Applied Time Series Analysis
FS 2012 – Week 10

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, April 30, 2012

Applied Time Series Analysis
FS 2012 – Week 10

Forecasting with Time Series

Goal:

Prediction of future observations with a measure of uncertainty (confidence interval)

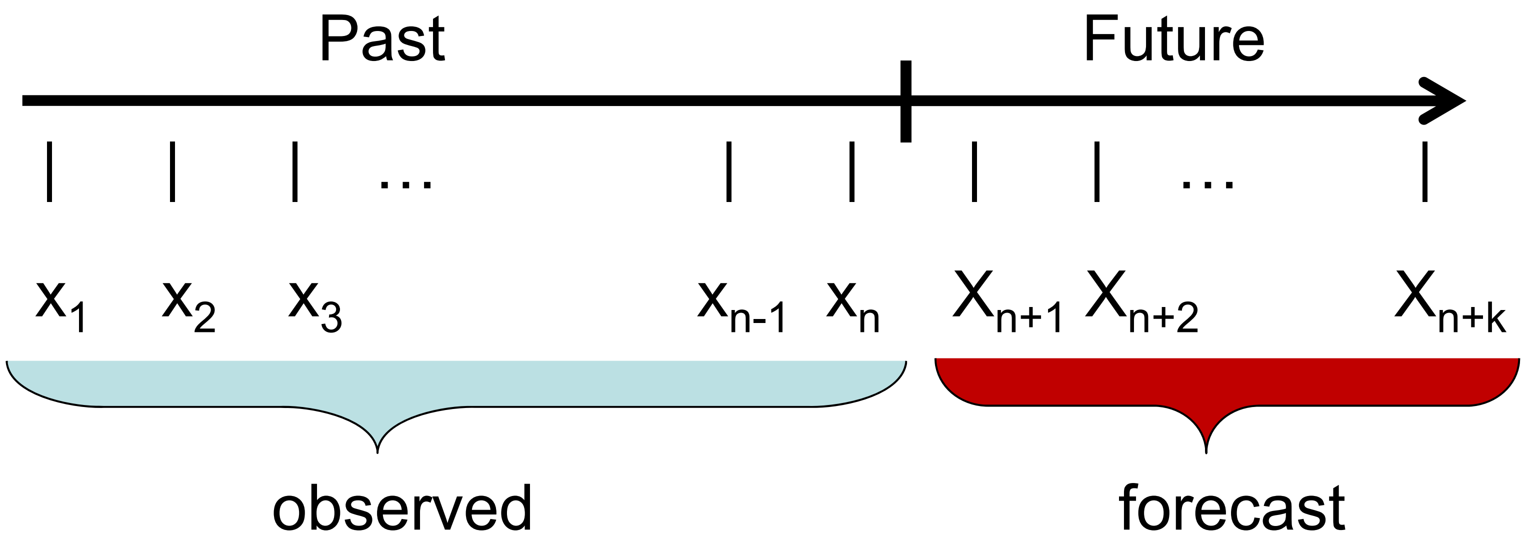
Important:

- will be based on a stochastic model
- builds on the dependency structure and past data
- is an extrapolation, thus to take with a grain of salt
- similar to driving a car by using the rear window mirror

Zurich University of Applied Sciences

Applied Time Series Analysis
FS 2012 – Week 10

Forecasting, More Technical



observations

$$\{x_1, \ldots, x_n\} = X_1^n$$

estimates

$$\left\{ \hat{X}_{n+1,n}, \ldots, \hat{X}_{n+k,n} \right\}$$

Sources of Uncertainty

There are 3 main sources of uncertainty:

1) Does the data generating model from the past also apply in the future?

2) Is the AR(p)-model we fitted to the data $\{x_1,\ldots,x_n\}$ correctly chosen?

3) Are the parameters $\alpha_1,\ldots,\alpha_p,\sigma_E^2$ and μ accurately estimated?

→ **we will here restrict to short-term forecasting!**

Zurich University of Applied Sciences

Exponential Smoothing

Simple exponential smoothing:

- works for stationary time series without trend & season
- is a heuristic, model-free approach
- further in the past -> less weight in the forecast

Turns out to yield these forecasts:

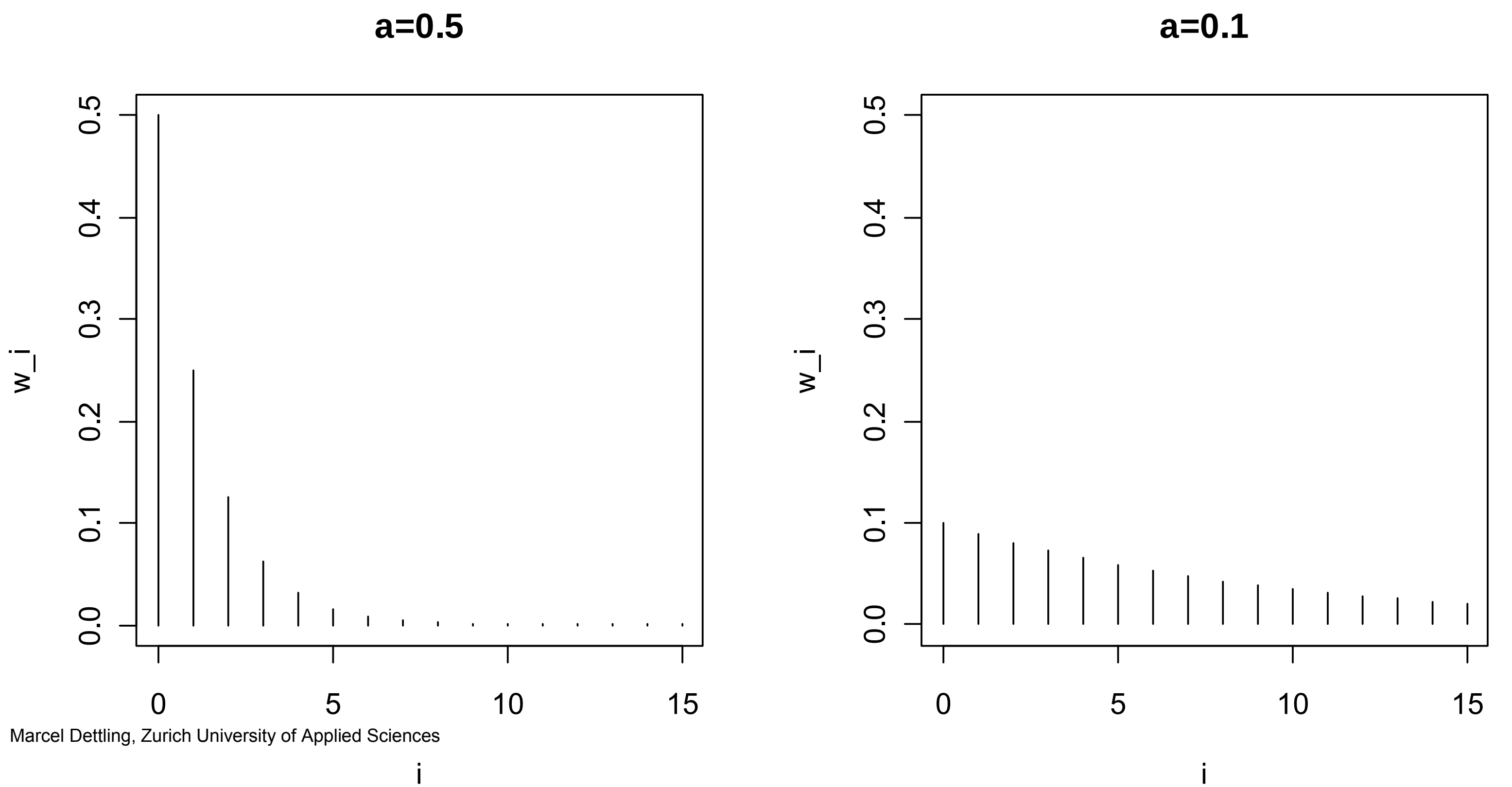
$$\hat{X}_{n+1,n} = \sum_{i=0}^{n-1} w_i x_{n-i} \quad \text{where } w_0 \geq w_1 \geq w_2 \geq \ldots \geq 0 \text{ and } \sum_{i=0}^{n-1} w_i = 1$$

→ **See the blackboard for the derivation...**

Choice of Weights

An usual choice are exponentially decaying weights:

$$w_i = \alpha(1-\alpha)^i \quad \text{where } \alpha \in (0,1)$$



Forecasting with Exponential Smoothing

The 1-step forecast is:

$$\underbrace{\hat{X}_{n+1,n} = \sum_{i=0}^{n-1} \alpha(1-\alpha)^i \cdot x_{n-i}}_{\text{General Formula}} = \underbrace{\alpha x_n + (1-\alpha)\hat{X}_{n,n-1}}_{\text{“Update”-Formula}}$$

Remarks:

- in real applications (finite sum), the weights do not add to 1.
- the update-formula is useful if “new” observations appear.
- the k-step forecast is identical to the 1-step forecast.

Exponential Smoothing: Remarks

- the parameter α can be determined by evaluating forecasts that were generated from different α. We then choose the one resulting in the lowest sum of squared residuals.

