

Applied Time Series Analysis

SS 2015 – Introduction

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, February-May 2015

Applied Time Series Analysis

SS 2015 – Introduction

Your Lecturer



Name: **Marcel Dettling**

Age: 40 Years

Civil Status: Married, 2 children

Education: Dr. Math. ETH

Position:

Lecturer @ ETH Zürich and @ ZHAW

Researcher in Applied Statistics @ ZHAW

Connection:

Research with industry: *hedge funds, insurance, ...*

Academic research: *high-frequency financial data*

Applied Time Series Analysis

SS 2015 – Introduction

A First Example

In 2006, Singapore Airlines decided to place an order for new aircraft. It contained the following jets:

- 20 Boeing 787
- 20 Airbus A350
- 9 Airbus A380

How was this decision taken?

It was based on a combination of time series analysis on airline passenger trends, plus knowing the corporate plans for maintaining or increasing the market share.

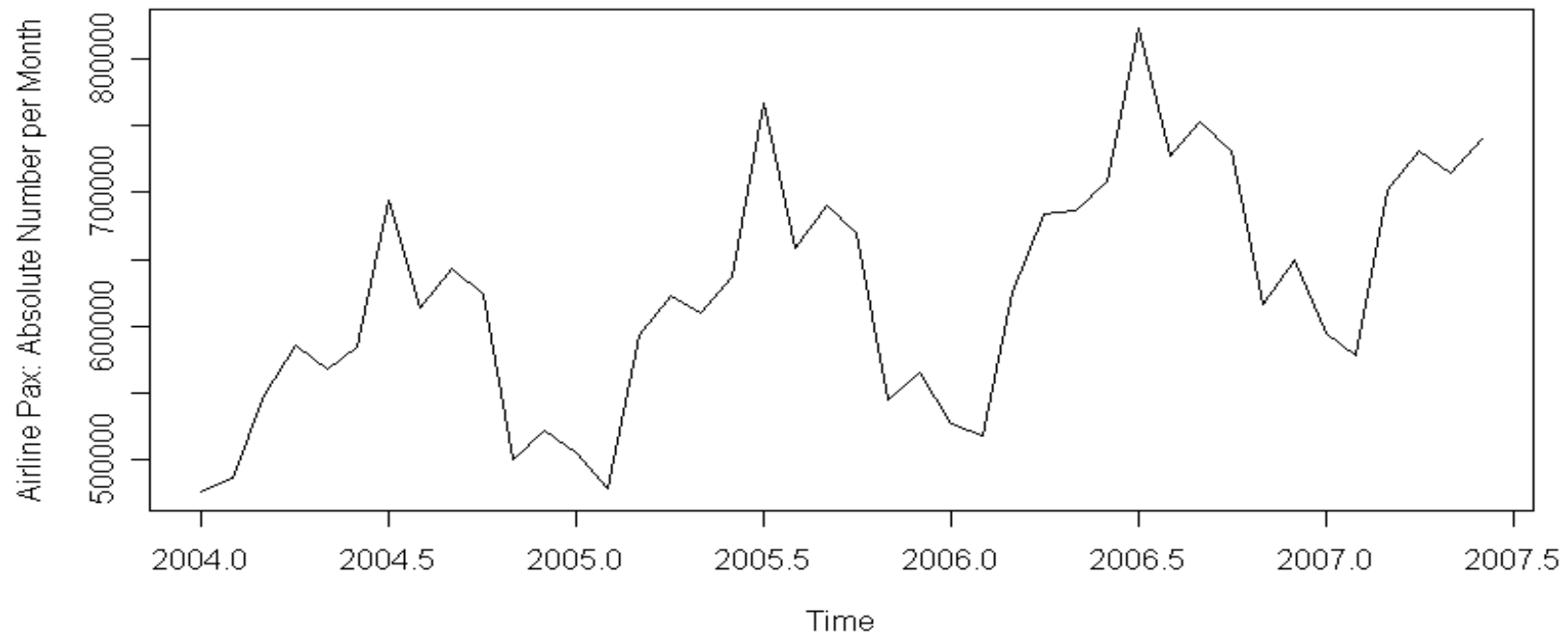
Applied Time Series Analysis

SS 2015 – Introduction

A Second Example

- Taken from a former research project @ ZHAW
- Airline business: # of checked-in passengers per month

Airline Pax: Absolute Number per Month



Applied Time Series Analysis

SS 2015 – Introduction

Some Properties of the Series

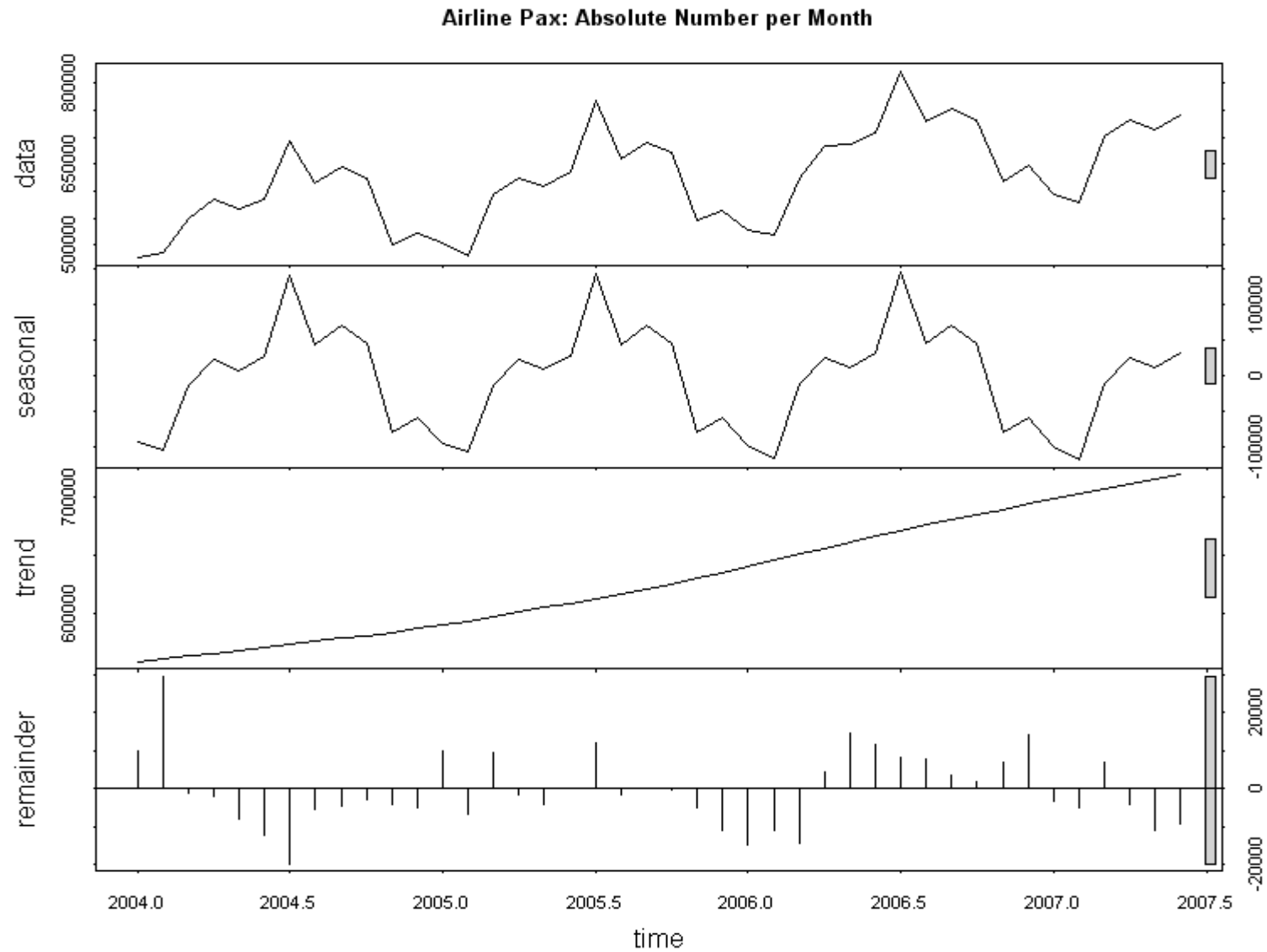
- Increasing trend (i.e. generally more passengers)
- Very prominent seasonal pattern (i.e. peaks/valleys)
- Hard to see details beyond the obvious

