6 Estimation of trend and seasonality

6.1 Moving averages

Consider a decomposition into trend, seasonal, cyclic and residual components.

$$X_t = T_t + I_t + C_t + E_t\,.$$

Thus far we have been concerned with modelling $\{E_t\}$. We have also seen that the periodogram can be useful for recognising the presence of $\{C_t\}$.

We can estimate trend using a **symmetric moving average**,

$$\hat{T}_t = \sum_{S=-k}^{k} a_s X_{t+s}\,,$$

where $a_s = a_{-s}$. In this case the transfer function is real-valued.

The choice of moving averages requires care. For example, we might try to estimate the trend with

$$\hat{T}_t = \tfrac{1}{3}\left(X_{t-1} + X_t + X_{t+1}\right)\,.$$

But suppose $X_t = T_t + \epsilon_t$, where trend is the quadratic $T_t = a + bt + ct^2$. Then

$$\hat{T}_t = T_t + \tfrac{2}{3}c + \tfrac{1}{3}(\epsilon_{t-1} + \epsilon_t + \epsilon_{t+1})\,,$$

so $\mathbb{E}\hat{T}_t = \mathbb{E}X_t + \frac{2}{3}c$ and thus $\hat{T}$ is a biased estimator of the trend.

This problem is avoided if we estimate trend by fitting a polynomial of sufficient degree, e.g., to find a cubic that best fits seven successive points we minimize

$$\sum_{t=-3}^{3}\left(X_t - b_0 - b_1 t - b_2 t^2 - b_3 t^3\right)^2\,.$$

So

$$\begin{array}{rcl}
\sum X_t &=& 7\hat{b}_0 + 28\hat{b}_2 \\
\sum tX_t &=& 28\hat{b}_1 + 196\hat{b}_3 \\
\sum t^2X_t &=& 28\hat{b}_0 + 196\hat{b}_2 \\
\sum t^3X_t &=& 196\hat{b}_1 + 1588\hat{b}_3
\end{array}$$

Then

$$\begin{aligned}
\hat{b}_0 &= \tfrac{1}{21}\left(7\textstyle\sum X_t - \sum t^2 X_t\right) \\
&= \tfrac{1}{21}\left(-2X_{-3} + 3X_{-2} + 6X_{-1} + 7X_0 + 6X_1 + 3X_2 - 2X_3\right)\,.
\end{aligned}$$

We estimate the trend at time 0 by $\hat{T}_0 = \hat{b}_0$, and similarly,

$$\hat{T}_t = \tfrac{1}{21}\left(-2X_{t-3} + 3X_{t-2} + 6X_{t-1} + 7X_t + 6X_{t+1} + 3X_{t+2} - 2X_{t+3}\right)\,.$$

A notation for this moving average is $\frac{1}{21}[-2,3,6,7,6,3,-2]$. Note that the weights sum to 1. In general, we can fit a polynomial of degree q to $2m+1$ points by applying a symmetric moving average. (We fit to an odd number of points so that the midpoint of fitted range coincides with a point in time at which data is measured.)

A value for q can be identified using the **variate difference method**: if $\{X_t\}$ is indeed a polynomial of degree q, plus residual error $\{\epsilon_t\}$, then the trend in $\Delta^r X_t$ is a polynomial of degree $q-r$ and

$$\Delta^q X_t = \text{constant} + \Delta^q \epsilon_t = \text{constant} + \epsilon_t - \binom{q}{1}\epsilon_{t-1} + \binom{q}{2}\epsilon_{t-2} - \cdots + (-1)^q \epsilon_{t-q}\,.$$

The variance of $\Delta^q X_t$ is therefore

$$\operatorname{var}(\Delta^q \epsilon_t) = \left[1 + \binom{q}{1}^2 + \binom{q}{2}^2 + \cdots + 1\right]\sigma^2 = \binom{2q}{q}\sigma^2\,,$$

where the simplification in the final line comes from looking at the coefficient of z^q in expansions of both sides of

$$(1+z)^q(1+z)^q = (1+z)^{2q}\,.$$

Define $V_r = \operatorname{var}(\Delta^r X_t)/\binom{2r}{r}$. The fact that the plot of V_r against r should flatten out at $r \geq q$ can be used to identify q.