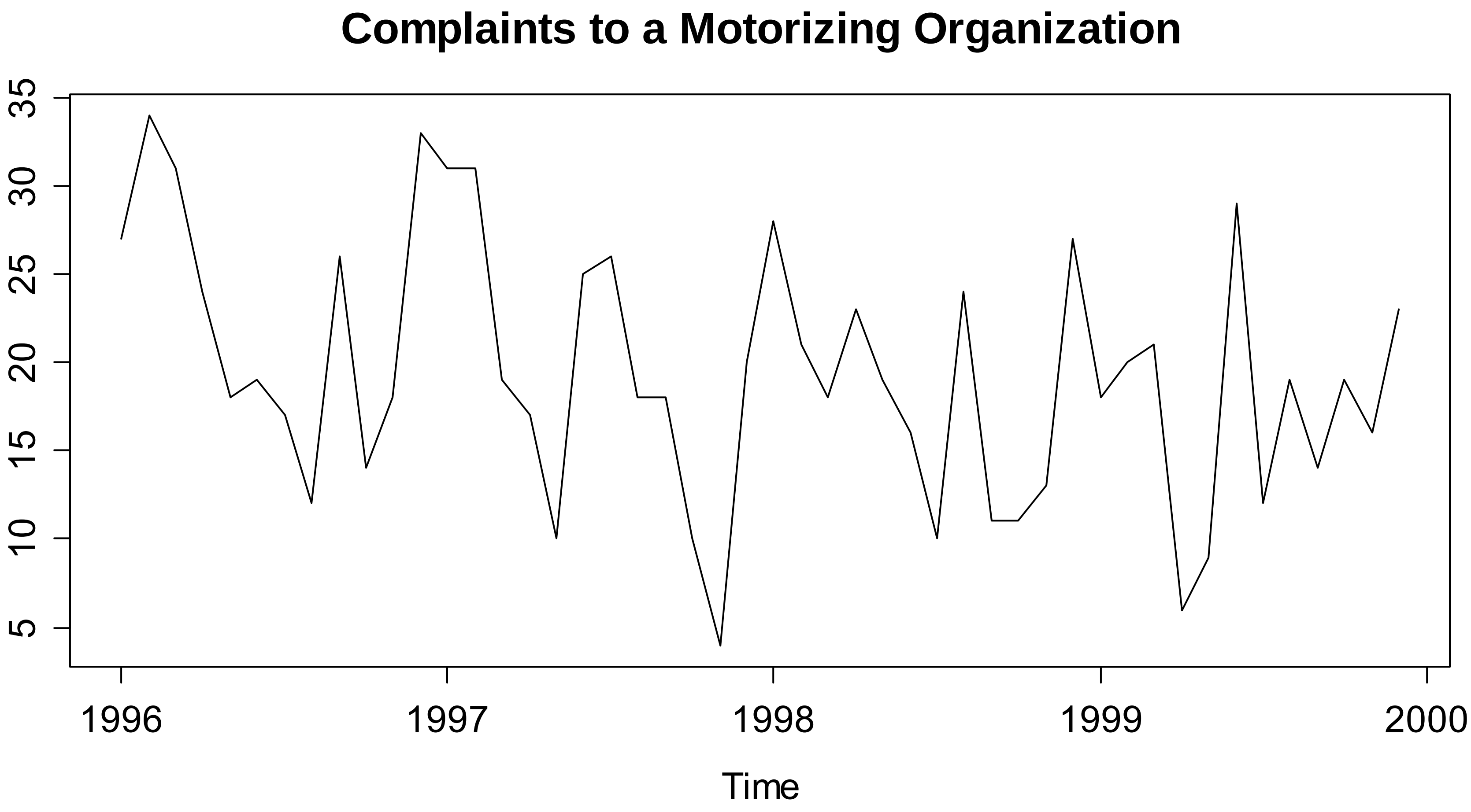
- exponential smoothing is fundamentally different from AR(p)-forecasting. All past values are regarded for the 1-step forecast, but all k-step forecasts are identical to the 1-step.

- It can be shown that exponential smoothing can be optimal for MA(1)-models.

- there are double/triple exponential smoothing approaches that can deal with linear/quadratic trends.

Applied Time Series Analysis
FS 2012 – Week 10

Exponential Smoothing: Example

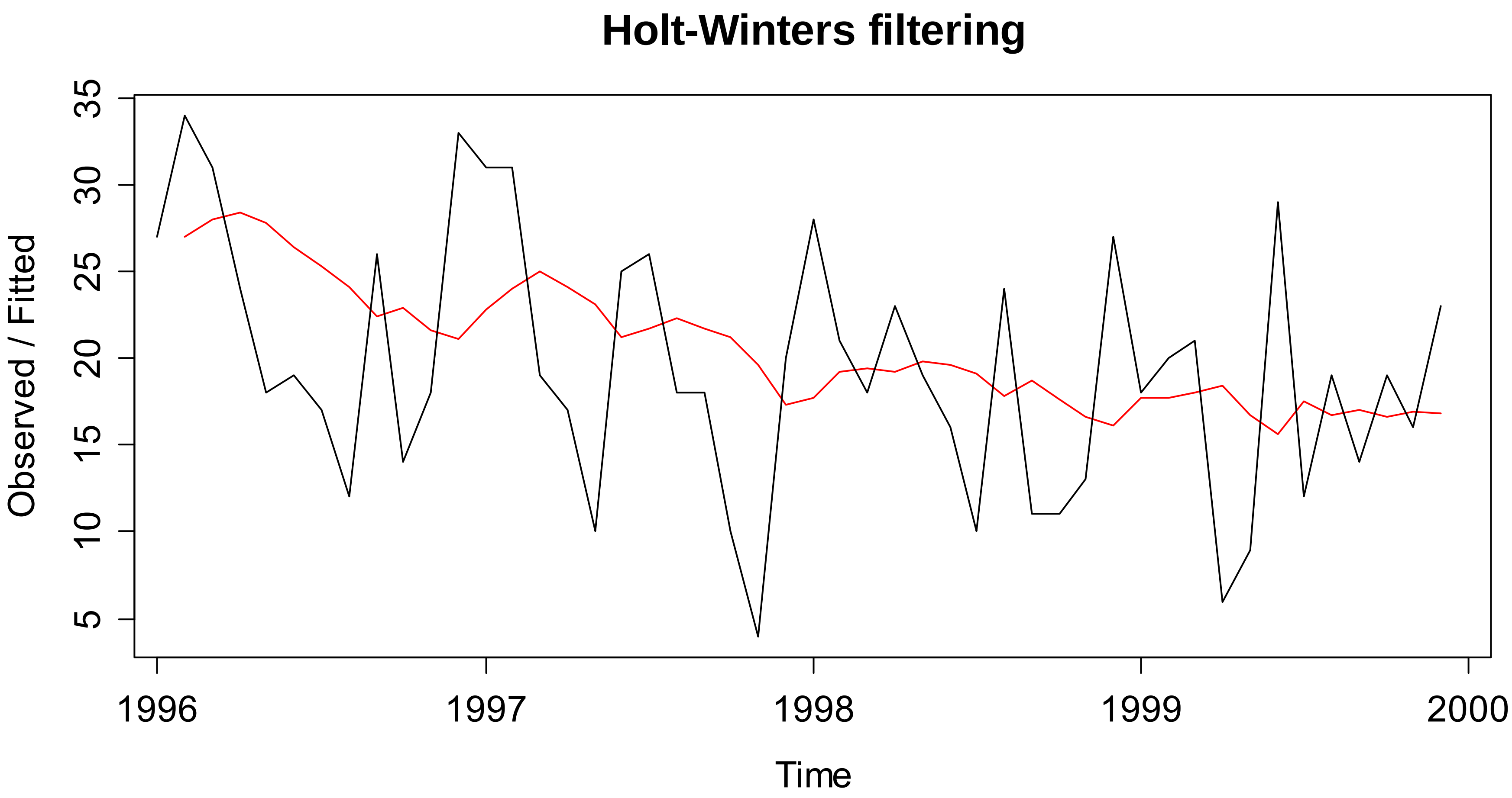


Exponential Smoothing: Example

```
> fit  <- HoltWinters(cmpl, beta=F, gamma=F)

Holt-Winters exponential smoothing without trend
and without seasonal component.

Smoothing parameters:
 alpha:  0.1429622
 beta :  FALSE
 gamma:  FALSE

Coefficients:
      [,1]
a 17.70343
```

Applied Time Series Analysis
FS 2012 – Week 10

Exponential Smoothing: Example



Applied Time Series Analysis
FS 2012 – Week 10

Holt-Winters Method

Purpose:

- is for time series with deterministic trend and/or seasonality
- is still a heuristic, model-free approach
- again based on weighted averaging

Is based on these 3 formulae:

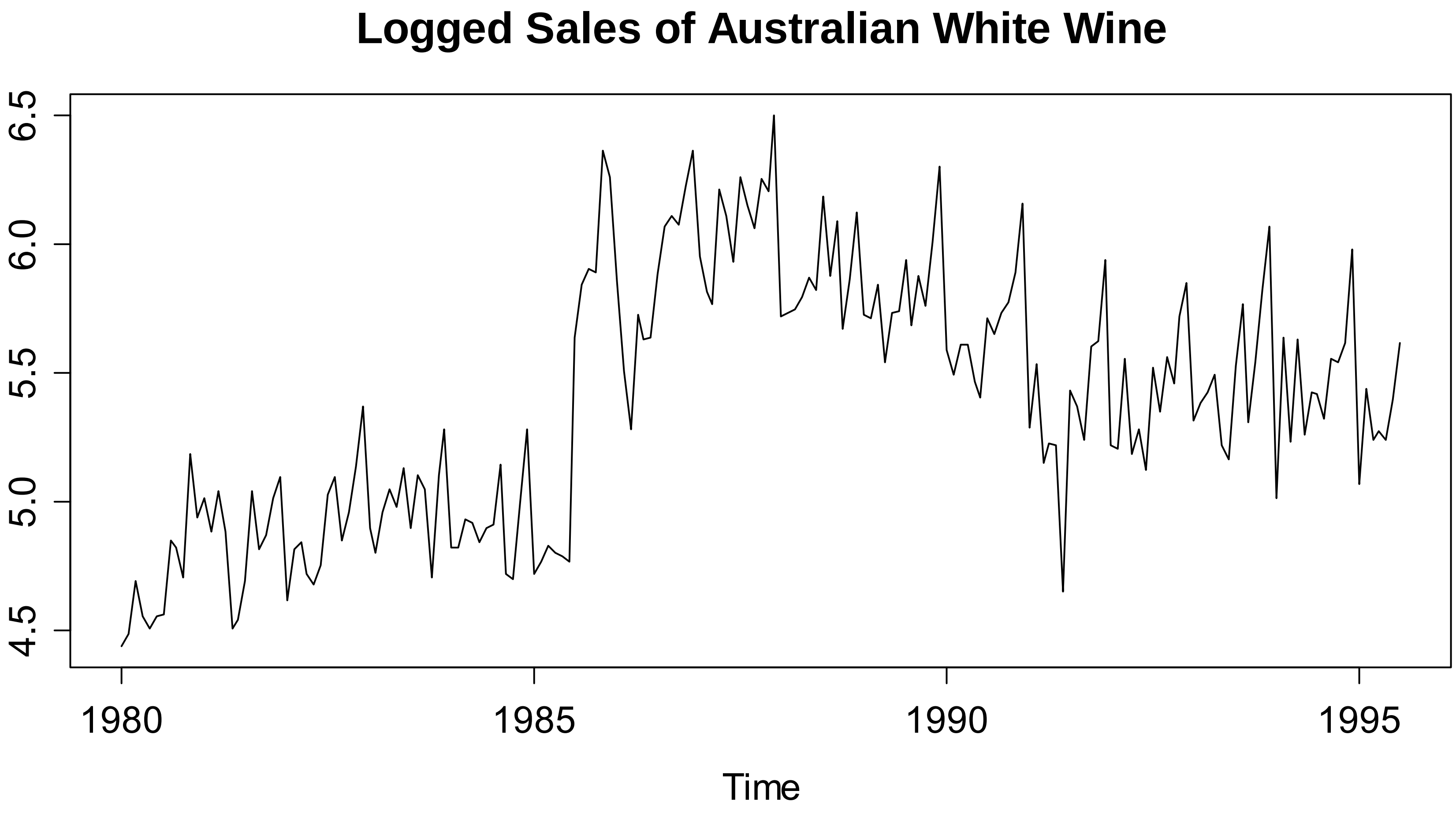
$$\begin{aligned} a_t &= \alpha(x_t - s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1}) \\ b_t &= \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(x_t - a_t) + (1-\gamma)s_{t-p} \end{aligned}$$

