Applied Time Series Analysis
FS 2012 – 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 5, 2012

Descriptive Decomposition

It is convenient to describe non-stationary time series with a simple decomposition model

$$X_t = m_t + s_t + E_t$$

= trend + seasonal effect + stationary remainder

The modelling can be done with:

1) ***taking differences with appropriate lag (=differencing)***
2) ***smoothing approaches (= filtering)***
3) ***parametric models (= curve fitting)***

Applied Time Series Analysis
FS 2012 – Week 03

Parametric Modelling

When to use?

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

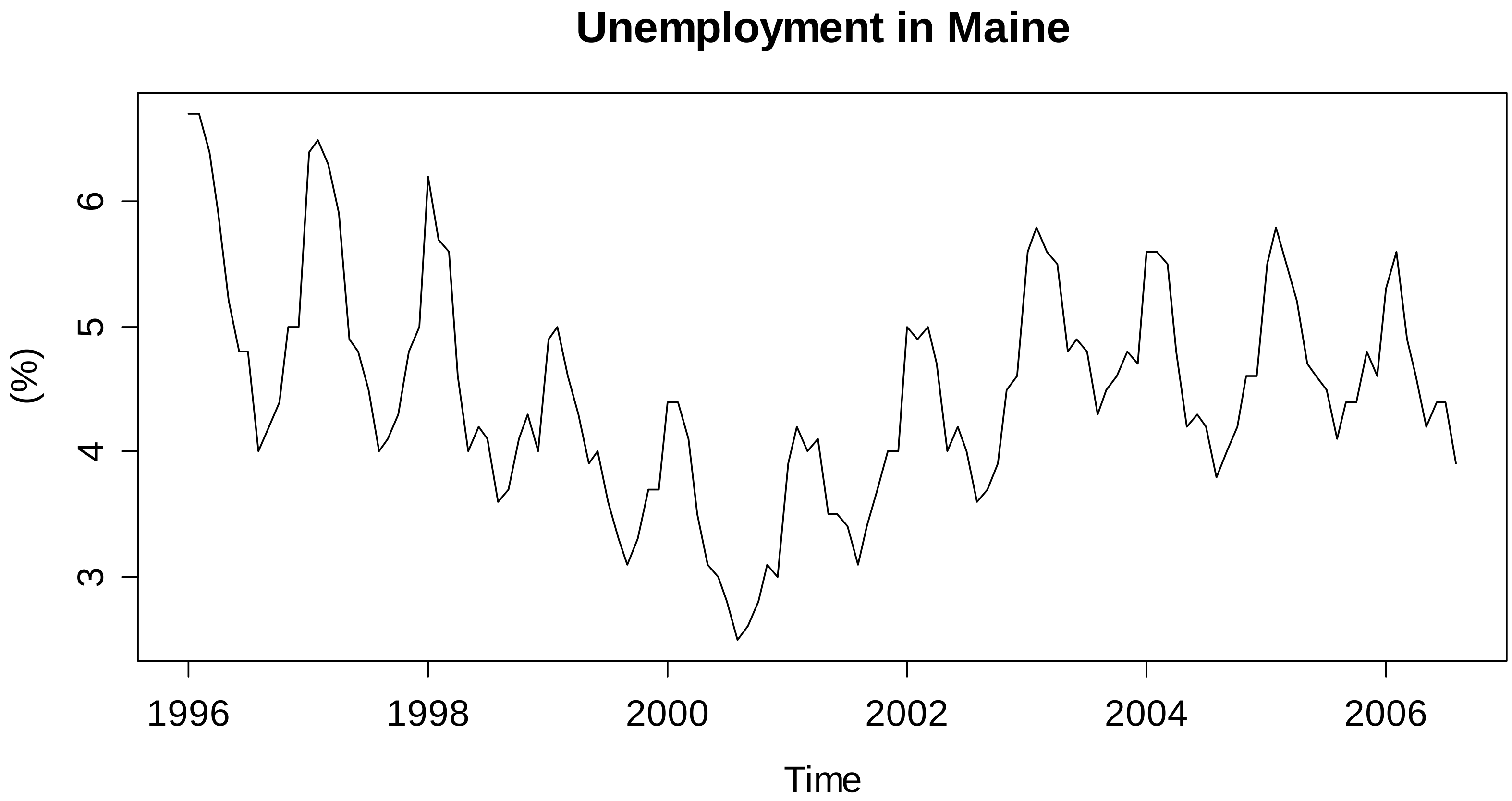
→ If the main goal of the analysis is forecasting, a trend in functional form may allow for easier extrapolation than a trend obtained via smoothing.

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

Applied Time Series Analysis
FS 2012 – Week 03

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?

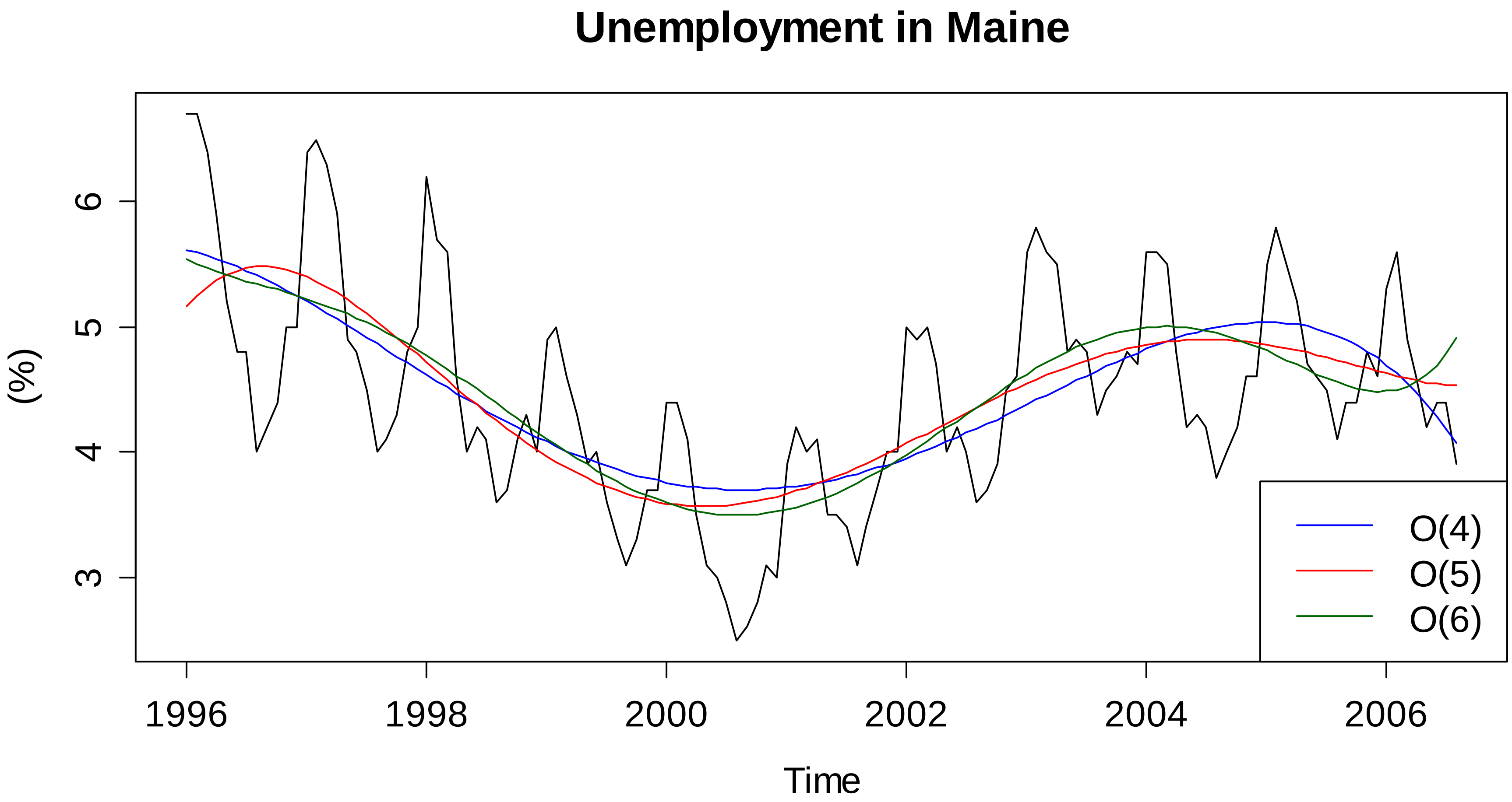
Eyeballing allows to determine the minimum grade that is required for the polynomial. It is at least the number of maxima the hypothesized trend has, plus one.

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!

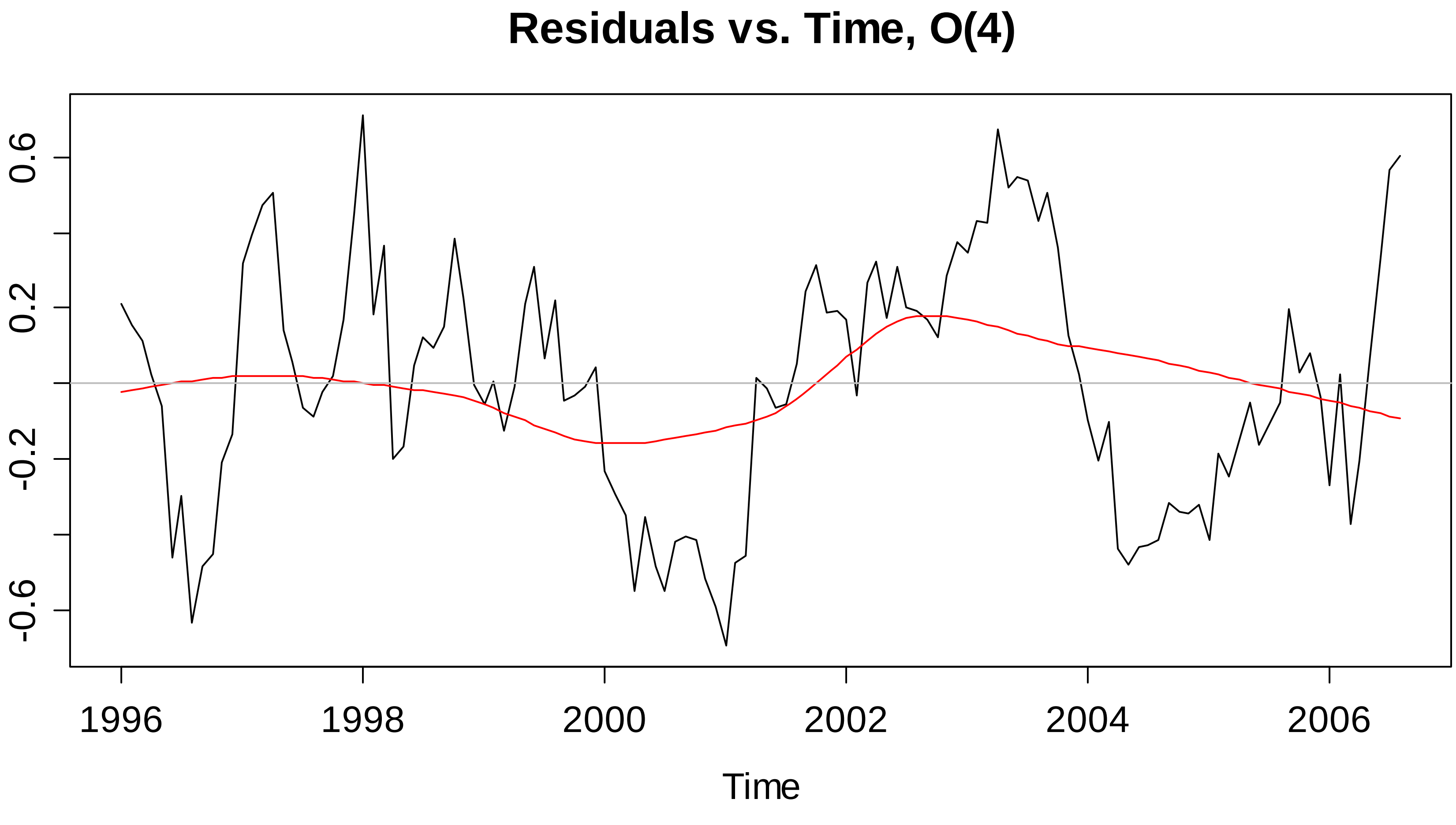
Applied Time Series Analysis
FS 2012 – Week 03

