# Week 3: Adaptive smoothing (cont.)

# Adaptive approaches: moving averages (linear filters)

$$\widehat{Tr}_t = \sum_{i=-m}^{m} w_i y_{t-i},$$

where $\sum_{i=-m}^{m} w_i = 1$ are weights

What did you read about

- derivation of $w_i$ under local polynomial trend of order $r$

# Adaptive approaches: moving averages (linear filters)

$$\widehat{Tr}_t = \sum_{i=-m}^{m} w_i y_{t-i},$$

where $\sum_{i=-m}^{m} w_i = 1$ are weights

What did you read about

- derivation of $w_i$ under local polynomial trend of order $r$

This week:

- criterion for selection of $r$
- how to compute the edges
- other linear filters:
  - simple $w_i = \frac{1}{2m+1}$
  - filter for seasonal data $\rightsquigarrow$ centered moving average
  - robust approach: take the median instead of the mean

See some R examples.

# Exponential smoothing

Assumption

$$Y_t = Tr_t + E_t = L_t + E_t$$

where $L_t$ is a level **locally constant**

# Exponential smoothing

Assumption

$$Y_t = Tr_t + E_t = L_t + E_t$$

where $L_t$ is a level **locally constant**

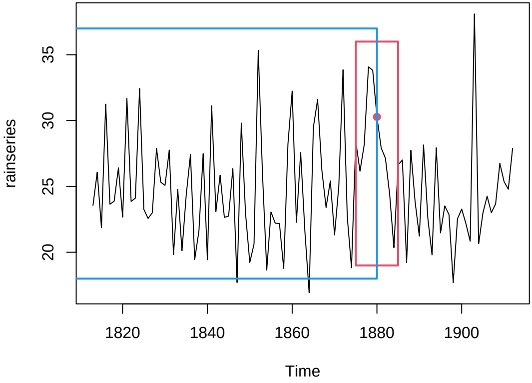
### Linear filters

- Aim: series decomposition (trend elimination)
- $\widehat{Tr}_t$ computed from $Y_{t-m}, \ldots, Y_{t+m}$

### Exponential smoothing

- Aim: **forecast** of future values
- $\widehat{L}_t$ computed only from the past $Y_{t-i}$

# Exponential smoothing



# Exponential smoothing: motivation

Series $Y_1, \ldots, Y_n$, locally constant

- Naive forecasts of future $Y_{n+h}$

  $$\widehat{Y}_{n+h}(n) = \frac{1}{n} \sum_{t=1}^{n} Y_t$$

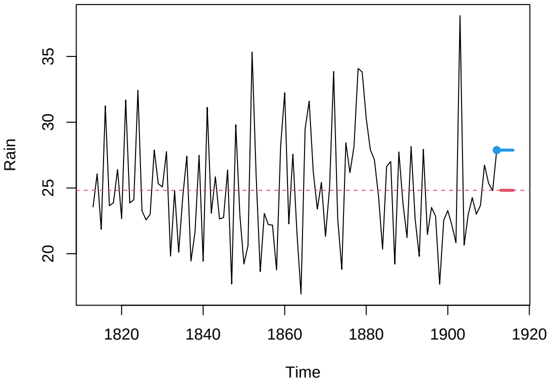
  gives the same weight to all observations. Or

  $$\widehat{Y}_{n+h}(n) = Y_n$$

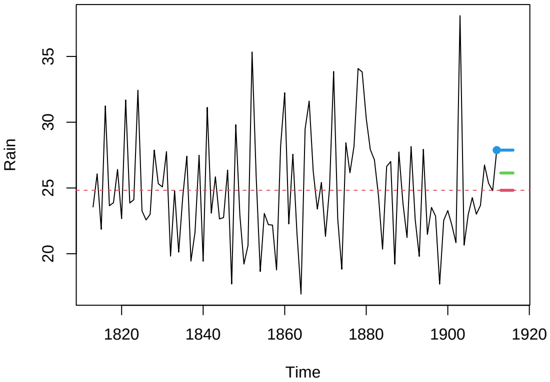
  gives weight 1 to the last observation
- we typically want something in between

  $$\widehat{Y}_{n+h}(n) = \widehat{L}_n,$$

  where $\widehat{L}$ gives more weight to recent observations, but at the same time takes into account all observations