Goals of the Project

- Visualize, or better, extract trend and seasonal pattern
- Quantify the amount of random variation/uncertainty
- Provide the basis for a man-made forecast after mid-2007
- Forecast (extrapolation) from mid-2007 until end of 2008
- How can we better organize/collect data?

Applied Time Series Analysis

SS 2015 – Introduction



Applied Time Series Analysis

SS 2015 – Introduction

Organization of the Course

Contents:

- Basics, Mathematical Concepts, Time Series in R
- Descriptive Analysis (Plots, Decomposition, Correlation)
- Models for Stationary Series (AR(p), MA(q), ARMA(p,q))
- Non-Stationary Models (SARIMA, GARCH, Long-Memory)
- Forecasting (Regression, Exponential Smoothing, ARMA)
- Miscellaneous (Multivariate, Spectral Analysis, State Space)

Goal:

The students acquire experience in analyzing time series problems, are able to work with the software package R, and can perform time series analyses correctly on their own.

Applied Time Series Analysis

SS 2015 – Introduction

Organization of the Course

Applied Time Series Analysis – SS 2013

People:

Lecturer: Dr. Marcel Detting (marcel.detting@stat.math.ethz.ch)
 Assistants: Patric Müller (patric.mueller@stat.math.ethz.ch)
 Preetam Nandy (nandy@stat.math.ethz.ch)

Organization:

This course will be visited by students from various Master and Doctoral Programs at ETH and elsewhere. It is the short version of the course which will be awarded with 4 ECTS credits. The extended version with 6 ECTS credits takes place in the even years.

Lectures:

Lectures will be held on Mondays from 10.15-11.55 at ETH Zentrum, room HG E1.2. Theory and examples will be shown on power point slides and the blackboard. Also, a scriptum is available. The tentative schedule is as follows:

Week	Date	L/L	Topics
01	18.02.2013	L/L	Introduction, Examples, Goals
02	25.02.2013	L/E	Mathematical Concepts, Stationarity
03	04.03.2013	L/L	Visualization, Transformations
04	11.03.2013	L/E	Descriptive Decomposition
05	18.03.2013	L/L	Autocorrelation, Partial Autocorrelation
06	25.03.2013	L/E	Stationary Time Series Models 1
07	08.04.2013	L/L	Stationary Time Series Models 2
08	15.04.2013	L/E	Time Series Regression
09	22.04.2013	L/L	Forecasting with Time Series
10	29.04.2013	L/E	Exponential Smoothing
11	06.05.2013	L/L	Multivariate Time Series Analysis
12	13.05.2013	L/E	Spectral Analysis
13	20.05.2013	-/-	-
14	27.05.2013	L/L	Miscellaneous, Outlook, Exam Information

Exercises:

Exercises will be held every second week in the lecture room HG E1.2, where an assistant will provide some background and useful hints on how to approach the problems. Solving the problems needs to be done autonomously and requires the use of the statistical software package R. The exercise schedule is as follows:

Series	Date	Topic	Hand-In	Solutions
01	25.02.2013	Time series in R	04.03.2013	11.03.2013
02	11.03.2013	Plotting and Decomposing	25.03.2013	18.03.2013
03	25.03.2013	Autocorrelation, Modelling	08.04.2013	15.04.2013
04	15.04.2013	ARMA-Models and Applications	22.04.2013	29.04.2013
05	29.04.2013	Forecasting with Time Series	06.05.2013	13.05.2013
06	13.05.2013	Miscellaneous Topics	21.05.2013	---

→ more details are given on the additional organization sheet

Applied Time Series Analysis

SS 2015 – Introduction

Introduction: What is a Time Series?

A time series is a set of observations x_t , where each of the observations was made at a specific time t .

- the set of times T is discrete and finite
- observations were made at fixed time intervals
- continuous and irregularly spaced time series are not covered

Rationale behind time series analysis:

The rationale in time series analysis is to understand the past of a series, and to be able to predict the future well.

Applied Time Series Analysis

SS 2015 – Introduction

Example 1: Air Passenger Bookings

```
> data(AirPassengers)
```

```
> AirPassengers
```