→ **See the blackboard for the derivation...**

Applied Time Series Analysis
FS 2012 – Week 10

Holt-Winters: Example

Holt-Winters: Example



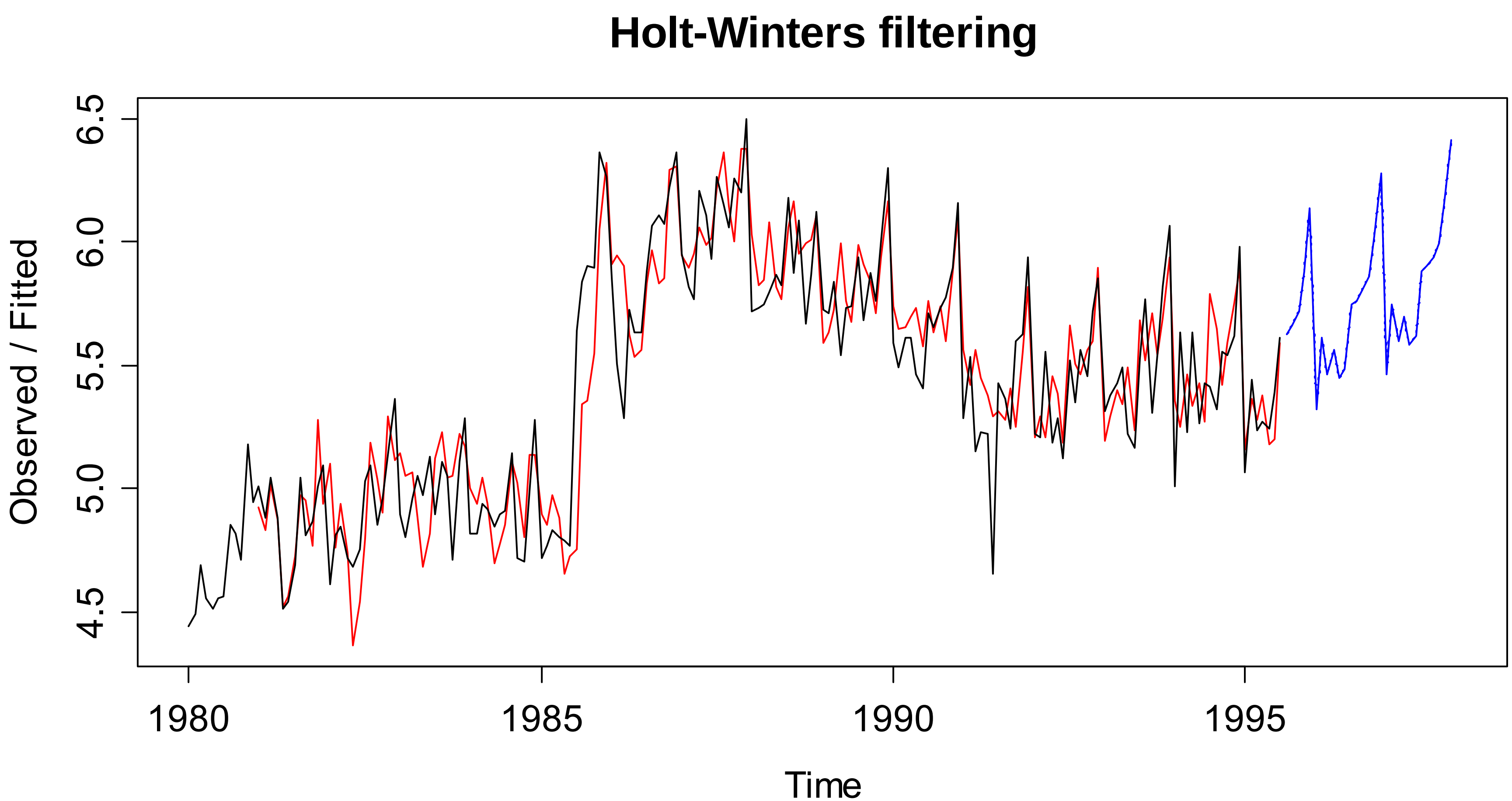
Applied Time Series Analysis
FS 2012 – Week 10

Holt-Winters: R-Code and Output

```
> HoltWinters(x = log(aww))

Holt-Winters exponential smoothing with trend and
additive seasonal component.

Smoothing parameters:
 alpha:  0.4148028; beta :  0; gamma:  0.4741967

Coefficients:
a     5.62591329; b    0.01148402
s1  -0.01230437; s2    0.01344762; s3    0.06000025
s4   0.20894897; s5    0.45515787; s6  -0.37315236
s7  -0.09709593; s8  -0.25718994; s9  -0.17107682
s10 -0.29304652; s11 -0.26986816; s12 -0.01984965
```

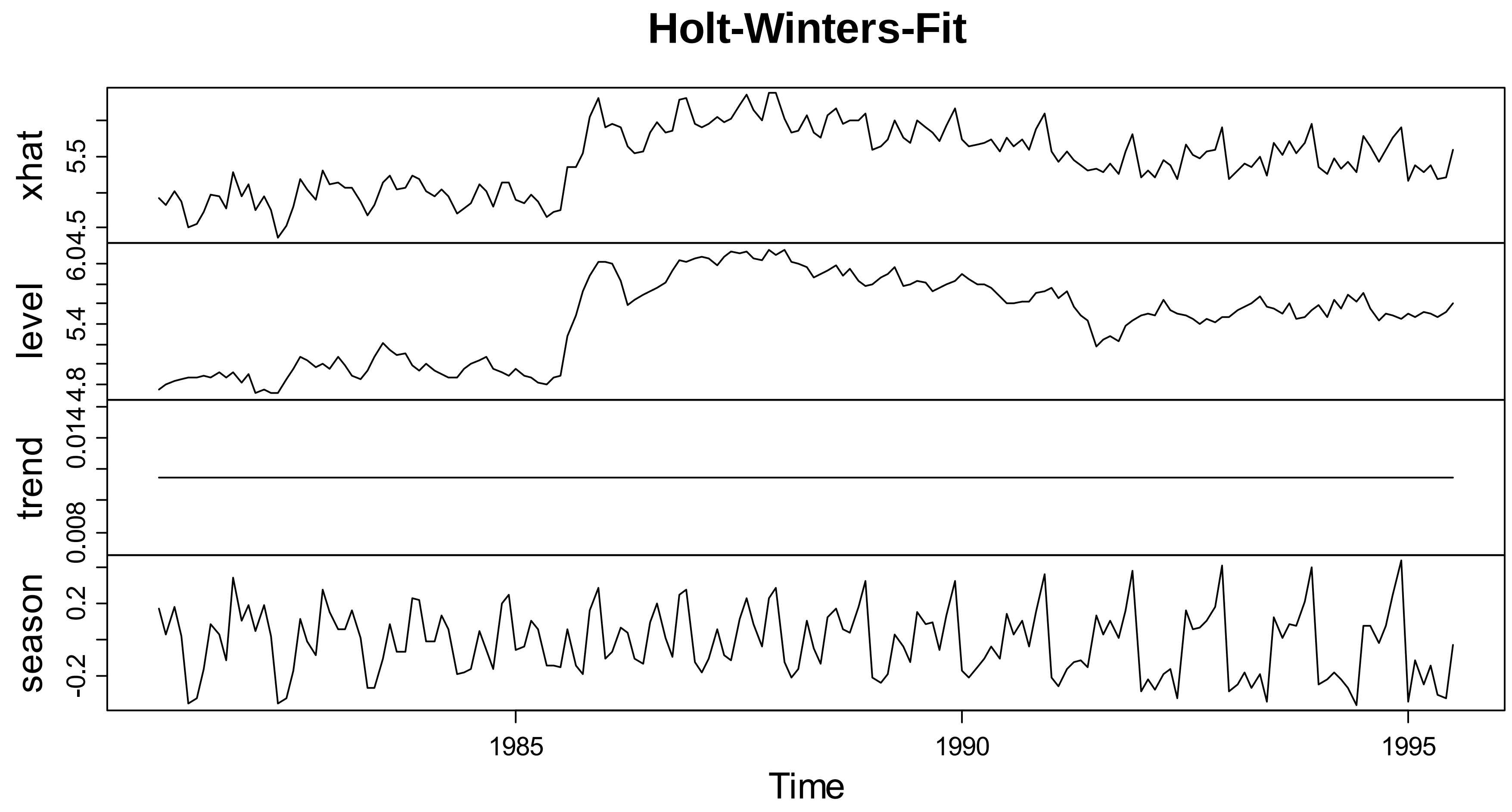
Applied Time Series Analysis
FS 2012 – Week 10

Holt-Winters: Fitted Values & Predictions



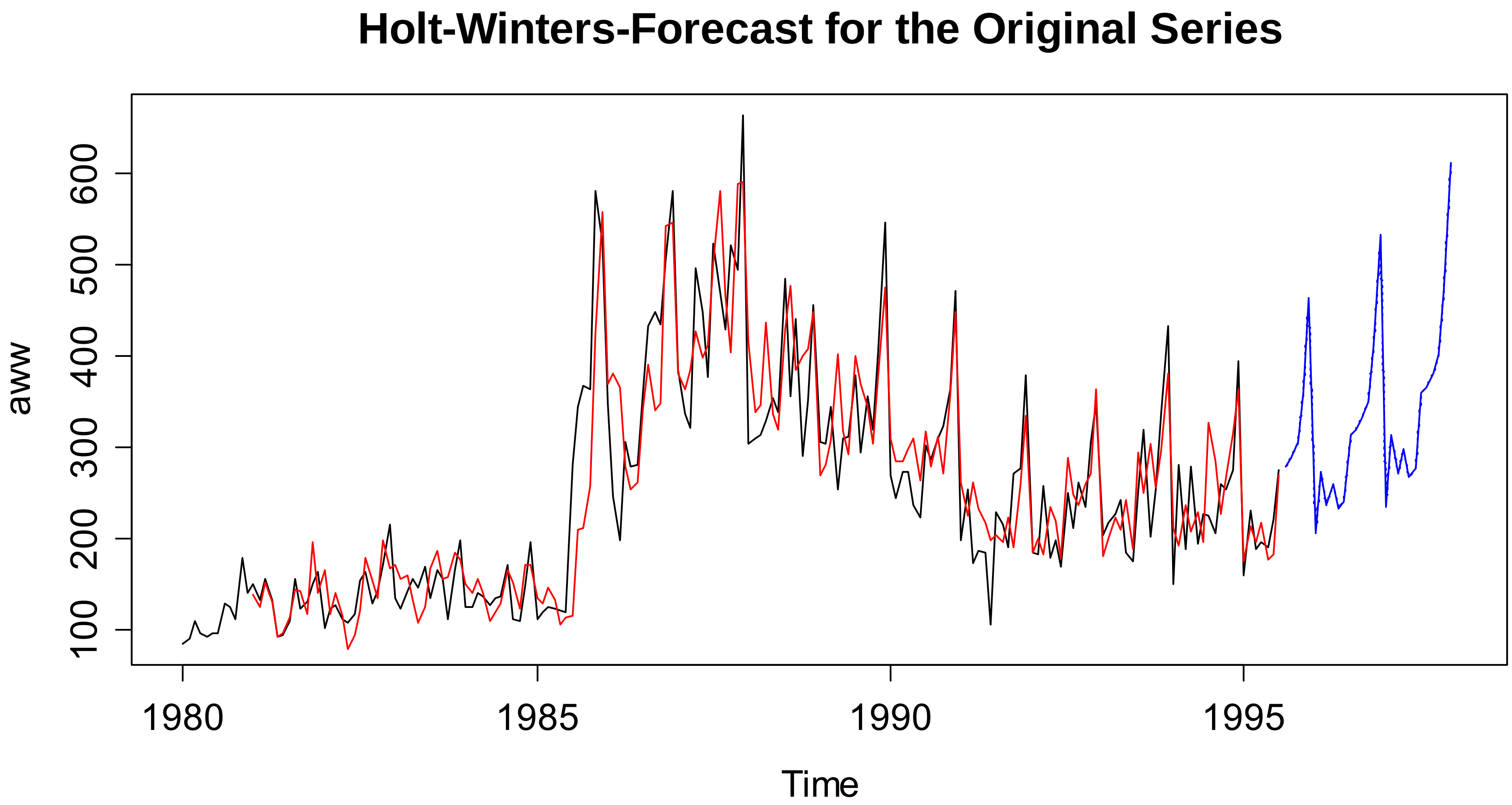
Applied Time Series Analysis
FS 2012 – Week 10