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

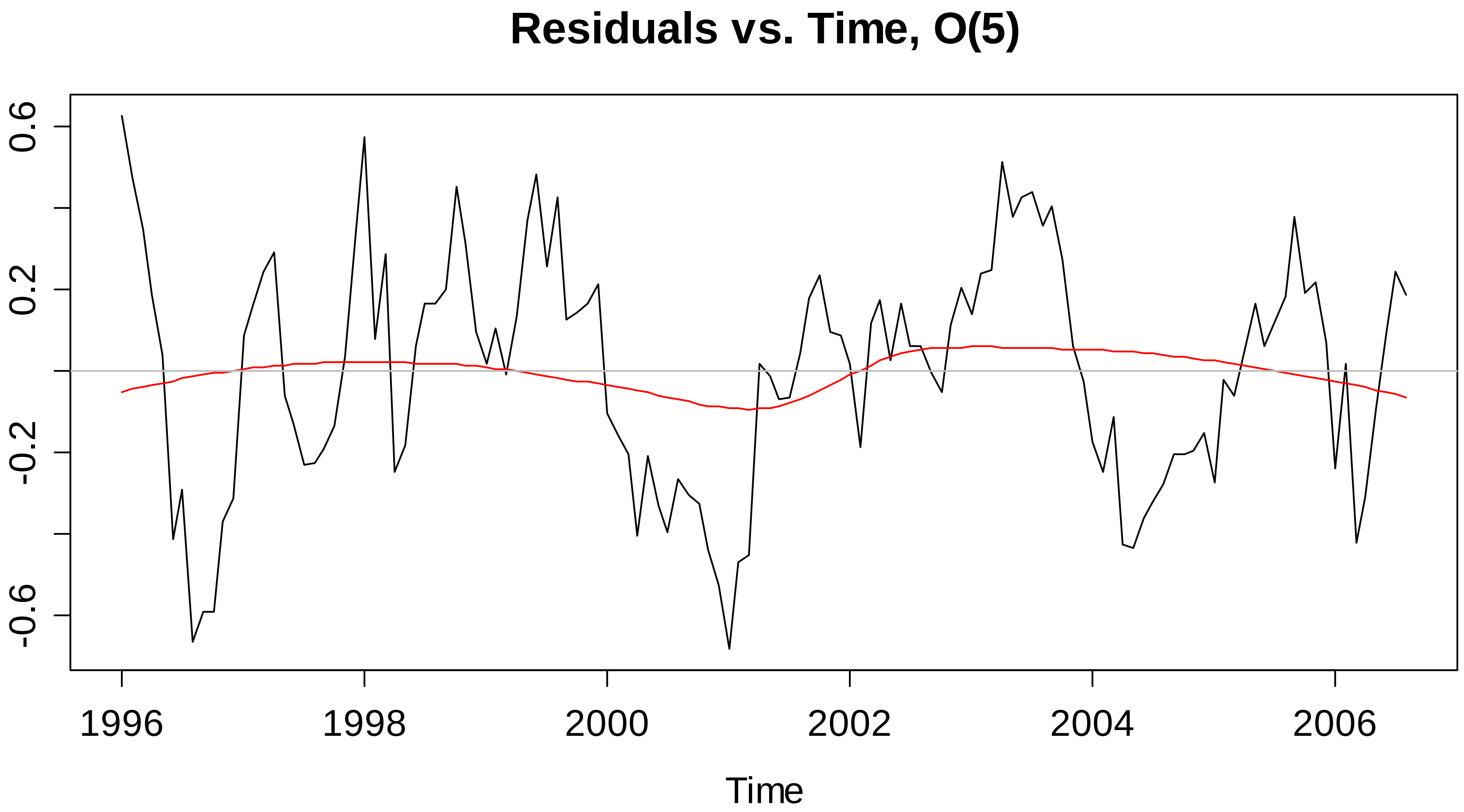


Applied Time Series Analysis
FS 2012 – Week 03

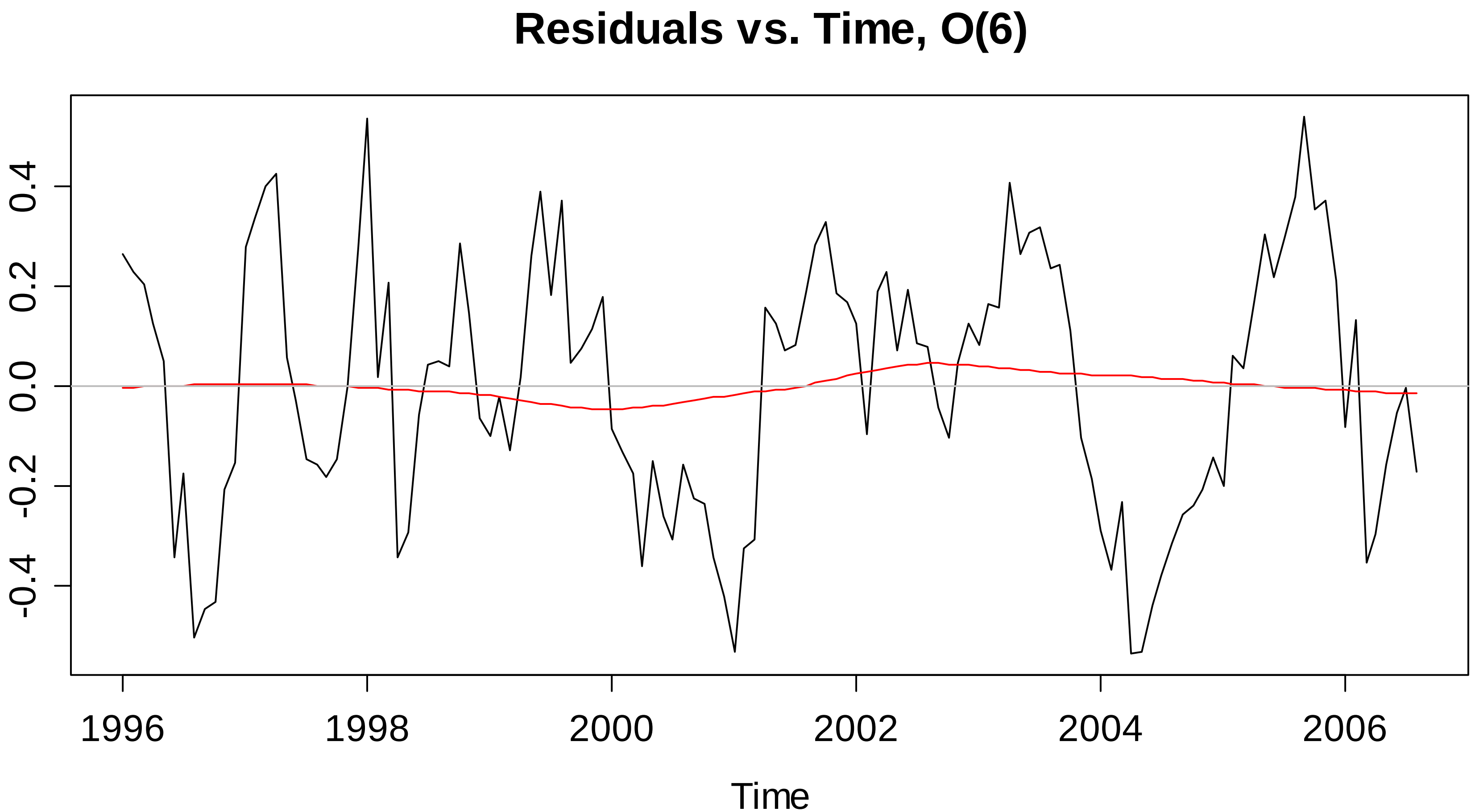
Residual Analysis: O(4)



Residual Analysis: O(5)



Residual Analysis: O(6)



Applied Time Series Analysis
FS 2012 – Week 03

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
FS 2012 – 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.

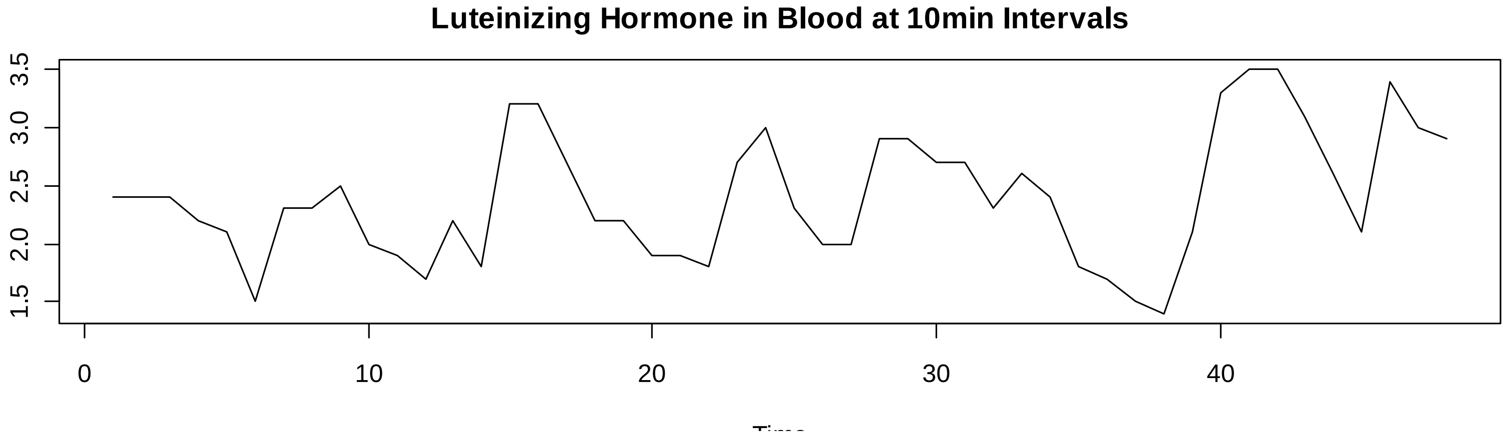
Def: **Autocorrelation**

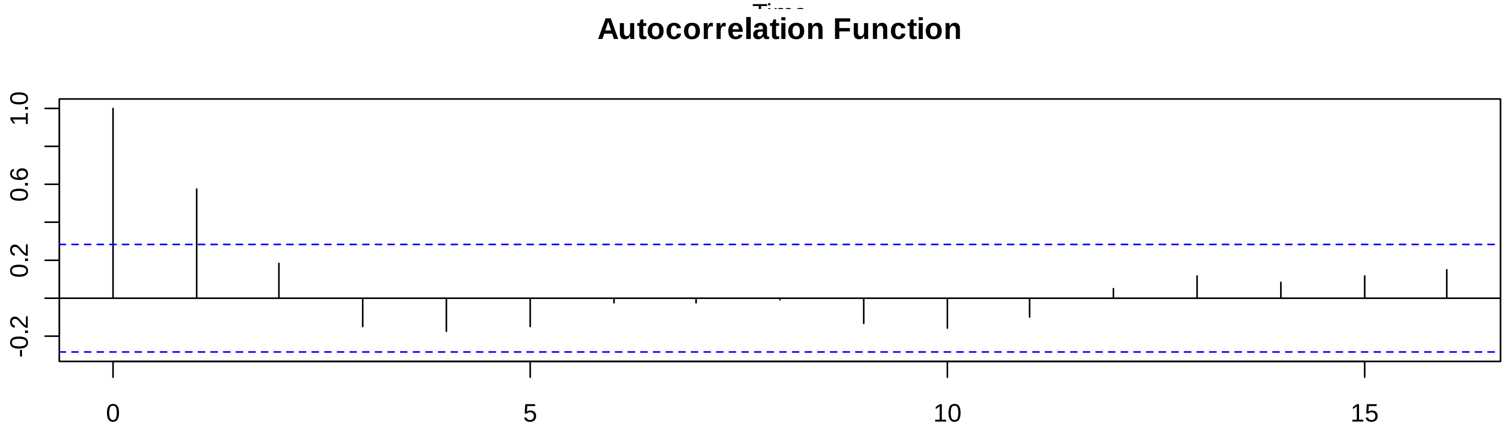
$$Cor(X_{t+k}, X_t) = \frac{Cov(X_{t+k}, X_t)}{\sqrt{Var(X_{t+k}) \cdot Var(X_t)}}$$

