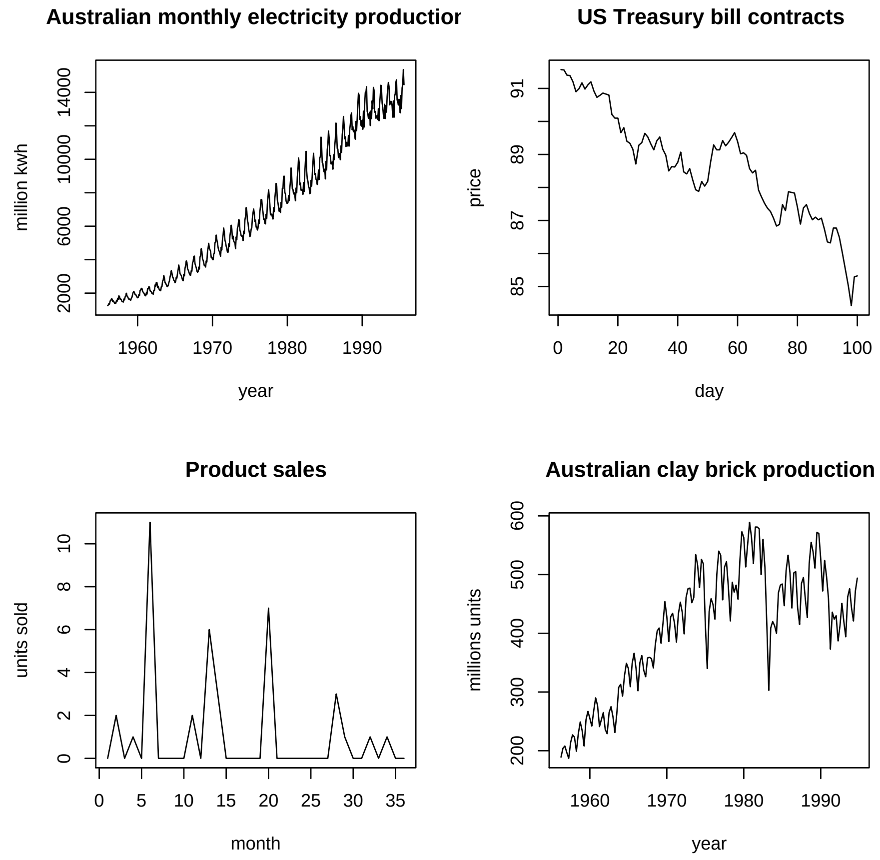
620-374 Notes: Time Series

1 Introduction

We suppose that we observe $\{X_k\}_{k=1}^t$, evenly spaced in time, and we wish to forecast X_{t+1}.

We model the *structural component* of the series by a deterministic function $f(k)$ and the *noise* by random variables ϵ_k with mean 0, so that $X_k = f(k) + \epsilon_k$. We project the mean f forwards, and use the noise to estimate the accuracy of our projection. Our ability to project f depends on its structure: trends and periodic components are tractable, cyclic components (with irregular period) are more difficult.



Australian monthly electricity productior

US Treasury bill contracts

Product sales

Australian clay brick production

Figure 1:

In Figure 1, the top two series show clear trends, the bottom-left has no notable trend and the bottom-right has an initial trend only. The top-left and bottom right examples