Holt-Winters: In-Sample Analysis



Applied Time Series Analysis
FS 2012 – Week 10

Holt-Winters: Predictions on Original Scale



Applied Time Series Analysis
FS 2012 – Week 10

Multivariate Time Series Analysis

Idea: Infer the relation between two time series
$X_1 = (X_{1,t})$ and $X_2 = (X_{2,t})$.

What is the difference to time series regression?

- Here, the two series arise „on an equal footing“, and we are interested in the correlation between them.
- In time series regression, the two (or more) series are causally related and we are interested in inferring that relation. There is an independent and several dependent variables.
- The difference is comparable to the difference between correlation and regression.

Gas Furnace Example

- given is a bivariate time series with 2*296 observations
- 2 measurements were made every 9 seconds
- series 1: methane input
- series 2: percentage of co2 in the furnace exhaust

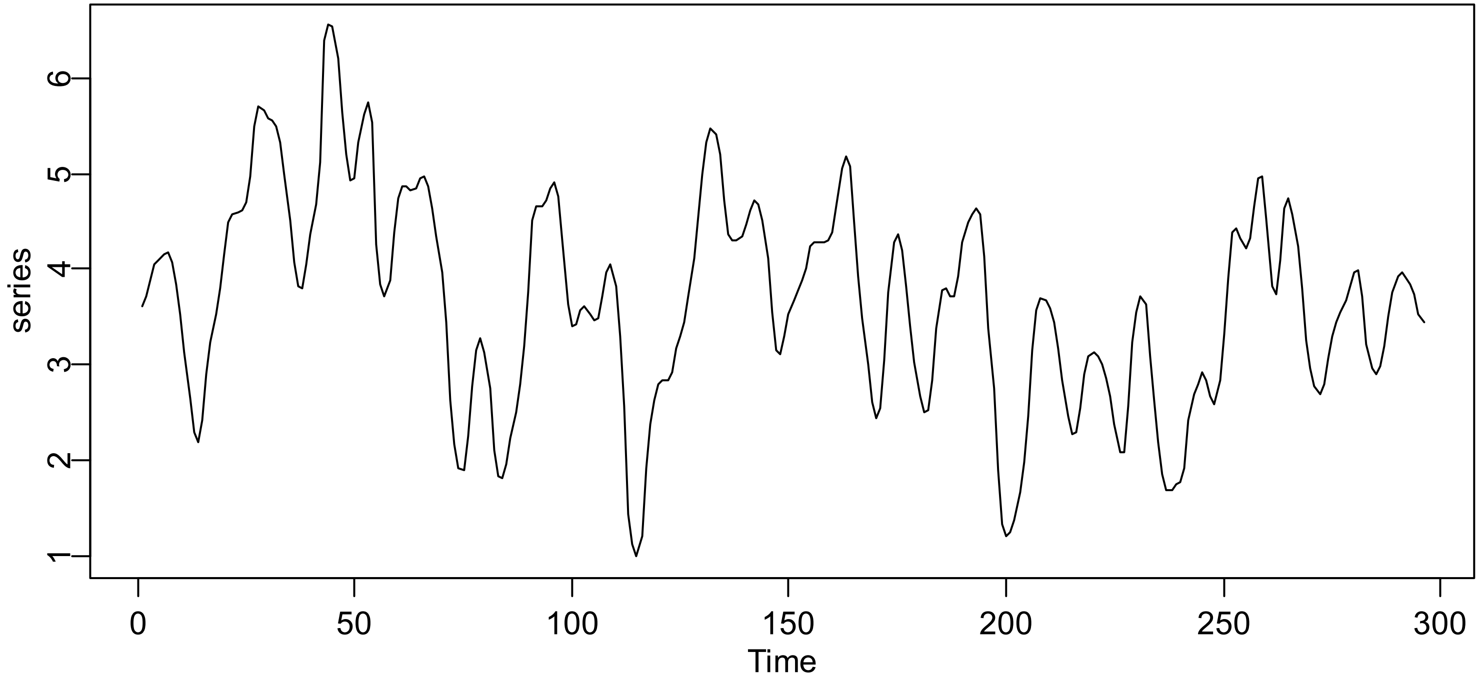
Goal of the analysis:

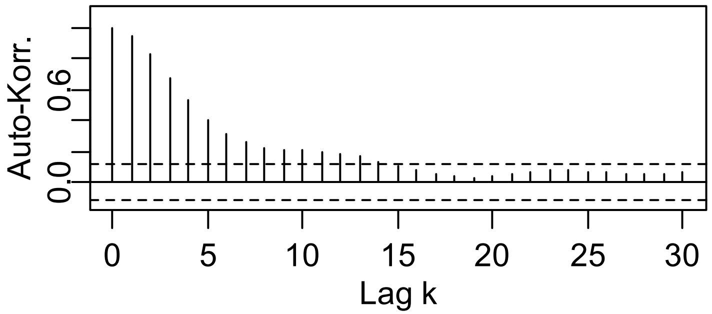
1) Answer whether changes in the co2 percentage are correlated with changes in the methane input.

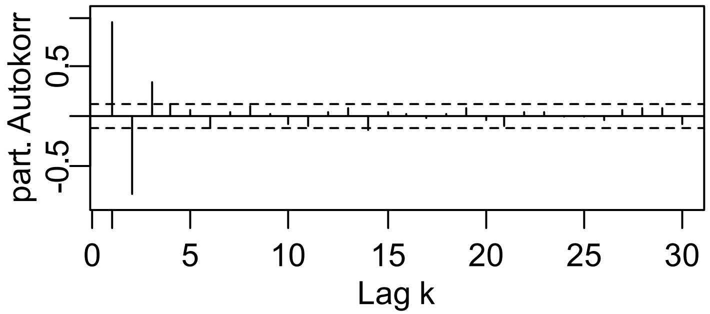
2) If a correlation is present, what is the lag?

