Applied Time Series Analysis
SS 2013 – Week 03

Marcel Dettling

Institute for Data Analysis and Process Design

Zurich University of Applied Sciences

marcel.dettling@zhaw.ch

http://stat.ethz.ch/~dettling

ETH Zürich, March 4, 2013

Decomposition

Stationarity is key for statistical learning, but real data often have trend/seasonality, and are non-stationary. We can (often) deal with that using the simple additive decomposition model:

$$X_t = m_t + s_t + R_t$$

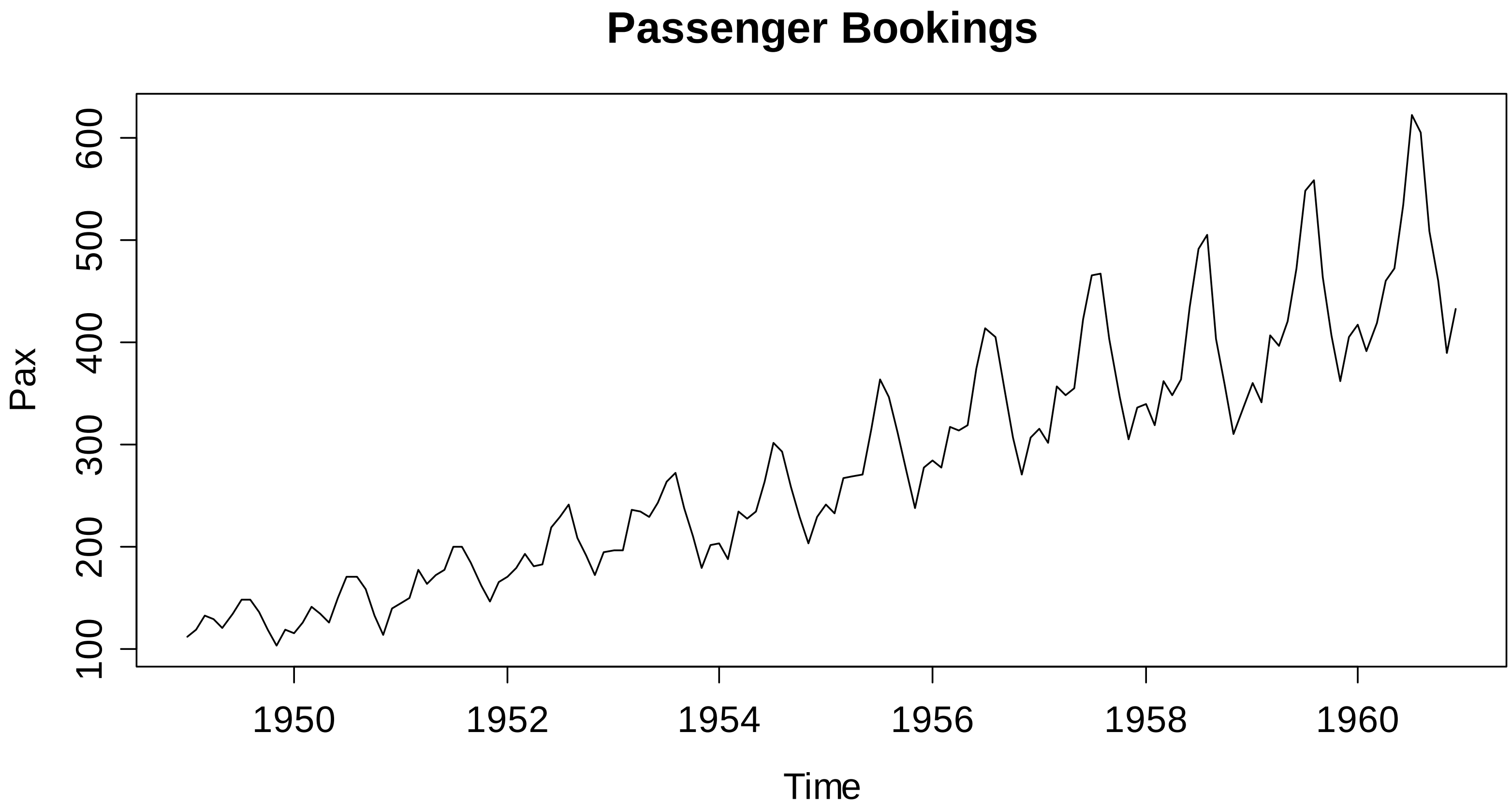
= trend + seasonal effect + stationary remainder

The goal is to find a remainder term R_t, as a sequence of correlated random variables with mean zero, i.e. a stationary ts.

We can employ: 1) *taking differences (=differencing)*
2) *smoothing approaches (= filtering)*
3) *parametric models (= curve fitting)*

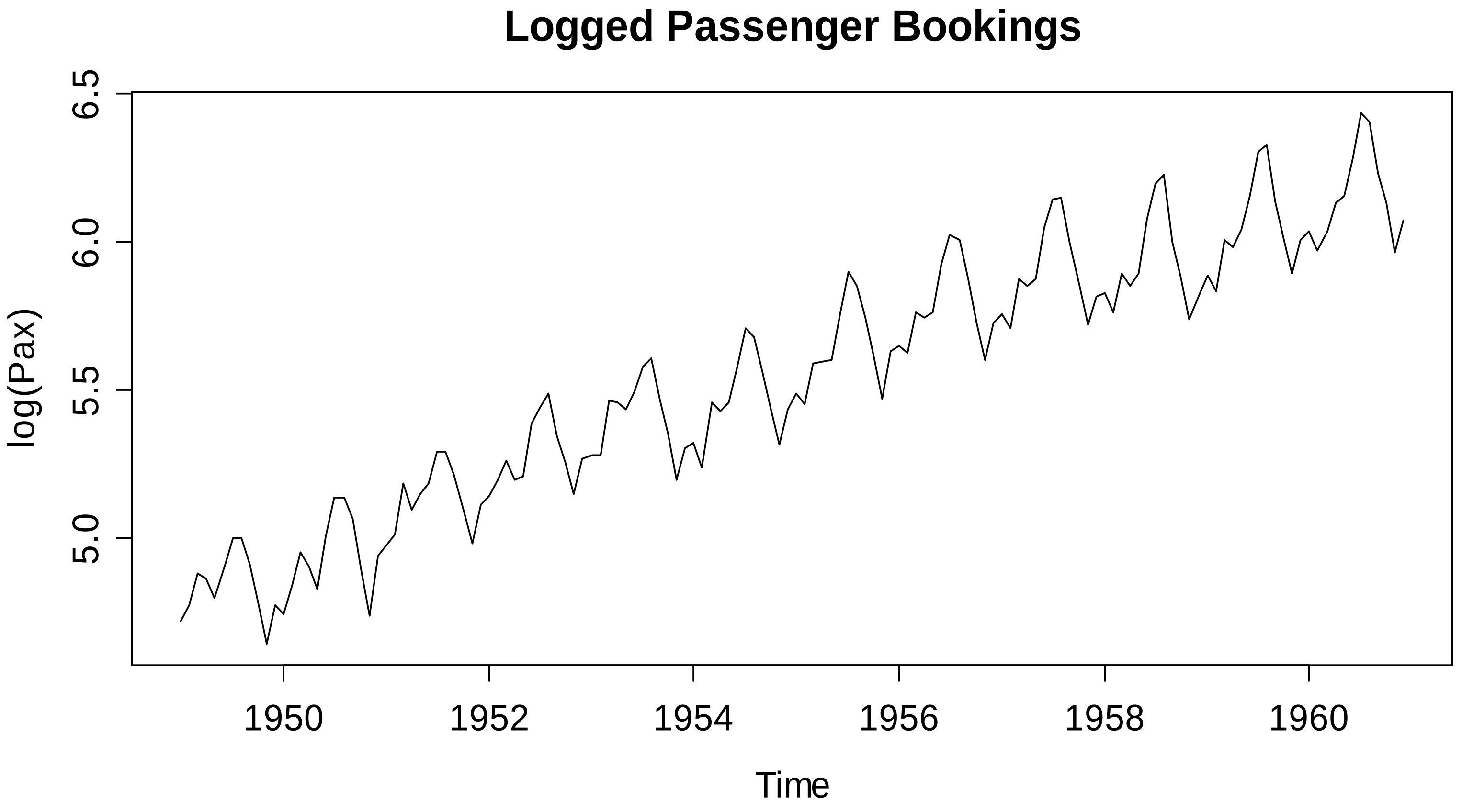
Multiplicative Decomposition

$X_t = m_t + s_t + R_t$ is not always a good model:



Multiplicative Decomposition

Better: $X_t = m_t \cdot s_t \cdot R_t$, respectively $\log(X_t) = m_t' + s_t' + R_t'$



Smoothing, Filtering: Part 1

In the absence of a seasonal effect, the trend of a non-stationary time series can be determined by applying any **additive, linear filter**. We obtain a new time series $\hat{m}_t$, representing the trend:

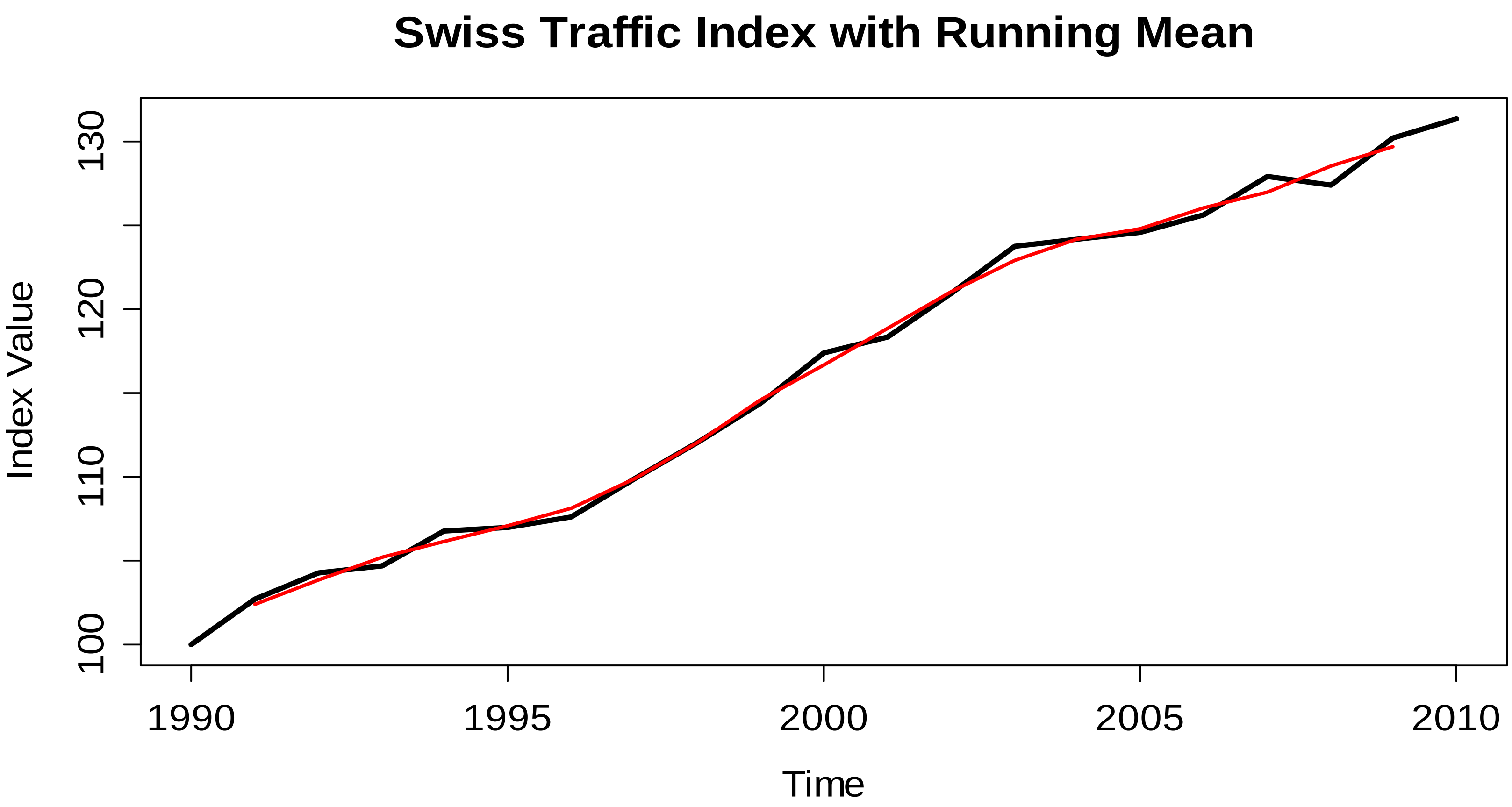
$$\hat{m}_t = \sum_{i=-p}^{q} a_i X_{t+i}$$

- the window, defined by p and q, can or can‘t be symmetric
- the weights, given by a_i, can or can‘t be uniformly distributed
- other smoothing procedures can be applied, too.