6.2 Centred moving averages

If there is a seasonal component then a **centred-moving average** is useful. Suppose data is measured quarterly, then applying twice the moving average $\frac{1}{4}[1,1,1,1]$ is equivalent to applying once the moving average $\frac{1}{8}[1,2,2,2,1]$. Notice that this so-called **centred average of fours** weights each quarter equally. Thus if $X_t = I_t + \epsilon_t$, where I_t has period 4, and $I_1 + I_2 + I_3 + I_4 = 0$, then $\hat{T}_t$ has no seasonal component. Similarly, if data were monthly we use a centred average of 12s, that is, $\frac{1}{24}[1,2,2,2,2,2,2,2,2,2,2,2,1]$.

6.3 The Slutzky-Yule effect

To remove both trend and seasonal components we might successively apply a number of moving averages, one or more to remove trend and another to remove seasonal effects. This is the procedure followed by some standard forecasting packages.

However, there is a danger that application of successive moving averages can introduce spurious effects. The **Slutzky-Yule effect** is concerned with the fact that a moving average repeatedly applied to a purely random series can introduce artificial cycles. Slutzky (1927) showed that some trade cycles of the nineteenth century were no more than artifacts of moving averages that had been used to smooth the data.

To illustrate this idea, suppose the moving average $\frac{1}{6}[-1, 2, 4, 2, -1]$ is applied k times to a white noise series. This moving average has transfer function, $a(\omega) = \frac{1}{6}(4 + 4\cos\omega - 2\cos 2\omega)$, which is maximal at $\omega = \pi/3$. The smoothed series has a spectral density, say $f_k(\omega)$, proportional to $a(\omega)^{2k}$, and hence for $\omega \neq \pi/3$, $f_k(\omega)/f_k(\pi/3) \to 0$ as $k \to \infty$. Thus in the limit the smoothed series is a periodic wave with period 6.

6.4 Exponential smoothing

Single exponential smoothing

Suppose the mean level of a series drifts slowly over time. A naive one-step-ahead forecast is $X_t(1) = X_t$. However, we might let all past observations play a part in the forecast, but give greater weights to those that are more recent. Choose weights to decrease exponentially and let

$$X_t(1) = \frac{1-\omega}{1-\omega^t}\left(X_t + \omega X_{t-1} + \omega^2 X_{t-2} + \cdots + \omega^{t-1} X_1\right),$$

where $0 < \omega < 1$. Define S_t as the right hand side of the above as $t \to \infty$, i.e.,

$$S_t = (1-\omega)\sum_{s=0}^{\infty} \omega^s X_{t-s}.$$

S_t can serve as a one-step-ahead forecast, $X_t(1)$. S_t is known as **simple exponential smoothing**. Let $\alpha = 1 - \omega$. Simple algebra gives

$$\begin{aligned} S_t &= \alpha X_t + (1-\alpha) S_{t-1} \\ X_t(1) &= X_{t-1}(1) + \alpha[X_t - X_{t-1}(1)]. \end{aligned}$$

This shows that the one-step-ahead forecast at time t is the one-step-ahead forecast at time $t-1$, modified by α times the forecasting error incurred at time $t-1$.

To get things started we might set S_0 equal to the average of the first few data points. We can play around with α, choosing it to minimize the mean square forecasting error. In practice, α in the range 0.25–0.5 usually works well.