Applied Time Series Analysis
FS 2012 – Week 10

Methane Input

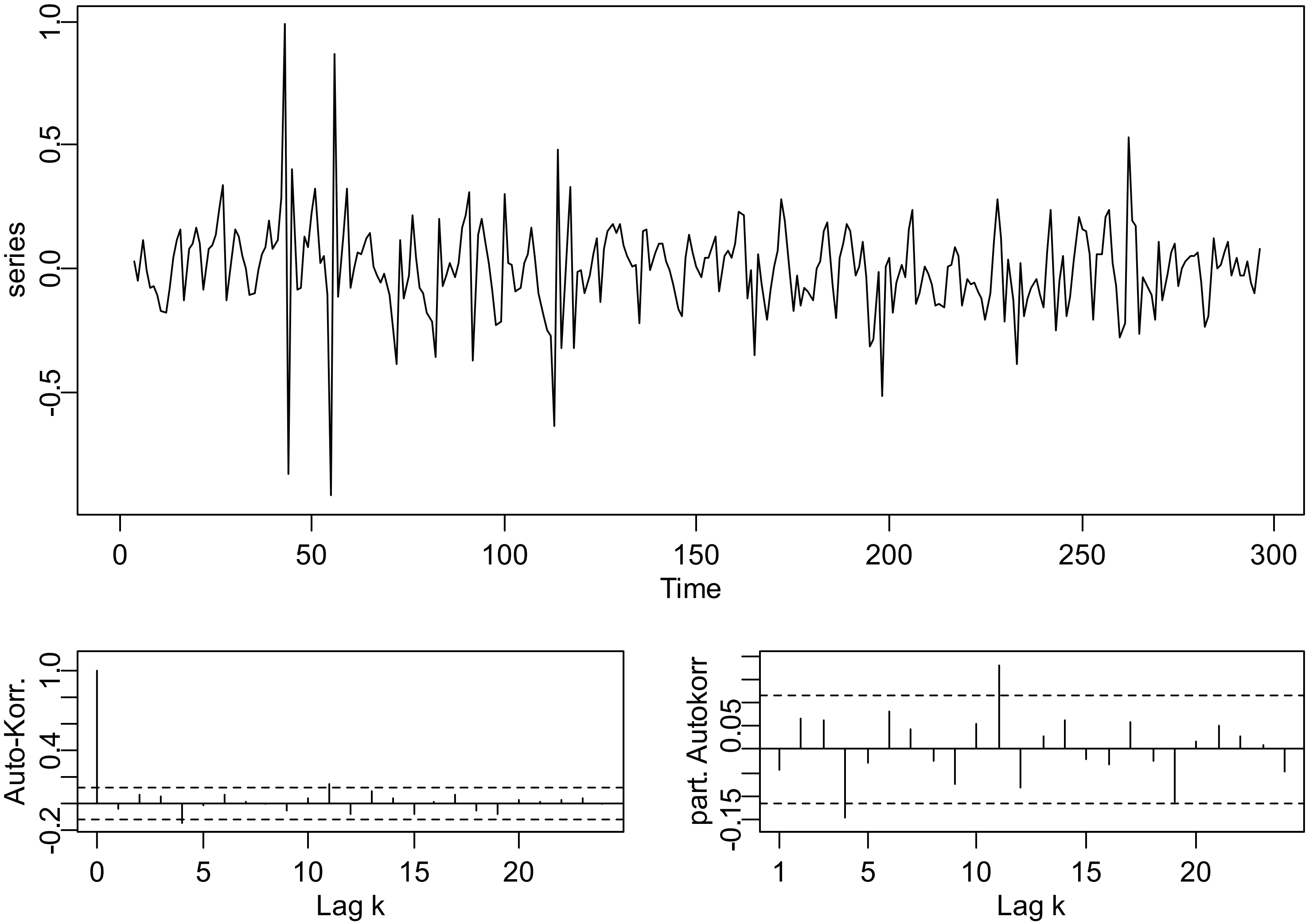






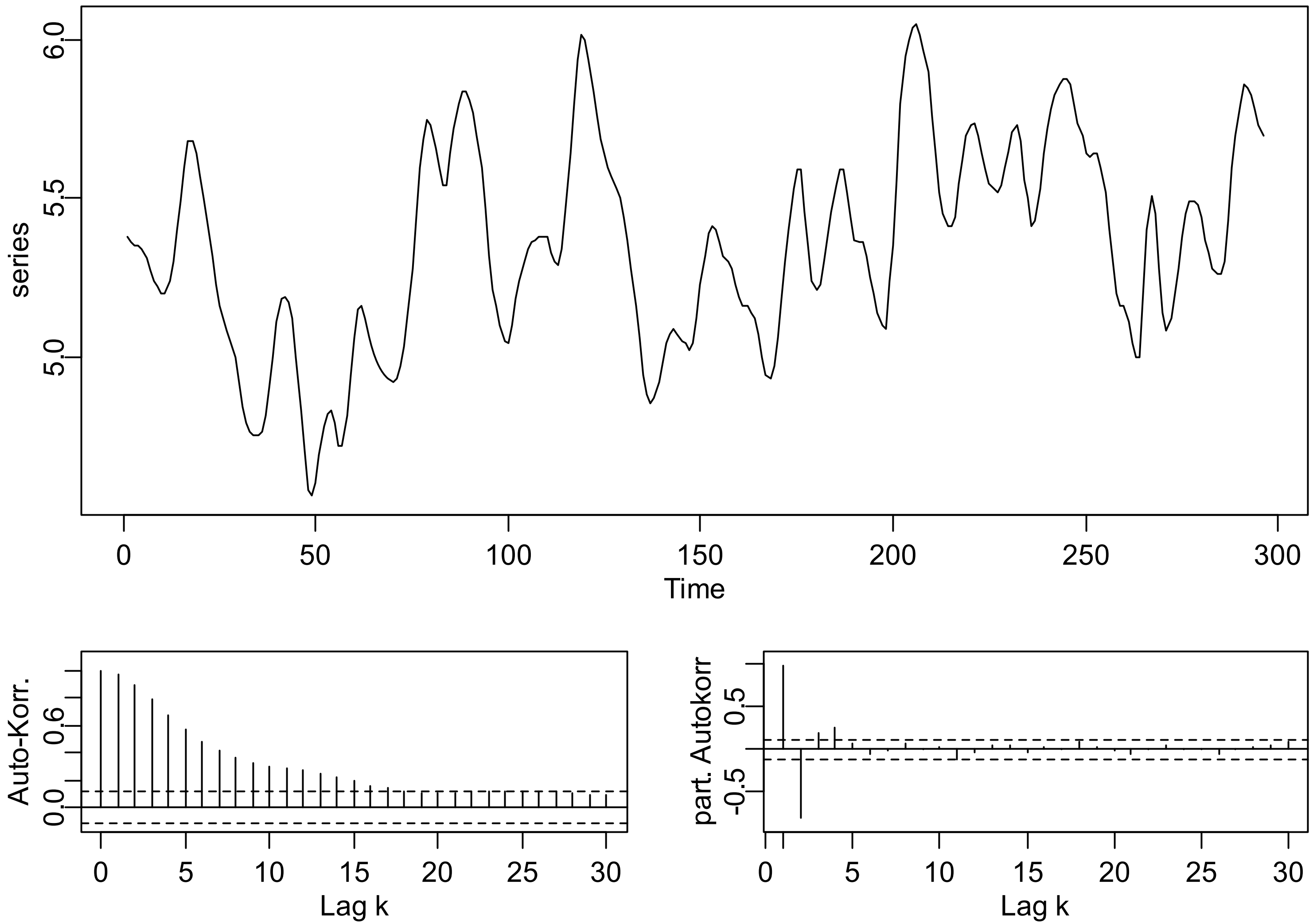
Applied Time Series Analysis
FS 2012 – Week 10

Residuals of an AR(3) on Methane Input

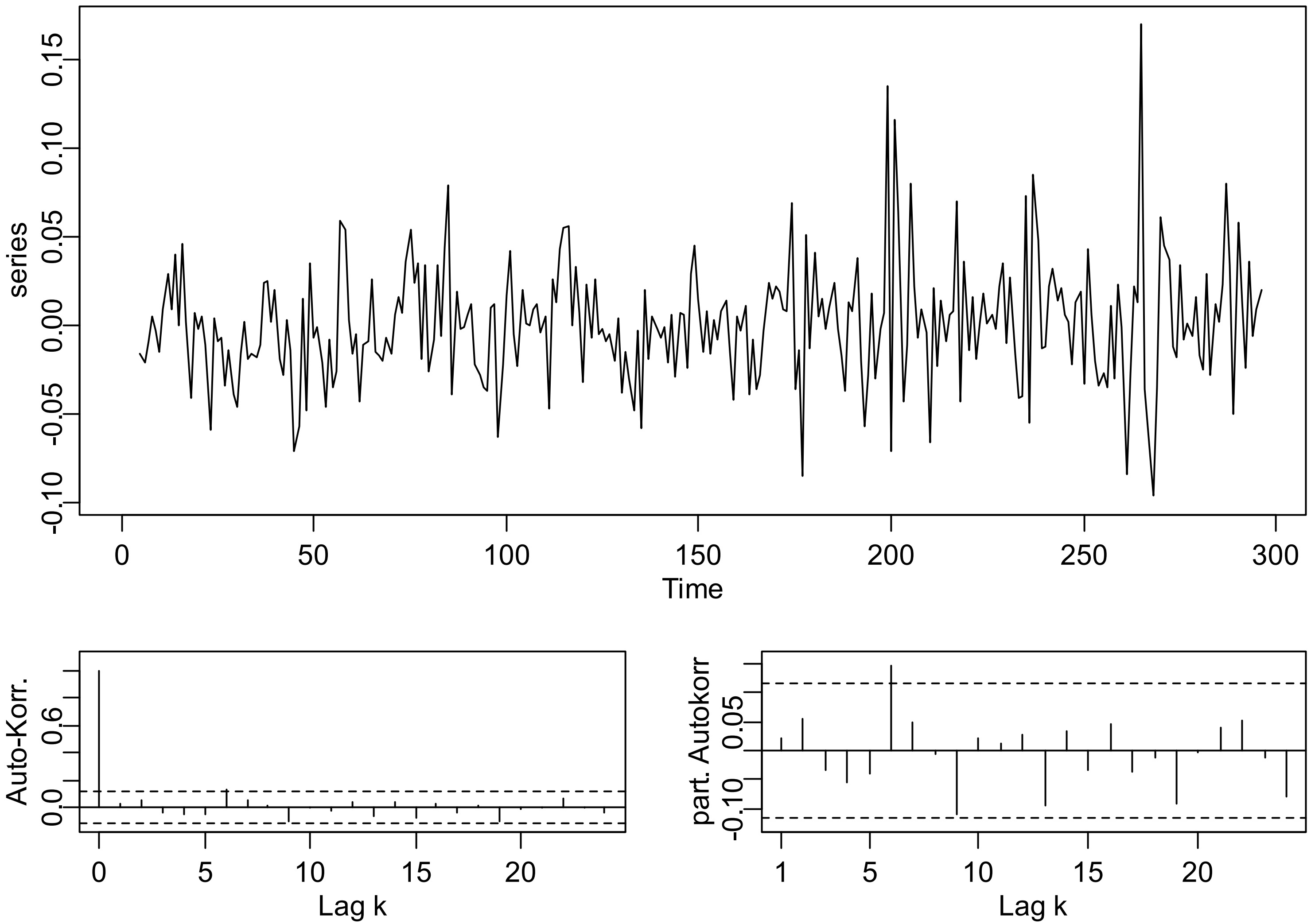


Applied Time Series Analysis
FS 2012 – Week 10

CO2 Exhaust Percentage



Residuals of an AR(4) on CO2 Exhaust



Zurich University of Applied Sciences

How to Proceed?

1) The series seem to have „long memory“

2) An AR(3)/AR(4) doesn‘t fit the data perfectly

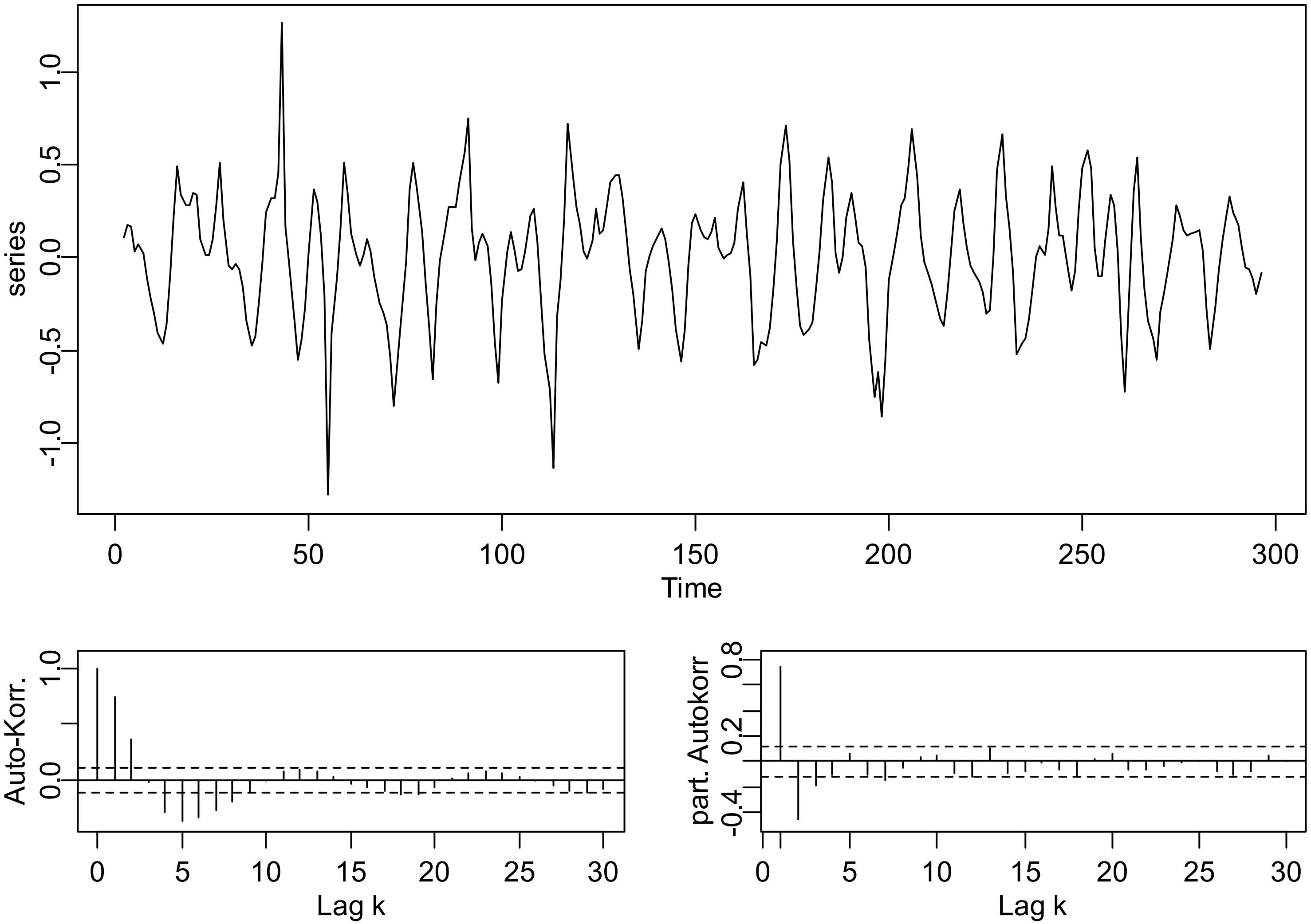
→ Differencing may help with this

Another advantage of taking differences:

→ we infer, whether there is a relation between the changes in the gas input, and the changes in the co2 exhaust.

