Time Series Analysis

Henrik Madsen

hm@imm.dtu.dk

Informatics and Mathematical Modelling
Technical University of Denmark
DK-2800 Kgs. Lyngby

Outline of the lecture

Regression based methods, 2nd part:

- Regression and exponential smoothing (Sec. 3.4)
- Time series with seasonal variations (Sec. 3.5)

Regression without explanatory variables

- During Lecture 2 we saw that assuming known independent variables $\boldsymbol{x}$ we can forecast the dependent variable Y
- To be able to do so we estimated $\boldsymbol{\theta}$ in

$$Y_t = f(\boldsymbol{x}_t, t; \boldsymbol{\theta}) + \varepsilon_t$$

- If we do not have access to $\boldsymbol{x}$ we may use:

$$Y_t = f(t; \boldsymbol{\theta}) + \varepsilon_t$$

- During this lecture we shall consider models of this (last) form and we shall consider how $\widehat{\boldsymbol{\theta}}$ can be updated as more information becomes available
- Only models linear in $\boldsymbol{\theta}$ will be considered

Model: Constant mean

- $Y_t = \mu + \varepsilon_t$, ε_t i.i.d. with mean zero and constant variance σ^2 (white noise).
- In vector form ($t = 1, \ldots, N$): $\boldsymbol{Y} = \mathbf{1}\mu + \boldsymbol{\varepsilon}$
- Estimate: $\hat{\mu} = (\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T\boldsymbol{Y} = N^{-1}\sum_{t=1}^{N} Y_t = \bar{y}$.
- Prediction (the conditional mean): $\widehat{Y}_{N+\ell|N} = \widehat{\mu} = \frac{1}{N}\sum_{t=1}^{N} Y_t$
- Variance of the prediction error:
  $V\left[Y_{N+\ell} - \widehat{Y}_{N+\ell|N}\right] = \sigma^2(1 + \frac{1}{N})$

Updating the estimate

- Based on $Y_1, Y_2, \ldots, Y_N$ we have $\hat{\mu}_N = \frac{1}{N} \sum_{t=1}^{N} Y_t$
- When we get one more observation Y_{N+1} the best estimate is $\hat{\mu}_{N+1} = \frac{1}{N+1} \sum_{t=1}^{N+1} Y_t$
- Recursive update:

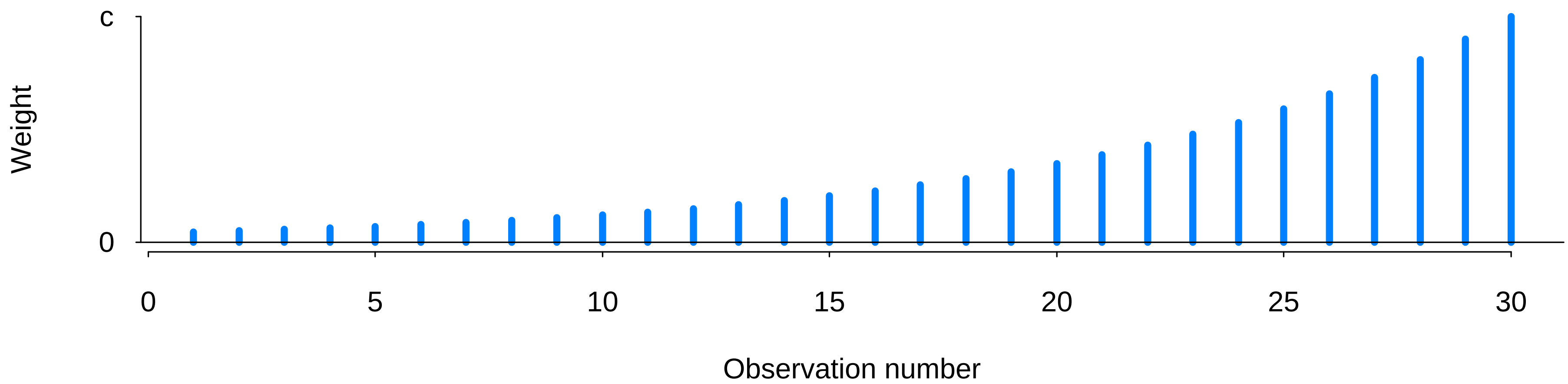
$$\hat{\mu}_{N+1} = \frac{1}{N+1} \sum_{t=1}^{N+1} Y_t = \frac{1}{N+1} Y_{N+1} + \frac{N}{N+1} \hat{\mu}_N$$

