Time Series Exercises 1: Solutions

1. (i) With five equal weights ($a_r = 1/5$ in our notation) the smoothed series is

0.88	1.08	1.24	1.44	1.26	1.48	1.34	1.72	1.62	2.36
2.50	2.78	2.88	2.66	2.58	2.7	2.94	2.84	3.16	3.14

and is somewhat smoother than the original series. To estimate the trend we can fit the regression line using R, to find

$$x(t) = 0.826 + 0.125t.$$

As a rough guide to the quality of fit of the line, we note that $R^2 = 0.89$. (ii) Now using weights which are in fact $\frac{1}{16}$, $\frac{4}{16}$, $\frac{6}{16}$, $\frac{4}{16}$ and $\frac{1}{16}$ respectively, the new smoothed series is

0.92	1.15	1.28	1.37	1.44	1.40	1.34	1.46	1.83	2.34
2.74	2.86	2.81	2.76	2.62	2.48	2.69	3.14	3.31	3.16

which is in fact rather more locally variable than that in (i). Fitting a line through these points we find regression line to be

$$x(t) = 0.837 + 0.126t,$$

similar to that in (i), and in fact $R^2 = 0.88$. Actually for the raw data the regression line is

$$x(t) = 0.64 + 0.119t$$

and $R^2 = 0.63$, and the fit of the line is poorer. On this evidence smoothing the series might give a more reliable estimate of an underlying trend.

2. Plotting the data, if the model really is the logistic equation given then the rate of convergence to the limit a is rather slow. Judging from the plot, we might estimate a to be about 1400, say. Now, note that the model can be re-written as

$$\frac{a - X(t)}{X(t)} = ce^{-abt},$$

so taking logarithms we have

$$\begin{aligned}\log\left(\frac{a-X(t)}{X(t)}\right) &= \log(c)+\log\left(e^{-abt}\right)\\ &= \log(c)-abt.\end{aligned}$$

Hence plotting $\log((1400-x(t))/x(t))$ against t should give a straight line graph (with approximately) intercept $\log(c)$ and gradient $-1400b$. Plotting $(4.14, 3.55, \ldots, -0.708)$ against time gives a good straight line of negative slope, the regression line being in fact $4.14 - 0.133t$, and $R^2 = 0.97$ suggesting an excellent fit. From this information we might estimate c to be 62.80 and b to be 9.5×10^{-5}. It is of interest to investigate here how the estimation of b and c depends on the estimate of a.

3. In general, the new series is

$$\begin{aligned} z(t) &= \sum_{j=-r}^{s} b_j y(t+j)\\ &= \sum_{j=-r}^{s} b_j\left(\sum_{k=-m}^{n} a_k x(t+j+k)\right)\\ &= \sum_{i=-r-m}^{s+n} c_i x(t+i)\end{aligned}$$

(putting $i = j + k$) where

$$c_i = \sum_{k=-m}^{n} a_k b_{i-k}$$

defines the so–called *convolution* of $\{a_j\}$ and $\{b_j\}$. For instance, applying the filter $\left\{\frac{1}{2}, \frac{1}{2}\right\}$ followed by $\left\{\frac{1}{3}, \frac{2}{3}\right\}$ we can describe the combined effect in the following table:

Series	$\left\{\frac{1}{2},\frac{1}{2}\right\}$	$\left\{\frac{1}{3},\frac{2}{3}\right\}$
$x(1)$	$\frac{x(1)+x(2)}{2}$	$\frac{x(1)+x(2)}{6}+\frac{2(x(2)+x(3))}{6}=\frac{x(1)+3x(2)+2x(3)}{6}$
$x(2)$	$\frac{x(2)+x(3)}{2}$	$\frac{x(2)+x(3)}{6}+\frac{2(x(3)+x(4))}{6}=\frac{x(2)+3x(3)+2x(4)}{6}$
$x(3)$	$\frac{x(3)+x(4)}{2}$	
$x(4)$		$\vdots$
$x(5)$	$\vdots$	
$\vdots$		

So the weights for the combined filter is $\left\{\frac{1}{6},\frac{1}{2},\frac{1}{3}\right\}$. Here $r=m=0$, $s=n=1$, so using our convolution result observe that

$$\begin{aligned} \frac{1}{2}\times\frac{1}{3} &= \frac{1}{6}, \\ \frac{1}{2}\times\frac{1}{3}+\frac{1}{2}\times\frac{2}{3} &= \frac{1}{2}, \\ \frac{1}{2}\times\frac{2}{3} &= \frac{1}{3}. \end{aligned}$$

4. .

(a) The plot shows a seasonal effect of period 4 as would be expected, along with an apparent upwards trend.