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

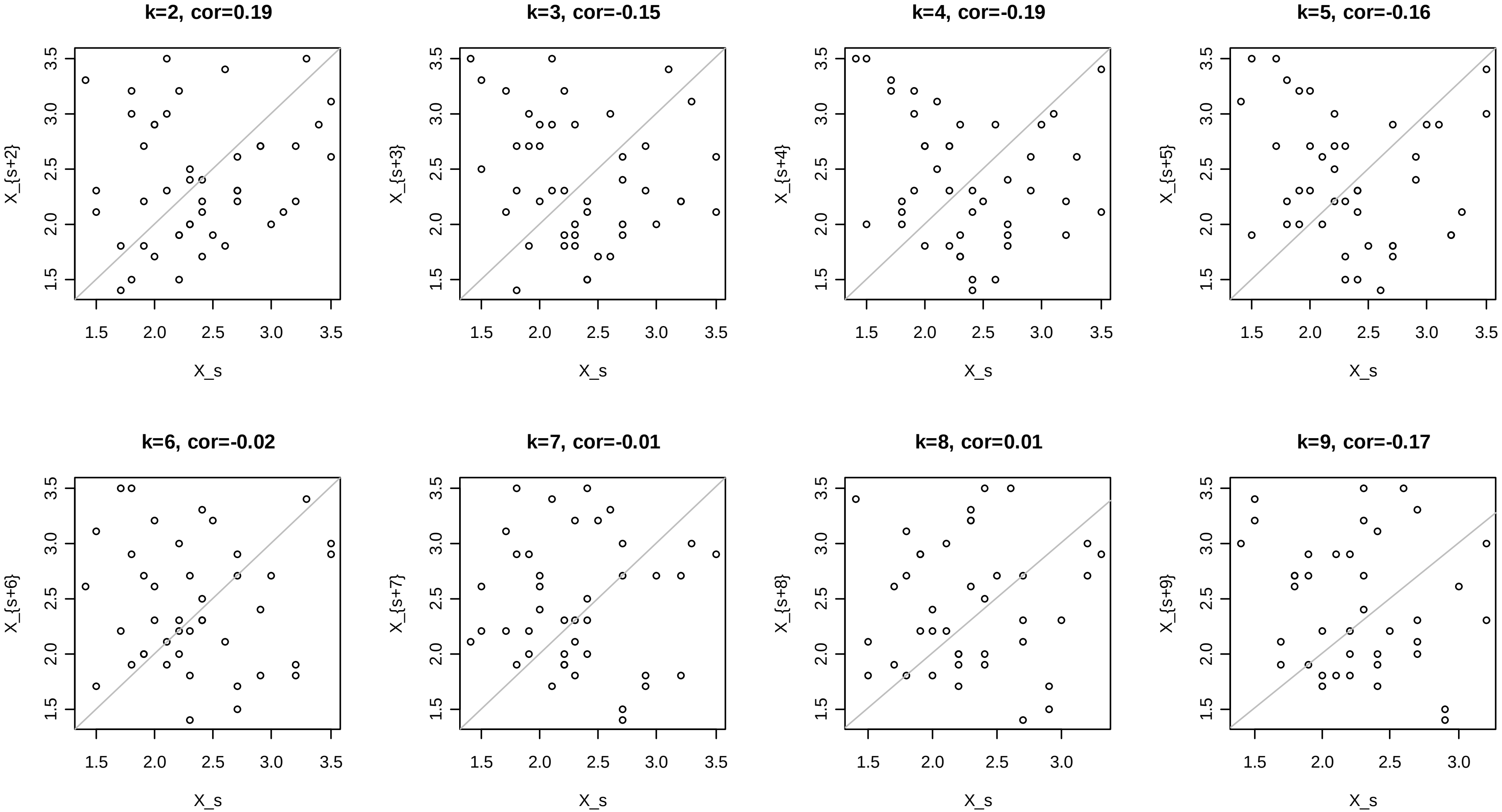
Our next goal is to estimate the autocorrelation function (acf) from a realization of weakly stationary time series.





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

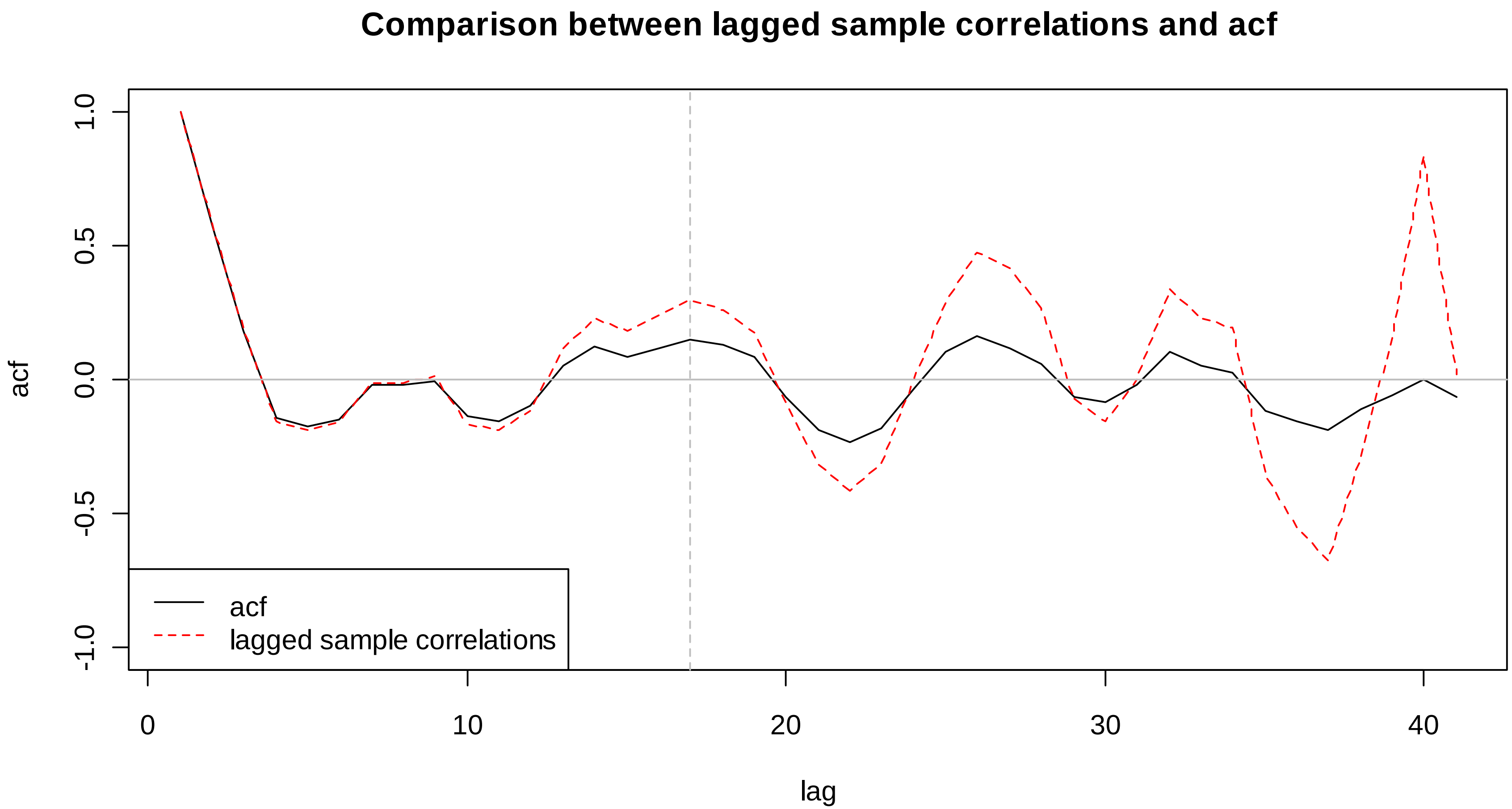
and $$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$$

Standard approach in time series analysis for computing the acf

Applied Time Series Analysis
FS 2012 – Week 03

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

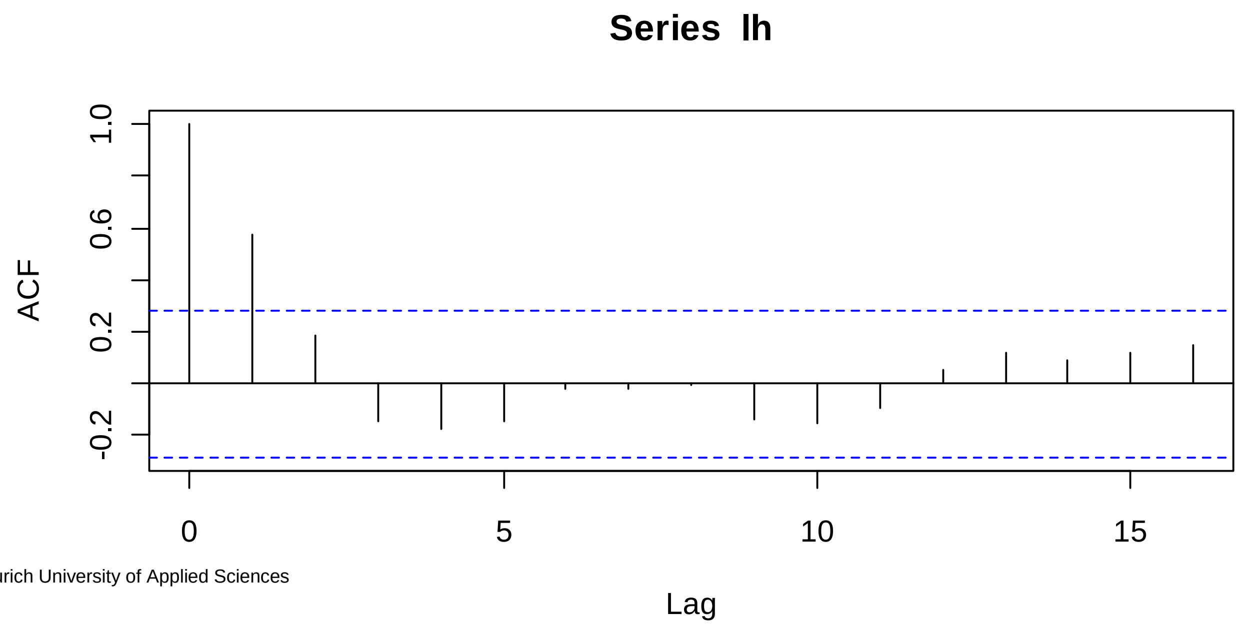
- The estimated sample ACs can be highly correlated.

- 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.

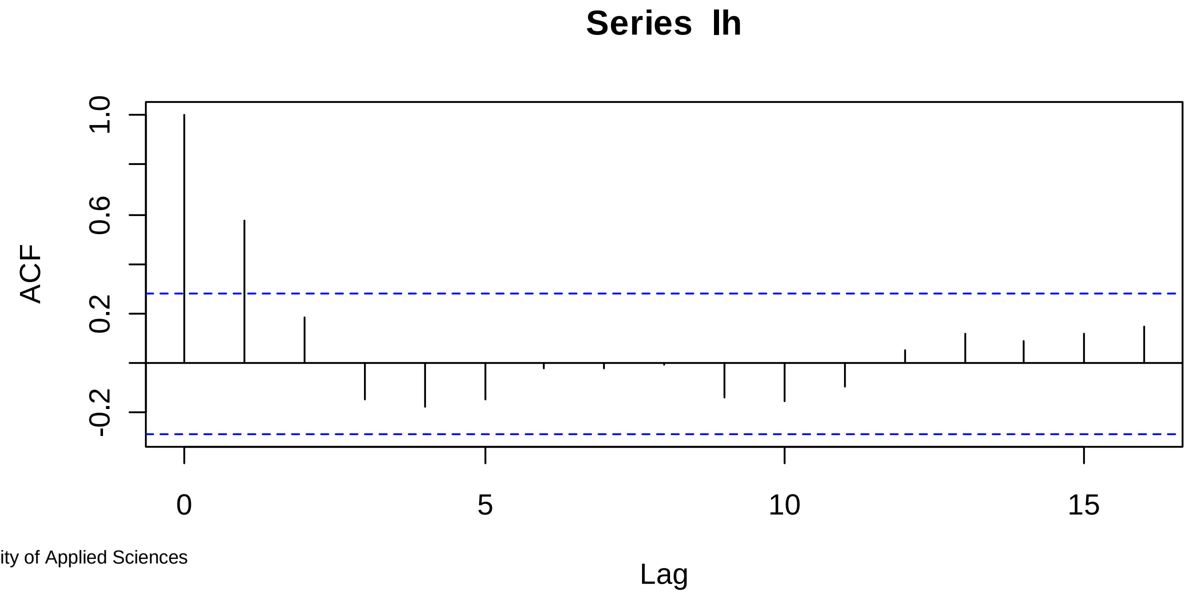
Interpreting the meaning of a set of autocorrelation coefficients is not always easy. The following slides offer some advice.



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.

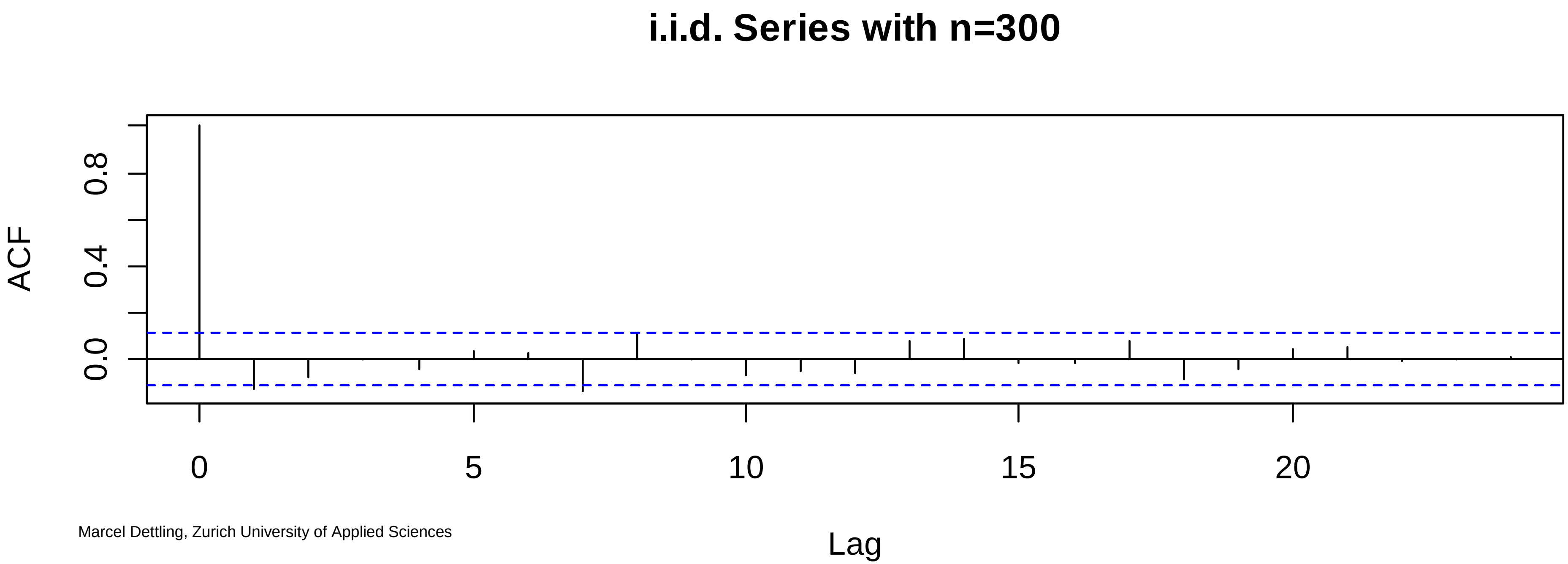
However, the estimated $\hat{\rho}(k)$ are not. We thus need to decide, whether an observed $\hat{\rho}(k) \neq 0$ is significantly so, or just appeared by chance. This is the idea behind the confidence bands.