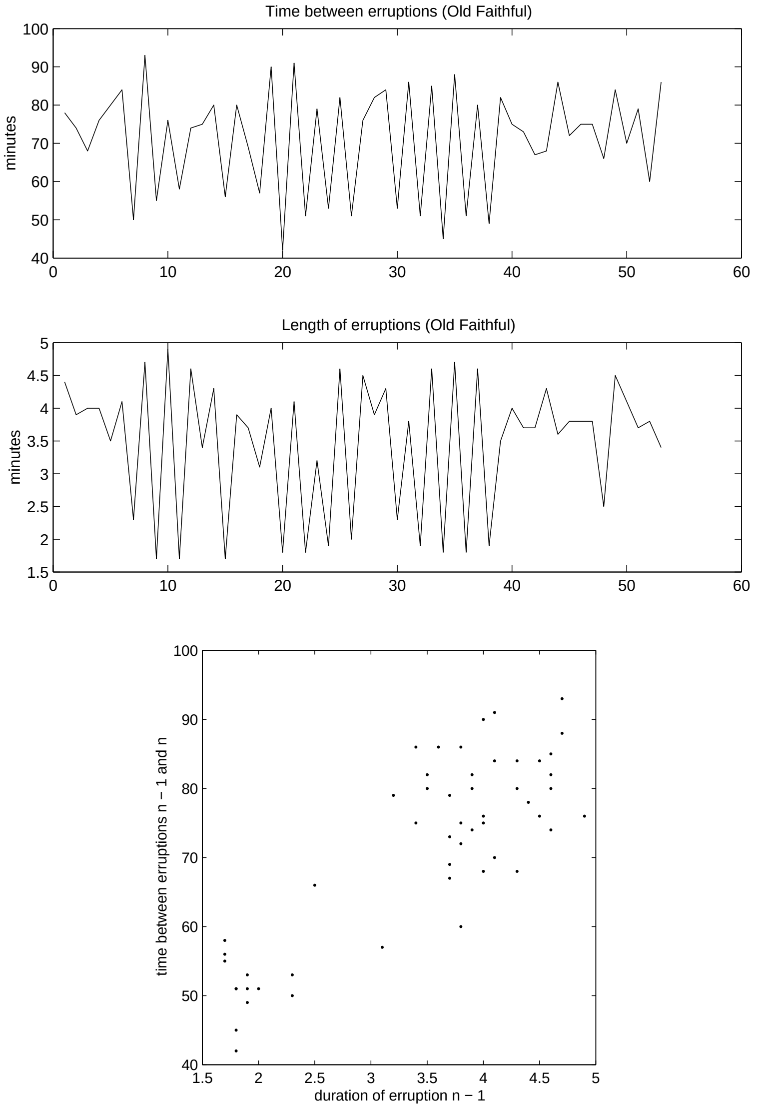


Figure 2: Old Faithful geyser, Yellowstone National Park

both exhibit seasonal periodic fluctuations, and the bottom-right example also exhibits cyclic behaviour (with irregular period). Both of the bottom examples would be difficult to forecast, the bottom-left because there is very little structure compared to noise, and the bottom-right because of the irregular cycles.

Auxiliary variables can be used to improve a forecast, and are one way of dealing with cycles, but we will not pursue this topic. The time between eruptions for the Old Faithful geyser in Yellowstone National Park is an example of a time-series whose forecasting can be improved using an auxiliary variable. In this case, forecasts of the time between eruptions can be improved if you know the length of eruptions, since there is a strong linear relation between the duration of an eruption and the time to the next one.