Rain
20
25
30
35
1820
1840
1860
1880
1900
1920
Time



Rain
35
30
25
20
1820
1840
1860
1880
1900
1920
Time

# Idea of exponential smoothing

- Compute

$$\widehat{L}_t = \widehat{Y}_{t+1}(t) = \sum_{i=0}^{\infty} w_i Y_{t-i}$$

as a weighted average of $Y_t, Y_{t-1}, \dots$ with geometrically decaying weights $w_i = const \cdot \beta^i \quad 0 < \beta < 1$.

# Idea of exponential smoothing

- Compute

$$\widehat{L}_t = \widehat{Y}_{t+1}(t) = \sum_{i=0}^{\infty} w_i Y_{t-i}$$

as a weighted average of $Y_t, Y_{t-1}, \ldots$ with geometrically decaying weights $w_i = const \cdot \beta^i \quad 0 < \beta < 1$.

- From $\sum_{i=0}^{\infty} w_i = 1$, we get

$$w_i = (1-\beta)\beta^i = \alpha(1-\alpha)^i$$

for $\alpha = 1 - \beta$ a smoothing constant

# Idea of exponential smoothing

- Compute

$$\widehat{L}_t = \widehat{Y}_{t+1}(t) = \sum_{i=0}^{\infty} w_i Y_{t-i}$$

as a weighted average of $Y_t, Y_{t-1}, \ldots$ with geometrically decaying weights $w_i = const \cdot \beta^i \quad 0 < \beta < 1$.

- From $\sum_{i=0}^{\infty} w_i = 1$, we get

$$w_i = (1-\beta)\beta^i = \alpha(1-\alpha)^i$$

for $\alpha = 1 - \beta$ a smoothing constant

- Then

$$\begin{aligned}
\widehat{L}_t &= \alpha Y_t + \alpha(1-\alpha) Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \ldots \\
&= \alpha Y_t + (1-\alpha) \underbrace{[\alpha Y_{t-1} + \alpha(1-\alpha) Y_{t-2} + \ldots]}_{\widehat{L}_{t-1}} \\
&= \alpha Y_t + (1-\alpha)\widehat{L}_{t-1}
\end{aligned}$$

a recursive formula

# Another point of view

Let say we want to find $\theta$ such that

$$\min_{\theta} \sum_{j=0}^{\infty} (Y_{t-j} - \theta)^2 w_j.$$

The solution is

$$\widehat{\theta} = \sum_{j=0}^{\infty} w_j Y_{t-j},$$

so for $w_j = \alpha(1-\alpha)^j$, we get $\widehat{L}_t$

# Another point of view

Let say we want to find $\theta$ such that

$$\min_{\theta} \sum_{j=0}^{\infty} (Y_{t-j} - \theta)^2 w_j.$$

The solution is

$$\widehat{\theta} = \sum_{j=0}^{\infty} w_j Y_{t-j},$$

so for $w_j = \alpha(1-\alpha)^j$, we get $\widehat{L}_t$

Also see

$$\widehat{Y}_{t+1}(t) = \widehat{L}_t = \widehat{L}_{t-1} + \alpha(Y_t - \widehat{L}_{t-1}) = \widehat{L}_{t-1} + \alpha \underbrace{(Y_t - \widehat{Y}_t(t-1))}_{e_t}$$

