 421.250
## 66  390 427.00 425.125
## 67  410 425.75 426.375
## 68  488 427.75 426.750
## 69  415 430.00 428.875
## 70  398 430.00 430.000
## 71  419 429.75 429.875
## 72  488 423.75 426.750
## 73  414     NA      NA
## 74  374     NA      NA
```

The notation *2 x 4-MA* in the last column means a 4-MA followed by a 2-MA. The values in the last column are obtained by taking a moving average of order 2 of the values in the previous column.

When a 2-MA follows a moving average of an even order (such as 4), it is called a "centred moving average of order 4". This is because the results are now symmetric. To see that this is the case, we can write the *2 x 4-MA* as follows:

$$
\begin{aligned}
\hat{T}_t &= \frac{1}{2}[\frac{1}{4}(y_{t-2} + y_{t-1} + y_t + y_{t+1}) + \frac{1}{4}(y_{t-1} + y_t + y_{t+1} + y_{t+2})] \\
&= \frac{1}{8}y_{t-2} + \frac{1}{4}y_{t-1} + \frac{1}{4}y_t + \frac{1}{4}y_{t+1} + \frac{1}{8}y_{t+2}
\end{aligned}
$$

It is now a weighted average of observations that is symmetric. By default, the ma() function in R will return a centred moving average for even orders (unless *center=FALSE* is specified).

Other combinations of moving averages are also possible. For example, a *3 × 3-MA* is often used, and consists of a moving average of order 3 followed by another moving average of order 3. In general, an even order MA should be followed by an even order MA to make it symmetric. Similarly, an odd order MA should be followed by an odd order MA.

Estimating the trend-cycle with seasonal data

The most common use of centred moving averages is for estimating the trend-cycle from seasonal data. Consider the *2 x 4-MA*:

$$\hat{T}_t = \frac{1}{8}y_{t-2} + \frac{1}{4}y_{t-1} + \frac{1}{4}y_t + \frac{1}{4}y_{t+1} + \frac{1}{8}y_{t+2}$$

When applied to quarterly data, each quarter of the year is given equal weight as the first and last terms apply to the same quarter in consecutive years. Consequently, the seasonal variation will be averaged out and the resulting values of $\hat{T}_t$ will have little or no seasonal variation remaining. A similar effect would be obtained using a *2 × 8-MA* or a *2 × 12-MA* to quarterly data.

In general, a *2 × m-MA* is equivalent to a weighted moving average of order $m+1$, where all observations take the weight $\frac{1}{m}$, except for the first and last terms which take weights $\frac{1}{2m}$. So, if the seasonal period is even and of order m, we use a *2 × m-MA* to estimate the trend-cycle. If the seasonal period is odd and of order m, we use a *m-MA* to estimate the trend-cycle. For example, a *2 × 12-MA* can be used to estimate the trend-cycle of monthly data and a *7-MA* can be used to estimate the trend-cycle of daily data with a weekly seasonality.

Other choices for the order of the MA will usually result in trend-cycle estimates being contaminated by the seasonality in the data.

Note that combinations of moving averages result in weighted moving averages. For example, the *2 x 4-MA* is equivalent to a weighted *5-MA* with weights given by $[\frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}]$