Random Series – Confidence Bands

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

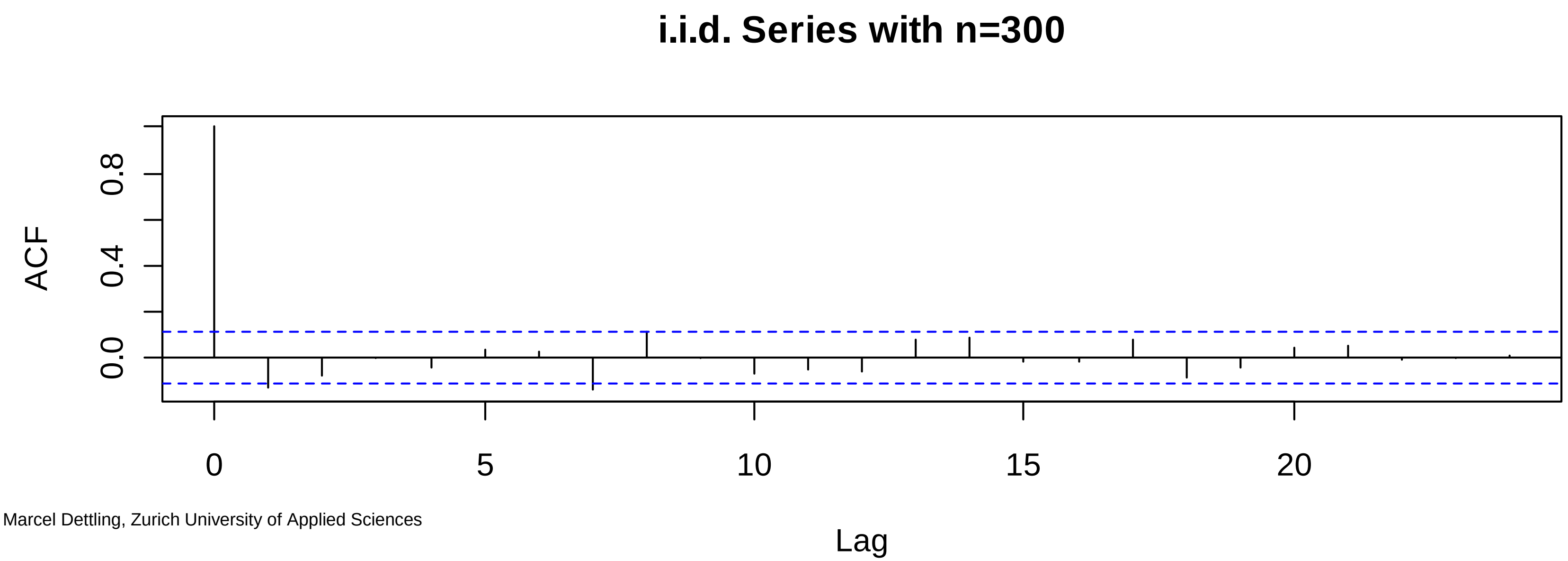
Thus, if a series is random, 95% of the estimated $\hat{\rho}(k)$ can be expected to lie within the interval $\pm 2/\sqrt{n}$



Random Series – Confidence Bands

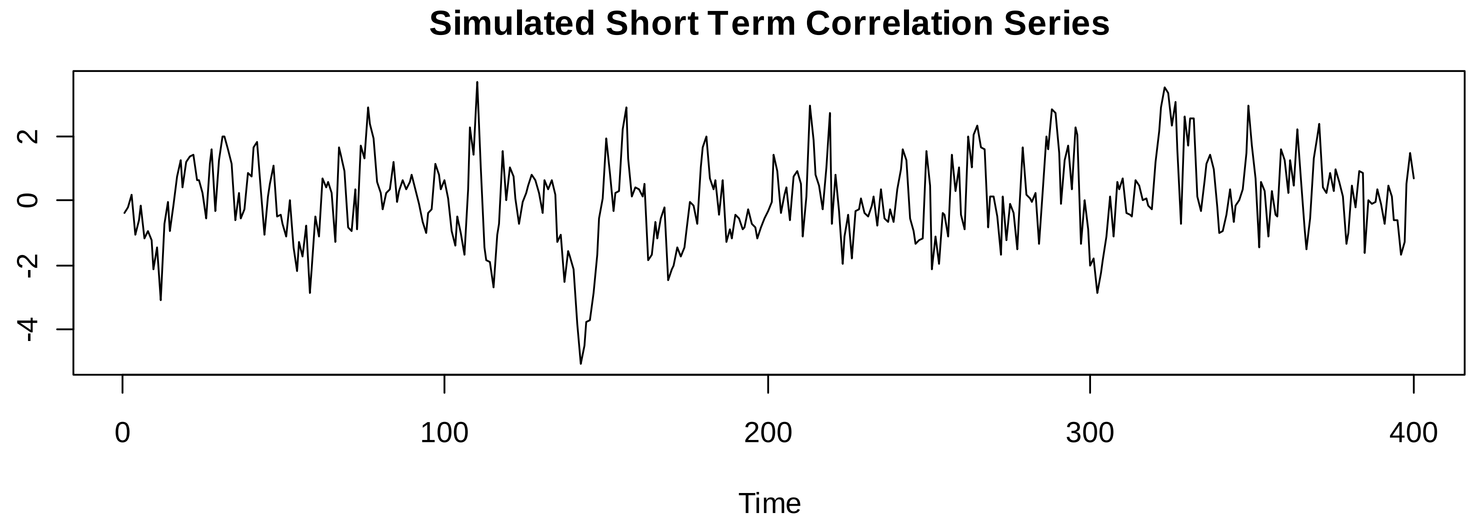
Thus, even for a (long) i.i.d. time series, we expect that 5% of the estimated autocorrelation coeffcients exceed the confidence bounds. They correspond to type I errors.

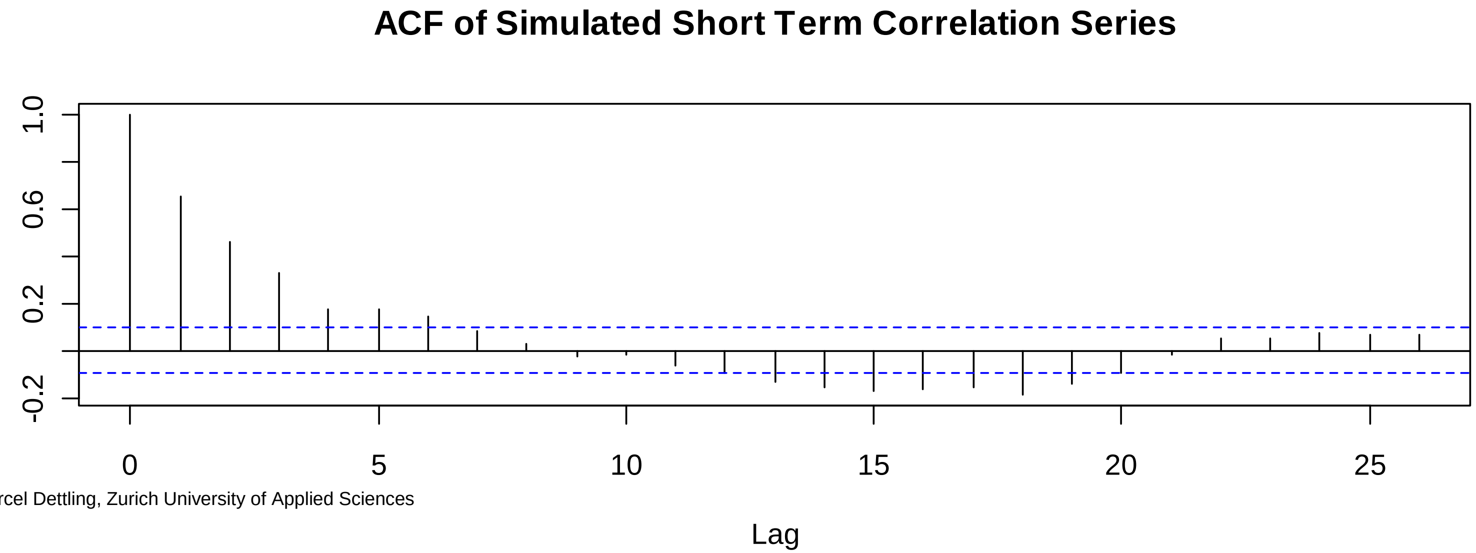
Note: the probabilistic properties of non-normal i.i.d series are much more difficult to derive.