Applied Time Series Analysis
SS 2013 – Week 03

Trend Estimation with the Running Mean

```
> trd <- filter(SwissTraffic, filter=c(1,1,1)/3)
```



Smoothing, Filtering: Part 2

In the presence a seasonal effect, smoothing approaches are still valid for estimating the trend. We have to make sure that the sum is taken over an entire season, i.e. for monthly data:

$$\hat{m}_t = \frac{1}{12}\left(\frac{1}{2}X_{t-6} + X_{t-5} + \ldots + X_{t+5} + \frac{1}{2}X_{t+6}\right) \ for \ t = 7,...,n-6$$

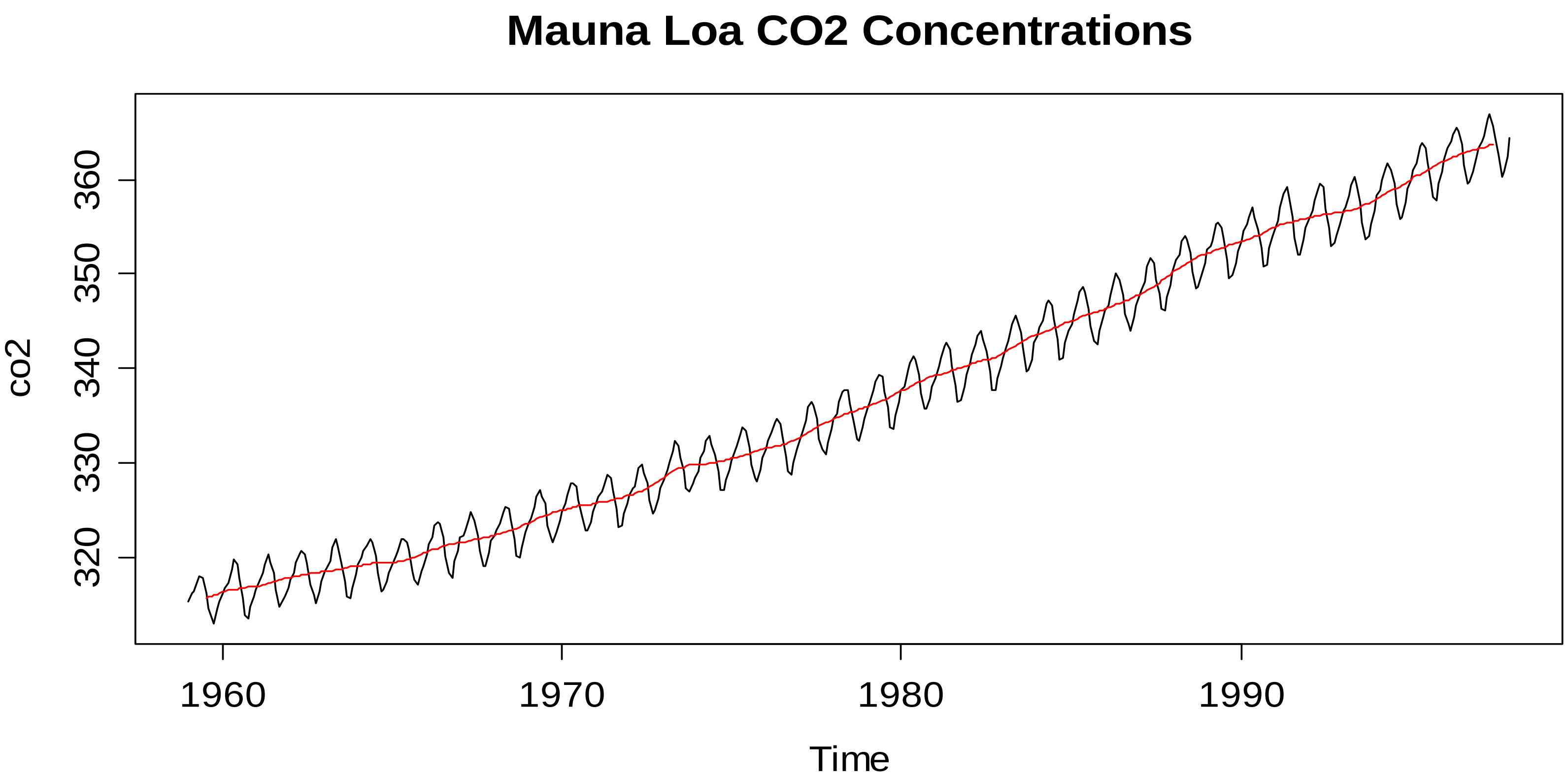
An estimate of the seasonal effect s_t at time t can be obtained by:

$$\hat{s}_t = x_t - \hat{m}_t$$

By averaging these estimates of the effects for each month, we obtain a single estimate of the effect for each month.

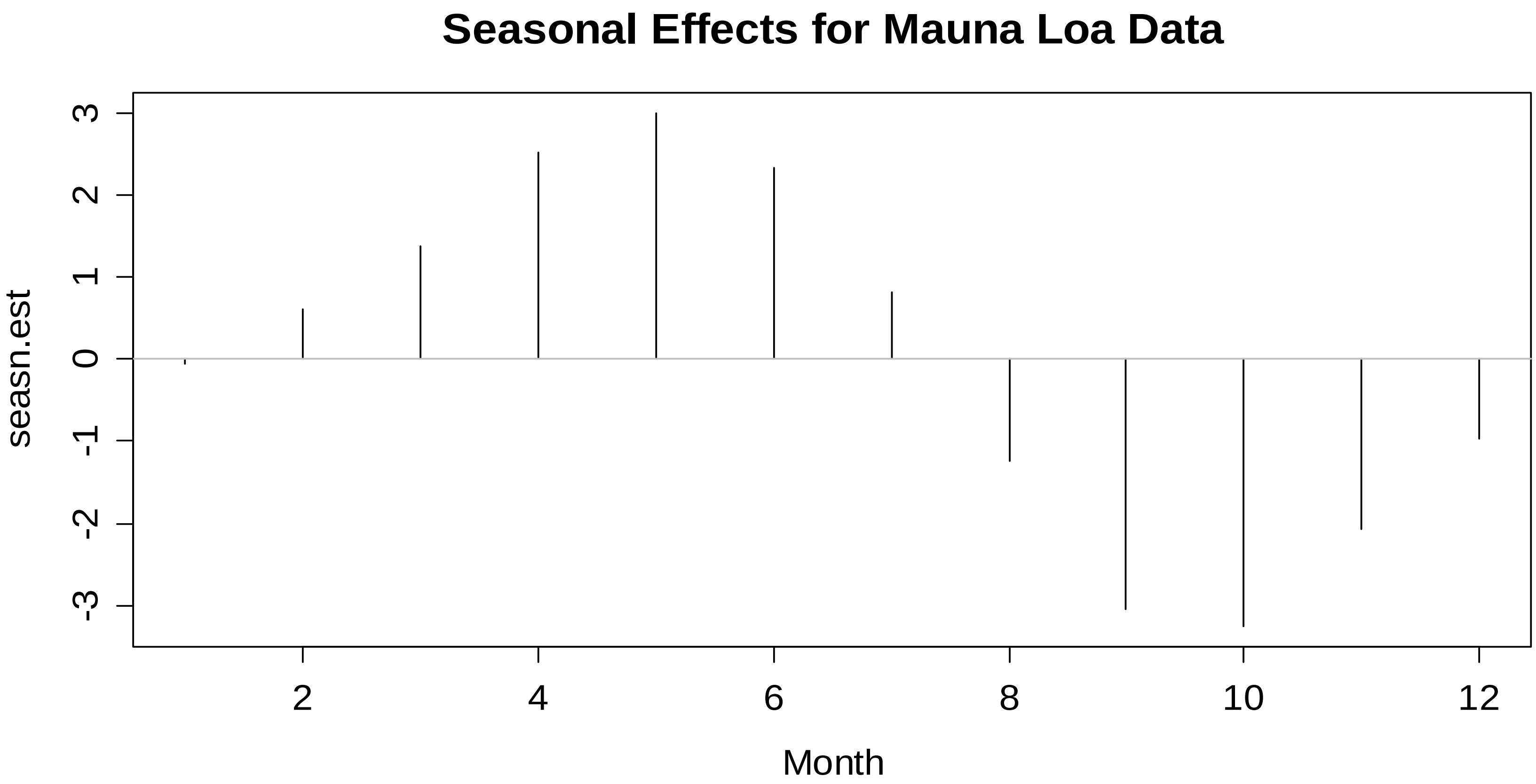
Trend Estimation for Mauna Loa Data

```
> wghts  <- c(.5,rep(1,11),.5)/12
> trd    <- filter(co2, filter=wghts, sides=2)
```



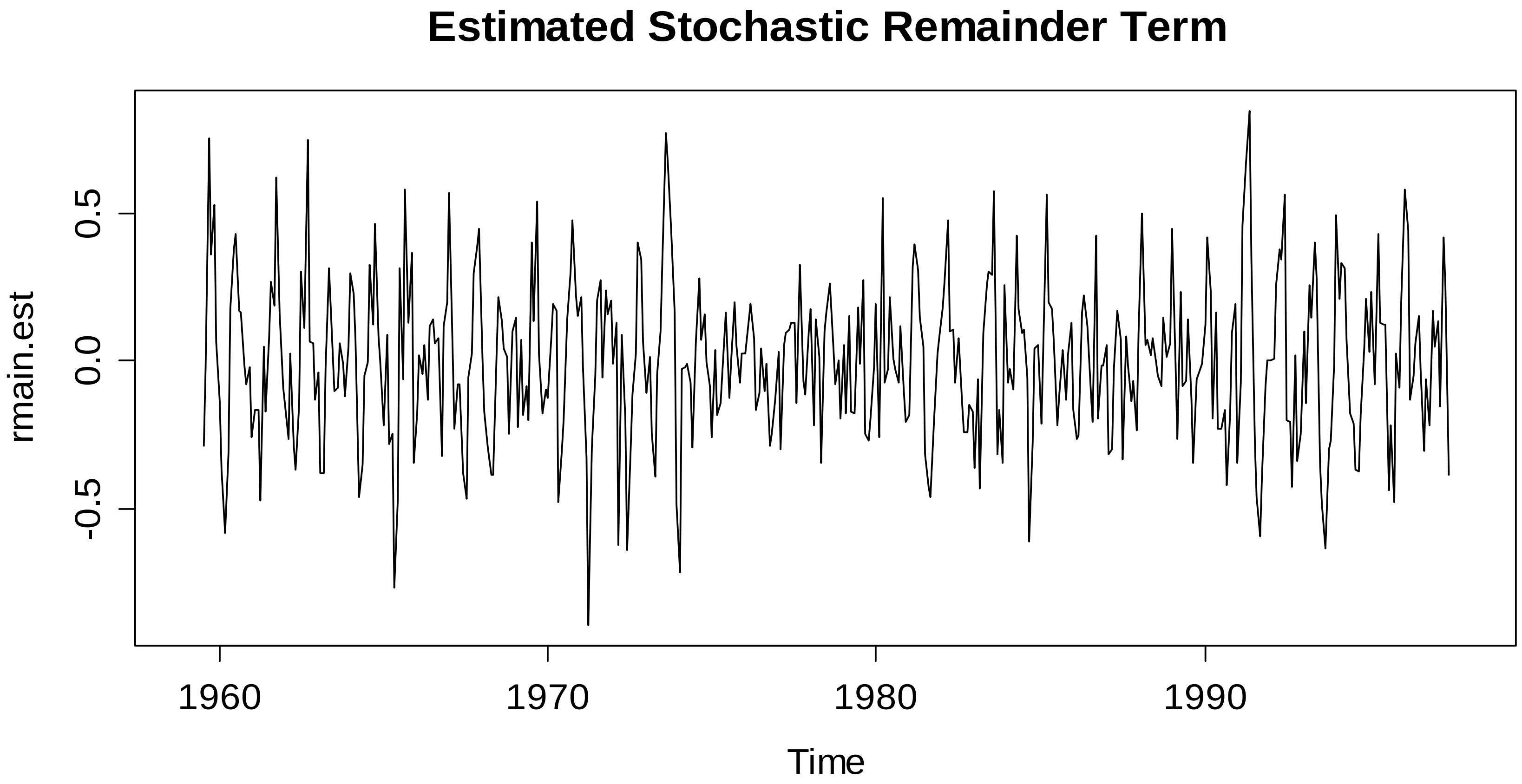
Estimating the Seasonal Effects

$$\hat{s}_{Jan} = \hat{s}_1 = \hat{s}_{13} = \ldots = \frac{1}{39} \cdot \sum_{j=0}^{38} (x_{12j+1} - \hat{m}_{12j+1})$$