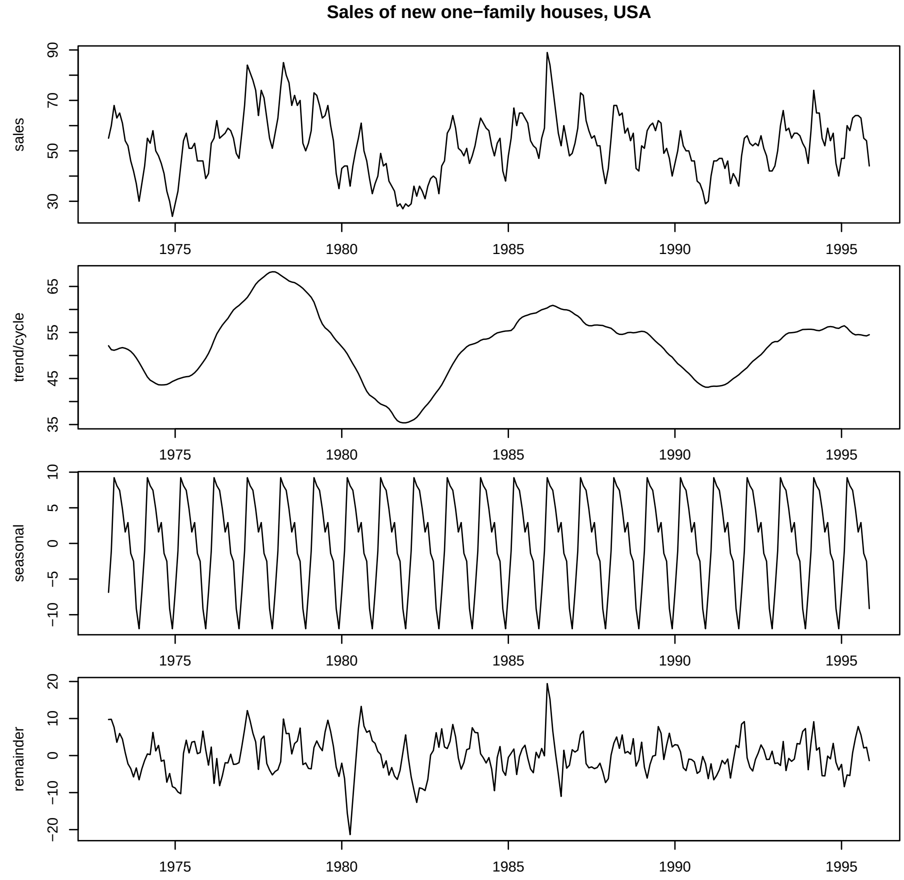


Figure 3: Decomposition of a time-series into trend/cycle, seasonal and noise components

2 Components of a time-series

Suppose that we have a time series of one of the following forms

$$\begin{aligned} X_t &= m_t + S_t + \epsilon_t \text{ (additive seasonality, additive noise)} \\ X_t &= m_t \times S_t \times \epsilon_t \text{ (multiplicative seasonality, multiplicative noise)} \\ X_t &= m_t \times S_t + \epsilon_t \text{ (multiplicative seasonality, additive noise)} \end{aligned}$$

where m_t is the trend/cycle component (deseasonalised mean), S_t is the seasonal component with period s, and ϵ_t the 'error' component.

In Figure 1 the top-left series exhibits multiplicative seasonality, and the bottom-right additive seasonality. Both series are monthly observations, and the seasonal component has period 12. Figure 3 shows a decomposition of a series with an additive seasonal component.

In the case of additive seasonality and additive noise (case 1 above) we assume that

$$\sum_{k=1}^{s} S_k = 0 \text{ and } \mathbb{E}\epsilon_t = 0.$$

When dealing with multiplicative seasonality and multiplicative error (case 2 above) we take logs and thus convert the problem to one of additive noise and additive error. Thus in this case we assume that

$$\prod_{k=1}^{s} S_k = 1,\ S_k > 0,\ \text{and } \mathbb{E}\log\epsilon_t = 0.$$

In the case of multiplicative seasonality and additive error (case 3 above) we again take logs, and assume $\prod_{k=1}^{s} S_k = 1$, $S_k > 0$. In this case let $\delta_t = (1 + \epsilon_t/(m_t S_t))$, so that $X_t = m_t \times S_t \times \delta_t$. If ϵ_t is small compared to $m_t S_t$ then, using a one-term Taylor series approximation, we have that $\log\delta_t \approx \epsilon_t/(m_t S_t)$. Thus the assumption $\mathbb{E}\log\delta_t = 0$ is roughly equivalent to $\mathbb{E}\epsilon_t = 0$, and so after taking logs case 3 is essentially equivalent to case 1. (If the errors ϵ_t are large compared to $m_t S_t$ then a different approach is required, which we do not pursue here.)

For the remainder of this section we assume that we have additive seasonality with additive noise. To estimate the trend/cycle component we use smoothers.

2.1 Moving Average smoothers

Moving Average (MA) smoothers take a weighted average of data from a *window* of fixed width, and use this to approximate m_k for k in the *centre* of the window. Let the window width be m, a positive integer, then the centre of the window $\{t+1, t+2, \ldots, t+m\}$ is the point $t+c$ where $c = (m+1)/2$. Note that if m is even then the centre is not integer valued.

The simplest MA smoothers just average across the window, with equal weights. That is, an m-MA smoother approximates m_{t+c} by

$$M_{t+c} = \frac{1}{m}\sum_{k=1}^{m} X_{t+k}.$$

More generally a *weighted* MA smoother approximates m_{t+c} by

$$M_{t+c} = \sum_{k=1}^{m} a_k X_{t+k}.$$

It is said to be centred if $a_k = a_{m+1-k}$, for $k = 1, \ldots \lfloor m/2 \rfloor$.