Model: Local constant mean

- In the constant mean model the variance of the forecast error decrease towards σ^2 as $1/N$
- Therefore, if N is sufficiently high (say 100) there is not much gained by increasing the number of observations
- If there is indications that the true (underlying) mean is actually changing slowly it can even be advantageous to “forget” old observations.
- One way of doing this is to base the estimate on a rolling window containing e.g. the 100 most recent observations
- An alternative is *exponential smoothening*

Exponential smoothening

$$\hat{\mu}_N = c \sum_{j=0}^{N-1} \lambda^j Y_{N-j} = c[Y_N + \lambda Y_{N-1} + \cdots + \lambda^{N-1} Y_1]$$



The constant c is chosen so that the weights sum to one, which implies that $c = (1-\lambda)/(1-\lambda^N)$. For large N:

$$\hat{\mu}_{N+1} = (1-\lambda) Y_{N+1} + \lambda \hat{\mu}_N \text{ or } \widehat{Y}_{N+\ell+1|N+1} = (1-\lambda) Y_{N+1} + \lambda \widehat{Y}_{N+\ell|N}$$

Choice of smoothing constant $\alpha = 1 - \lambda$

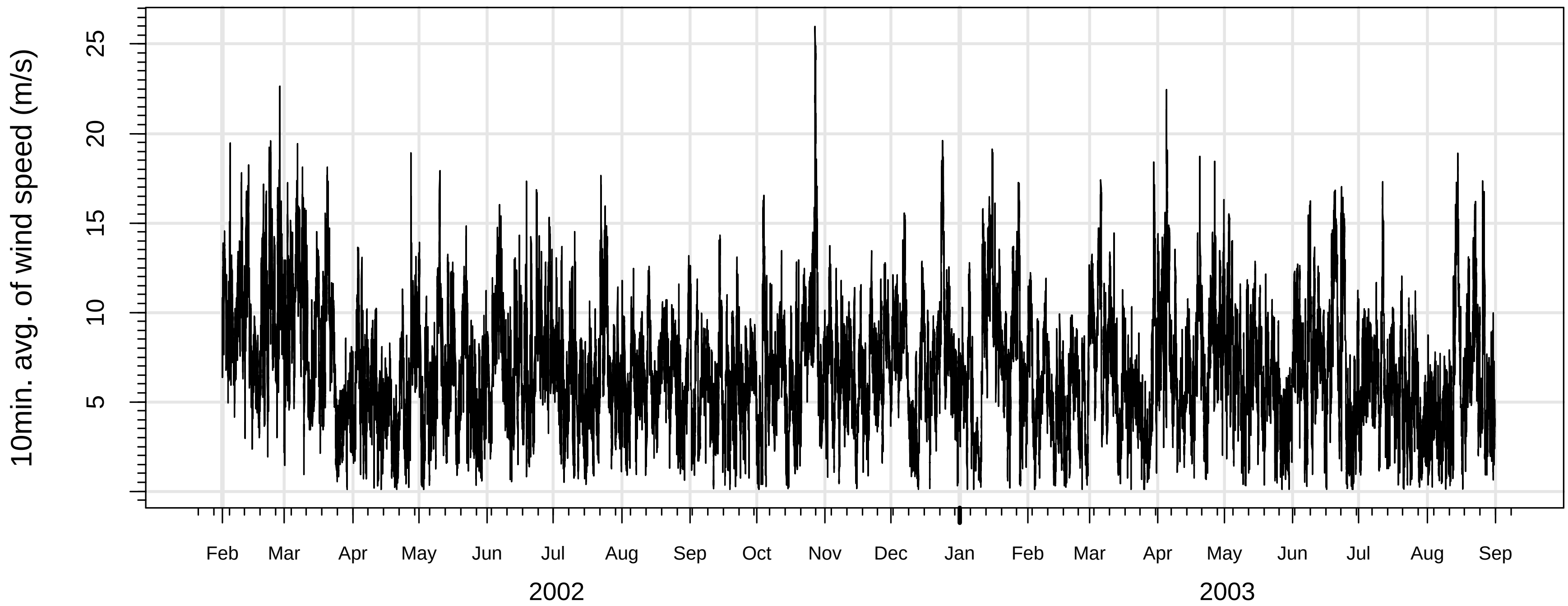
- The smoothing constant $\alpha = 1 - \lambda$ determines how much the latest observation influence the prediction
- Given a data set $t = 1, \ldots, N$ we can try different values before implementing the method on-line