Applied Time Series Analysis
FS 2012 – Week 03

Short Term Correlation





Applied Time Series Analysis
FS 2012 – Week 03

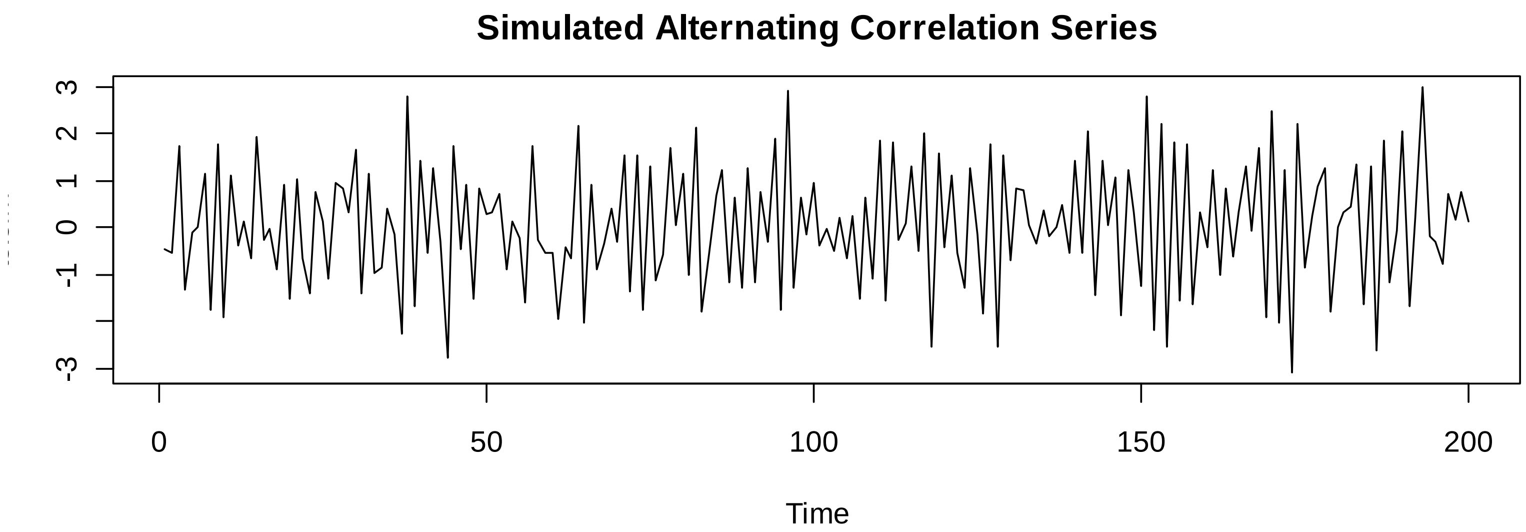
Short Term Correlation

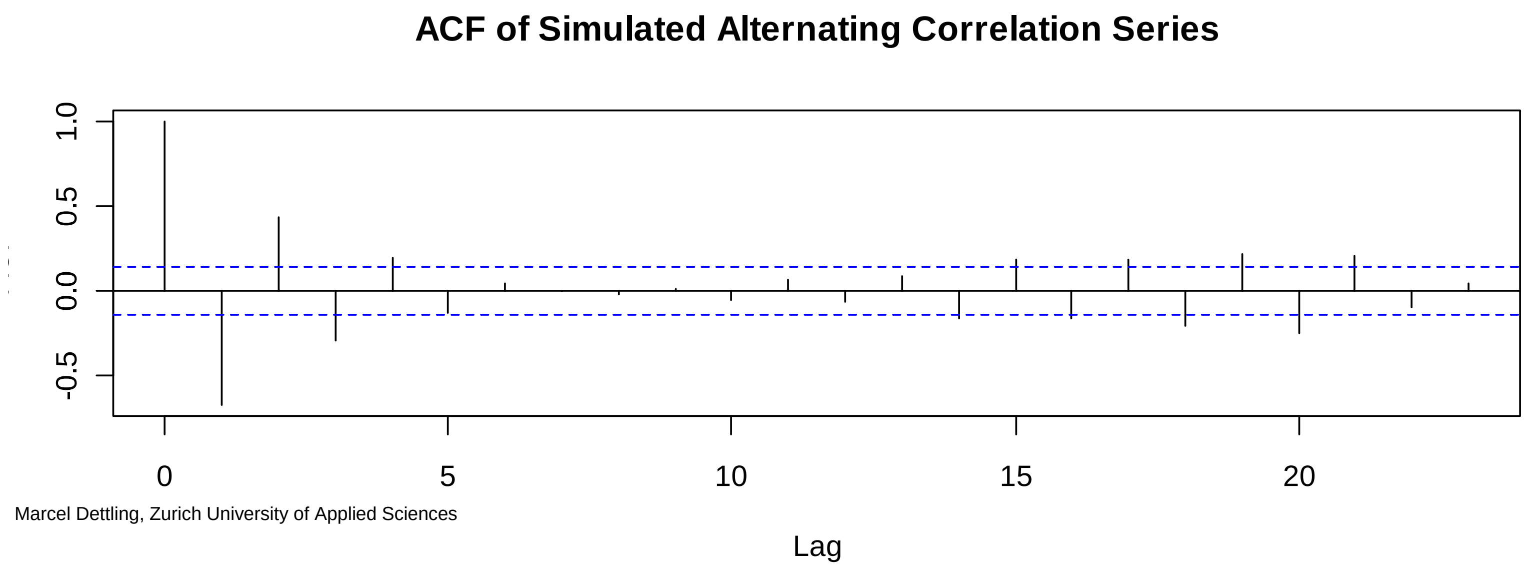
Stationary series often exhibit short-term correlation, characterized by a fairly large value of $\hat{\rho}(1)$, followed by a few more coefficients which, while significantly greater than zero, tend to get successively smaller. For longer lags k, they are close to 0.

A time series which gives rise to such a correlogram, is one for which an observation above the mean tends to be followed by one or more further observations above the mean, and similarly for observations below the mean.

A model called an autoregressive model may be appropriate for series of this type.

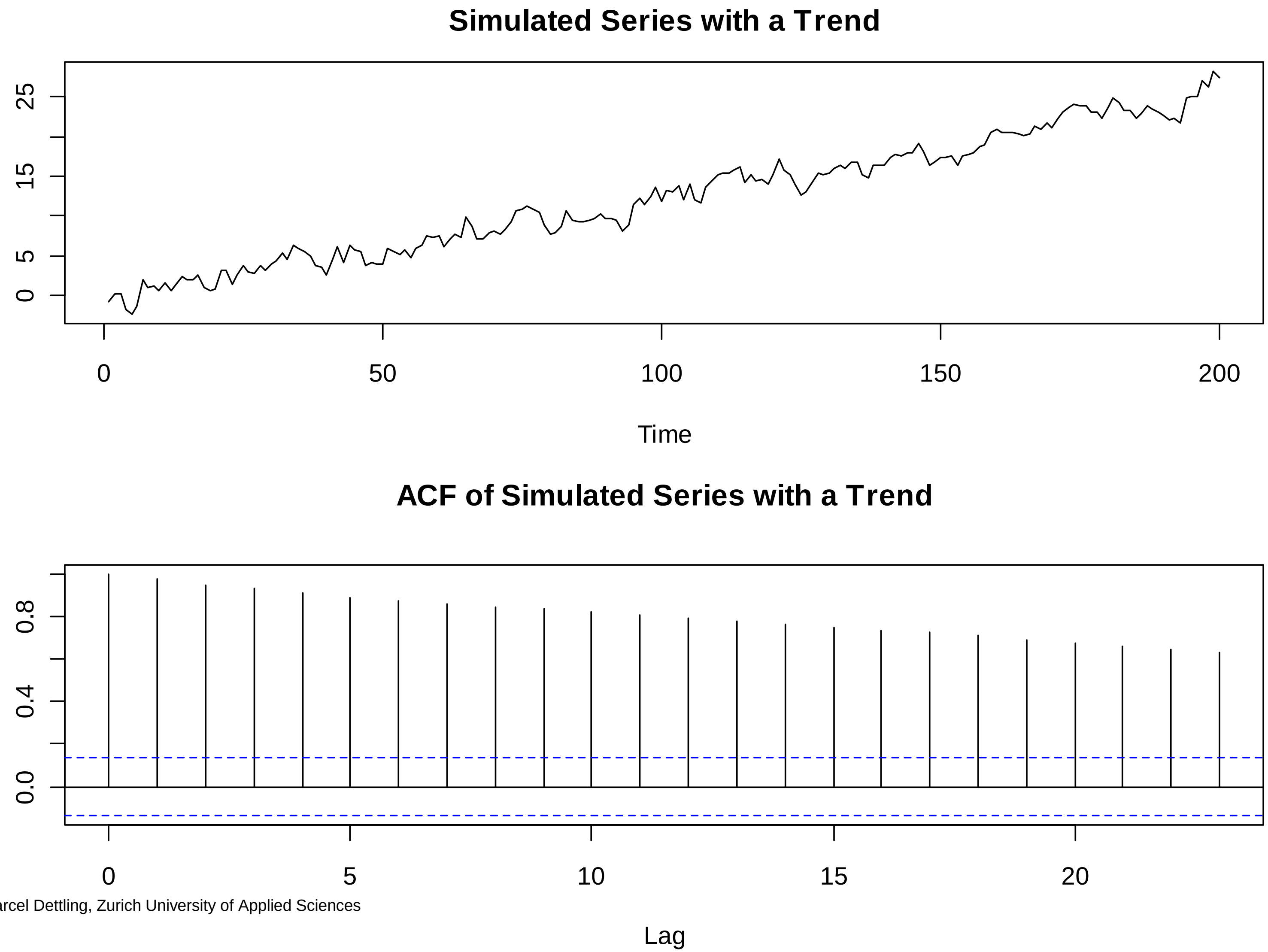
Alternating Time Series



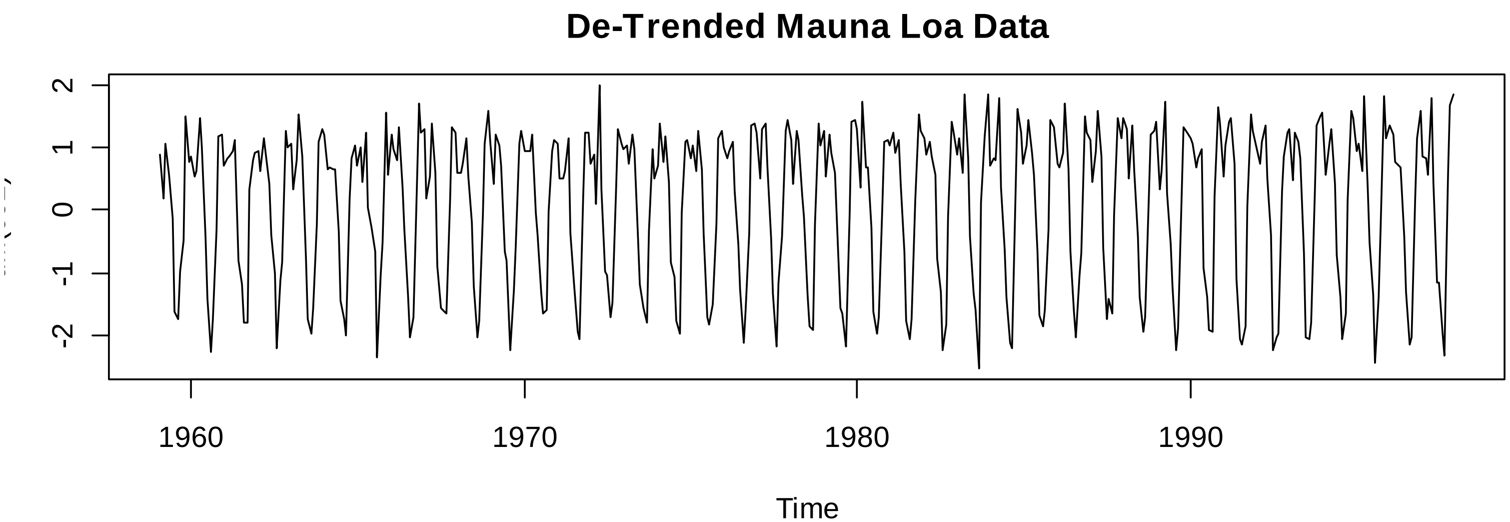


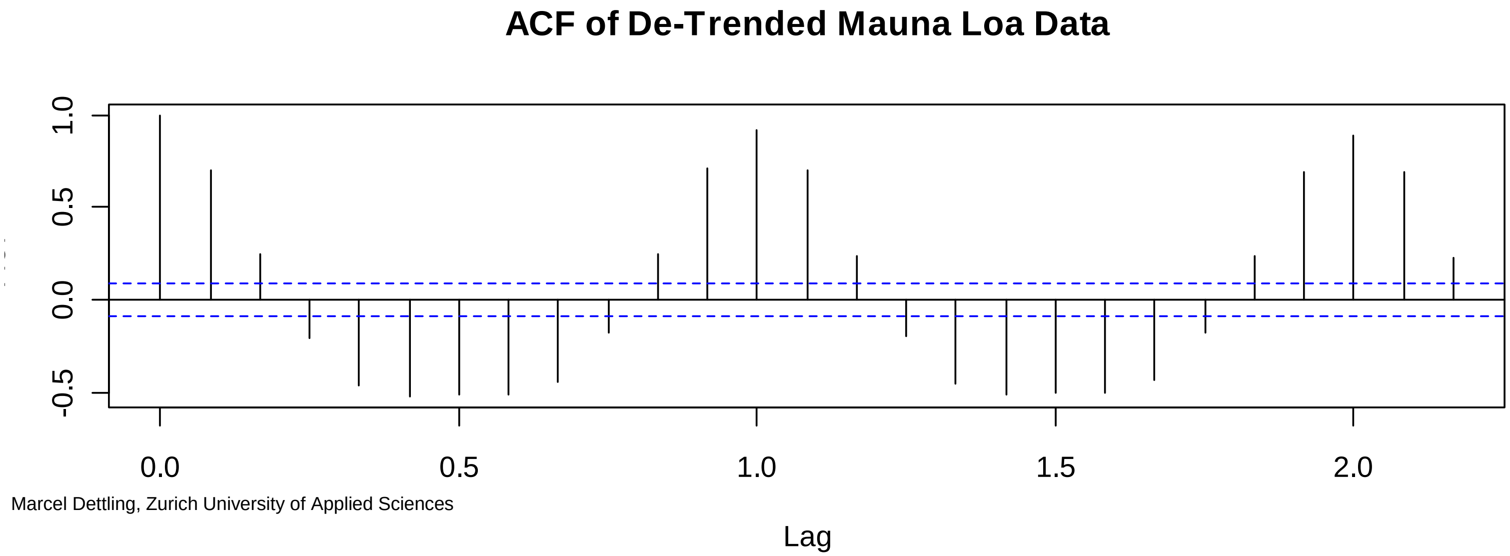
Applied Time Series Analysis
FS 2012 – Week 03