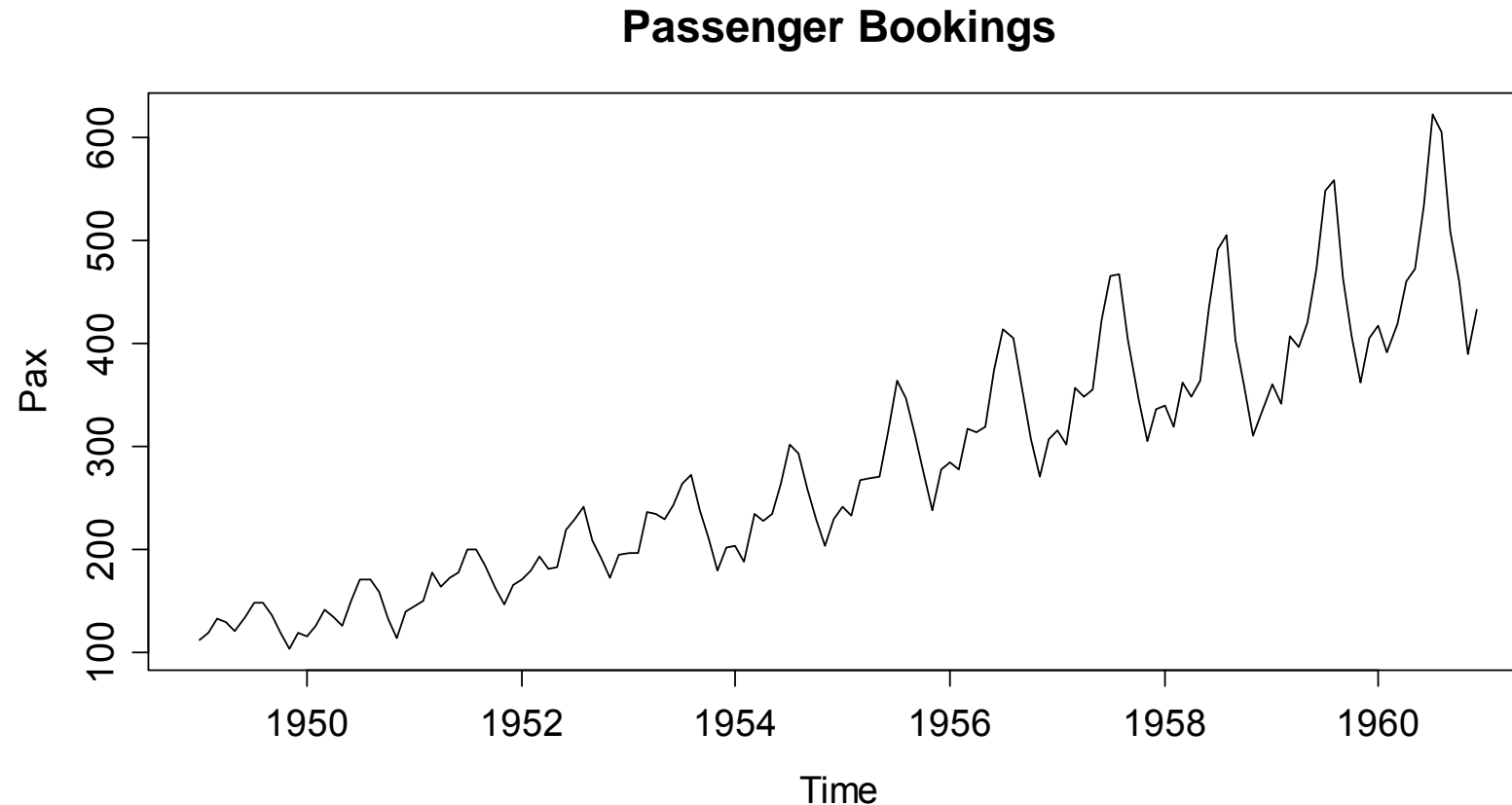
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1949	112	118	132	129	121	135	148	148	136	119	104	118
1950	115	126	141	135	125	149	170	170	158	133	114	140
1951	145	150	178	163	172	178	199	199	184	162	146	166
1952	171	180	193	181	183	218	230	242	209	191	172	194
1953	196	196	236	235	229	243	264	272	237	211	180	201
1954	204	188	235	227	234	264	302	293	259	229	203	229
1955	242	233	267	269	270	315	364	347	312	274	237	278
1956	284	277	317	313	318	374	413	405	355	306	271	306
1957	315	301	356	348	355	422	465	467	404	347	305	336
1958	340	318	362	348	363	435	491	505	404	359	310	337
1959	360	342	406	396	420	472	548	559	463	407	362	405
1960	417	391	419	461	472	535	622	606	508	461	390	432

Applied Time Series Analysis

SS 2015 – Introduction

Example 1: Air Passenger Bookings

```
> plot(AirPassengers, ylab="Pax", main="Pax Bookings")
```