Estimating the Remainder Term

$$\hat{R}_t = x_t - \hat{m}_t - \hat{s}_t$$



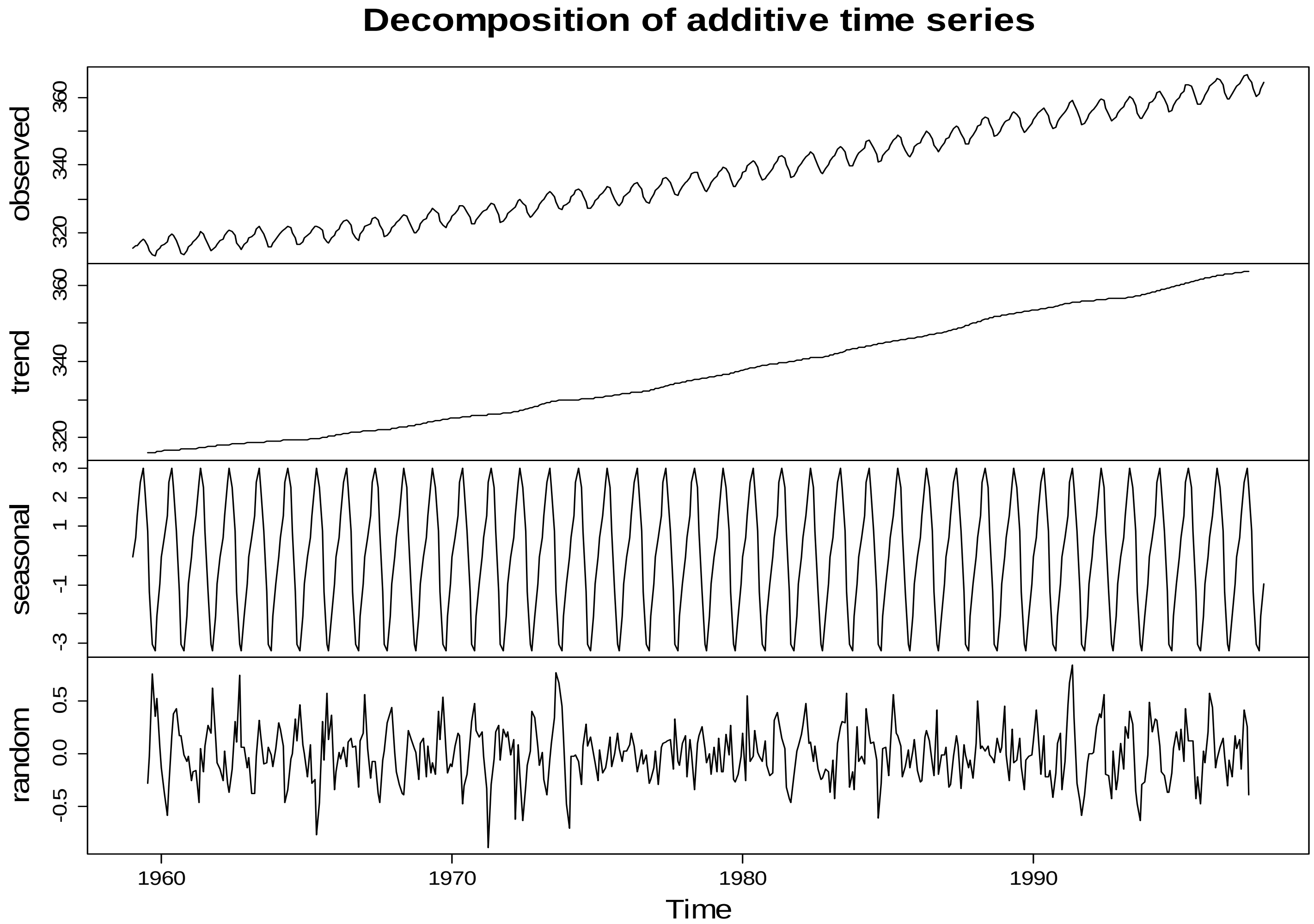
Smoothing, Filtering: Part 3

- The smoothing approach is based on estimating the trend first, and then the seasonality.
- The generalization to other periods than $p = 12$, i.e. monthly data is straighforward. Just choose a symmetric window and use uniformly distributed coefficients that sum up to 1.
- The sum over all seasonal effects will be close to zero. Usually, it is centered to be exactly there.
- This procedure is implemented in R with function: **`decompose()`**

Applied Time Series Analysis
SS 2013 – Week 03

Estimating the Remainder Term

```
> plot(decompose(co2))
```



Smoothing, Filtering: STL-Decomposition

The Seasonal-Trend Decomposition Procedure by Loess

- is an iterative, non-parametric smoothing algorithm
- yields a simultaneous estimation of trend and seasonal effect

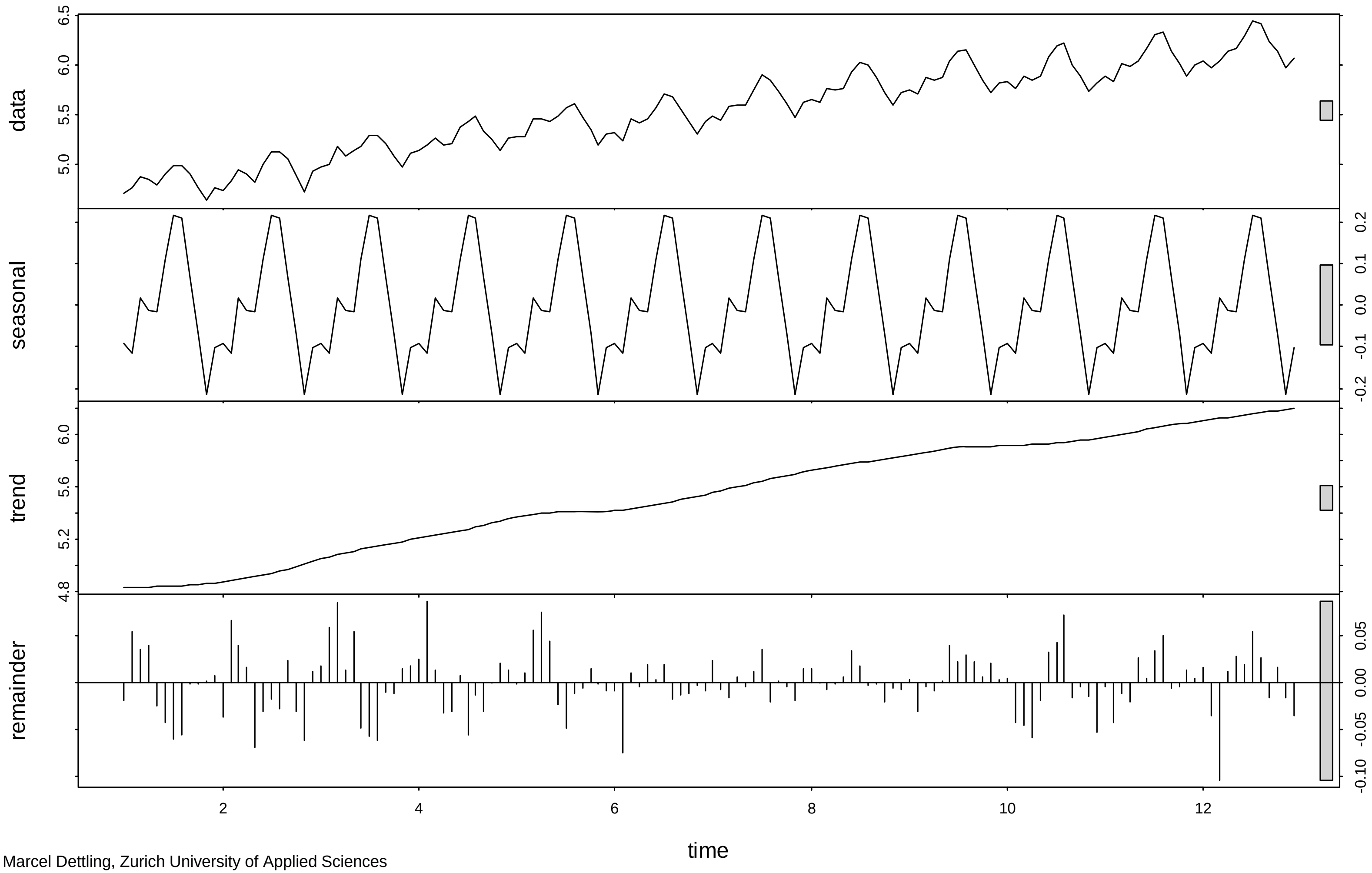
→ similar to what was presented above, but more robust!