Double exponential smoothing

Suppose the series is approximately linear, but with a slowly varying trend. If it were true that $X_t = b_0 + b_1 t + \epsilon_t$, then

$$\begin{aligned} S_t &= (1-\omega)\sum_{s=0}^{\infty} \omega^s \left(b_0 + b_1(t-s) + \epsilon_t\right) \\ &= b_0 + b_1 t - b_1(1-\omega)\sum_{s=0}^{\infty}\omega^s s + b_1(1-\omega)\sum_{s=0}^{\infty}\omega^s \epsilon_{t-s}, \end{aligned}$$

and hence

$$\mathbb{E}S_t = b_0 + b_1 t - b_1\omega/(1-\omega) = \mathbb{E}X_{t+1} - b_1/(1-\omega).$$

Thus the forecast has a bias of $-b_1/(1-\omega)$. To eliminate this bias let $S_t^1 = S_t$ be the first smoothing, and $S_t^2 = \alpha S_t^1 + (1-\alpha)S_{t-1}^2$ be the simple exponential smoothing of S_t^1. Then

$$\begin{aligned} \mathbb{E}S_t^2 &= \mathbb{E}S_t^1 - b_1\omega/(1-\omega) = \mathbb{E}X_t - 2b_1\omega/(1-\omega), \\ \mathbb{E}(2S_t^1 - S_t^2) &= b_0 + b_1 t, \qquad \mathbb{E}(S_t^1 - S_t^2) = b_1(1-\alpha)/\alpha. \end{aligned}$$

This suggests the estimates $\hat{b}_0 + \hat{b}_1 t = 2S_t^1 - S_t^2$ and $\hat{b}_1 = \alpha(S_t^1 - S_t^2)/(1-\alpha)$. The forecasting equation is then

$$X_t(s) = \hat{b}_0 + \hat{b}_1(t+s) = (2S_t^1 - S_t^2) + s\alpha(S_t^1 - S_t^2)/(1-\alpha).$$

As with single exponential smoothing we can experiment with choices of α and find S_0^1 and S_0^2 by fitting a regression line, $X_t = \hat{\beta}_0 + \hat{\beta}_1 t$, to the first few points of the series and solving

$$S_0^1 = \hat{\beta}_0 - (1-\alpha)\hat{\beta}_1/\alpha, \qquad S_0^2 = \hat{\beta}_0 - 2(1-\alpha)\hat{\beta}_1/\alpha.$$

6.5 Calculation of seasonal indices

Suppose data is quarterly and we want to fit an additive model. Let $\hat{I}_1$ be the average of $X_1, X_5, X_9, \ldots$, let $\hat{I}_2$ be the average of $X_2, X_6, X_{10}, \ldots$, and so on for $\hat{I}_3$ and $\hat{I}_4$. The cumulative seasonal effects over the course of year should cancel, so that if $X_t = a + I_t$, then $X_t + X_{t+1} + X_{t+2} + X_{t+3} = 4a$. To ensure this we take our final estimates of the seasonal indices as $I_t^* = \hat{I}_t - \frac{1}{4}(\hat{I}_1 + \cdots + \hat{I}_4)$.

If the model is multiplicative and $X_t = aI_t$, we again wish to see the cumulative effects over a year cancel, so that $X_t + X_{t+1} + X_{t+2} + X_{t+3} = 4a$. This means that we should take $I_t^* = \hat{I}_t - \frac{1}{4}(\hat{I}_1 + \cdots + \hat{I}_4) + 1$, adjusting so the mean of $I_1^*, I_2^*, I_3^*, I_4^*$ is 1.

When both trend and seasonality are to be extracted a two-stage procedure is recommended:

(a) Make a first estimate of trend, say $\hat{T}_t^1$.

Subtract this from $\{X_t\}$ and calculate first estimates of the seasonal indices, say I_t^1, from $X_t - \hat{T}_t^1$.

The first estimate of the deseasonalized series is $Y_t^1 = X_t - I_t^1$.

(b) Make a second estimate of the trend by smoothing Y_t^1, say $\hat{T}_t^2$.

Subtract this from $\{X_t\}$ and calculate second estimates of the seasonal indices, say I_t^2, from $X_t - \hat{T}_t^2$.

The second estimate of the deseasonalized series is $Y_t^2 = X_t - I_t^2$.