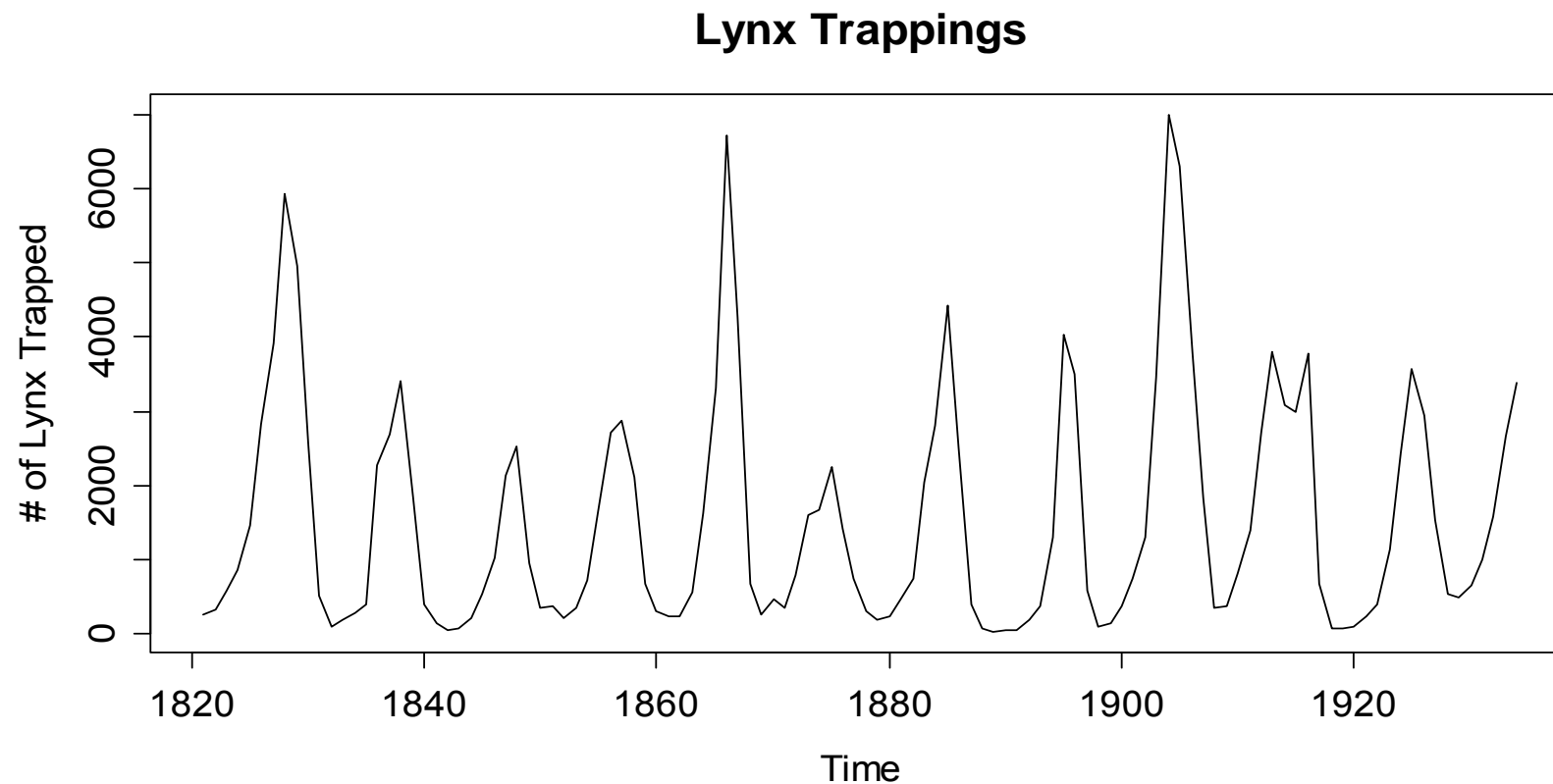


Applied Time Series Analysis

SS 2015 – Introduction

Example 2: Lynx Trappings

```
> data(lynx)
> plot(lynx, ylab="# of Lynx", main="Lynx Trappings")
```

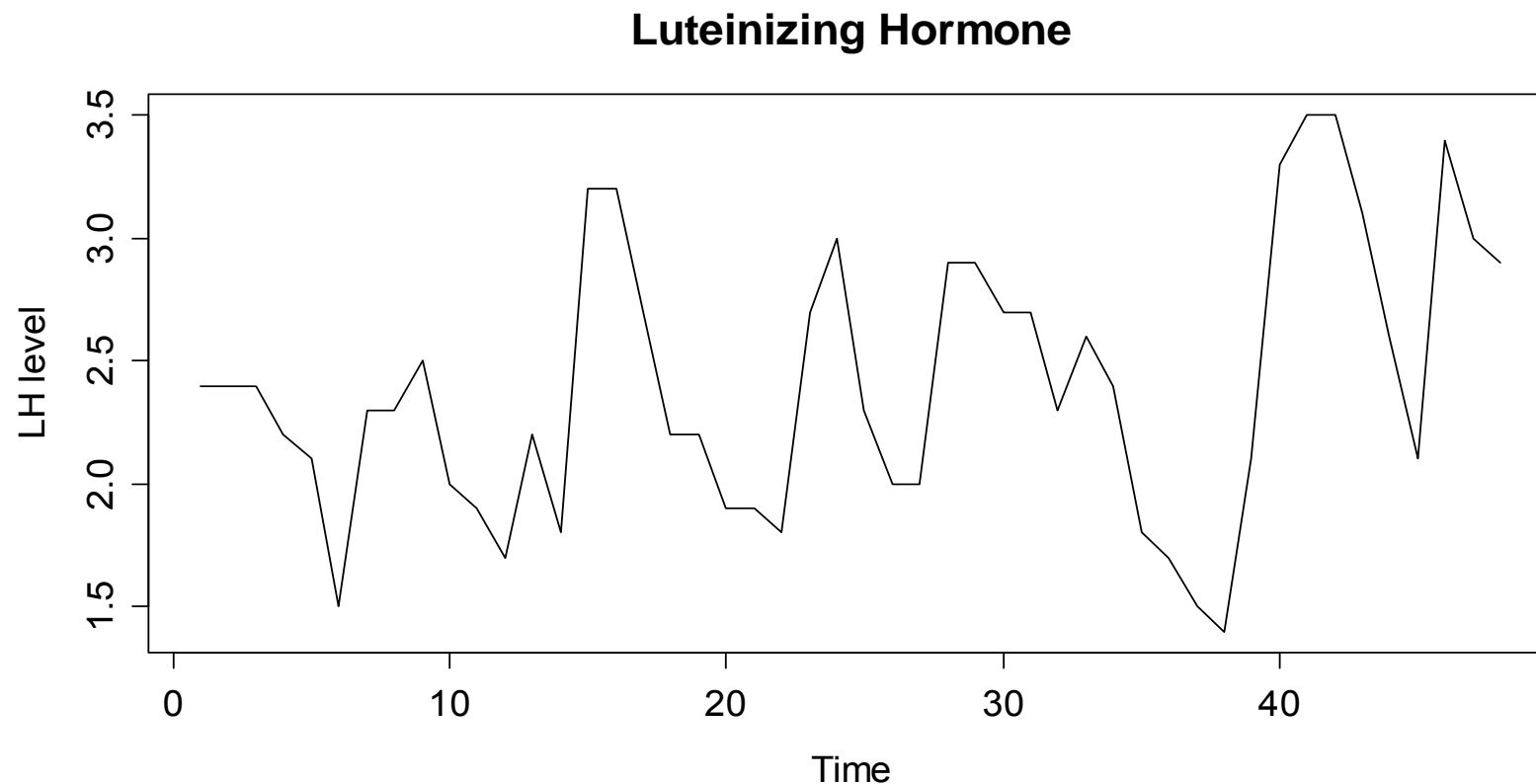


Applied Time Series Analysis

SS 2015 – Introduction

Example 3: Luteinizing Hormone

```
> data(lh)  
> plot(lh, ylab="LH level", main="Luteinizing Hormone")
```