$$S(\alpha) = \sum_{t=1}^{N} (Y_t - \widehat{Y}_{t|t-1}(\alpha))^2$$

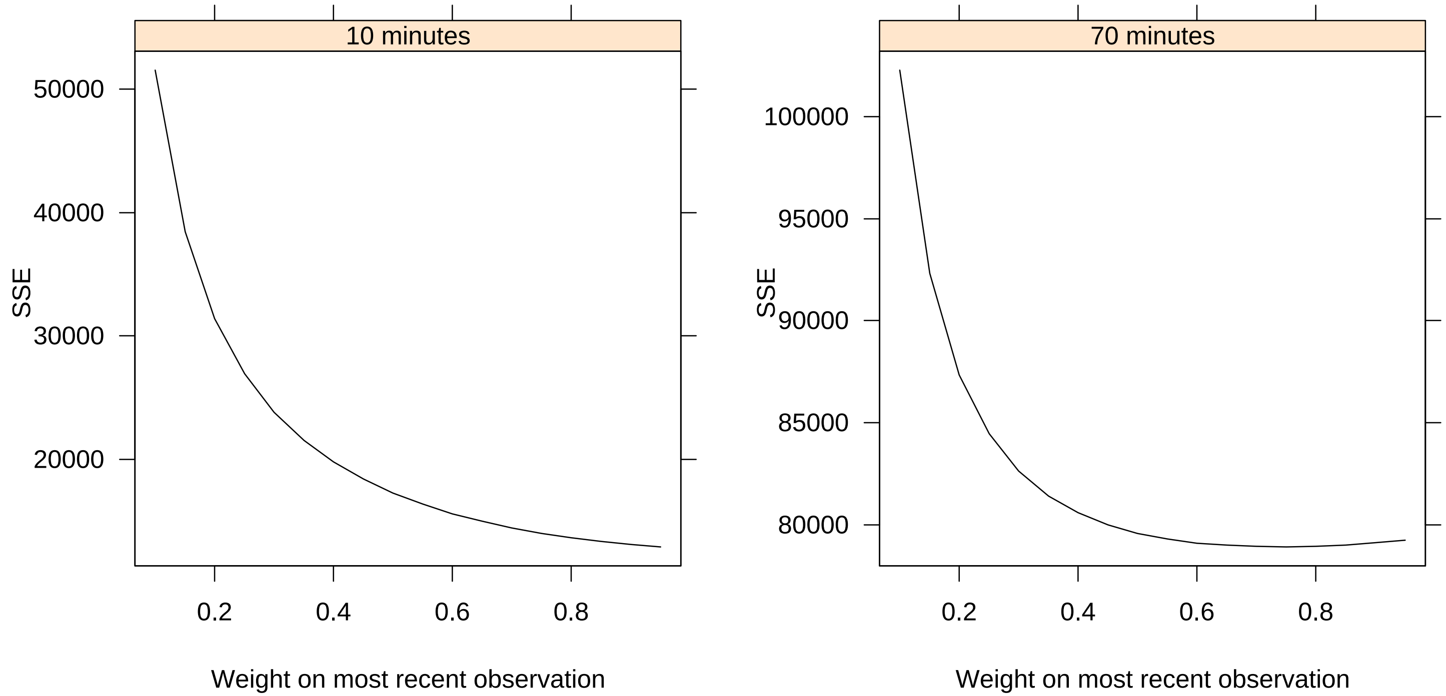
- If the data set is large we eliminate the influence of the initial estimate by dropping the first part of the errors when evaluating $S(\alpha)$