+ very simple to apply
+ very illustrative and quick
+ seasonal effect can be constant or smoothly varying
- model free, extrapolation and forecasting is difficult

→ **Good method for „having a quick look at the data“**

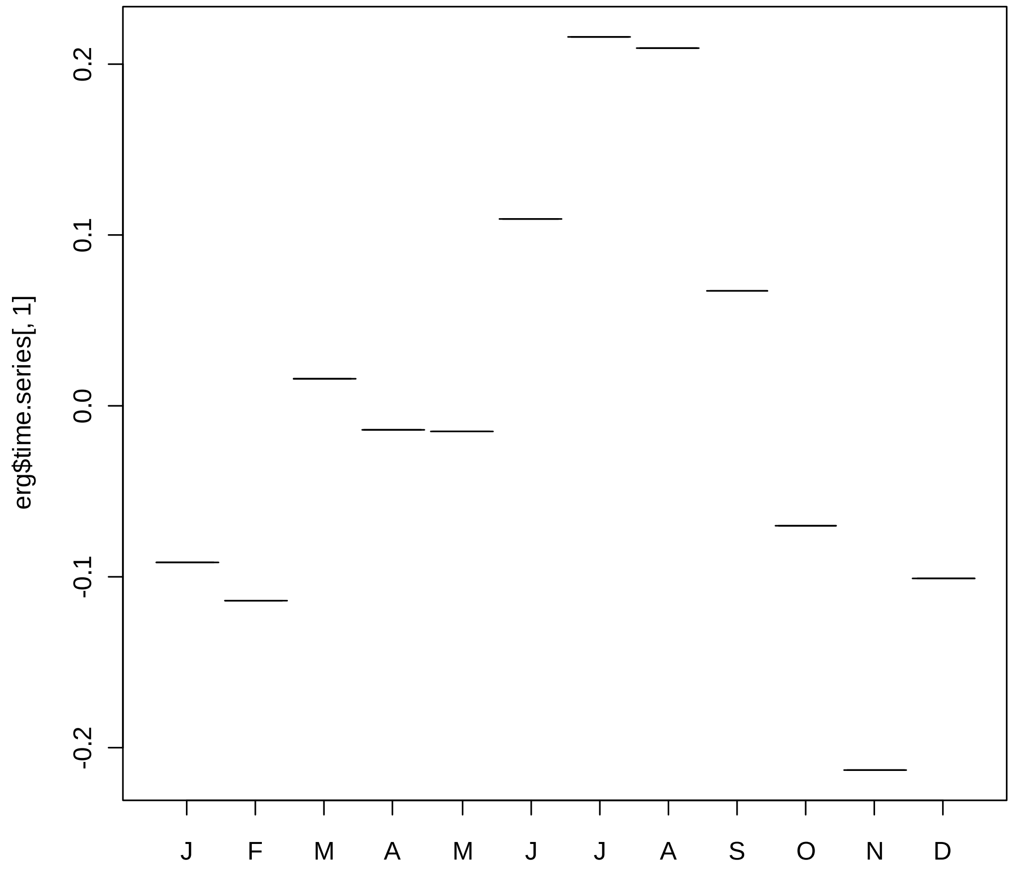
STL-Decomposition: Constant Season

`stl(log(ts(airline,freq=12)),s.window=„periodic“)`



STL-Decomposition: Constant Season

```
stl(log(ts(airline,freq=12)),s.window=„periodic“)
```

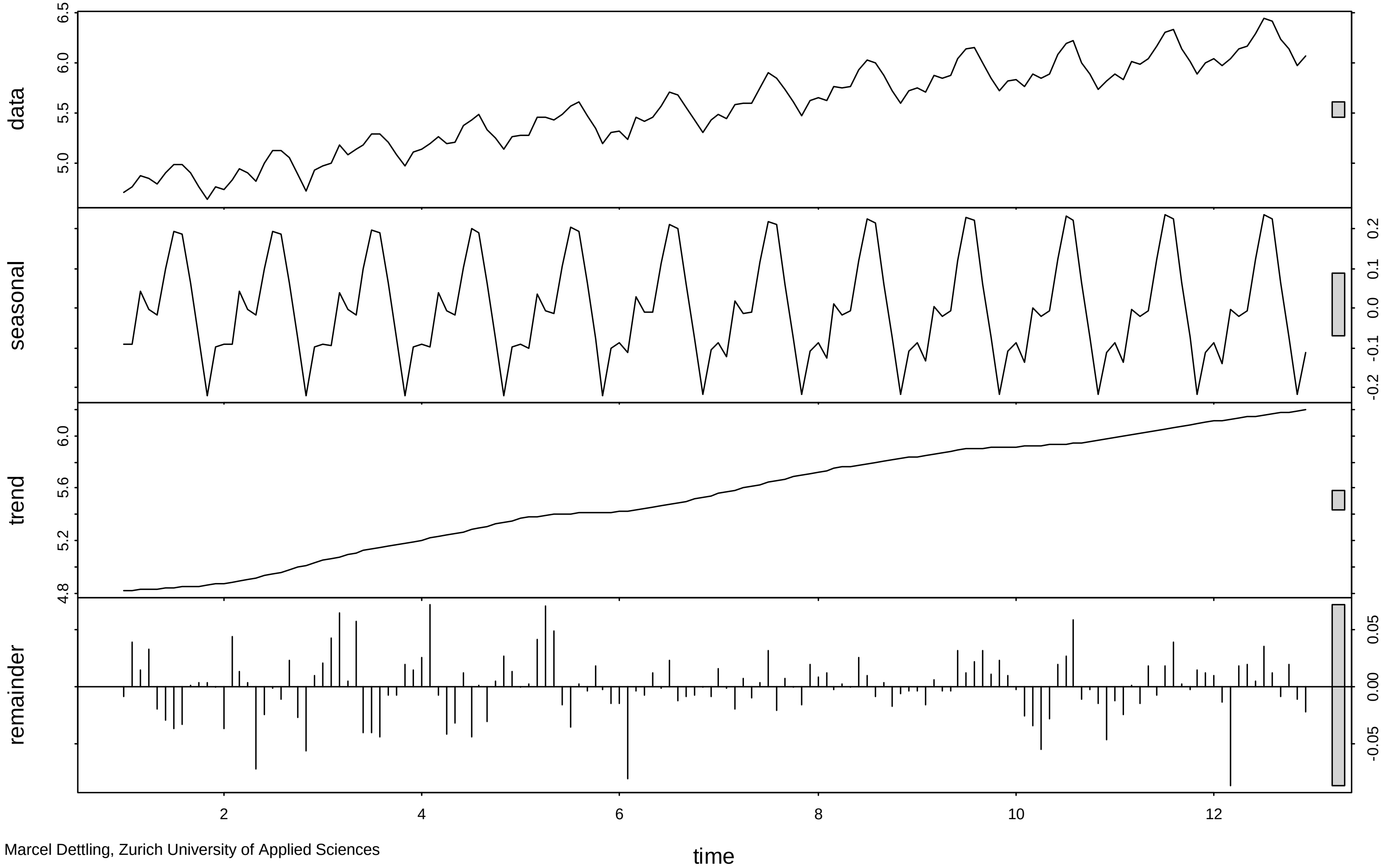


the seasonal effect here is not time varying

Applied Time Series Analysis
SS 2013 – Week 03

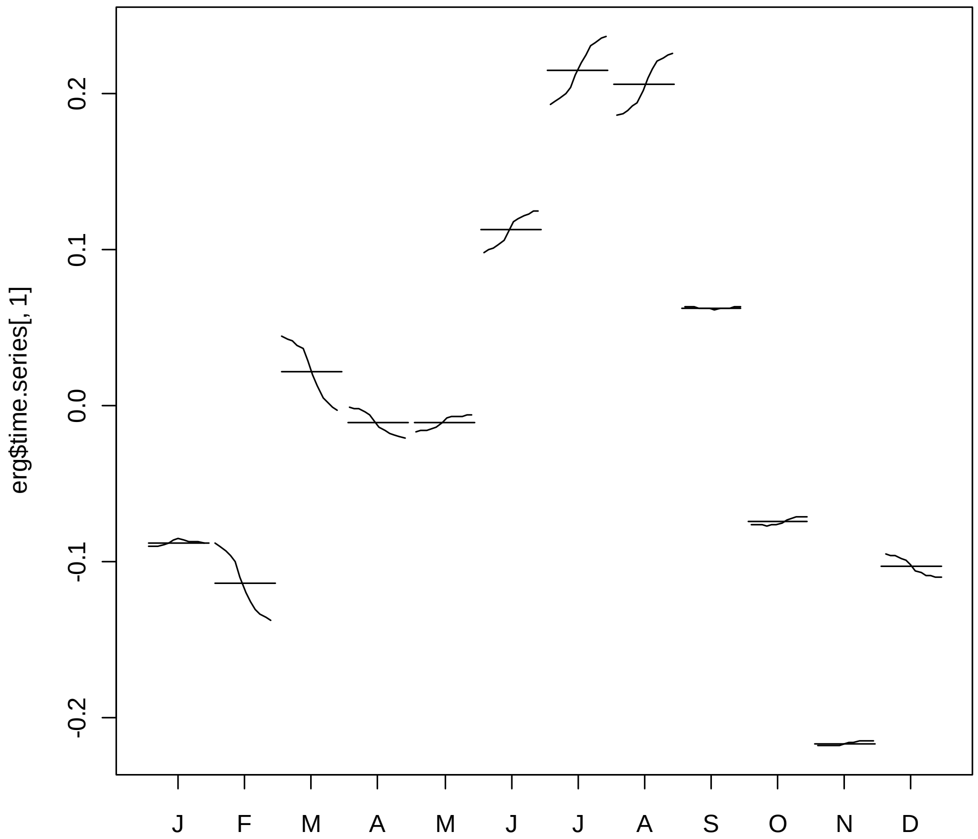
STL-Decomposition: Evolving Season

```
stl(log(ts(airline,freq=12)),s.window=15)
```



STL-Decomposition: Evolving Season

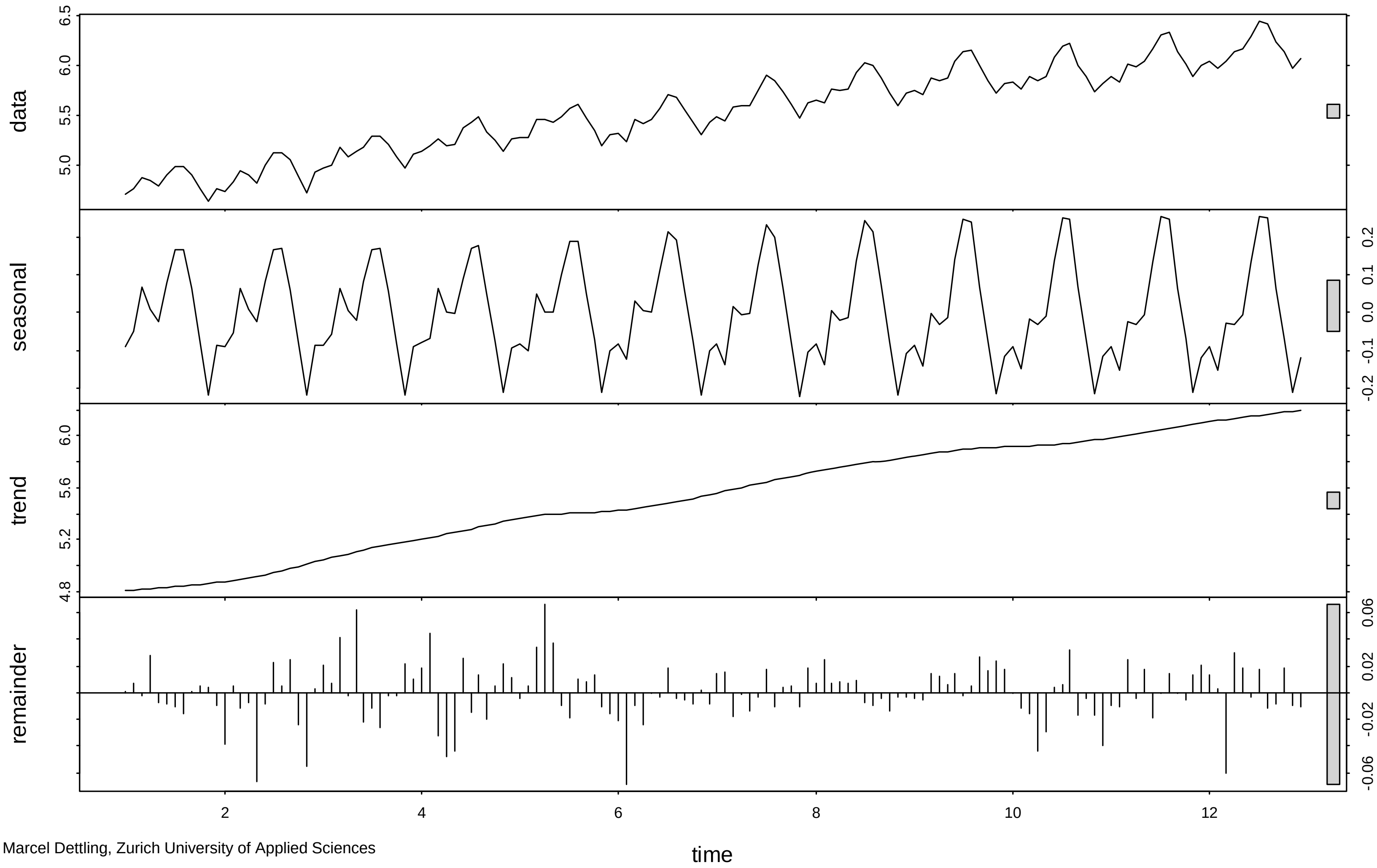
`stl(log(ts(airline,freq=12)),s.window=15)`