Applied Time Series Analysis
FS 2012 – Week 10

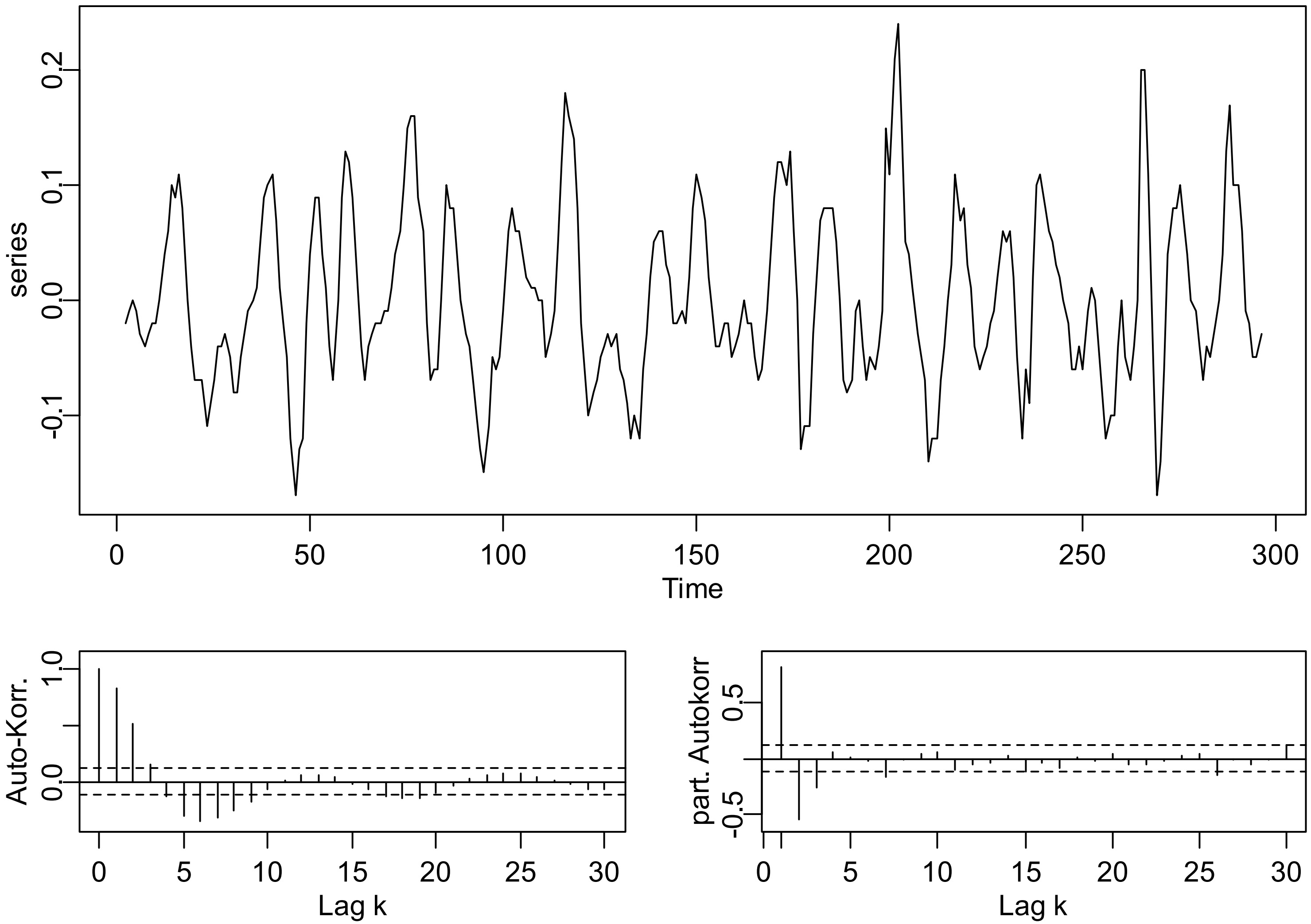
Differenced Methane Input



Zurich University of Applied Sciences

Applied Time Series Analysis
FS 2012 – Week 10

Differenced CO2 Exhaust



Cross Covariance

The cross correlations describe the relation between two time series. However, note that the interpretation is quite tricky!

$$\left.\begin{aligned}\gamma_{11}(k) &= Cov(X_{1,t+k}, X_{1,t}) \\ \gamma_{22}(k) &= Cov(X_{2,t+k}, X_{2,t})\end{aligned}\right\}$$

usual „within series“ covariance

$$\left.\begin{aligned}\gamma_{12}(k) &= Cov(X_{1,t+k}, X_{2,t}) \\ \gamma_{21}(k) &= Cov(X_{2,t+k}, X_{1,t})\end{aligned}\right\}$$

cross covariance, independent from t

Also, we have: $\gamma_{12}(-k) = Cov(X_{1,t-k}, X_{2,t}) = Cov(X_{2,t+k}, X_{1,t}) = \gamma_{21}(k)$

Cross Correlations

It suffices to analyze $\gamma_{12}(k)$, and neglect $\gamma_{21}(k)$, but we have to regard both positive and negative lags k.

We again prefer to work with correlations:

$$\rho_{12}(k) = \frac{\gamma_{12}(k)}{\sqrt{\gamma_{11}(0)\gamma_{22}(0)}}$$

which describe the linear relation between two values of X_1 and X_2, when the series X_1 is *k* time units ahead.

Applied Time Series Analysis
FS 2012 – Week 10

Estimation

Cross covariances and correlations are estimated as follows:

$$\hat{\gamma}_{12}(k) = \frac{1}{n} \sum_t (x_{1,t+k} - \overline{x}_1)(x_{2,t} - \overline{x}_2)$$

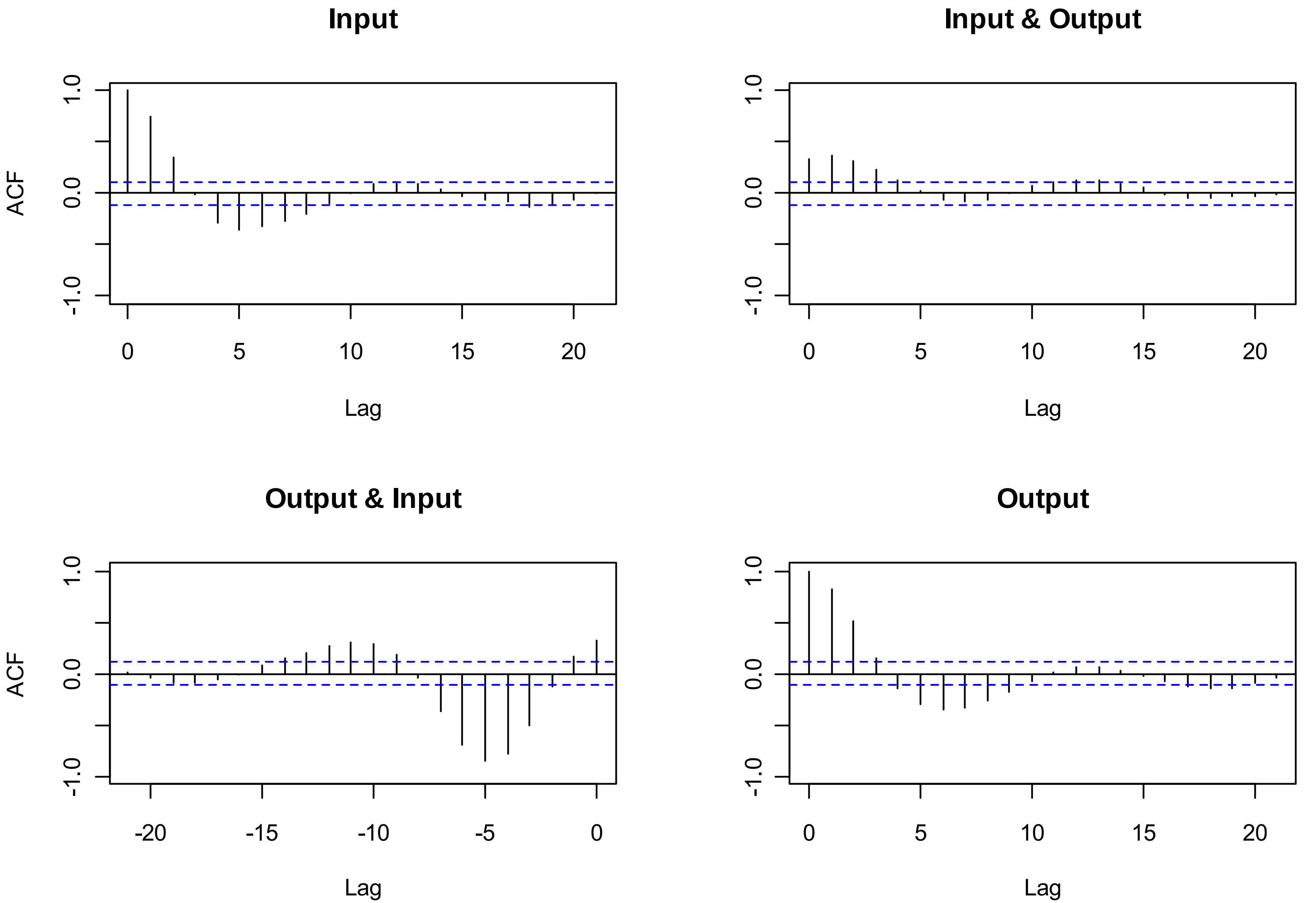
and

$$\hat{\rho}_{12}(k) = \frac{\hat{\gamma}_{12}(k)}{\sqrt{\hat{\gamma}_{11}(0)\hat{\gamma}_{22}(0)}}, \quad \text{respectively.}$$

The plot of $\hat{\rho}_{12}(k)$ versus the lag *k* is called the cross correlogram. It has to be inspected for both + and – *k*.

Applied Time Series Analysis
FS 2012 – Week 10

Sample Cross Correlation



Interpreting the Sample Cross Correlation

The confidence bounds in the sample cross correlation are only valid in some special cases, i.e. if there is no cross correlation and at least one of the series is uncorrelated.