A basic requirement is $\sum a_k = 1$. To see why, suppose $m_t = m$ and $S_t = 0$, and consider $\mathbb{E}M_t$.

MA smoothers can be composed by applying one after the other. For example, an $m_2 \times m_1$-MA smoother applies an m_1-MA smoother first and then and m_2-MA smoother to the smoothed sequence. The composed smoother has window width $m_1 + m_2 - 1$, thus if both m_1 and m_2 are even, the combined smoother has an odd window width (so its centre will be integer valued). The composition of a 2-MA smoother with an m-MA smoother is called a centred MA smoother.

A centred m-MA smoother has window width $m+1$, centre $(m/2)+1$ and weights $(\frac{1}{2m}, \frac{1}{m}, \ldots, \frac{1}{m}, \frac{1}{2m})$.

Additive seasonality is removed by a weighted MA smoother if

$$\sum_{k=j \bmod s} a_k = \frac{1}{s} \text{ for each } j = 1, \ldots, s.$$

It is left as an exercise to show that this condition is necessary, in the sense that if it fails we can always find a time series such that M_t retains a seasonal component.

In Figure 4 we illustrate the removal (or not) of an additive seasonal effect using an MA smoother.

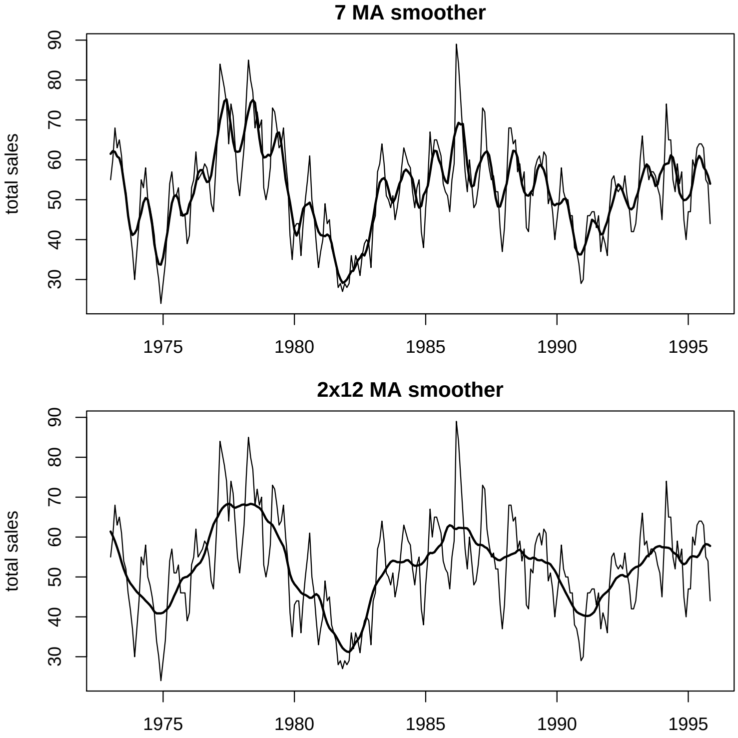


Figure 4: Two MA smoothers applied to the US house sales data, which has a seasonal component of period 12

Example: Spencer's 15 point moving average Consider

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
a_k	-0.009	-0.019	-0.016	0.009	0.066	0.144	0.209	0.231	0.209	0.144	0.066	0.009	-0.016	-0.019	-0.009

Show that a MA smoother with these weights removes a period 4 seasonal effect.

Example: m-MA smoother Show that if s is a divisor of m then an m-MA smoother will remove a period s seasonal effect. Moreover, any composition of such an m-MA smoother with another smoother will have the same effect.

In choosing the window width we have a trade off between the error of our estimate and over-smoothing. Consider a simple m-MA smoother, where m is a multiple of s, the seasonal period. Applying it to $X_t = m_t + S_t + \epsilon_t$ we get

$$M_{t+c} = \sum_{k=1}^{m} \frac{m_{t+k}}{m} + \sum_{k=1}^{m} \frac{\epsilon_{t+k}}{m}.$$

If the errors ϵ_t are i.i.d. with finite variance, then the central limit theorem tells us that the last term has size $O(1/\sqrt{m})$, decreasing with m. However, if m_t has any non-linear component then $\sum_{k=1}^{m} m_{t+k}/m$ will diverge from m_{t+c} as m increases.