correct amount of smoothing on the time varying seasonal effect

STL-Decomposition: Evolving Season

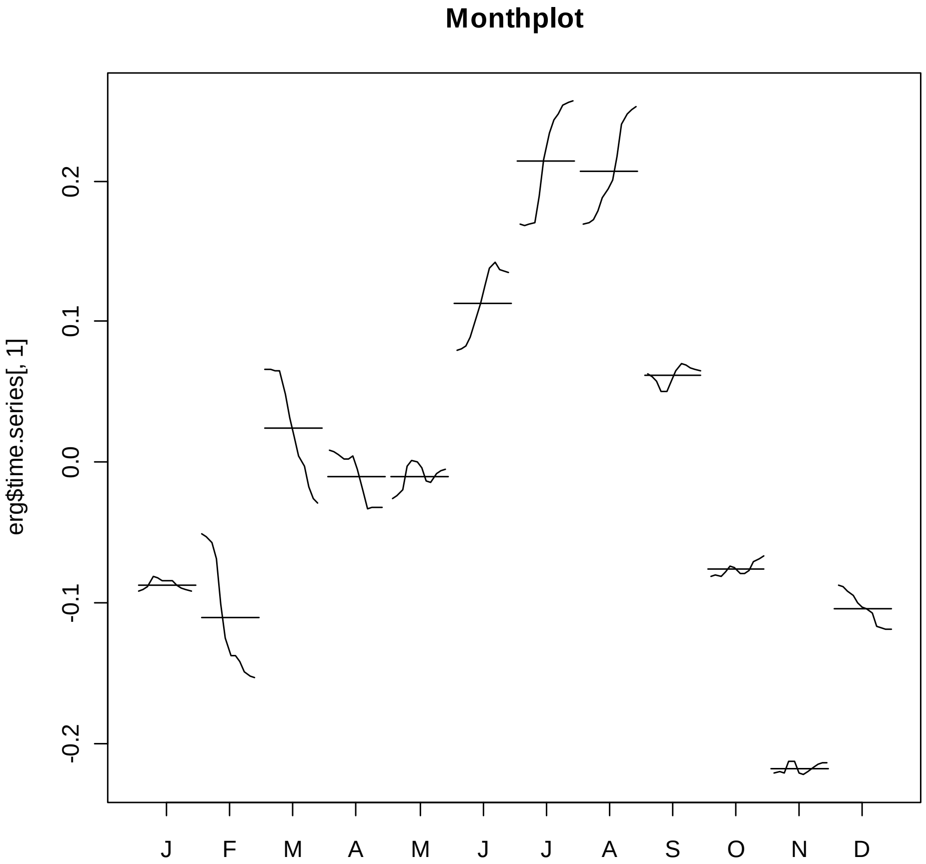
```
stl(log(ts(airline,freq=12)),s.window=7)
```



Applied Time Series Analysis
SS 2013 – Week 03

STL-Decomposition: Evolving Season

`stl(log(ts(airline,freq=12)),s.window=7)`



not enough smoothing on the time varying seasonal effect

Smoothing, Filtering: Remarks

Some advantages and disadvantages:

+ trend and seasonal effect can be estimated
+ $\hat{m}_t$, $\hat{s}_t$ and $\hat{R}_t$ are explicitly known, can be visualised
+ procedure is transparent, and simple to implement

- resulting time series will be shorter than the original
- the running mean is not the very best smoother
- extrapolation of $\hat{m}_t$, $\hat{s}_t$ are not entirely obvious

Parametric Modelling

When to use?

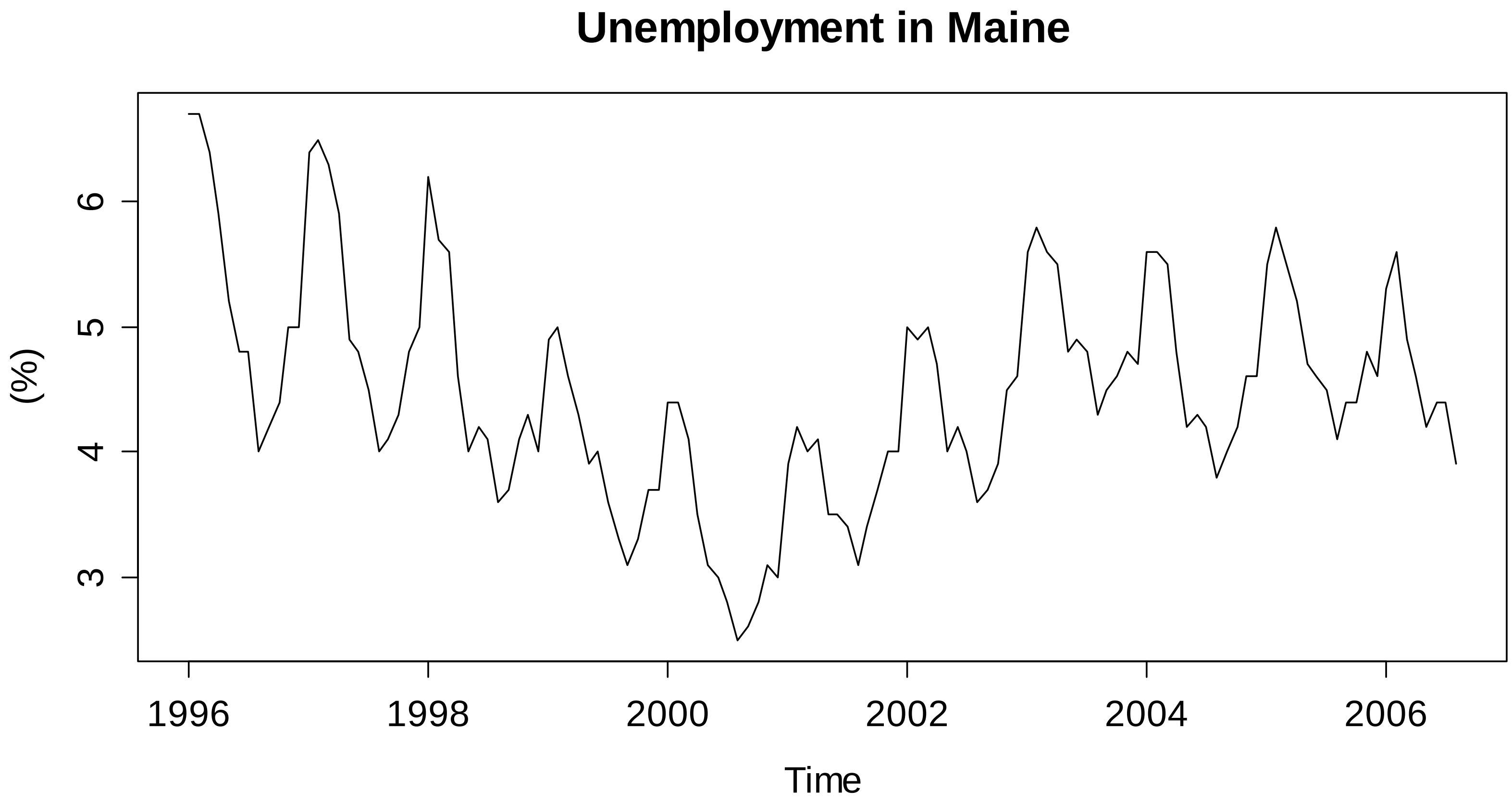
→ Parametric modelling is often used if we have previous knowledge about the trend following a functional form.

→ If the main goal of the analysis is forecasting, a trend in functional form may allow for easier extrapolation than a trend obtained via smoothing.

→ It can also be useful if we have a specific model in mind and want to infer it. Caution: correlated errors!

Parametric Modelling: Example

Maine unemployment data: Jan/1996 – Aug/2006



Modeling the Unemployment Data

Most often, time series are parametrically decomposed by using regression models. For the trend, polynomial functions are widely used, whereas the seasonal effect is modelled with dummy variables (= a factor).

$$X_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + \beta_4 \cdot t^4 + \alpha_{i(t)} + E_t$$

where $t \in \{1, 2, ..., 128\}$

$i(t) \in \{1, 2, ..., 12\}$

Remark: choice of the polynomial degree is crucial!

Polynomial Order / OLS Fitting

Estimation of the coefficients will be done in a regression context. We can use the ordinary least squares algorithm, but:

- we have violated assumptions, E_t is not uncorrelated
- the estimated coefficients are still unbiased
- standard errors (tests, CIs) can be wrong

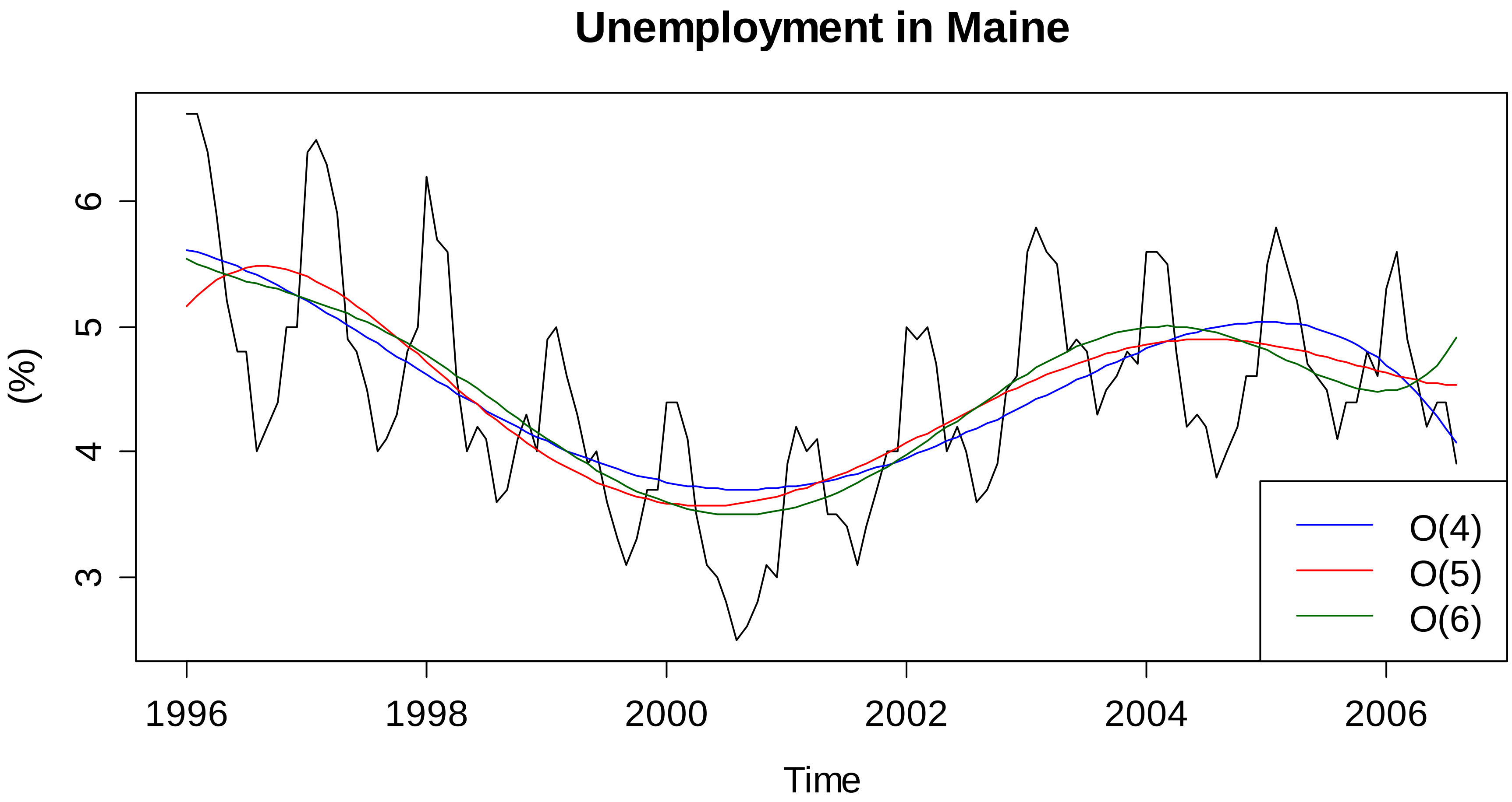
Which polynomial order is required?