(b) An additive seasonal effect seems appropriate here, so

$$X(t)=\mu(t)+S(t)+\varepsilon$$

seems reasonable, with $\mu(t)$ the trend at time t, $S(t)$ the seasonal index at time t and ε the error term. Difficult to say from such a short series, but a linear trend

$$\mu(t)=\beta_0+\beta_1 t$$

appears suitable.

(c)

		2003	2004	2005	Average
	1	*	–190.62	–206.25	–198.44
Quarter	2	*	–1107.50	–991.88	–1049.69
	3	672.50	903.75	*	788.13
	4	538.75	332.50	*	435.63

Adding the final column gives –24.37, so we can adjust by

$$-\frac{24.37}{4} = -6.06.$$

This gives seasonal index estimates as

$$\begin{aligned} S(1) &= -192.35 \\ S(2) &= -1043.60 \\ S(3) &= 794.22 \\ S(4) &= 441.72. \end{aligned}$$

(d) Residuals would be found using

$$z(t) = x(t) - \hat{\mu}(t) - S(t).$$

These would be useful for determining whether the model was valid.

(e) The values for the sales in the next two quarters could be taken as

$$\begin{aligned} \hat{x}(13) &= b_0 + 13b_1 - 192.35, \\ \hat{x}(14) &= b_0 + 14b_1 - 1043.60, \end{aligned}$$

where $\mu(t)$ is estimated by the line $b_0 + b_1 t$.

5. Using the `lag` command in R, a plot of $x(t)$ against $x(t+1)$ suggests some evidence of negative correlation at lag one, with consecutive values tending to lie opposite side of the mean. Finding the acf for the series, we have

k	1	2	3	4	5	6
r_k	–0.516	0.258	–0.276	0.268	–0.038	–0.021

6. We would expect about one in twenty observed values of r_k to be outside the range

$$-\frac{1}{400} \pm \frac{1.96}{\sqrt{400}} = (-1.005, 0.0955)$$

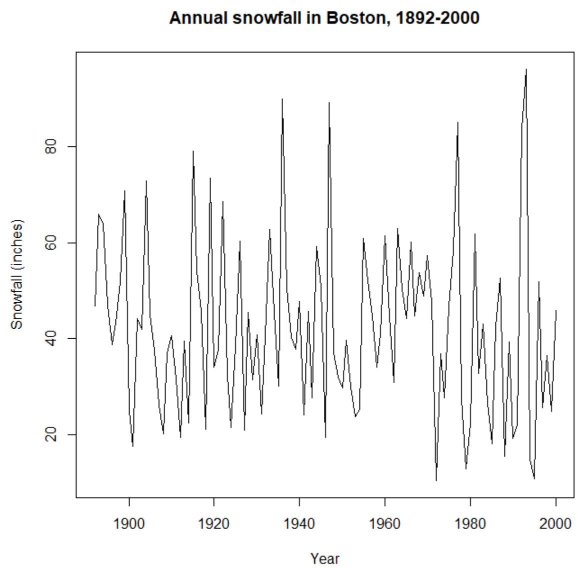
if the series is purely random. Now r_7 is just "significant" by this measure, but without any reasonable physical explanation as to why there should be correlation at lag 7, it is sensible to assume that this figure has simply been a chance occurrence, and suggest that the series in question appears to be purely random from the information given.

7. It is easy, and informative, to repeat this exercise several times using R. Note that we would expect all but about 5% of values of r_k to lie in the range
$$-\frac{1}{100} \pm \frac{1.96}{10}$$
for $k \geq 1$. What do you notice, if anything, about the behaviour of consecutive values of r_k?

8. In R, we can enter
   `> plot(bev, xlab="Year", ylab="Wheat price index", main="Beveridge wheat price index, 1500-1859")`
   Adjusting the aspect ratio is done by inserting `asp=y/x` in the above command, where y/x is the desired aspect ratio. We see that a low aspect ratio tends to compress the variation in the y–axis and can, for instance, visually downplay a trend. The opposite is true for larger values of the aspect ratio.
   In order to select the best aspect ratio, some authors (notably William Cleveland in works written in the 90s) suggest "banking to 45°", meaning ensuring that the average (or perhaps median) line segment elevation is 45°. However, selecting the aspect ratio to meet such a criterion in not straightforward.

9. The `read.table` command here requires we indicate that the file being read has headers, and the items are separated by tabs (hence `sep="\t"`). Enter `summary(yearlysnow)`for details about what has been read into R. We can coerce the snowfall data into a time series via
   `> snowfall <- ts(yearlysnow[,"Snowfall"], start=1892, end=2000)`
   then to adjust the data into inches use
   `> snowfall <- 10*snowfall`

The plot of the data is given below:



There is no obvious trend, but a high level of variability is evident, possibly of a cyclical nature. Examine the correlogram for the series – are there significant autocorrelations above lag 0?

BD