STA 137—Applied Time Series Analysis — Solutions to Homework 5

Seasonality removal

1. The left part of Figure 1 displays the accidental death data that already includes moving average estimates of the drift component.

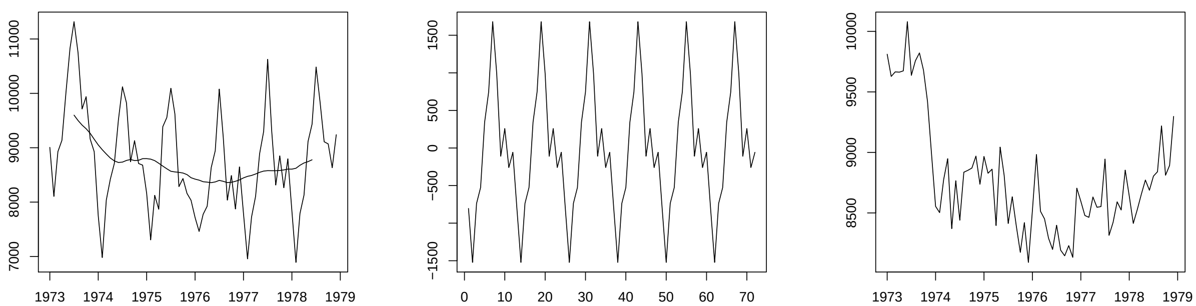

Figure 1: The accidental death data and a trend fit by moving average smoothing (left), the estimated seasonal component (middle) and the deseasonalized data (right).

2. We apply in the following the moving average method to estimate the seasonal components. Consequently, we fit first an estimated trend function $\hat{m}_t$ using the moving average filter with $q = d/2 = 6$. This means that there are no estimates for the first 6 and the last 6 observations. To perform the analysis in R, you may use the `filter` function

```
> hatm = filter(accidents, sides=2, c(.5,rep(1,11),.5)/12)
```

This produces the trend already mentioned in part (a). Now the seasonal components can be estimated with the following procedure. Note that we have six years of data, so $72 = n = Nd = 6 \cdot 12$. According to the class notes, we transform the data array `accidents` into a matrix. We do the same also for the estimated trend `hatm`.

```
> A = matrix(accidents, ncol=12, byrow="TRUE")
> M = matrix(hatm, ncol=12, byrow="TRUE")
> mu = array(0,12)
> for (k in 1:6) mu[k] = sum(A[2:6,k]-M[2:6,k])/5
> for (k in 7:12) mu[k] = sum(A[1:5,k]-M[1:5,k])/5
> hats = rep(mu-mean(mu),6)
```

With this small program, we compute first the quantities (the commands in lines 3–5)

$$\mu_k = \frac{1}{5}\sum_{j=2}^{6}(x_{k+12(j-1)} - \hat{m}_{k+12(j-1)}), \qquad k = 1, \ldots, 6,$$

$$\mu_k = \frac{1}{5}\sum_{j=1}^{5}(x_{k+12(j-1)} - \hat{m}_{k+12(j-1)}), \qquad k = 7, \ldots, 12.$$

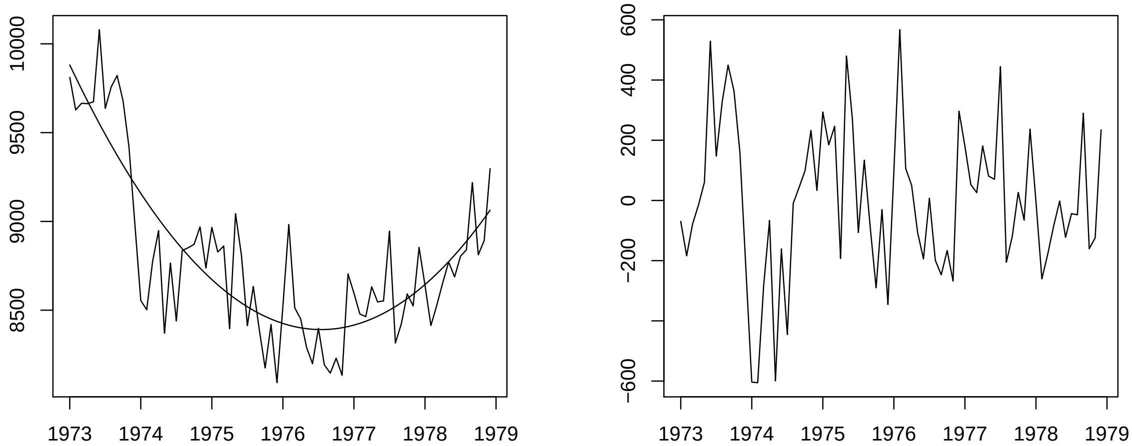

Figure 2: The fitted parabola (left) and the residuals (right) for the accidental death data.

Finally, the seasonality components can be obtained from the last line of code. These seasonal estimates $\hat{s}_t$, $t = 1, \ldots, 12$, are provided in the following table. The corresponding plot is shown in the middle part of Figure 1.

t	1	2	3	4	5	6
$\hat{s}_t$	-804.32	-1521.74	-737.47	-525.81	343.42	746.41
t	7	8	9	10	11	12
$\hat{s}_t$	1679.96	986.84	-108.77	258.31	-259.38	-57.46

3. The deseasonalized data is plotted in the right panel of Figure 1. It can be seen that there is still a substantial amount of trend left in the transformed observations.

4. The parabola can be fit with the following R code:

```
> desea = accidents-hats
> t = 1:72; t2 = t*t
> deseafit = lm(desea~t+t2)
```

With `summary(deseafit)` you will find that the fitted parabola is given by the estimated values

$$\hat{b}_0 = 9951.822, \qquad \hat{b}_1 = -71.817, \qquad \hat{b}_2 = 0.826.$$

The fitted trend is shown in Figure 2 (left panel). It is stored as `deseafit$fit`.

5. The corresponding residuals are given in the right panel of Figure 2. They can be accessed with the command `deseafit$resid`. In the plot, one can see a potential anomaly around the year 1974. It is probably due to a new seatbelt legislation in the US.