Important: the confidence bounds are often too small!

For computing them, we need: $Var(\hat{\rho}_{12}(k))$

This is a difficult problem. We are going to discuss a few special cases and then show how the problem can be circumvented.

Applied Time Series Analysis
FS 2012 – Week 10

Special Case 1

We assume that there is no cross correlation for large lags k:

If $\rho_{12}(j)=0$ for $|j|\geq m$, we have for $|k|\geq m$:

$$Var(\hat{\rho}_{12}(k)) \approx \frac{1}{n}\sum_{j=-\infty}^{\infty}\left(\rho_{11}(j)\rho_{22}(j)+\rho_{12}(j+k)\rho_{12}(j-k)\right)$$

This goes to zero for large k and we thus have consistency. For giving statements about the confidence bounds, we would have to know more about the cross correlations, though.

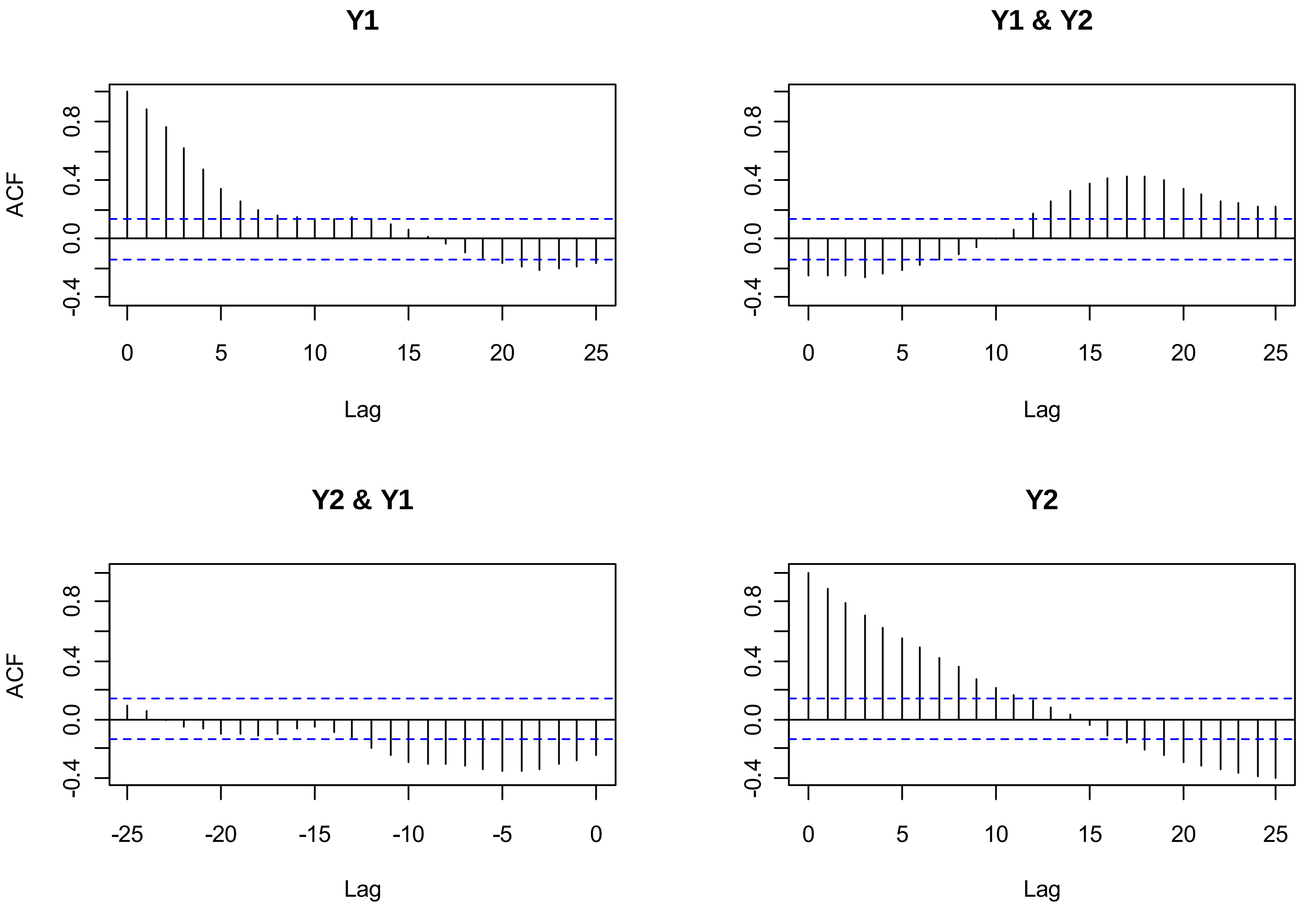
Special Case 2

There is no cross correlation, but X_1 and X_2 are time series that show correlation „within“:

$$Var(\hat{\rho}_{12}(k)) \approx \frac{1}{n} \sum_{j=-\infty}^{\infty} \rho_{11}(j)\rho_{22}(j)$$

See the blackboard… for the important example showing that the cross correlation estimations can be arbitrarily bad!

Special Case 2: Simulation Example



Special Case 3

There is no cross correlation, and X_1 is a white noise series that is independent from X_2. Then, the estimation variance simplifies to:

$$Var(\hat{\rho}_{12}(k)) \approx \frac{1}{n}$$

Thus, the confidence bounds are valid in this case.

However, we introduced the concept of cross correlation to infer the relation between correlated series. The trick of the so-called „prewhitening“ helps.

Prewhitening

Prewhitening means that the time series is transformed such that it becomes a white noise process, i.e. is uncorrelated.

We assume that both stationary processes X_1 and X_2 can be rewritten as follows:

$$U_t = \sum_{i=0}^{\infty} a_i X_{1,t-i} \quad \text{and} \quad V_t = \sum_{i=0}^{\infty} b_i X_{2,t-i},$$

with uncorrelated U_t and V_t. Note that this is possible for ARMA(p,q) processes by writing them as an AR(∞). The left hand side of the equation then is the innovation.

Cross Correlation of Prewhitened Series

The cross correlation between U_t and V_t can be derived from the one between $X_{1,t}$ and $X_{2,t}$:

$$\rho_{UV}(k) = \sum_{j=0}^{\infty}\sum_{j=0}^{\infty} a_i b_i \rho_{X_1 X_2}(k+i-j)$$

Thus we have:

$$\rho_{UV}(k) = 0 \quad \text{for all k} \quad \Leftrightarrow \quad \rho_{X_1 X_2}(k) = 0 \quad \text{for all k}$$

Now: generate U,V, estimate cross correlations and, by using the confidence bands, check whether they are signficant

Simulation Example

Since we are dealing with simulated series, we know that:

$$X_{i,t} = 0.9 \cdot X_{i,t-1} + E_t \text{ , thus } \quad E_t = X_{i,t} - 0.9 \cdot X_{i,t-1}$$

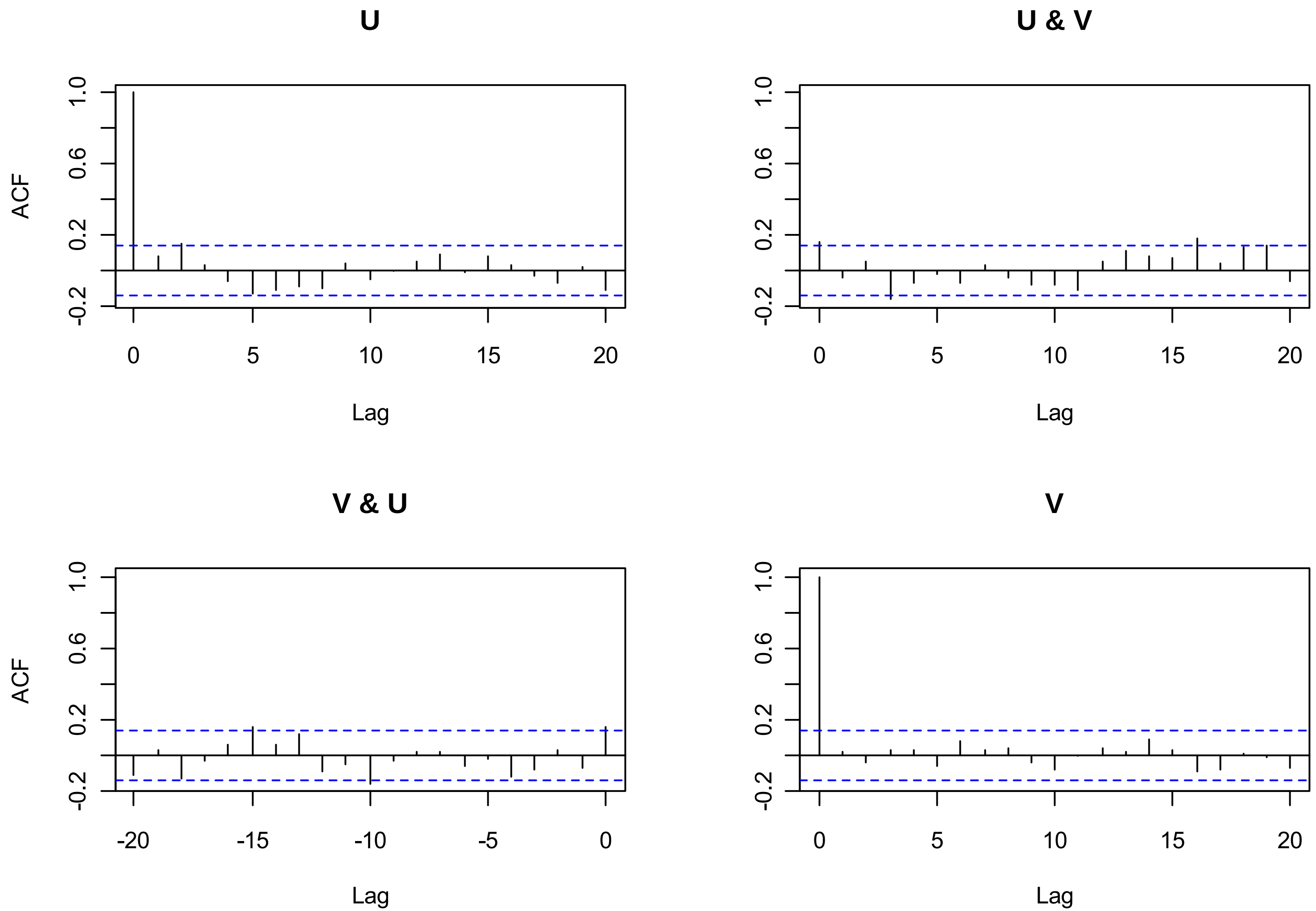
In practice, we don‘t know the AR-coefficients, but plug-in the respective estimates:

$$U_t = X_{1,t} - \hat{\alpha}_{1,1} X_{1,t-1} \quad \text{with} \quad \hat{\alpha}_{1,1} = 0.911$$

$$V_t = X_{2,t} - \hat{\alpha}_{2,1} X_{2,t-1} \quad \text{with} \quad \hat{\alpha}_{2,1} = 0.822$$

We will now analyse the sample cross correlation of U_t and V_t, which will also allow to draw conclusions about X_1 and X_2.