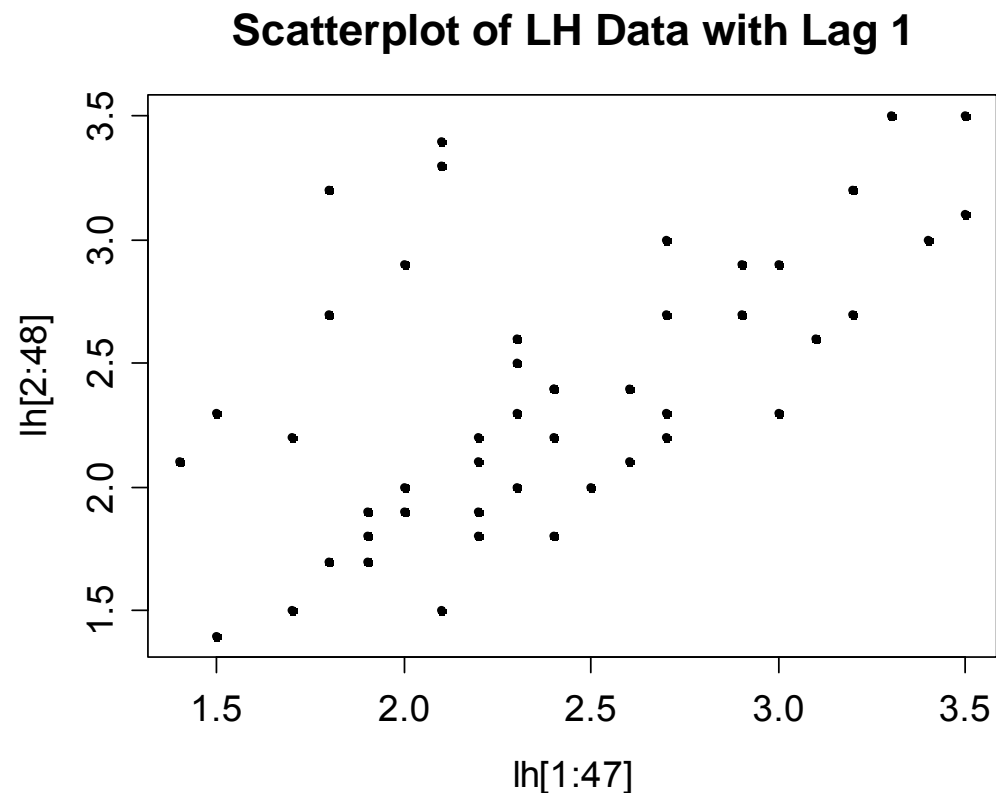


Applied Time Series Analysis

SS 2015 – Introduction

Example 3: Lagged Scatterplot

```
> plot(lh[1:47], lh[2:48], pch=20)  
> title("Scatterplot of LH Data with Lag 1")
```



Applied Time Series Analysis

SS 2015 – Introduction

Example 4: Swiss Market Index

We have a multiple time series object:

```
> data(EuStockMarkets)
> EuStockMarkets
Time Series:
Start = c(1991, 130)
End = c(1998, 169)
Frequency = 260
```

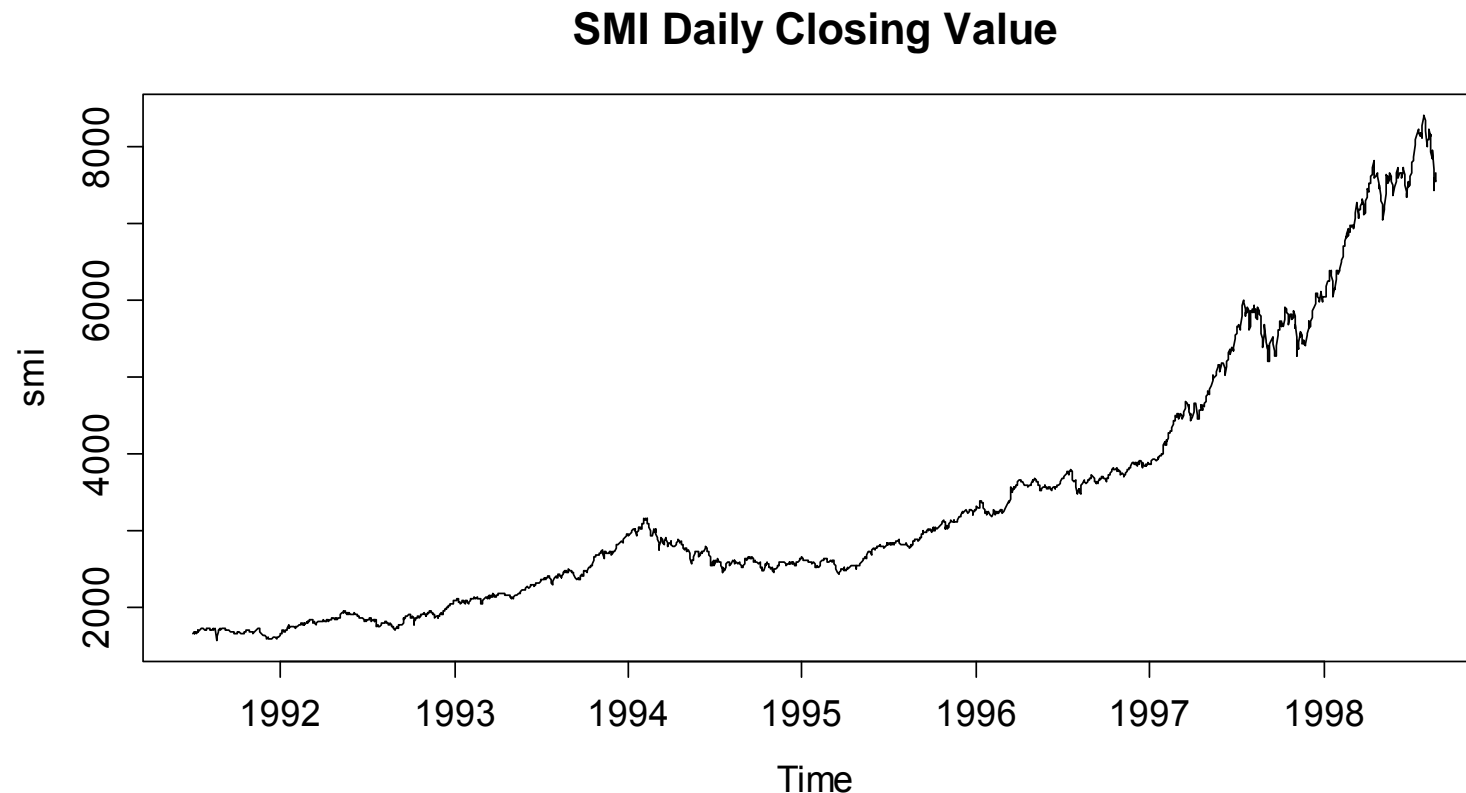
	DAX	SMI	CAC	FTSE
1991.496	1628.75	1678.1	1772.8	2443.6
1991.500	1613.63	1688.5	1750.5	2460.2
1991.504	1606.51	1678.6	1718.0	2448.2
1991.508	1621.04	1684.1	1708.1	2470.4
1991.512	1618.16	1686.6	1723.1	2484.7
1991.515	1610.61	1671.6	1714.3	2466.8

Applied Time Series Analysis

SS 2015 – Introduction

Example 4: Swiss Market Index

```
> smi <- ts(tmp, start=start(esm), freq=frequency(esm))  
> plot(smi, main="SMI Daily Closing Value")
```