For $t < 1 + c$ and $t > n - c$ (where we observe $X_1, \ldots, X_n$) we need to adjust M_{t+c}. We put

$$M_{t+c} = \frac{\sum_{k=1,\ldots,m \text{ s.t. } 1 \leq t+k \leq n} a_k X_{t+k}}{\sum_{k=1,\ldots,m \text{ s.t. } 1 \leq t+k \leq n} a_k}.$$

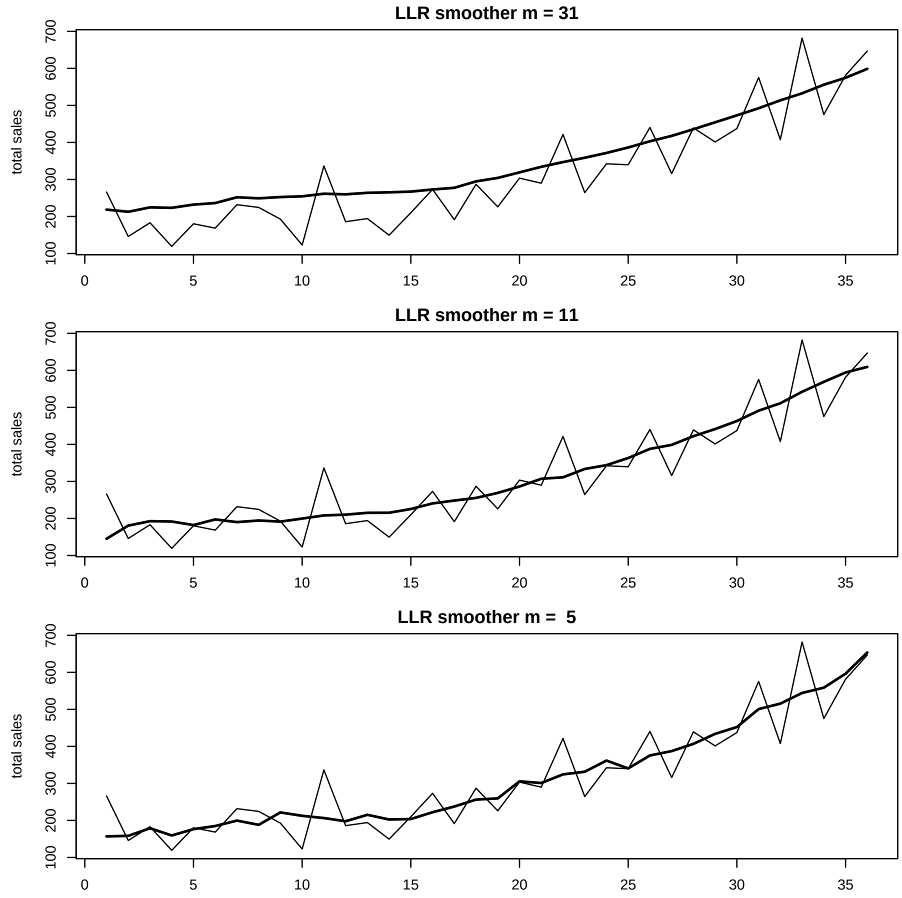


Figure 5: Local linear regression smooths with different window widths m

That is, we ignore the ‘missing’ values of X_t, and renormalise the remaining a_k. If m_t has a linear or higher order component then this procedure will introduce edge effects.

2.2 Local regression smoothing

Local regression smoothers also use a window of width m and centre $c = (m+1)/2$. Given positive weights $a_1, \ldots, a_m$ choose $\hat{a}_{t+c}$ and $\hat{b}_{t+c}$ to minimise

$$\sum_{k=1}^{m} a_k (X_{t+k} - a - b(t+k))^2$$

then estimate m_{t+c} with $M_{t+c} = \hat{a}_{t+c} + \hat{b}_{t+c}(t+c)$. If we do not have an observation of X_{t+k} (if $t+k < 1$ or $> n$) then we just omit that term. Typically we choose a_k so that $a_k = a_{m+1-k}$ and a_k is largest at the centre of the window and decreases out to the edges.