Example – wind speed 76 m a.g.l. at Risø

- Measurements of wind speed every 10th minute
- Task: Forecast up to approximately 3 hours ahead using exponential smoothing

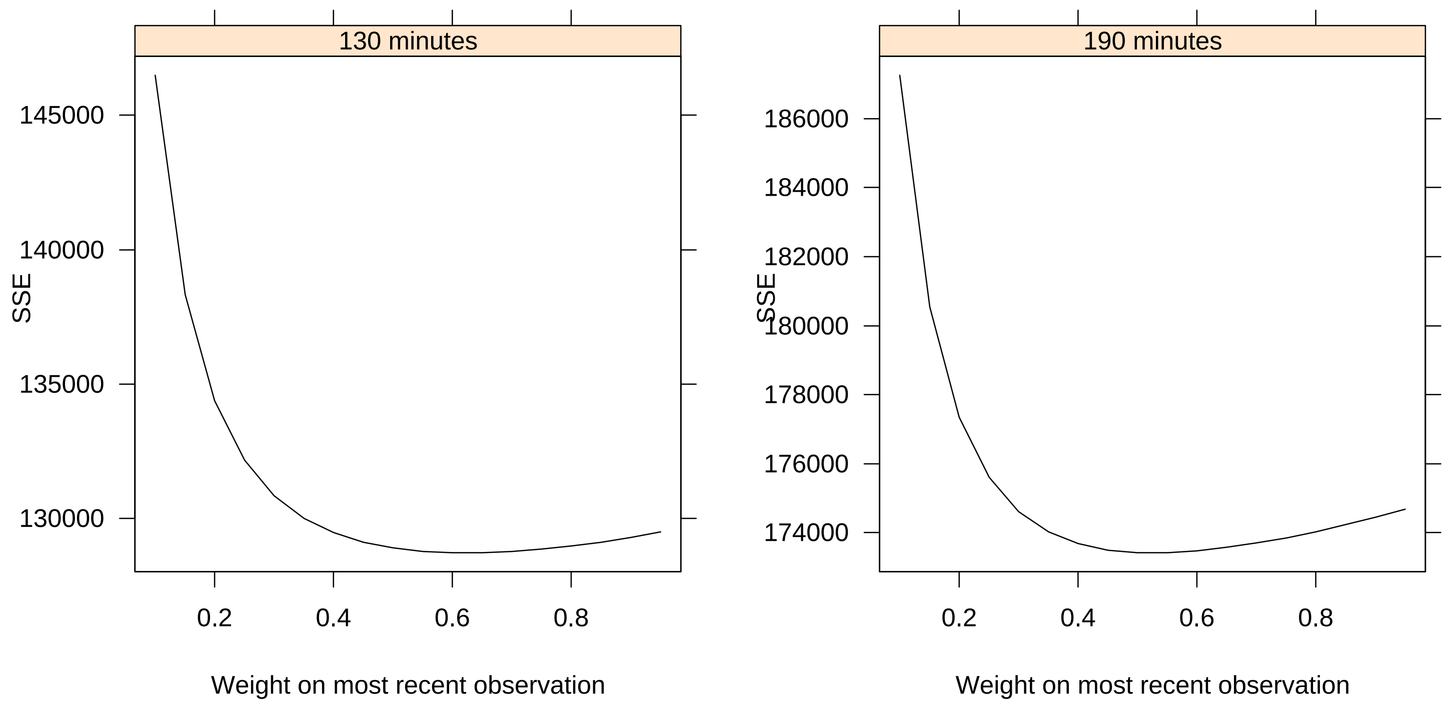