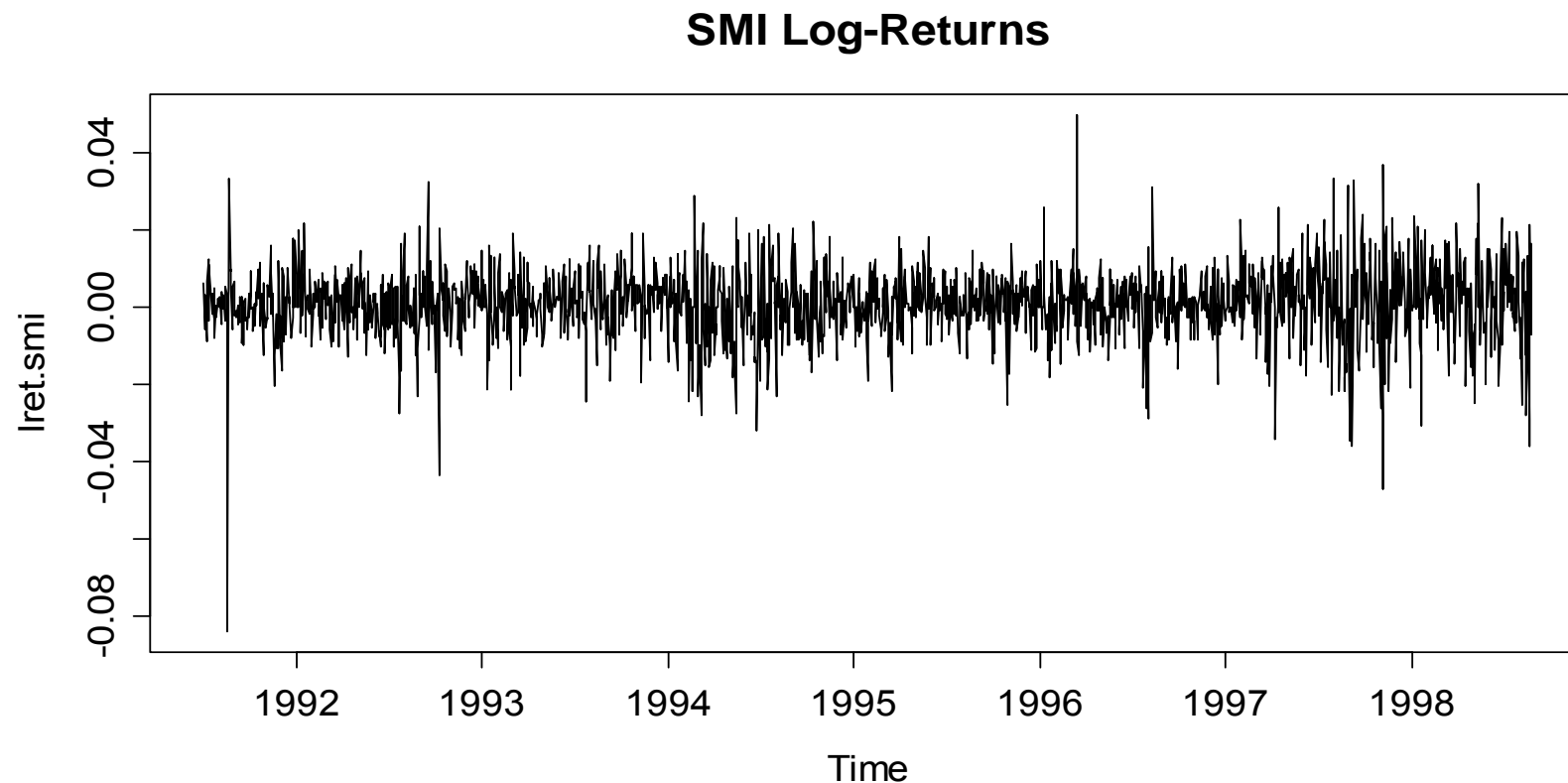


Applied Time Series Analysis

SS 2015 – Introduction

Example 4: Swiss Market Index

```
> lret.smi <- log(smi[2:1860]/smi[1:1859])  
> plot(lret.smi, main="SMI Log-Returns")
```



Applied Time Series Analysis

SS 2015 – Introduction

Goals in Time Series Analysis

1) Exploratory Analysis

Visualization of the properties of the series

- *time series plot*
- *decomposition into trend/seasonal pattern/random error*
- *correlogram for understanding the dependency structure*

2) Modeling

Fitting a stochastic model to the data that represents and reflects the most important properties of the series

- *done exploratory or with previous knowledge*
- *model choice and parameter estimation is crucial*
- *inference: how well does the model fit the data?*

Applied Time Series Analysis

SS 2015 – Introduction

Goals in Time Series Analysis

3) Forecasting

Prediction of future observations with measure of uncertainty

- *mostly model based, uses dependency and past data*
- *is an extrapolation, thus often to take with a grain of salt*
- *similar to driving a car by looking in the rear window mirror*

4) Process Control

The output of a (physical) process defines a time series

- *a stochastic model is fitted to observed data*
- *this allows understanding both signal and noise*
- *it is feasible to monitor normal/abnormal fluctuations*

Applied Time Series Analysis

SS 2015 – Introduction

Goals in Time Series Analysis

5) Time Series Regression

Modeling response time series using 1 or more input series

$$Y_t = \beta_0 + \beta_1 u_t + \beta_2 v_t + E_t$$

where E_t is independent of u_t and v_t , but not i.i.d.

Example: $(\text{Ozone})_t = (\text{Wind})_t + (\text{Temperature})_t + E_t$

Fitting this model under i.i.d error assumption:

- leads to unbiased estimates, but...
- often grossly wrong standard errors
- thus, confidence intervals and tests are misleading

Applied Time Series Analysis

SS 2015 – Mathematical Concepts

Stochastic Model for Time Series

Def: A *time series process* is a set $\{X_t, t \in T\}$ of random variables, where T is the set of times. Each of the random variables $X_t, t \in T$ has a univariate probability distribution F_t .

- If we exclusively consider time series processes with equidistant time intervals, we can enumerate $\{T = 1, 2, 3, \dots\}$
- An observed time series is a realization of $X = (X_1, \dots, X_n)$, and is denoted with small letters as $x = (x_1, \dots, x_n)$.
- We have a multivariate distribution, but only 1 observation (i.e. 1 realization from this distribution) is available. In order to perform “statistics”, we require some additional structure.

Applied Time Series Analysis

SS 2015 – Mathematical Concepts

Stationarity

For being able to do statistics with time series, we require that the series “doesn’t change its probabilistic character” over time. This is mathematically formulated by **strict stationarity**.

Def: A time series $\{X_t, t \in \mathbb{T}\}$ is strictly stationary, if the joint distribution of the random vector (X_t, \dots, X_{t+k}) is equal to the one of (X_s, \dots, X_{s+k}) for all combinations of t, s and k .

→

$X_t \sim F$	all X_t are identically distributed
$E[X_t] = \mu$	all X_t have identical expected value
$Var(X_t) = \sigma^2$	all X_t have identical variance
$Cov(X_t, X_{t+h}) = \gamma_h$	the autocov depends only on the lag h

Applied Time Series Analysis

SS 2015 – Mathematical Concepts

Stationarity

It is impossible to „prove“ the theoretical concept of stationarity from data. We can only search for evidence in favor or against it.

However, with strict stationarity, even finding evidence only is too difficult. We thus resort to the concept of *weak stationarity*.