Eyeballing allows to determine the minimum grade that is required for the polynomial. It is at least the number of maxima the hypothesized trend has, plus one.

Important Hints for Fitting

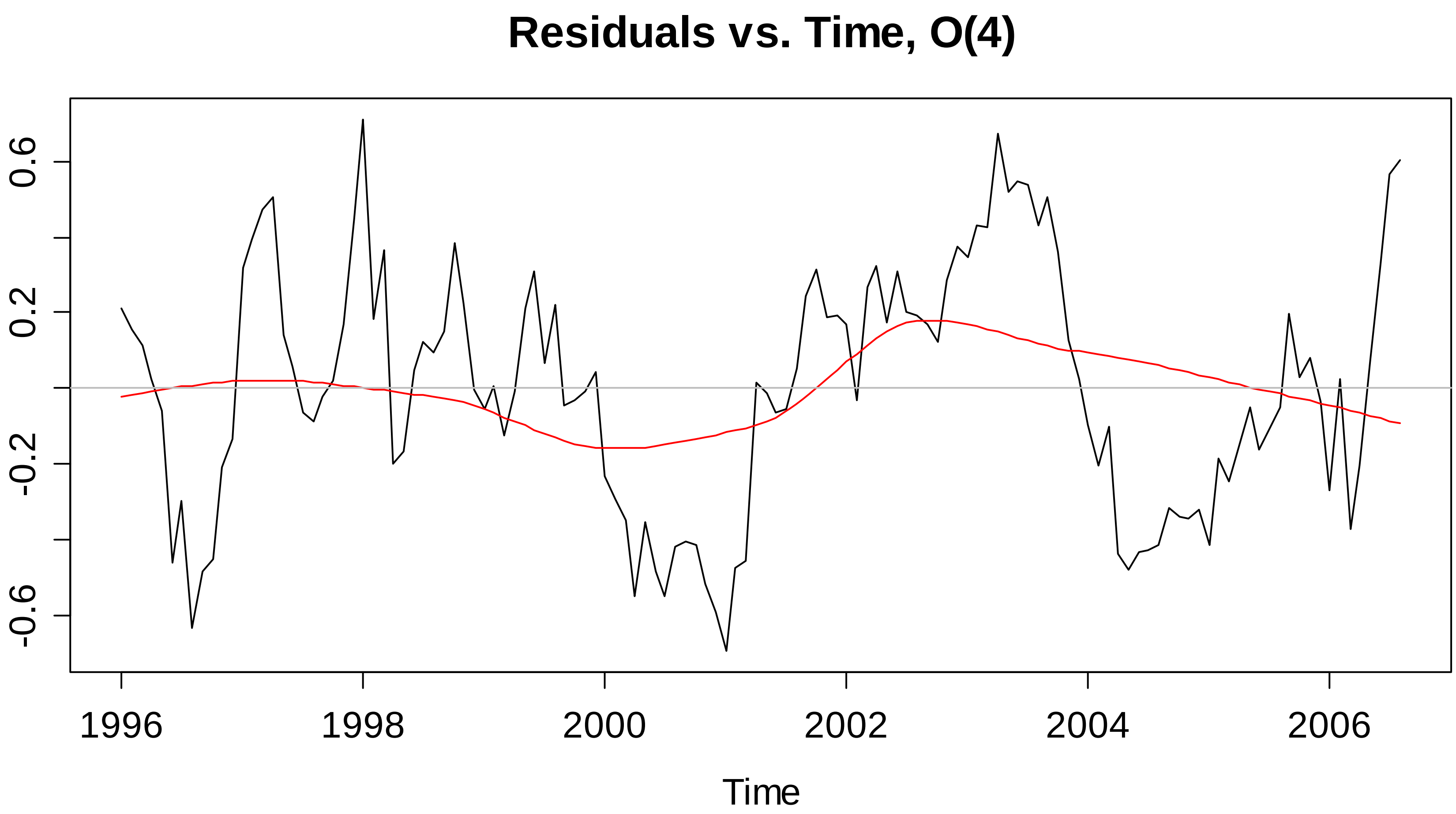
- The main predictor used in polynomial parametric modeling is the time of the observations. It can be obtained by typing `time(maine)`.
- For avoiding numerical and collinearity problems, it is essential to center the time/predictors!
- R sets the first factor level to 0, seasonality is thus expressed as surplus to the January value.
- For visualization: when the trend must fit the data, we have to adjust, because the mean for the seasonal effect is usually different from zero!

Trend of O(4), O(5) and O(6)

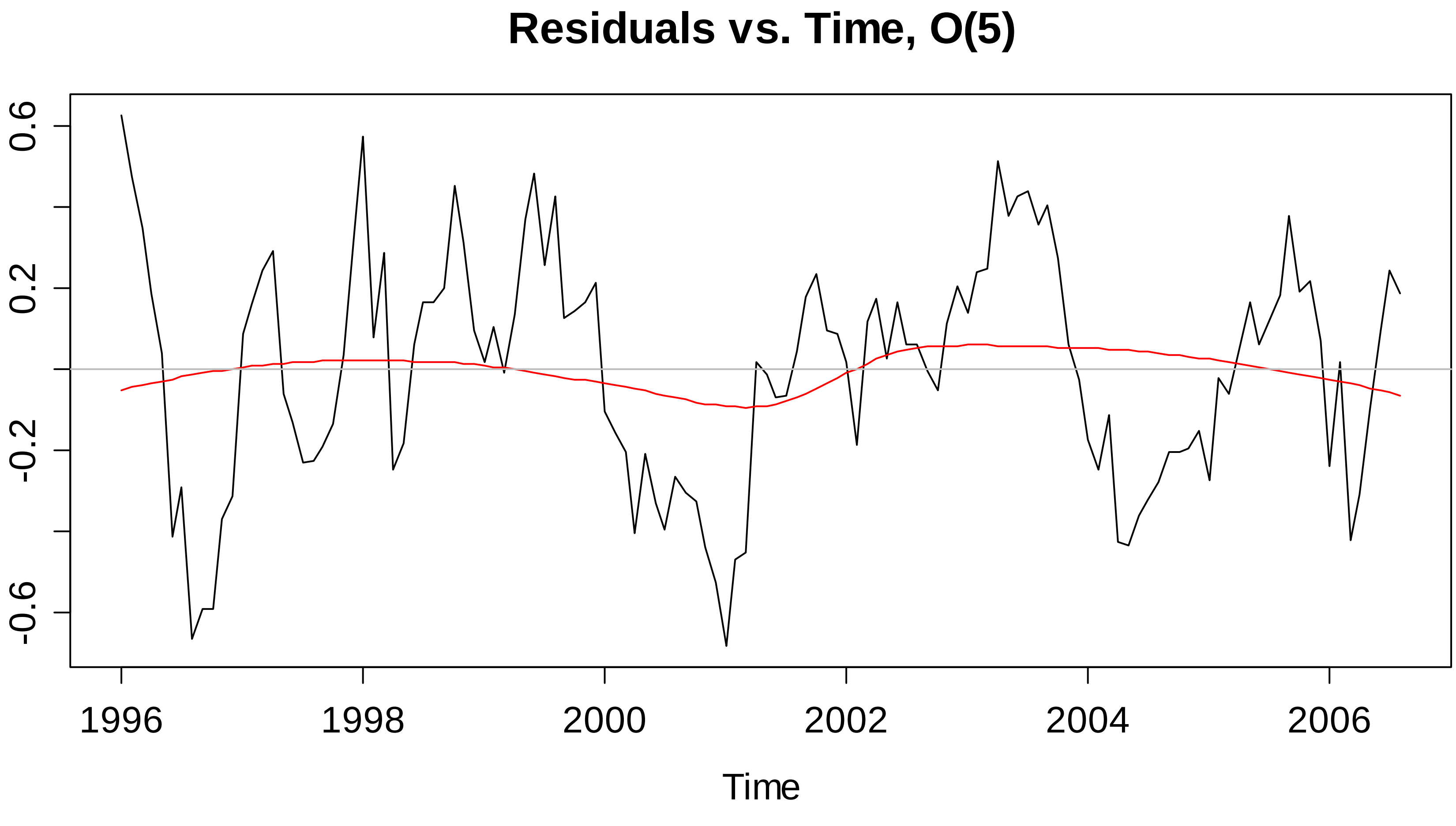


Applied Time Series Analysis
SS 2013 – Week 03

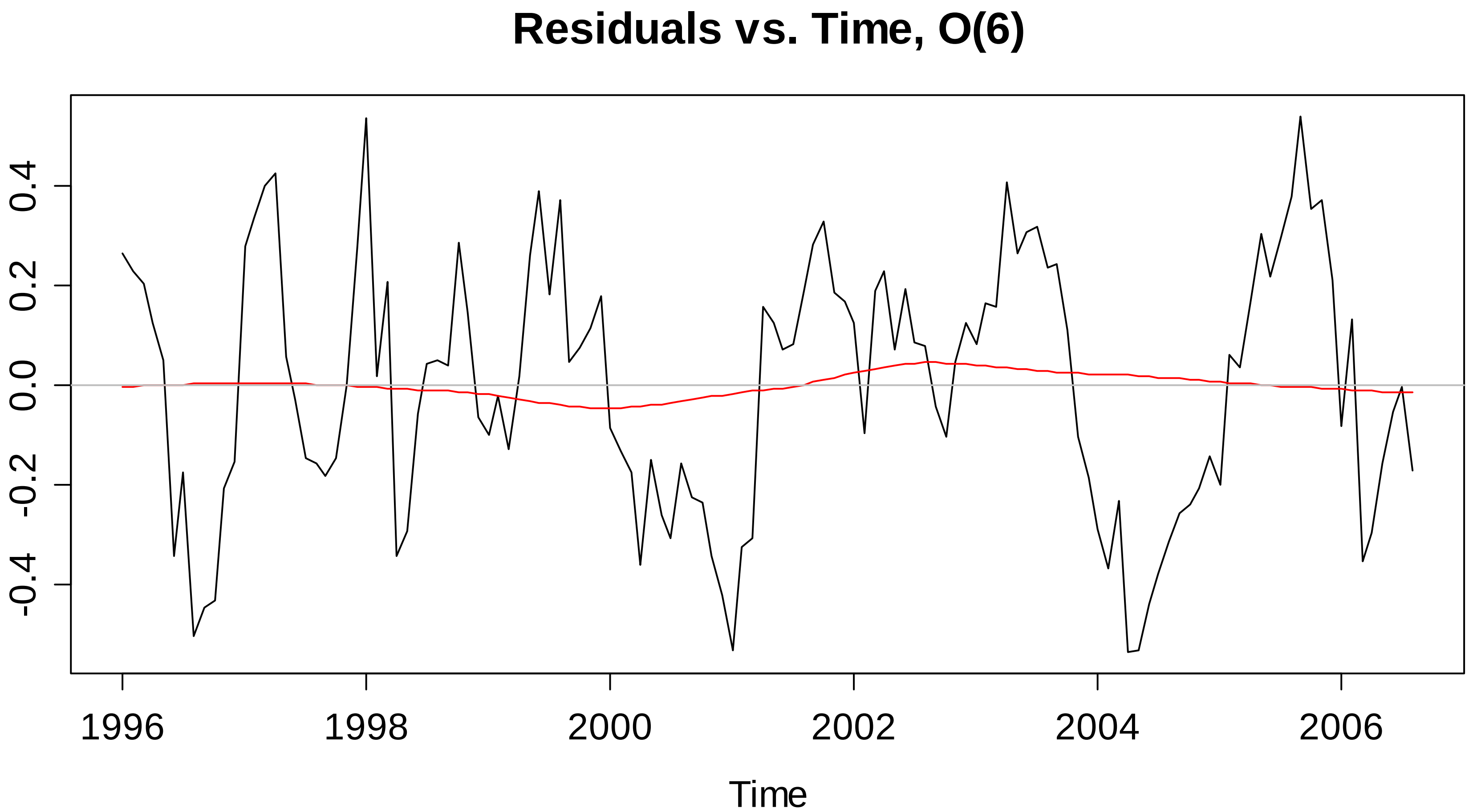
Residual Analysis: O(4)



Residual Analysis: O(5)



Residual Analysis: O(6)



Parametric Modeling: Remarks

Some advantages and disadvantages:

- \+ trend and seasonal effect can be estimated
- \+ $\hat{m}_t$ and $\hat{s}_t$ are explicitly known, can be visualised
- \+ even some inference on trend/season is possible
- \+ time series keeps the original length

- \- choice of a/the correct model is necessary/difficult
- \- residuals are correlated: this is a model violation!
- \- extrapolation of $\hat{m}_t$, $\hat{s}_t$ are not entirely obvious

Applied Time Series Analysis
SS 2013 – Week 03

Where are we?