$S(\alpha)$ for horizons 10 and 70 minutes



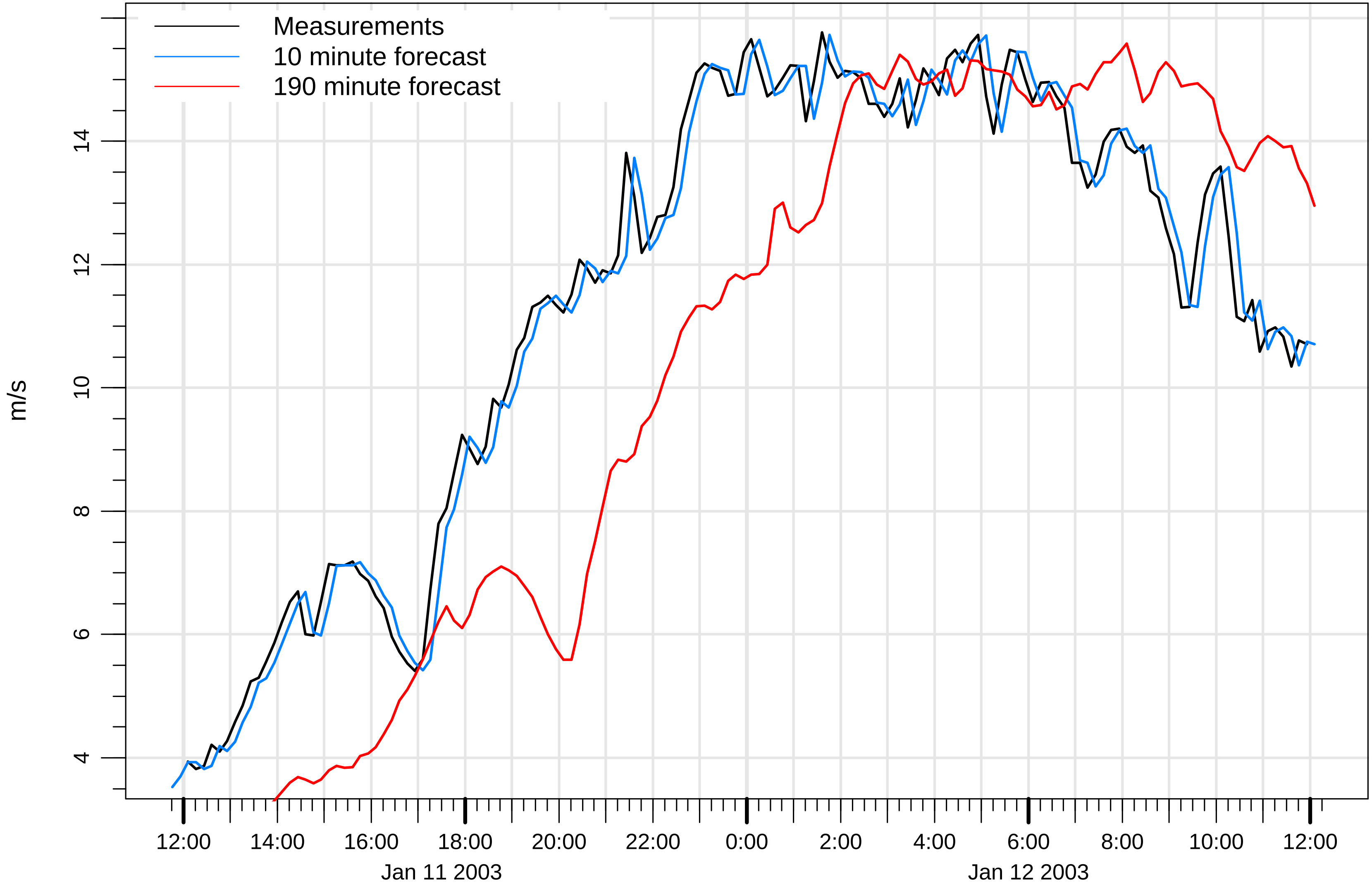
- 10 minutes (1-step): Use $\alpha = 0.95$ or higher
- 70 minutes (7-step): Use $\alpha \approx 0.7$