Def: A time series $\{X_t, t \in \mathbb{T}\}$ is said to be *weakly stationary*, if

$$E[X_t] = \mu$$

$$\text{Cov}(X_t, X_{t+h}) = \gamma_h \quad \text{for all lags } h$$

$$\text{and thus also: } \text{Var}(X_t) = \sigma^2$$

Note that weak stationarity is sufficient for „practical purposes“.

Applied Time Series Analysis

SS 2015 – Mathematical Concepts

Testing Stationarity

- In time series analysis, we need to verify whether the series has arisen from a stationary process or not. Be careful: stationarity is a property of the process, and not of the data.
- Treat stationarity as a hypothesis! We may be able to reject it when the data strongly speak against it. However, we can never prove stationarity with data. At best, it is plausible.
- Formal tests for stationarity do exist (→ see scriptum). We discourage their use due to their low power for detecting general non-stationarity, as well as their complexity.

→ **Use the time series plot for deciding on stationarity!**

Applied Time Series Analysis

SS 2015 – Mathematical Concepts

Evidence for Non-Stationarity

- **Trend**, i.e. non-constant expected value
- **Seasonality**, i.e. deterministic, periodical oscillations
- **Non-constant variance**, i.e. multiplicative error
- **Non-constant dependency structure**

Remark:

Note that some periodical oscillations, as for example in the lynx trappings data, can be stochastic and thus, the underlying process is assumed to be stationary. However, the boundary between the two is fuzzy.

Applied Time Series Analysis

SS 2015 – Mathematical Concepts

Strategies for Detecting Non-Stationarity

1) Time series plot

- non-constant expected value (trend/seasonal effect)
- changes in the dependency structure
- non-constant variance

2) Correlogram (presented later...)

- non-constant expected value (trend/seasonal effect)
- changes in the dependency structure

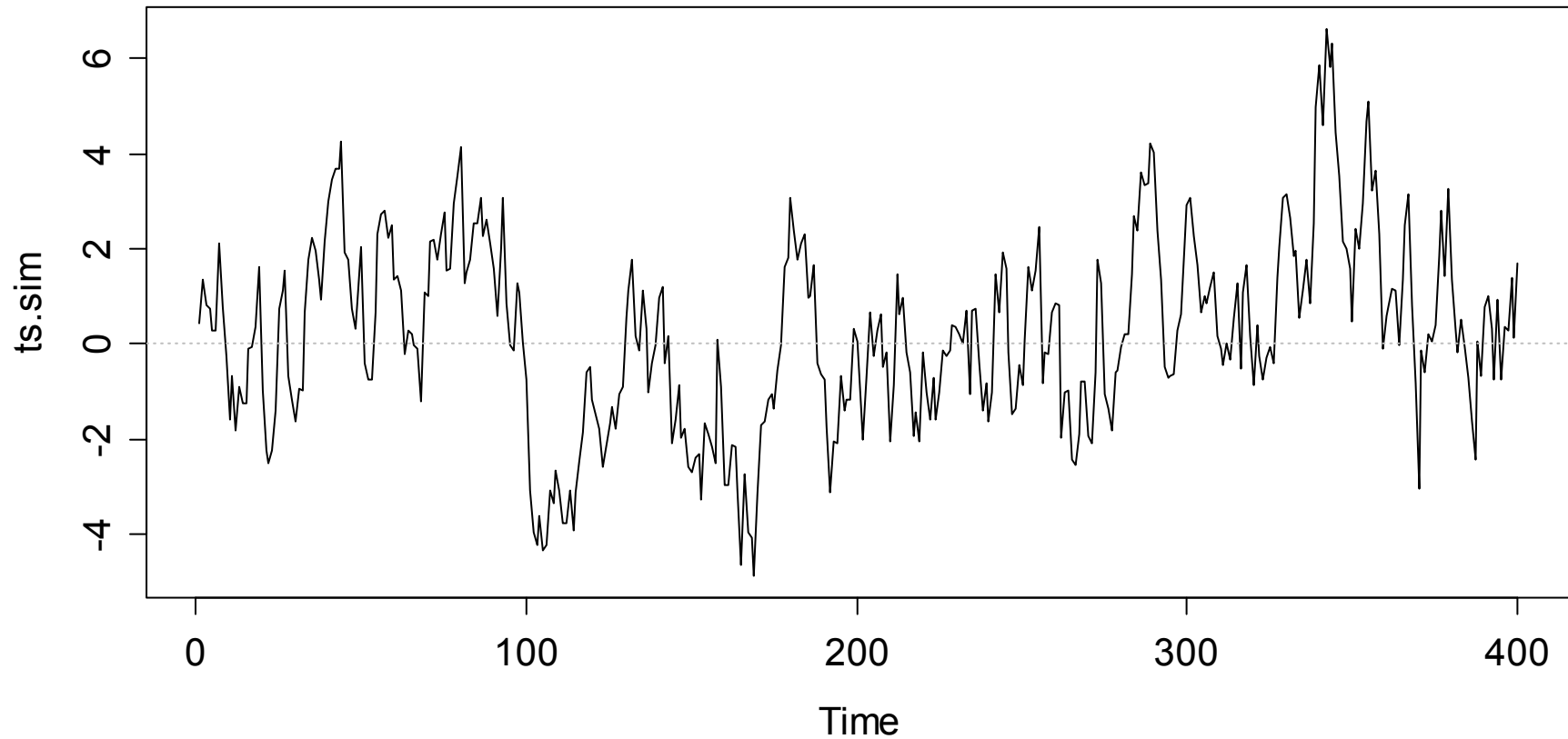
A (sometimes) useful trick, especially when working with the correlogram, is to split up the series in two or more parts, and producing plots for each of the pieces separately.