Non-Stationarity in the ACF: Trend



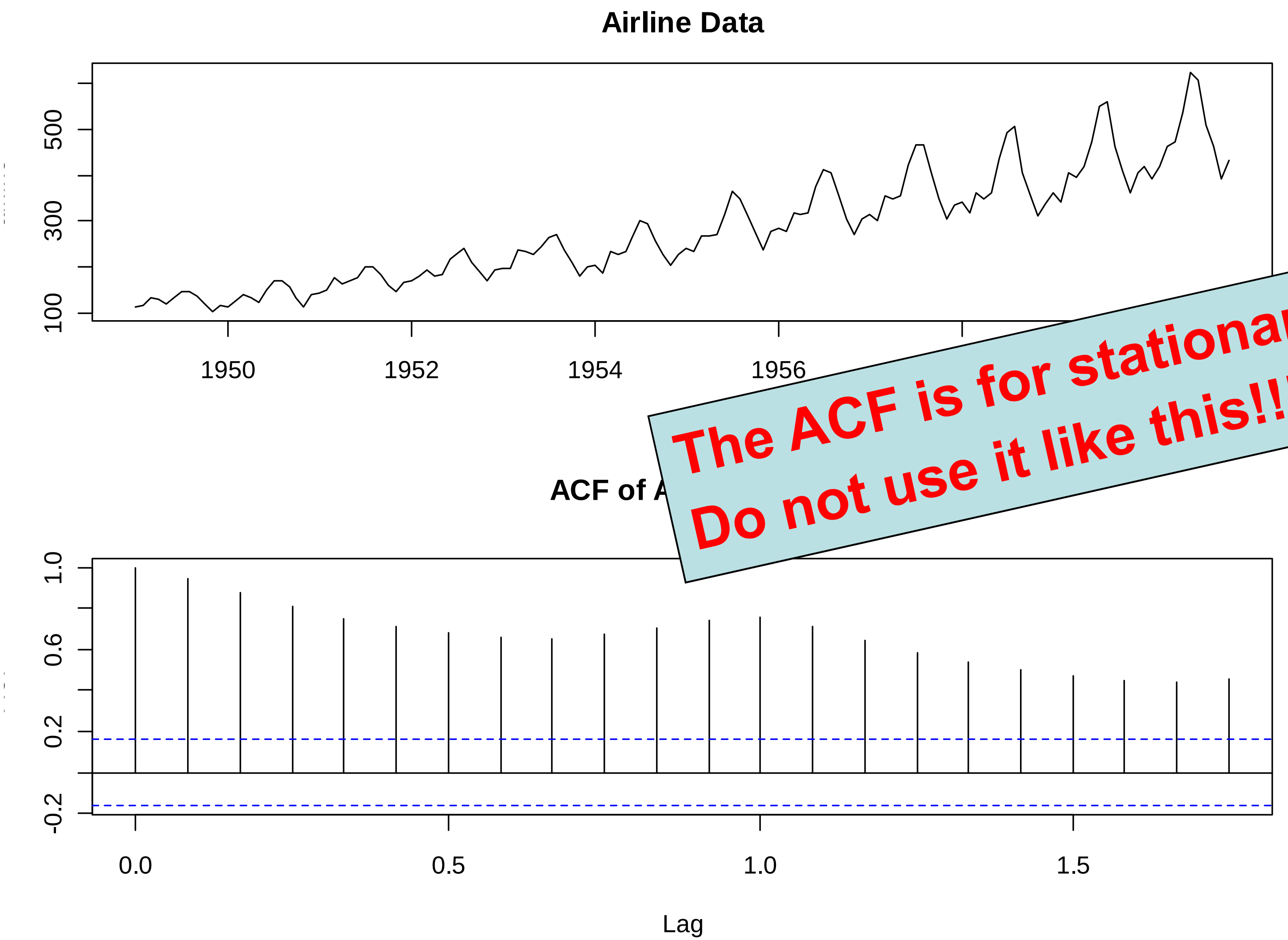
Non-Stationarity in the ACF: Seasonal Pattern





Applied Time Series Analysis
FS 2012 – Week 03

ACF of the Raw Airline Data

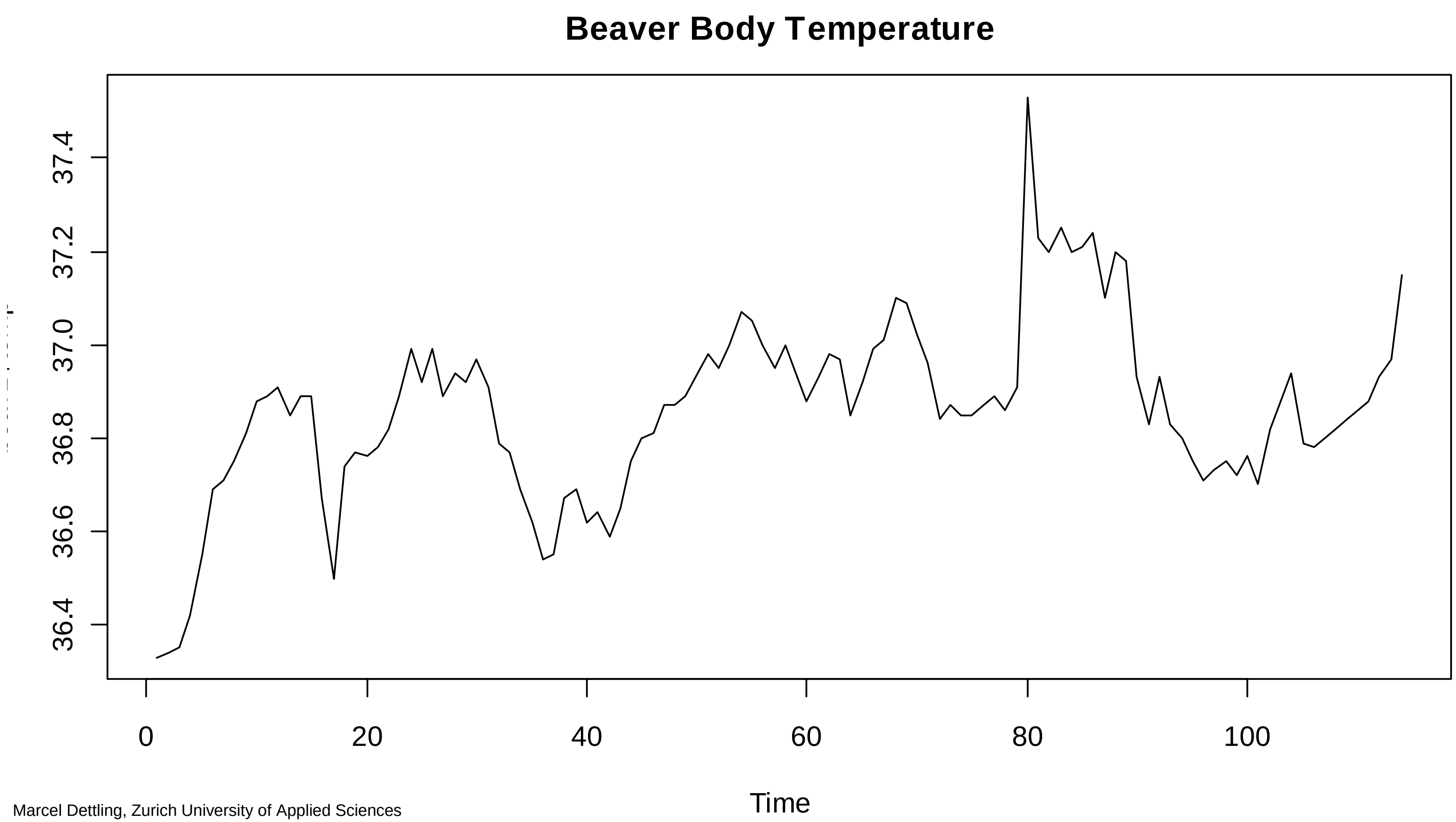


The ACF is for stationary series only!
Do not use it like this!!!

Applied Time Series Analysis
FS 2012 – Week 03

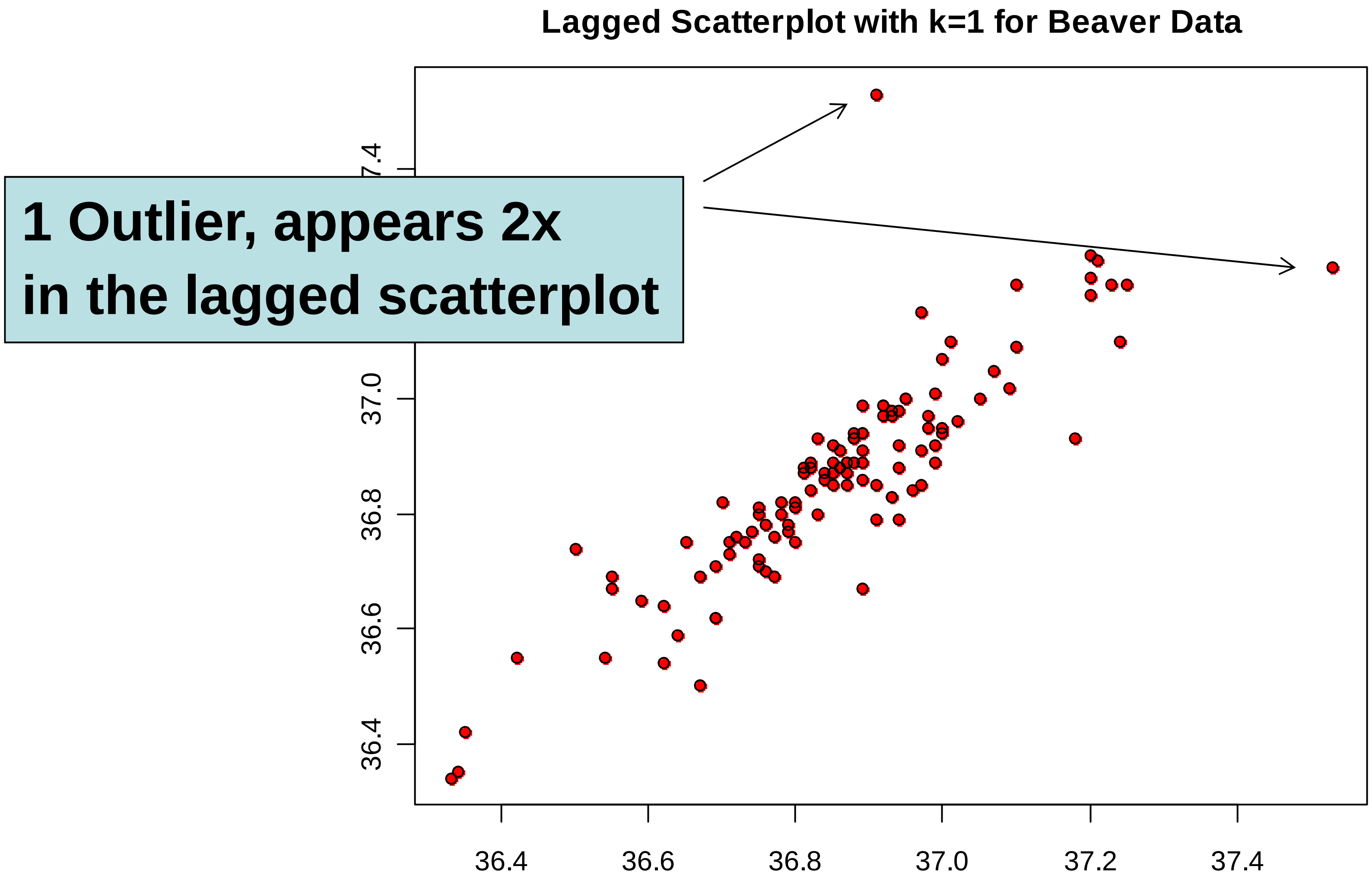
Outliers and the ACF

Outliers in the time series strongly affect the ACF estimation!



Applied Time Series Analysis
FS 2012 – Week 03

Outliers and the ACF



Outliers and the ACF

The estimates $\hat{\rho}(k)$ are very sensitive to outliers. They can be diagnosed using the lagged scatterplot, where every single outlier appears twice.

Strategy for dealing with outliers:

- if it is an outlier: delete the observation

- replace the now missing observations by either:

  a) global mean of the series
  b) local mean of the series, e.g. +/- 3 observations
  c) fit a time series model and predict the missing value

General Remarks about the ACF

a) Appearance of the series => Appearance of the ACF
Appearance of the series ~~<=~~ Appearance of the ACF