Cross Correlation in the Simulation Example



Cross Correlation in the Simulation Example

We observe that:

- U_t and V_t are white noise processes
- There are no (strongly) significant cross correlations

We conjecture that:

- X_1 and X_2 are not cross correlated either.

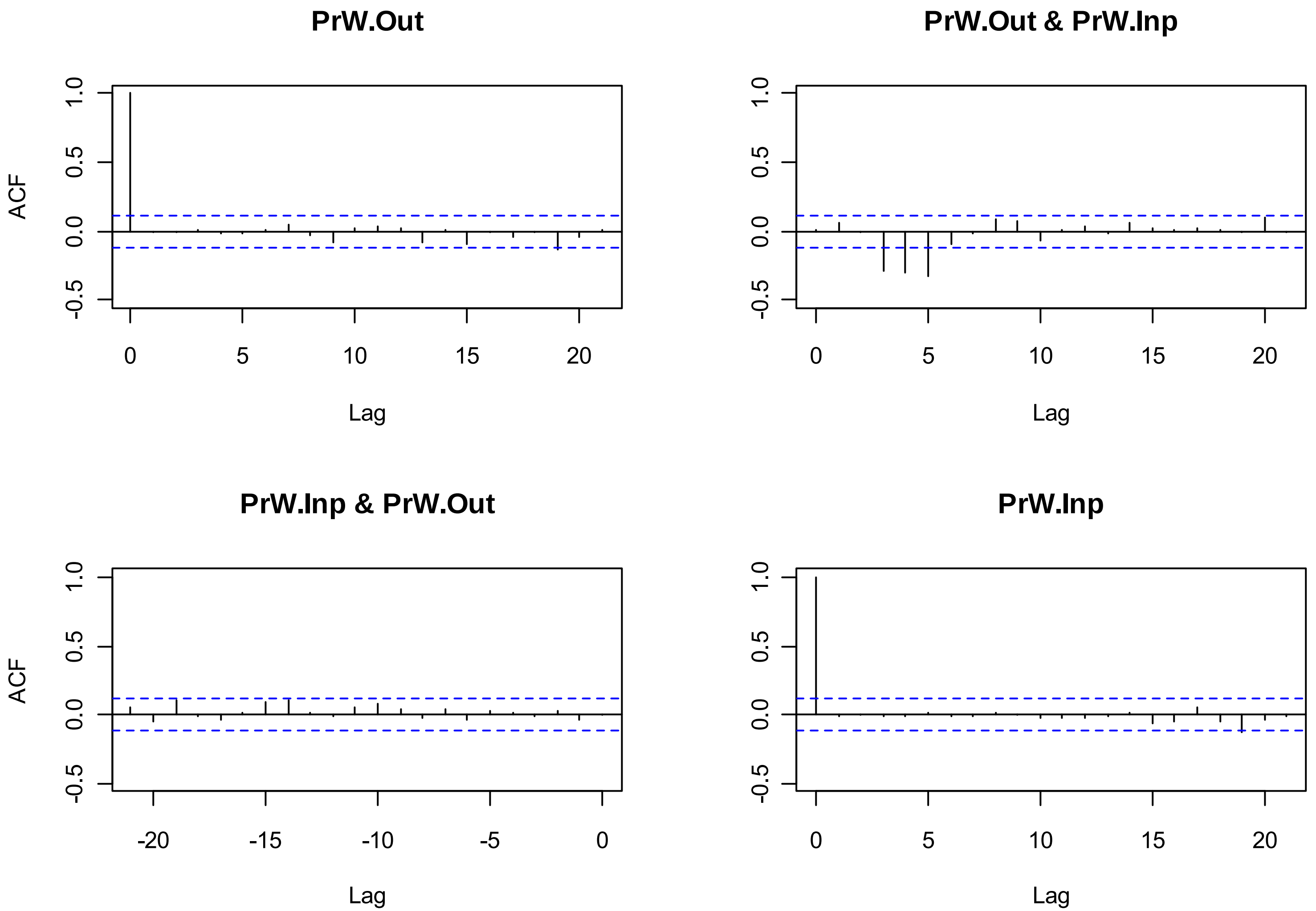
→ This matches our „expectations“, or better, true process.

Prewhitening the Gas Furnace Data

What to do:

- AR(p)-models are fitted to the differenced series
- The residual time series are U_t and V_t, white noise
- Check the sample cross correlation (see next slide)
- Model the output as a linear combination of past input values: transfer function model.

Prewhitening the Gas Furnace Data



Applied Time Series Analysis
FS 2012 – Week 10

Transfer Function Models

Properties:

- Transfer function models are an option to describe the dependency between two time series.

- The first (input) series influences the second (output) one, but there is no feedback from output to input.

- The influence from input to output only goes „forward".

The model is:

$$X_{2,t} - \mu_2 = \sum_{j=0}^{\infty} \nu_j (X_{1,t-j} - \mu_1) + E_t$$

Applied Time Series Analysis
FS 2012 – Week 10

Transfer Function Models

The model is:

$$X_{2,t} - \mu_2 = \sum_{j=0}^{\infty} \nu_j (X_{1,t-j} - \mu_1) + E_t$$

- $E[E_t]=0$.
- E_t and $X_{1,s}$ are uncorrelated for all t and s.
- E_t and E_s are usually correlated.
- For simplicity of notation, we here assume that the series have been mean centered.

Applied Time Series Analysis
FS 2012 – Week 10

Cross Covariance

When plugging-in, we obtain for the cross covariance:

$$\gamma_{21}(k) = Cov(X_{2,t+k}, X_{1,t}) = Cov\left(\sum_{j=0}^{\infty} \nu_j X_{1,t+k-j}, X_{1,t}\right) = \sum_{j=0}^{\infty} \nu_j \gamma_{11}(k-j)$$

- If only finitely many coefficients are different from zero, we could generate a linear equation system, plug-in $\hat{\gamma}_{11}$ and $\hat{\gamma}_{21}$ to obtain the estimates $\hat{\nu}_j$.

→ **This is not a statistically efficient estimation method.**

Special Case: $X_{1,t}$ Uncorrelated

If $X_{1,t}$ was an uncorrelated series, we would obtain the coefficients of the transfer function model quite easily:

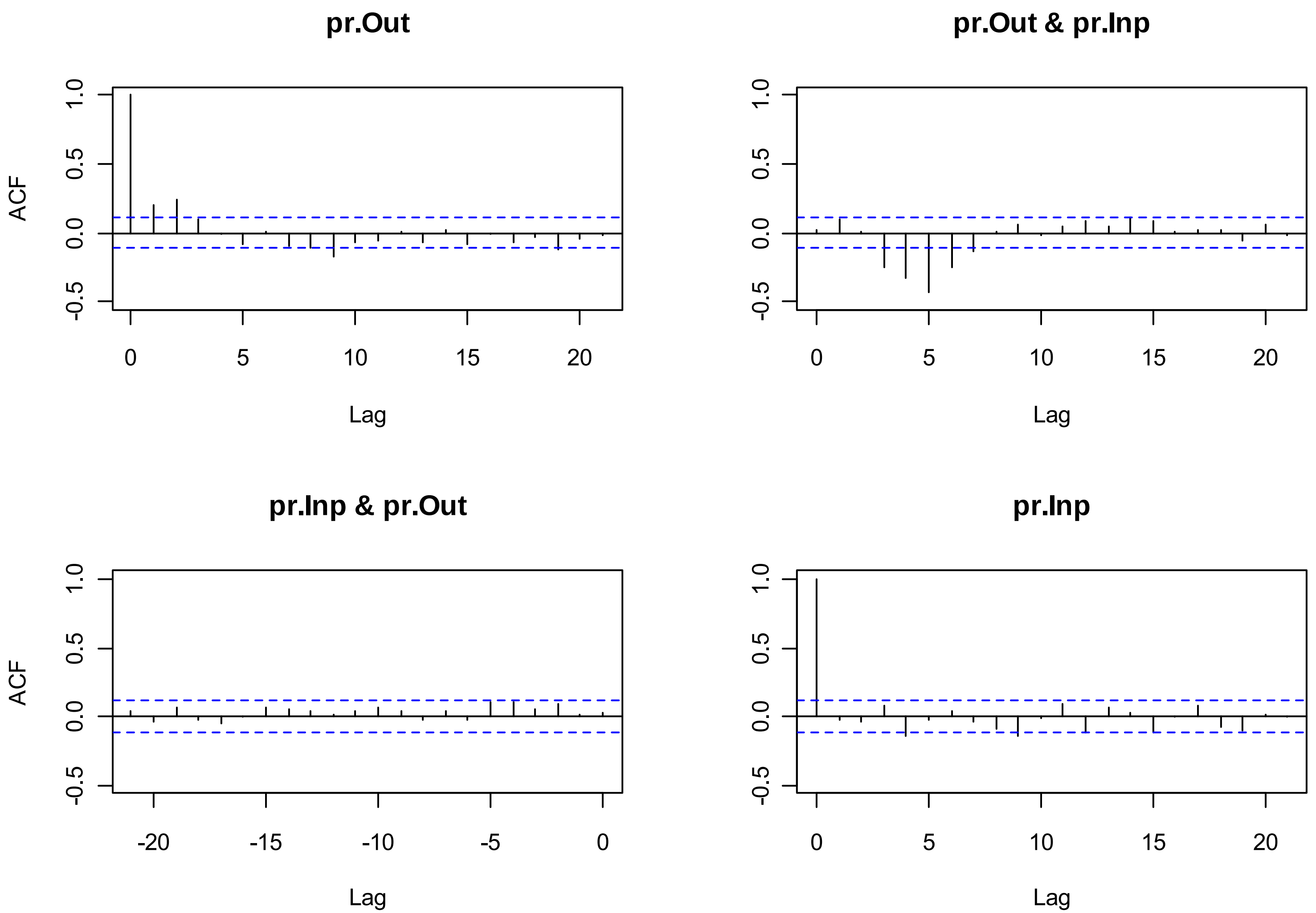
$$\nu_k = \frac{\gamma_{21}(k)}{\gamma_{11}(0)}$$

However, this is usually not the case. We can then:

- transform all series in a clever way
- the transfer function model has identical coefficients
- the new, transformed input series is uncorrelated

→ see blackboard for the transformation

Gas Furnace Transformed



Gas Furnace: Final Remarks

- In the previous slide, we see the empirical cross correlations $\hat{\rho}_{21}(k)$ of the two series D_t and Z_t.
- The coefficients $\hat{v}_k$ from the transfer function model will be proportional to the empirical cross correlations. We can already now conjecture that the output is delayed by 3-7 times, i.e. 27-63 seconds.
- The formula for the transfer function model coefficients is:

$$\hat{v}_k = \frac{\hat{\sigma}_Z}{\hat{\sigma}_D} \hat{\rho}_{21}(k)$$