$S(\alpha)$ for horizons 130 and 190 minutes



- 130 minutes (13-step): Use $\alpha \approx 0.6$
- 190 minutes (19-step): Use $\alpha \approx 0.5$

Example of forecasts with optimal α



Trend models

- Linear regression model
- Functions of time are taken as the independent variables

Linear trend

- Observations for $t = 1, \ldots, N$
- Naive formulation of the model: $Y_t = \phi_0 + \phi_1 \, t + \varepsilon_t$
- If we want to forecast Y_{N+j} given information up to N we use $\widehat{Y}_{N+j|N} = \hat{\phi}_0 + \hat{\phi}_1 \, (N + j)$
- However, for on-line applications $N + j$ can be arbitrary large
- The problem arise because ϕ_0 and ϕ_1 is defined w.r.t. the origin 0
- Defining the parameters w.r.t. the origin n we obtain the model: $Y_t = \theta_0 + \theta_1 \, (t - N) + \varepsilon_t$
- Using this formulation we get: $\widehat{Y}_{N+j|N} = \hat{\theta}_0 + \hat{\theta}_1 \, j$

Linear trend in a general setting

- The general trend model:

$$Y_{N+j} = \boldsymbol{f}^T(j)\boldsymbol{\theta} + \varepsilon_{N+j}$$

- The linear trend model is obtained when: $\boldsymbol{f}(j) = \begin{pmatrix} 1 \\ j \end{pmatrix}$
- It follows that for $N + 1 + j$:

$$Y_{N+1+j} = \begin{pmatrix} 1 \\ j+1 \end{pmatrix}^T \boldsymbol{\theta} + \varepsilon_{N+1+j} = \left(\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ j \end{pmatrix} \right)^T \boldsymbol{\theta} + \varepsilon_{N+1+j}$$

- The 2×2 matrix $\boldsymbol{L}$ defines the transition from $\boldsymbol{f}(j)$ to $\boldsymbol{f}(j+1)$

Trend models in general

- Model: $Y_{N+j} = \boldsymbol{f}^T(j)\boldsymbol{\theta} + \varepsilon_{N+j}$
- Requirement: $\boldsymbol{f}(j+1) = \boldsymbol{L}\boldsymbol{f}(j)$
- Initial value: $\boldsymbol{f}(0)$
- In Section 3.4 some trend models which fulfill the requirement above are listed.
  - Constant mean: $Y_{N+j} = \theta_0 + \varepsilon_{N+j}$
  - Linear trend: $Y_{N+j} = \theta_0 + \theta_1 j + \varepsilon_{N+j}$
  - Quadratic trend: $Y_{N+j} = \theta_0 + \theta_1 j + \theta_2 \frac{j^2}{2} + \varepsilon_{n+j}$
  - k'th order polynomial trend:
    $Y_{n+j} = \theta_0 + \theta_1 j + \theta_2 \frac{j^2}{2} + \cdots + \theta_k \frac{j^k}{k!} + \varepsilon_{N+j}$
  - Harmonic model with the period p:
    $Y_{N+j} = \theta_0 + \theta_1 \sin \frac{2\pi}{p} j + \theta_2 \cos \frac{2\pi}{p} j + \varepsilon_{N+j}$

Estimation

- Model equations written for all observations $Y_1, \ldots, Y_N$

$$\boldsymbol{Y} = \boldsymbol{x}_N \boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}^T(-N+1) \\ \boldsymbol{f}^T(-N+2) \\ \vdots \\ \boldsymbol{f}^T(0) \end{bmatrix} \boldsymbol{\theta} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