For most of the rest of this course, we will deal with (weakly) stationary time series. They have the following properties:

- $E[X_t] = \mu$
- $Var(X_t) = \sigma^2$
- $Cov(X_t, X_{t+h}) = \gamma_h$

If a time series is non-stationary, we know how to decompose into deterministic and stationary, random part.

Our forthcoming goals are:

- understanding the dependency in a stationary series
- modeling this dependency and generate forecasts

Autocorrelation

The aim of this section is to explore the dependency structure within a time series.

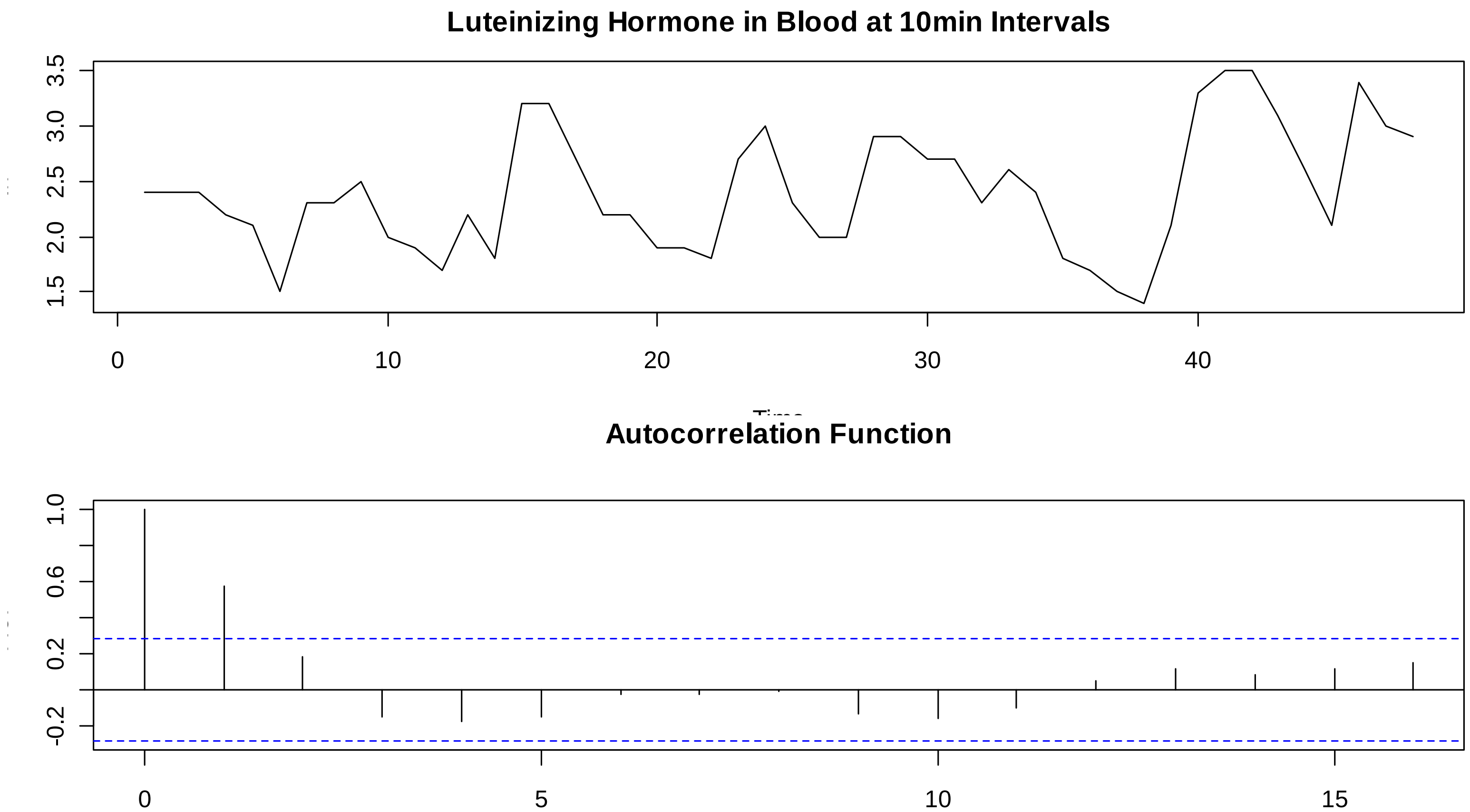
Def: **Autocorrelation**

$$Cor(X_{t+k}, X_t) = \frac{Cov(X_{t+k}, X_t)}{\sqrt{Var(X_{t+k}) \cdot Var(X_t)}}$$

The autocorrelation is a dimensionless measure for the amount of linear association between the random variables collinearity between the random variables X_{t+k} and X_t.

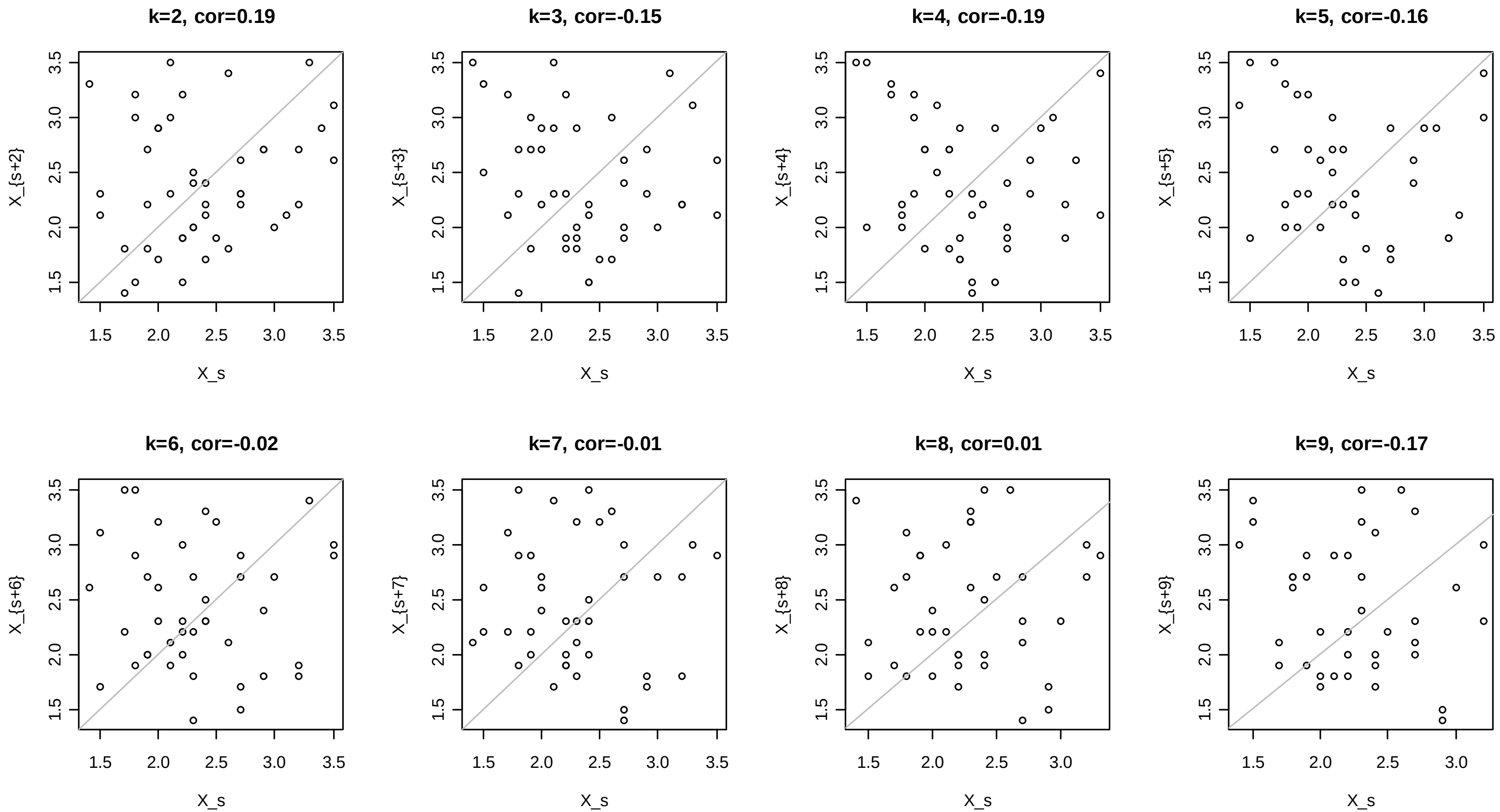
Autocorrelation Estimation

Our next goal is to estimate the autocorrelation function (acf) from a realization of weakly stationary time series.



Autocorrelation Estimation: lag k>1

Idea 1: Compute the sample correlation for all pairs (x_s, x_{s+k})



Autocorrelation Estimation: lag k

Idea 2: Plug-in estimate with sample covariance

How does it work?

→ see blackboard…

Autocorrelation Estimation: lag k

Idea 2: Plug-in estimate with sample covariance

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} = \frac{Cov(X_t, X_{t+k})}{Var(X_t)}$$