$e_t$ one-step-ahead error

# Practical issues

Recall $\widehat{L}_t = \widehat{Y}_{t+1}(t)$ and $\widehat{L}_t = \alpha Y_t + (1-\alpha)\widehat{L}_{t-1} \rightsquigarrow$

$$\begin{aligned}
\widehat{L}_1 &= \alpha Y_1 + (1-\alpha)\widehat{L}_0, \\
\widehat{L}_2 &= \alpha Y_2 + (1-\alpha)\widehat{L}_1, \\
&\vdots \\
\widehat{L}_n &= \alpha Y_n + (1-\alpha)\widehat{L}_{n-1},
\end{aligned}$$

and $\widehat{L}_n$ will be used for predictions

We need: choose $\widehat{L}_0$, $\alpha$

See R examples

# Practical issues

Recall $\widehat{L}_t = \widehat{Y}_{t+1}(t)$ and $\widehat{L}_t = \alpha Y_t + (1-\alpha)\widehat{L}_{t-1} \rightsquigarrow$

$$\begin{aligned}
\widehat{L}_1 &= \alpha Y_1 + (1-\alpha)\widehat{L}_0, \\
\widehat{L}_2 &= \alpha Y_2 + (1-\alpha)\widehat{L}_1, \\
&\vdots \\
\widehat{L}_n &= \alpha Y_n + (1-\alpha)\widehat{L}_{n-1},
\end{aligned}$$

and $\widehat{L}_n$ will be used for predictions

We need: choose $\widehat{L}_0$, $\alpha$

- e.g. $\widehat{L}_0 = \frac{1}{K}\sum_{t=1}^{K} Y_t$ for small $K$, e.g. $K = 6$

See R examples

# Practical issues

Recall $\widehat{L}_t = \widehat{Y}_{t+1}(t)$ and $\widehat{L}_t = \alpha Y_t + (1-\alpha)\widehat{L}_{t-1} \rightsquigarrow$

$$\begin{aligned}
\widehat{L}_1 &= \alpha Y_1 + (1-\alpha)\widehat{L}_0, \\
\widehat{L}_2 &= \alpha Y_2 + (1-\alpha)\widehat{L}_1, \\
&\vdots \\
\widehat{L}_n &= \alpha Y_n + (1-\alpha)\widehat{L}_{n-1},
\end{aligned}$$

and $\widehat{L}_n$ will be used for predictions

We need: choose $\widehat{L}_0$, $\alpha$

- e.g. $\widehat{L}_0 = \frac{1}{K}\sum_{t=1}^{K} Y_t$ for small $K$, e.g. $K = 6$
- $\alpha$ chosen to minimize sum of squared errors

$$SSE = \sum_{t=2}^{n} e_t^2 = \sum_{t=2}^{n}(Y_t - \widehat{Y}_t(t-1))^2 = \sum_{t=2}^{n}(Y_t - \widehat{L}_{t-1})^2$$

See R examples

# Prediction intervals

Recall that $e_t = Y_t - \widehat{L}_{t-1}$ and $\widehat{L}_t = \widehat{L}_{t-1} + \alpha e_t$. Assume that

$$Y_t = L_{t-1} + \varepsilon_t,$$
$$L_t = L_{t-1} + \alpha \varepsilon_t.$$

So called state space model.

Prediction intervals

- can be constructed under **normality** and **independence** assumptions for $\varepsilon_t$
- modern approach: use bootstrap

# Double exponential smoothing

Locally linear trend

$$Tr_{t+j} = L_t + B_t \cdot j$$

$L_t$ level, $B_t$ slope

Then

$$\widehat{Y}_{t+h}(t) = \widehat{L}_t + \widehat{B}_t h$$

# Double exponential smoothing

Locally linear trend

$$Tr_{t+j} = L_t + B_t \cdot j$$

$L_t$ level, $B_t$ slope

Then

$$\widehat{Y}_{t+h}(t) = \widehat{L}_t + \widehat{B}_t h$$

Basic idea:

$$\min_{\beta_{0t},\beta_{1t}} \sum_{j=0}^{\infty} [Y_{t-j} - \beta_{0t} - \beta_{1t}(-j)]^2 \cdot (1-\alpha)^j$$

$\rightsquigarrow$ recursive formulas (see the book)

More general approach: Holt's linear trend method (next week)