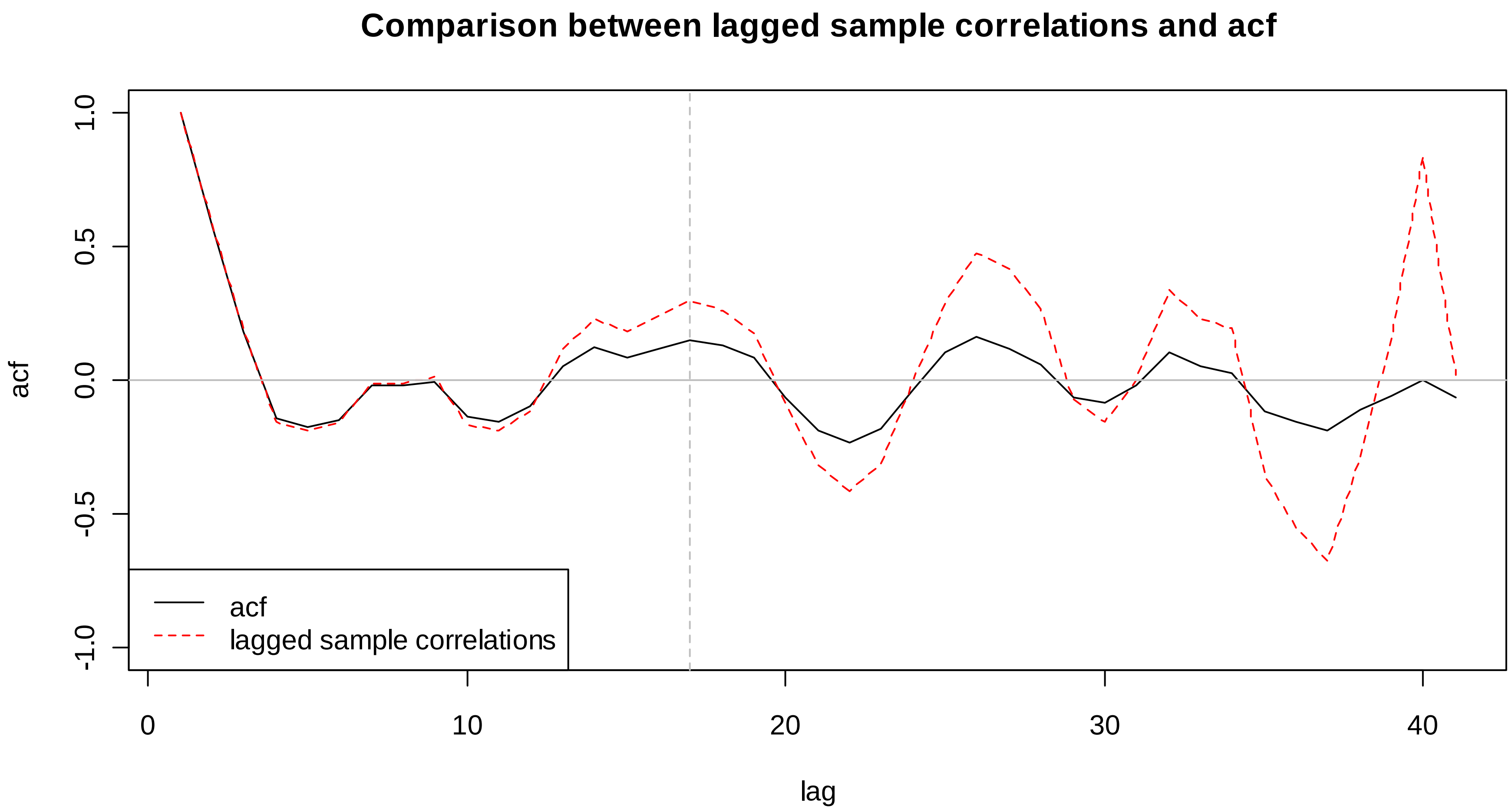
b) Compensation

$$\sum_{k=1}^{n-1} \hat{\rho}(k) = -\frac{1}{2}$$

All autocorrelation coefficients sum up to -1/2. For large lags k, they can thus not be trusted, but are at least damped. This is a reason for using the rule of the thumb.

Applied Time Series Analysis
FS 2012 – Week 03

ACF vs. Lagged Sample Correlations



How Well Can We Estimate the ACF?

What do we know already?

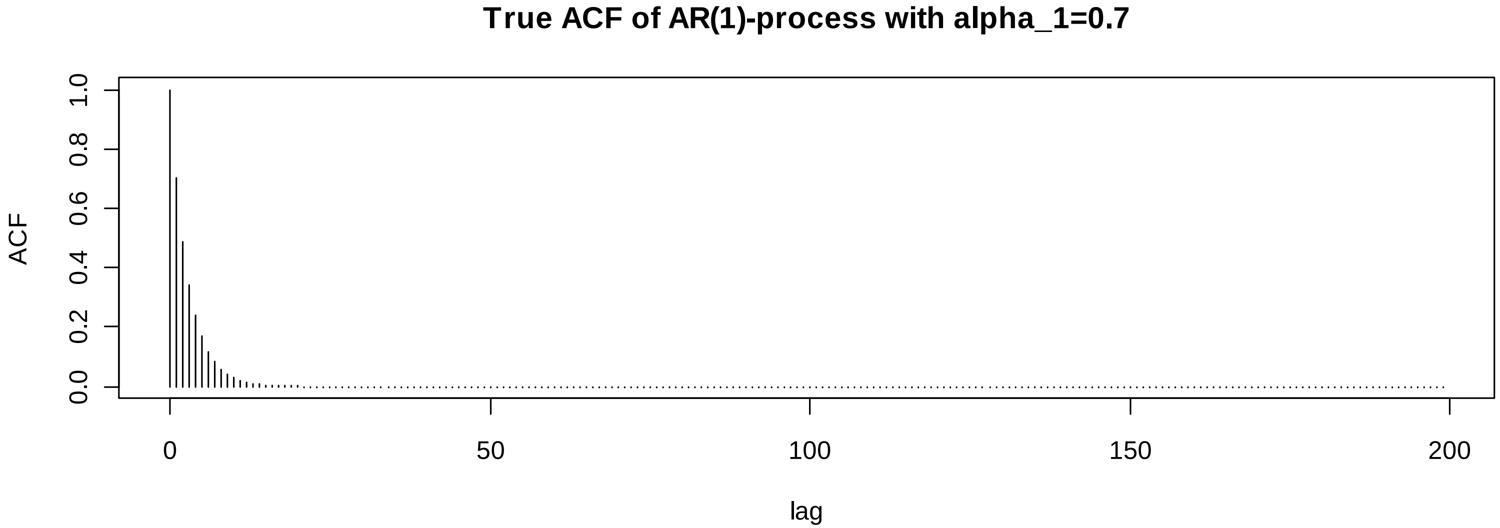
- The ACF estimates are biased
- At higher lags, we have few observations, and thus variability
- There also is the compensation problem…

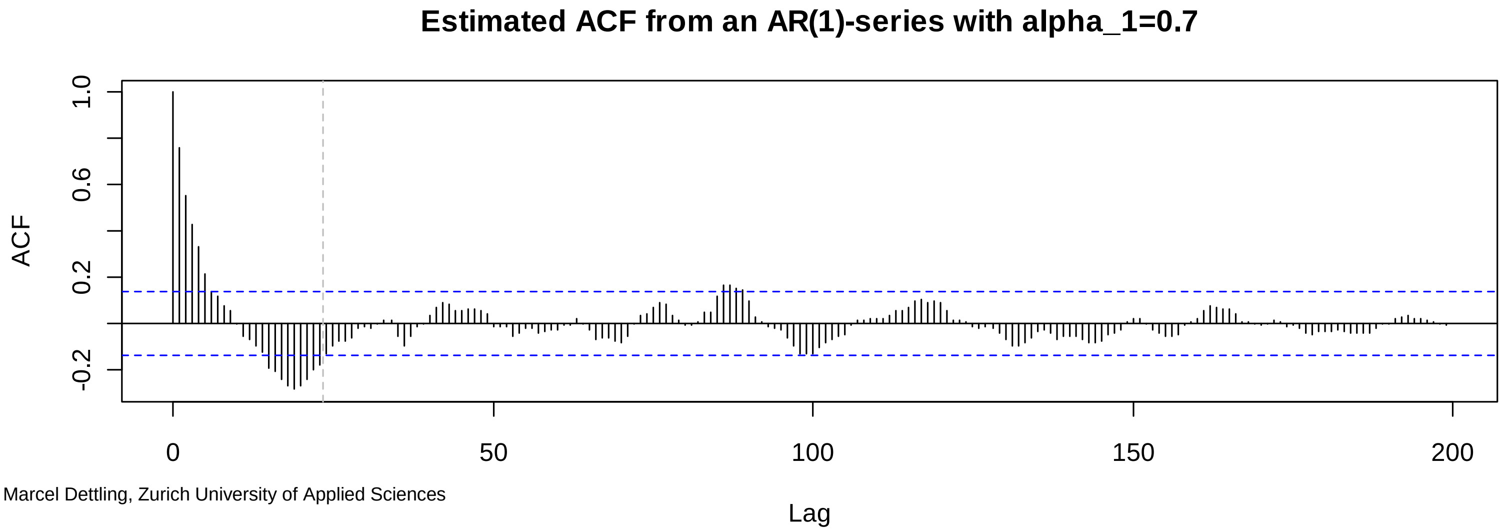
→ ACF estimation is not easy, and interpretation is tricky.

For answering the question above:

- For an AR(1) time series process, we know the true ACF
- We generate a number of realizations from this process
- We record the ACF estimates and compare to the truth

Theoretical vs. Estimated ACF





How Well Can We Estimate the ACF?