- OLS-estimates: $\widehat{\boldsymbol{\theta}}_N = (\boldsymbol{x}_N^T \boldsymbol{x}_N)^{-1} \boldsymbol{x}_N^T \boldsymbol{Y}$ or

$$\widehat{\boldsymbol{\theta}}_N = \boldsymbol{F}_N^{-1} \boldsymbol{h}_N \qquad \boldsymbol{F}_N = \sum_{j=0}^{N-1} \boldsymbol{f}(-j) \boldsymbol{f}^T(-j) \qquad \boldsymbol{h}_N = \sum_{j=0}^{N-1} \boldsymbol{f}(-j) Y_{N-j}$$

ℓ-step prediction

- Prediction:

$$\widehat{Y}_{N+\ell|N} = \boldsymbol{f}^T(\ell)\widehat{\boldsymbol{\theta}}_N$$

- Variance of the prediction error:

$$V[Y_{N+\ell} - \widehat{Y}_{N+\ell|N}] = \sigma^2 \left[1 + \boldsymbol{f}^T(\ell)\boldsymbol{F}_N^{-1}\boldsymbol{f}(\ell)\right]$$

- $100(1-\alpha)\%$ prediction interval:

$$\widehat{Y}_{N+\ell|N} \pm \mathrm{t}_{\alpha/2}(N-p)\sqrt{V[e_N(\ell)]} =$$
$$\widehat{Y}_{N+\ell|N} \pm \mathrm{t}_{\alpha/2}(N-p)\widehat{\sigma}\sqrt{1 + \boldsymbol{f}^T(\ell)\boldsymbol{F}_N^{-1}\boldsymbol{f}(\ell)}$$

where $\hat{\sigma}^2 = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}/(N-p)$ (p is the number of estimated parameters)

Updating the estimates when Y_{N+1} is available

- Task:
  - Going from estimates based on $t = 1, \ldots, N$, i.e. $\widehat{\boldsymbol{\theta}}_N$ to
  - estimates based on $t = 1, \ldots, N, N+1$, i.e. $\widehat{\boldsymbol{\theta}}_{N+1}$
  - without redoing everything...
- Solution:

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_{N+1} &= \boldsymbol{F}_{N+1}^{-1} \boldsymbol{h}_{N+1} \\
\boldsymbol{F}_{N+1} &= \boldsymbol{F}_N + \boldsymbol{f}(-N) \boldsymbol{f}^T(-N) \\
\boldsymbol{h}_{N+1} &= \boldsymbol{L}^{-1} \boldsymbol{h}_N + \boldsymbol{f}(0) Y_{N+1}
\end{aligned}$$

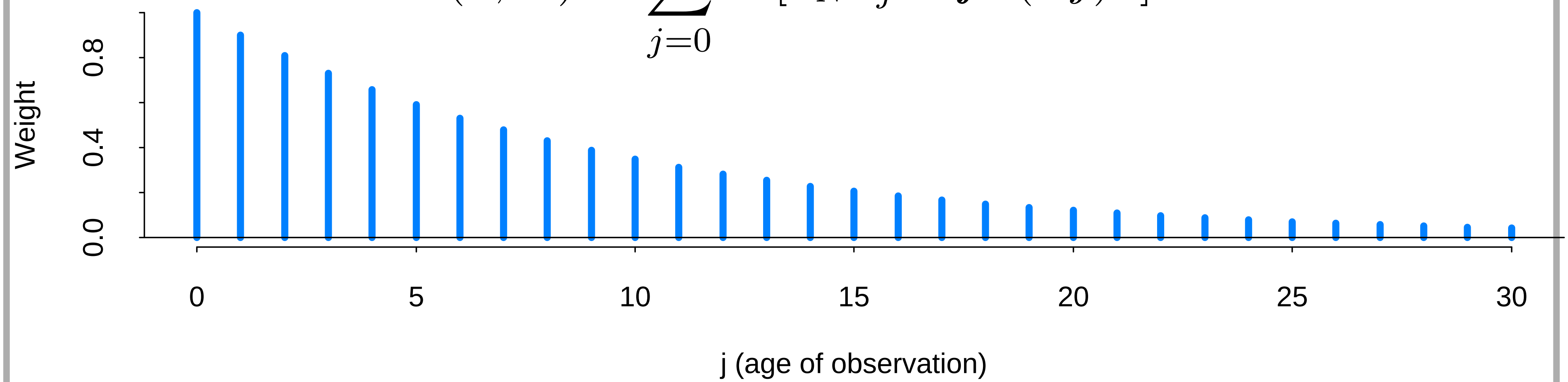
Local trend models

We forget old observations in an exponential manner:

$$\widehat{\boldsymbol{\theta}}_N = \arg\min_{\boldsymbol{\theta}} S(\boldsymbol{\theta}; N)$$

where for $0 < \lambda < 1$

$$S(\boldsymbol{\theta}; N) = \sum_{j=0}^{N-1} \lambda^j [Y_{N-j} - \boldsymbol{f}^T(-j)\boldsymbol{\theta}]^2$$



WLS formulation

The criterion:

$$S(\boldsymbol{\theta}; N) = \sum_{j=0}^{N-1} \lambda^j [Y_{N-j} - \boldsymbol{f}^T(-j)\boldsymbol{\theta}]^2$$

can be written as:

$$\begin{bmatrix} Y_1 - \boldsymbol{f}^T(N-1)\boldsymbol{\theta} \\ Y_2 - \boldsymbol{f}^T(N-2)\boldsymbol{\theta} \\ \vdots \\ Y_N - \boldsymbol{f}^T(0)\boldsymbol{\theta} \end{bmatrix}^T \begin{bmatrix} \lambda^{N-1} & 0 & \cdots & 0 \\ 0 & \lambda^{N-2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Y_1 - \boldsymbol{f}^T(N-1)\boldsymbol{\theta} \\ Y_2 - \boldsymbol{f}^T(N-2)\boldsymbol{\theta} \\ \vdots \\ Y_N - \boldsymbol{f}^T(0)\boldsymbol{\theta} \end{bmatrix}$$

which is a WLS criterion with $\boldsymbol{\Sigma} = \text{diag}[1/\lambda^{N-1}, \ldots, 1/\lambda, 1]$

WLS solution

$$\widehat{\boldsymbol{\theta}}_N = (\boldsymbol{x}_N^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_N)^{-1} \boldsymbol{x}_N^T \boldsymbol{\Sigma}^{-1} \boldsymbol{Y}$$

or

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_N &= \boldsymbol{F}_N^{-1} \boldsymbol{h}_N \\
\boldsymbol{F}_N &= \sum_{j=0}^{N-1} \lambda^j \boldsymbol{f}(-j) \boldsymbol{f}^T(-j) \\
\boldsymbol{h}_N &= \sum_{j=0}^{N-1} \lambda^j \boldsymbol{f}(-j) Y_{N-j}
\end{aligned}$$

Updating the estimates when Y_{N+1} is available

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{N+1} &= \boldsymbol{F}_{N+1}^{-1}\boldsymbol{h}_{N+1} \\ \boldsymbol{F}_{N+1} &= \boldsymbol{F}_N + \lambda^N \boldsymbol{f}(-N)\boldsymbol{f}^T(-N) \\ \boldsymbol{h}_{N+1} &= \lambda \boldsymbol{L}^{-1}\boldsymbol{h}_N + \boldsymbol{f}(0)Y_{N+1}\end{aligned}$$

When no data is available we can use $\boldsymbol{h}_0 = \mathbf{0}$ and $\boldsymbol{F}_0 = \mathbf{0}$

For many functions $\lambda^N \boldsymbol{f}(-N)\boldsymbol{f}^T(-N) \to 0$ for $N \to \infty$ and we get the stationary result $\boldsymbol{F}_{N+1} = \boldsymbol{F}_N = \boldsymbol{F}$. Hence:

$$\widehat{\boldsymbol{\theta}}_{N+1} = \boldsymbol{L}^T\widehat{\boldsymbol{\theta}}_N + \boldsymbol{F}^{-1}\boldsymbol{f}(0)[Y_{N+1} - \widehat{Y}_{N+1|N}]$$