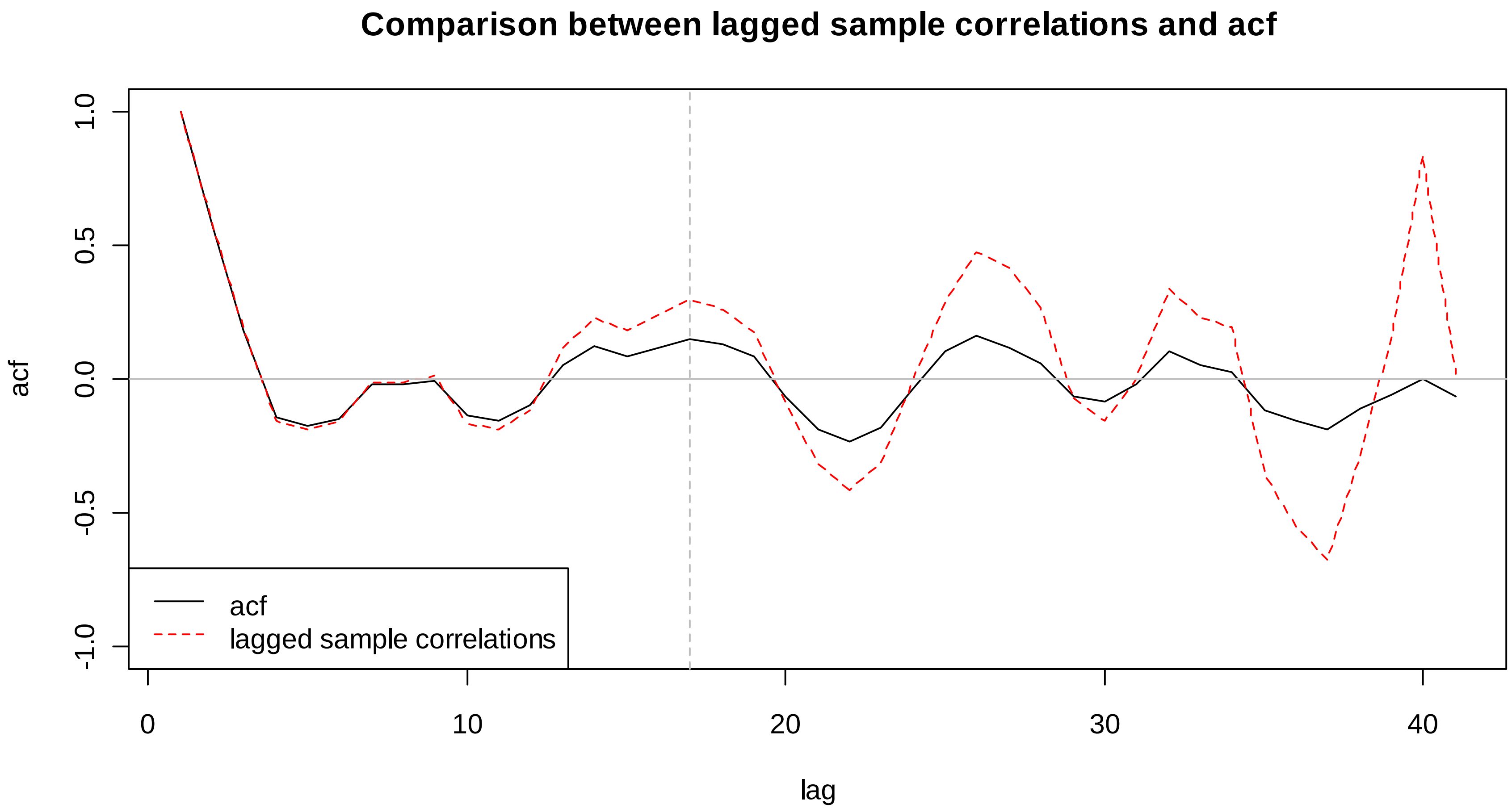
where $$\hat{\gamma}(k) = \frac{1}{n} \sum_{s=1}^{n-k} (x_{s+k} - \overline{x})(x_s - \overline{x})$$

and $$\overline{x} = \frac{1}{n} \sum_{t=1}^{n} x_t$$

Standard approach in time series analysis for computing the acf

Comparison Idea 1 vs. Idea 2

→ **see blackboard for some more information**



What is important about ACF estimation?

- Correlations are never to be trusted without a visual inspection with a scatterplot.

- The bigger the lag k, the fewer data pairs remain for estimating the acf at lag k.

- Rule of the thumb: the acf is only meaningful up to about

  a) lag $10 \cdot \log_{10}(n)$
  b) lag $n/4$

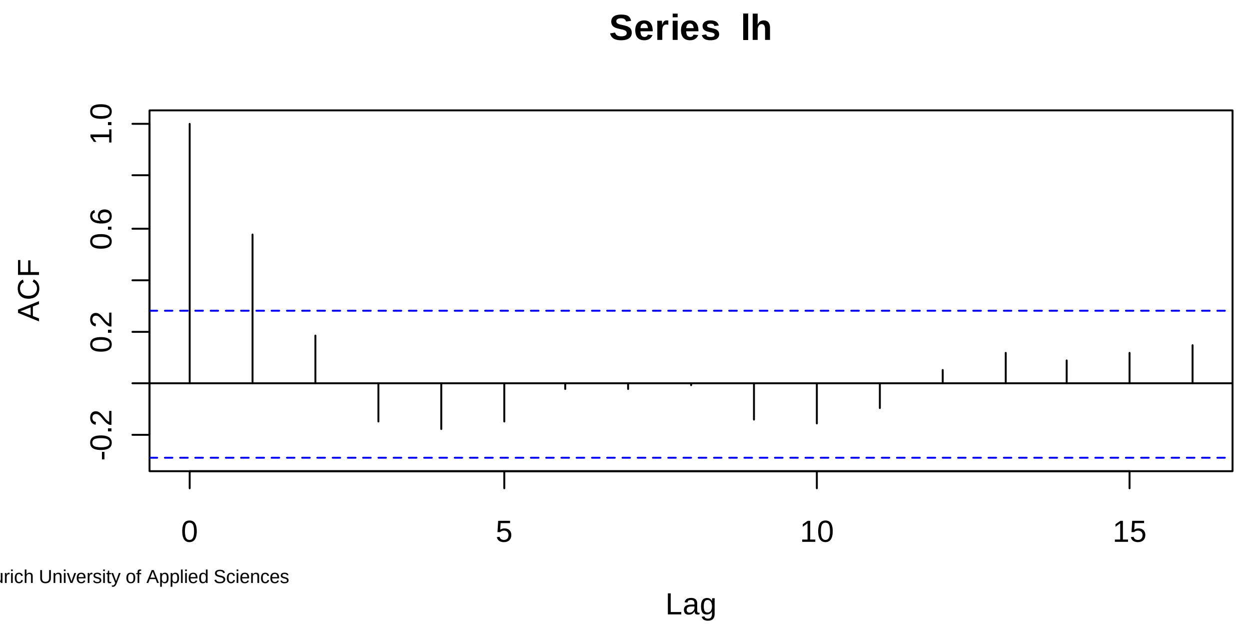
- The estimated sample ACs can be highly correlated.

- The correlogram is only meaningful for stationary series!!!

Correlogram

A useful aid in interpreting a set of autocorrelation coefficients is the graph called correlogram, where the $\hat{\rho}(k)$ are plotted against the lag k.

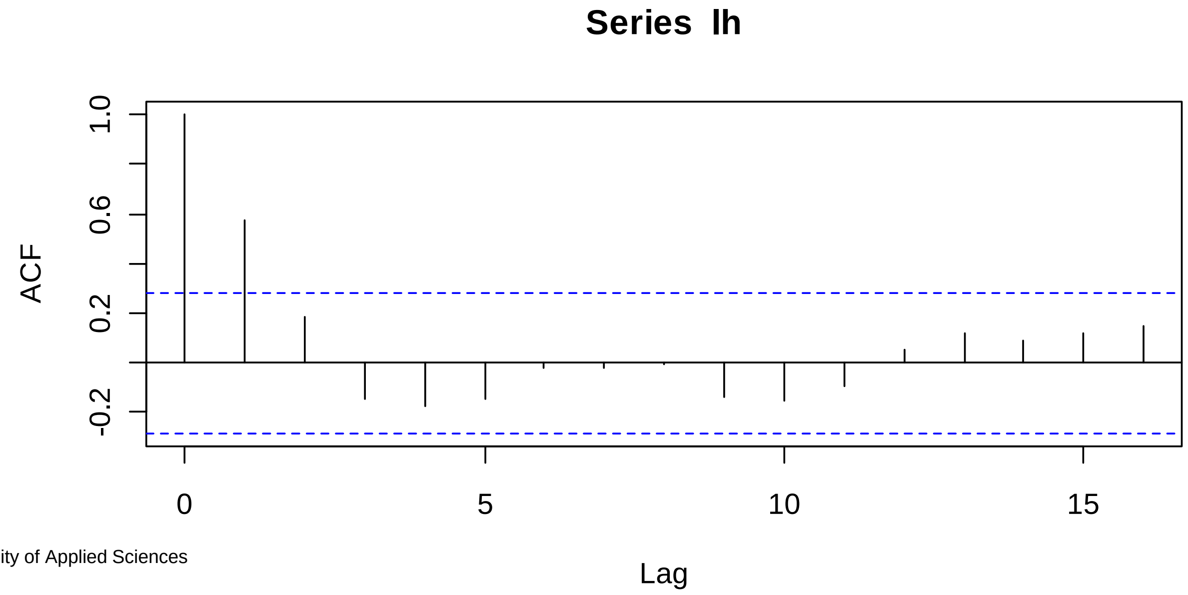
Interpreting the meaning of a set of autocorrelation coefficients is not always easy. The following slides offer some advice.



Random Series – Confidence Bands

If a time series is completely random, i.e. consists of i.i.d. random variables X_t, the (theoretical) autocorrelations $\rho(k)$ are equal to 0.

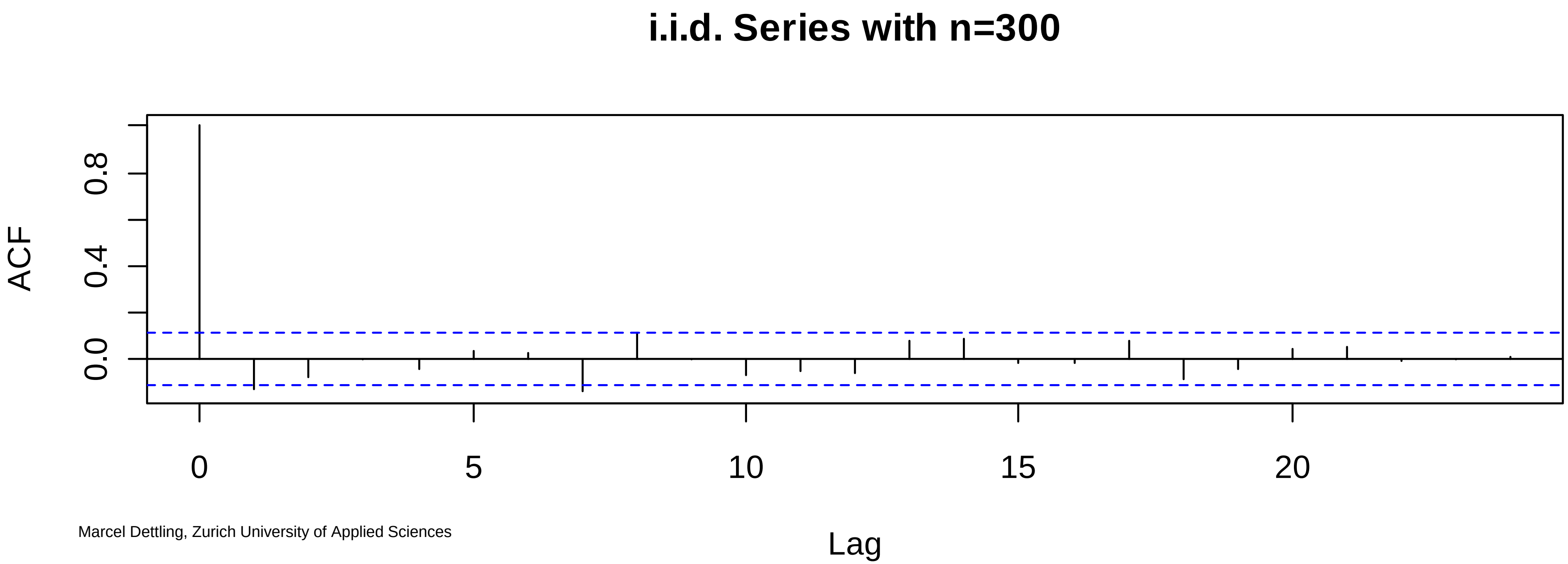
However, the estimated $\hat{\rho}(k)$ are not. We thus need to decide, whether an observed $\hat{\rho}(k) \neq 0$ is significantly so, or just appeared by chance. This is the idea behind the confidence bands.



Random Series – Confidence Bands

For long i.i.d. time series, it can be shown that the $\hat{\rho}(k)$ are approximately $N(0,1/n)$ distributed.

Thus, if a series is random, 95% of the estimated $\hat{\rho}(k)$ can be expected to lie within the interval $\pm 2/\sqrt{n}$



Random Series – Confidence Bands

Thus, even for a (long) i.i.d. time series, we expect that 5% of the estimated autocorrelation coeffcients exceed the confidence bounds. They correspond to type I errors.

Note: the probabilistic properties of non-normal i.i.d series are much more difficult to derive.