A) For AR(1)-processes we understand the theoretical ACF

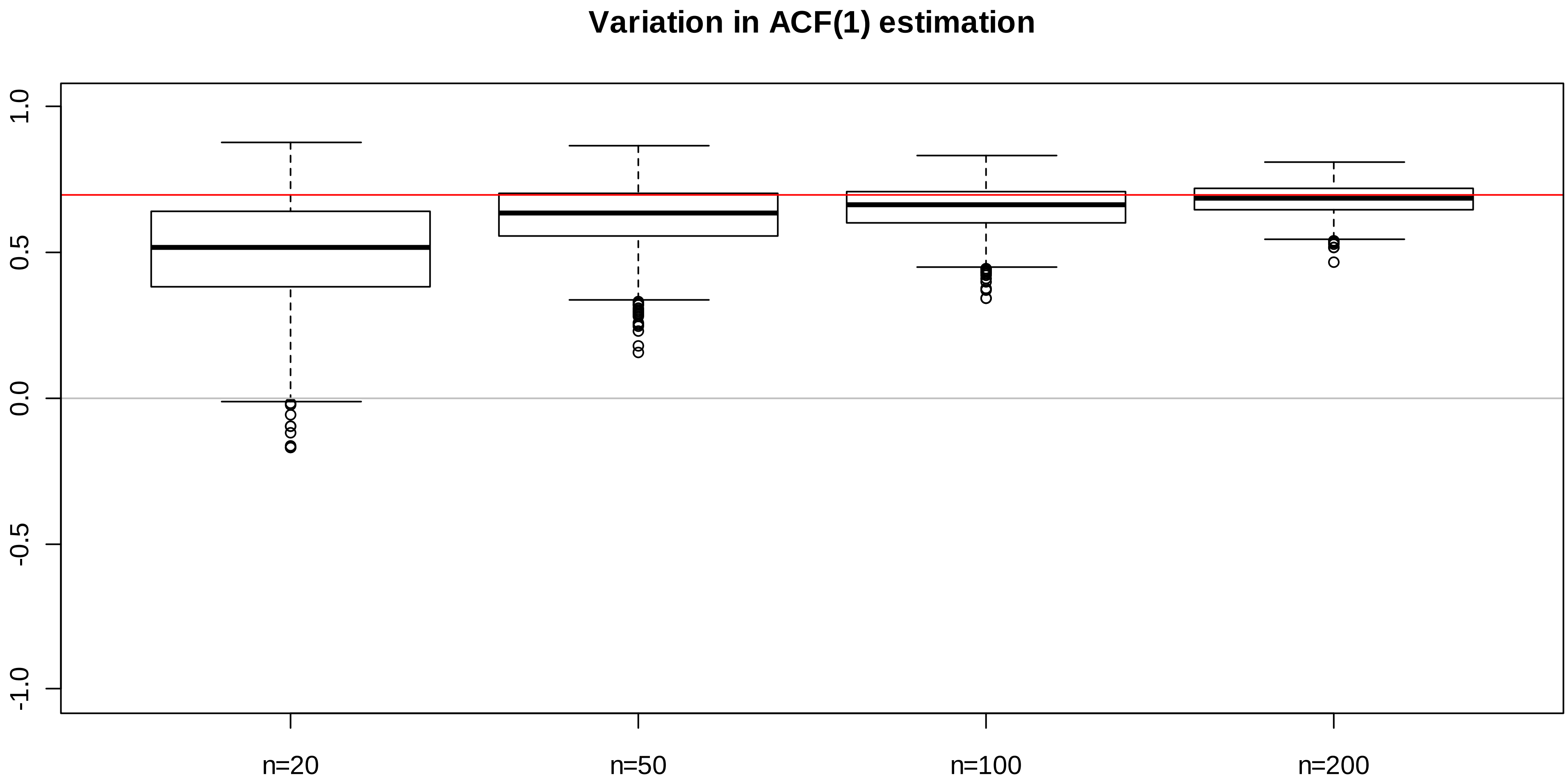
B) Repeat for i=1, …, 1000

Simulate a **length n** AR(1)-process
Estimate the ACF from that realization

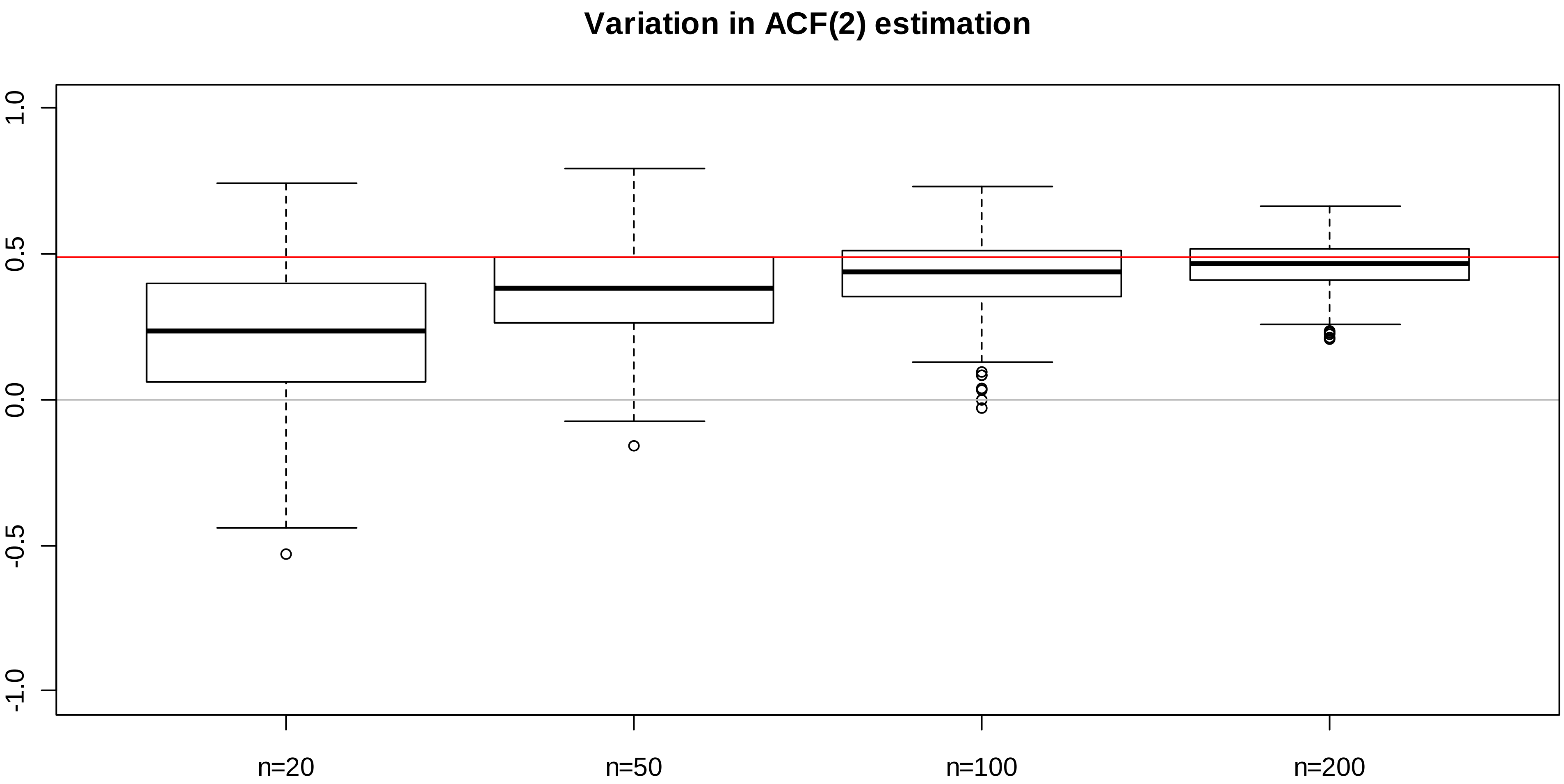
End for

C) Boxplot the (bootstrap) sample distribution of ACF-estimates
Do so for different **lags k** and different series **length n**

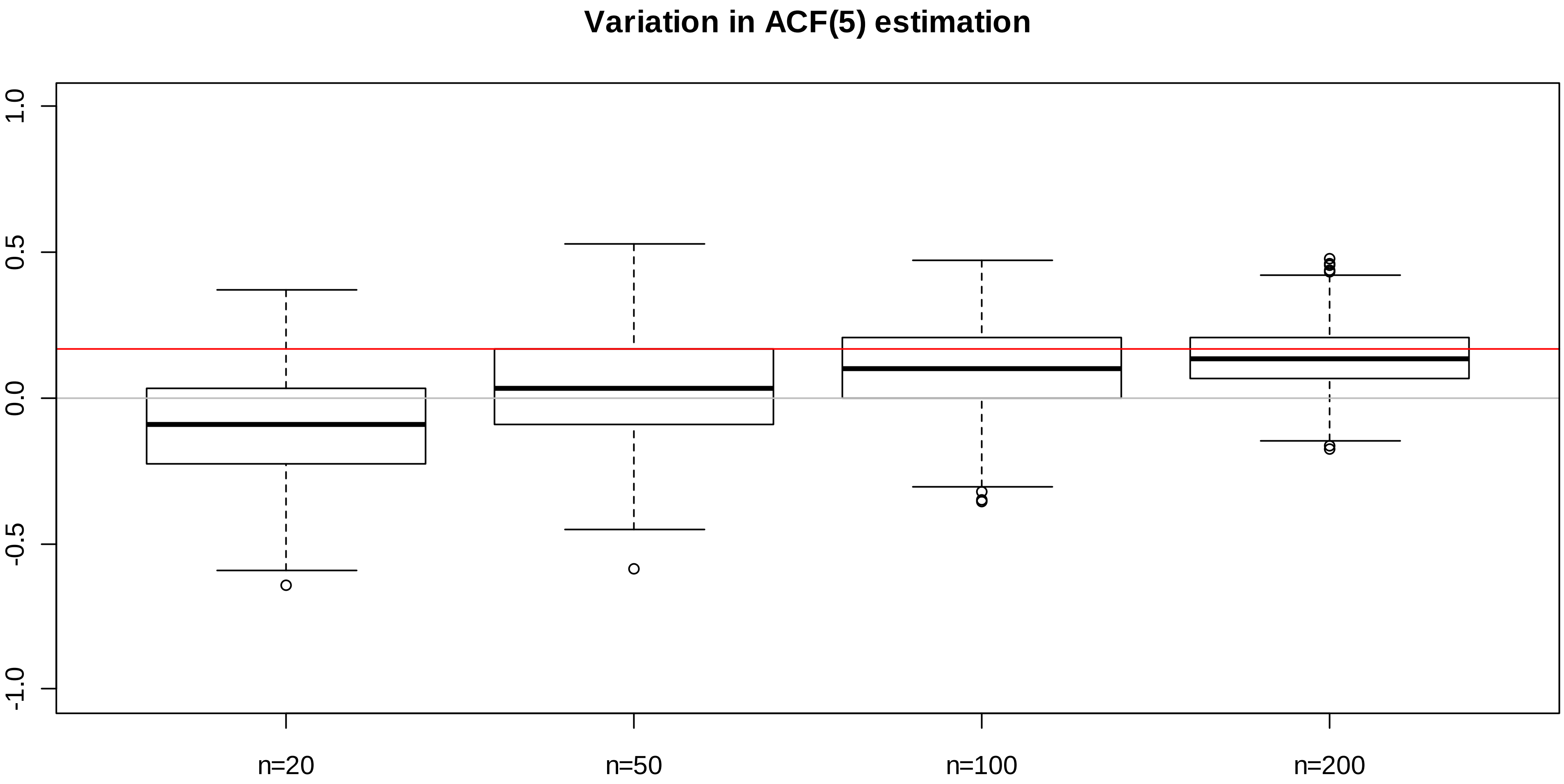
How Well Can We Estimate the ACF?



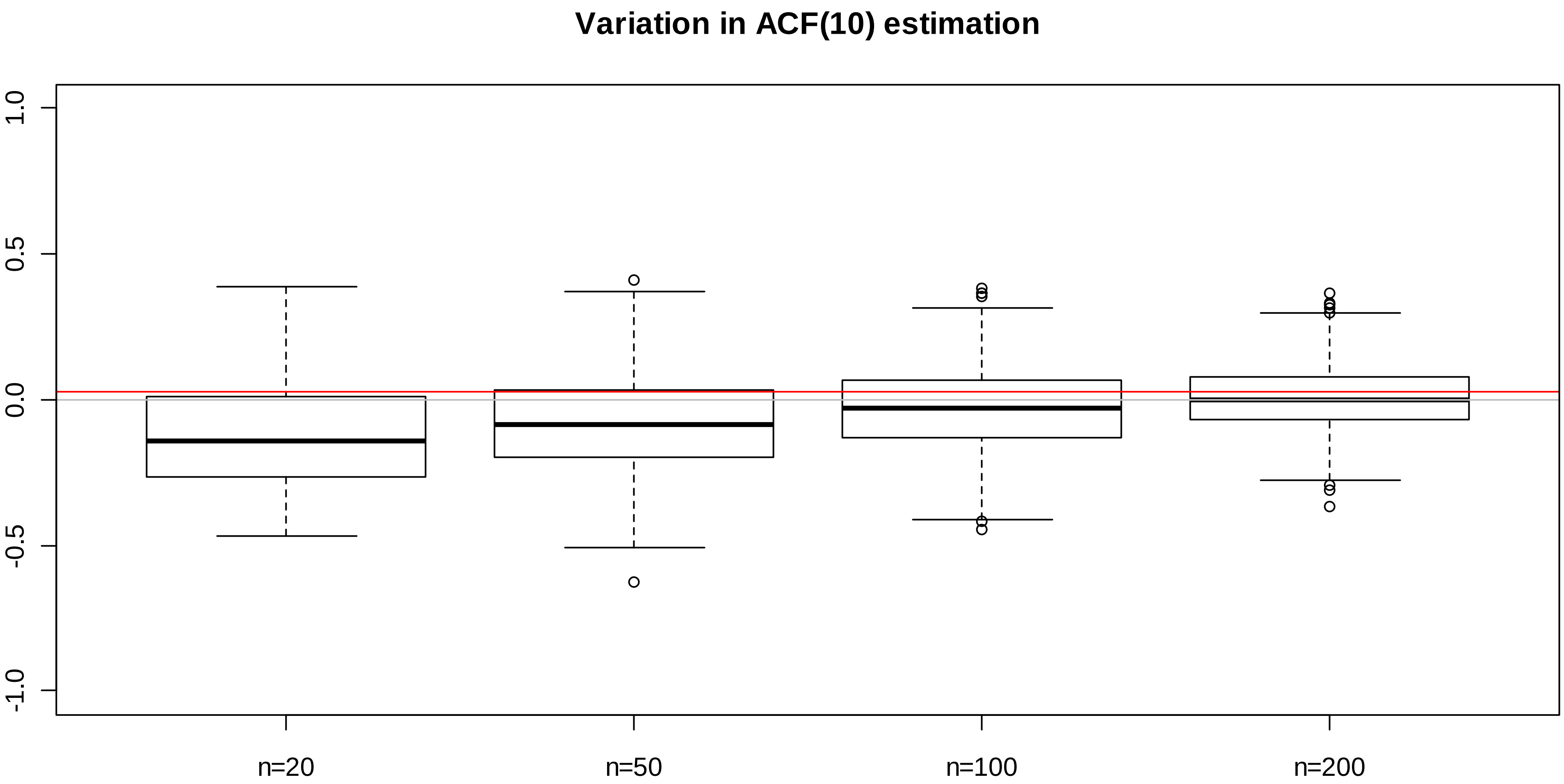
How Well Can We Estimate the ACF?



How Well Can We Estimate the ACF?



How Well Can We Estimate the ACF?



Trivia ACF Estimation

- In short series, the ACF is strongly biased. The consistency kicks in and kills the bias only after ~100 observations.
- The variability in ACF estimation is considerable. We observe that we need at least 50, or better, 100 observations.
- For higher lags k, the bias seems a little less problematic, but the variability remains large even with many observations n.
- The confidence bounds, derived under independence, are not very accurate for (dependent) time series.

→ ***Interpreting the ACF is tricky!***

Application: Variance of the Arithmetic Mean

Practical problem: we need to estimate the mean of a realized/ observed time series. We would like to attach a standard error.

- If we estimate the mean of a time series without taking into account the dependency, the standard error will be flawed.
- This leads to misinterpretation of tests and confidence intervals and therefore needs to be corrected.
- The standard error of the mean can both be over-, but also underestimated. This depends on the ACF of the series.

→ **For the derivation, see the blackboard…**

Partial Autocorrelation Function (PACF)

The k[th] partial autocorrelation coefficient $\rho_{part}(k)$ is defined as the correlation between the random variables X_{t+k} and X_t, given all the values in between.

$$\rho_{part}(k) = Cor(X_{t+k}, X_t \mid X_{t+1} = x_{t+1}, \dots, X_{t+k-1} = x_{t+k-1})$$

Their meaning is best understood by drawing an analogy to simple and multiple linear regression. The ACF measures the „simple“ dependence between X_{t+k} and X_t, whereas the PACF measures that dependence in a „multiple“ fashion.

Facts about the PACF

- Estimation of the PACF is complicated and will not be discussed in the course. R can do it ;-)
- The first PACF coefficient is equal to the first ACF coefficient. Subsequent coefficients are not equal, but can be derived from each other.
- For a time series generated by an AR(p)-process, the p^{th} PACF coefficient is equal to the p^{th} AR-coefficient. All PACF coefficients for lags $k>p$ are equal to 0.
- Confidence bounds also exist for the PACF.

Outlook to AR(p)-Models

Suppose that Z_t is an i.i.d random process with zero mean and variance σ_Z^2. Then a random process X_t is said to be an auto-regressive process of order p if

$$X_t = \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + Z_t$$

This is similar to a multiple regression model, but X_t is regressed not on independent variables, but on past values of itself. Hence the term auto-regressive.

We use the abbreviation AR(p).