It is left as exercise to find conditions on the a_k which remove an additive seasonal component.

Local regression smoothers are less prone to edge effects compared to moving average smoothers. Like MA smoothers the choice of window width is a trade off between sensitivity to noise and over smoothing, as illustrated in Figure 5

2.3 Seasonal effect

If M_t is our estimate of m_t then for additive seasonality we estimate S_k with

$$\hat{S}_k = \frac{1}{m}((X_k - M_k) + (X_{s+k} - M_{s+k}) + (X_{2s+k} - M_{2s+k}) + \cdots)$$

where $m = \lfloor (t-k)/s \rfloor$. If we have multiplicative seasonality with multiplicative noise then we should estimate the seasonality for the log-transformed process. If we have multiplicative seasonality with additive noise then we estimate S_k with

$$\hat{S}_k = \frac{1}{m}((X_k/M_k) + (X_{s+k}/M_{s+k}) + (X_{2s+k}/M_{2s+k}) + \cdots).$$

Note that if we have aggregated data, such as total sales per month, the differences in the length of period being aggregated can cause seasonal effects. For example, total sales in February are only over 28 days whereas total sales in January are over 31 days. Seasonality due to aggregation can be removed by transforming the data to rates, by dividing each observation by the length of the aggregation period. For example, instead of forecasting total monthly sales, we should forecast average daily sales for each month.

3 Forecasting with exponential smoothers

We suppose that we observe X_t for $t = 1, \ldots, n$, and we wish to forecast X_{n+1}.

3.1 Simple exponential smoother

Let F_t be the exponential smoothing forecast of X_t then for some $\alpha \in [0, 1]$

$$F_{t+1} = \alpha X_t + (1-\alpha) F_t.$$

Prime with $F_1 = X_1$ (for example).

3.2 Assessing Goodness of Fit

Let $E_t = F_t - X_t$, where F_t is based on $\{X_k\}_{k=1}^{t-1}$. We define

Mean absolute error (MAE)

$$\frac{1}{n}\sum_{k=1}^{n} |E_k|.$$

Mean square error (MSE)

$$\frac{1}{n}\sum_{k=1}^{n} E_k^2.$$

Mean absolute percentage error (MAPE)

$$\frac{1}{n}\sum_{k=1}^{n} \left|\frac{E_k}{X_k}\right|.$$

3.3 Holt-Winters forecasts

Let L_t be the estimated level, B_t the estimated slope and F_{t+m} the forecast of X_{t+m}, based on $\{X_k\}_{k=1}^{t}$. For some $\alpha, \beta \in [0,1]$ we have

$$\begin{aligned}
L_t &= \alpha X_t + (1-\alpha)(L_{t-1} + B_{t-1}) \\
B_t &= \beta(L_t - L_{t-1}) + (1-\beta)B_{t-1} \\
F_{t+m} &= L_t + m \cdot B_t
\end{aligned}$$

Initialise with $L_1 = X_1$, $B_1 = X_2 - X_1$ (for example).

If there is an *additive* periodic component of known period s then we include S_t, an estimate of the periodic effect at time t. For some $\alpha, \beta, \gamma \in [0,1]$ we have

$$\begin{aligned}
L_t &= \alpha(X_t - S_{t-s}) + (1-\alpha)(L_{t-1} + B_{t-1}) \\
B_t &= \beta(L_t - L_{t-1}) + (1-\beta)B_{t-1} \\
S_t &= \gamma(X_t - L_t) + (1-\gamma)S_{t-s} \\
F_{t+m} &= L_t + m \cdot B_t + S_{t+m-s}
\end{aligned}$$

Initialise with (for example)

$$\begin{aligned}
L_s &= \frac{1}{s}(X_1 + \cdots + X_s) \\
B_s &= \frac{1}{s}\left(\frac{X_{s+1} - X_1}{s} + \frac{X_{s+2} - X_2}{s} + \cdots + \frac{X_{s+s} - X_s}{s}\right) \\
S_1 &= X_1 - L_s \\
S_2 &= X_2 - L_s \\
&\vdots \\
S_s &= X_s - L_s
\end{aligned}$$