Applied Time Series Analysis

SS 2015 – Mathematical Concepts

Example: Simulated Time Series 1

Simulated Time Series Example

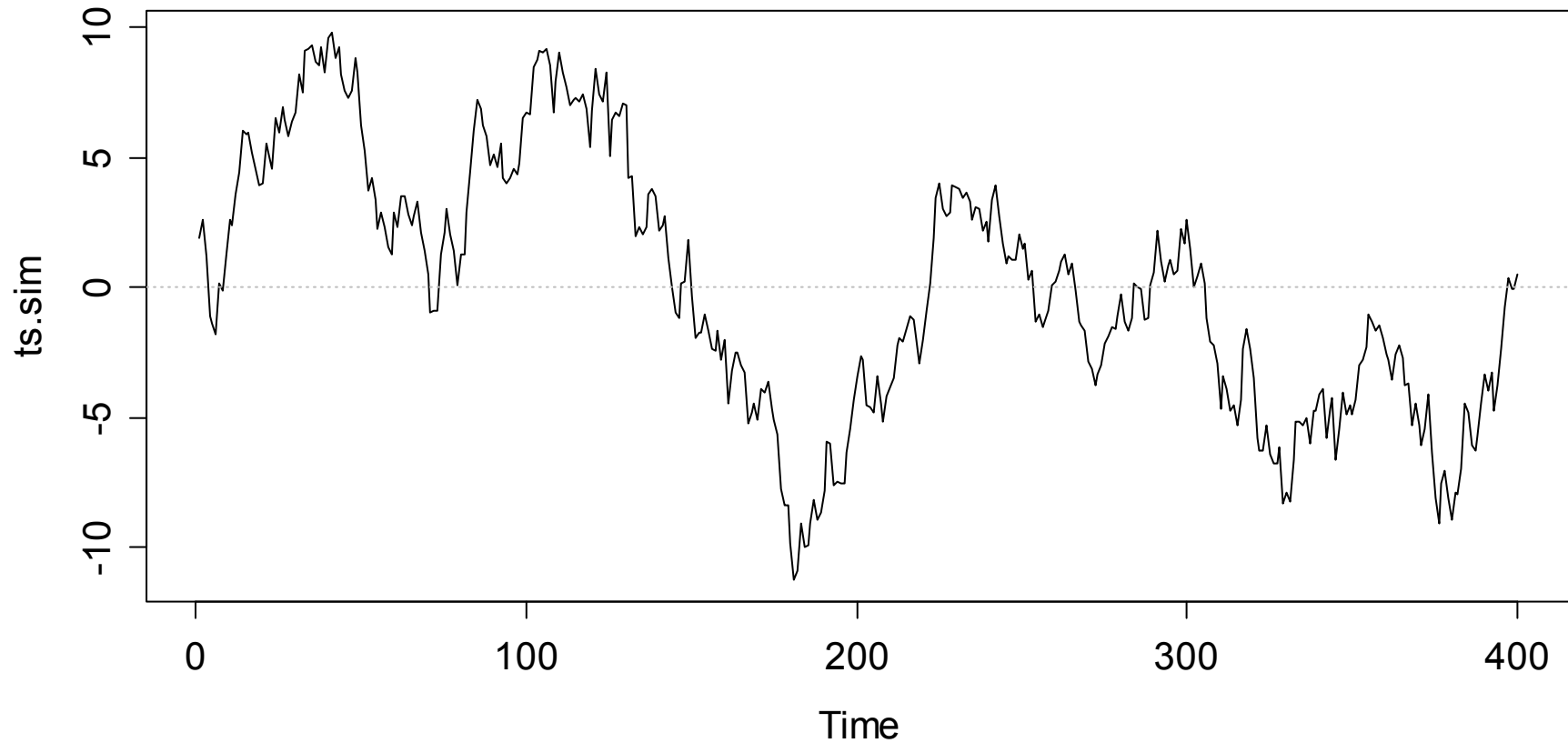


Applied Time Series Analysis

SS 2015 – Mathematical Concepts

Example: Simulated Time Series 2

Simulated Time Series Example

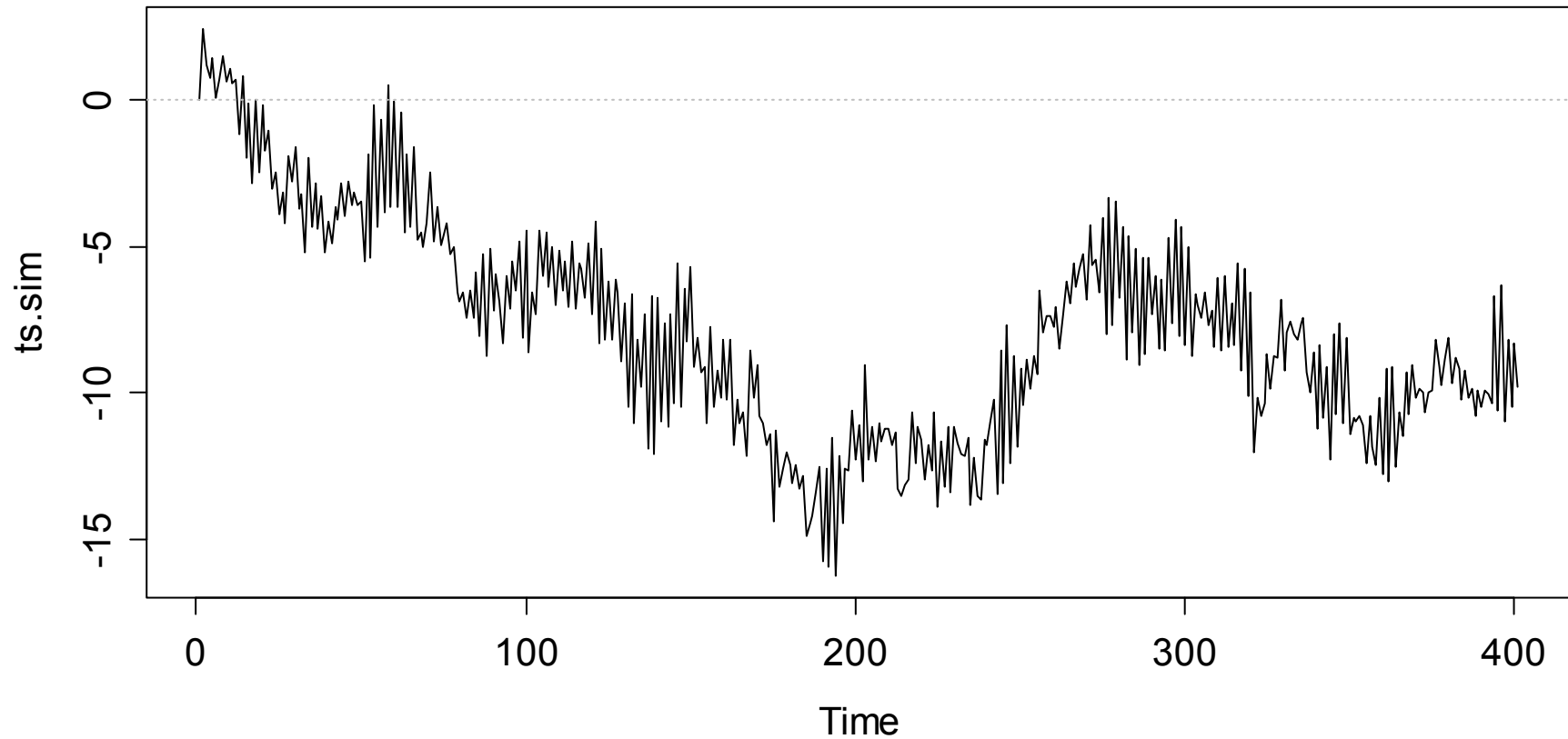


Applied Time Series Analysis

SS 2015 – Mathematical Concepts

Example: Simulated Time Series 3

Simulated Time Series Example

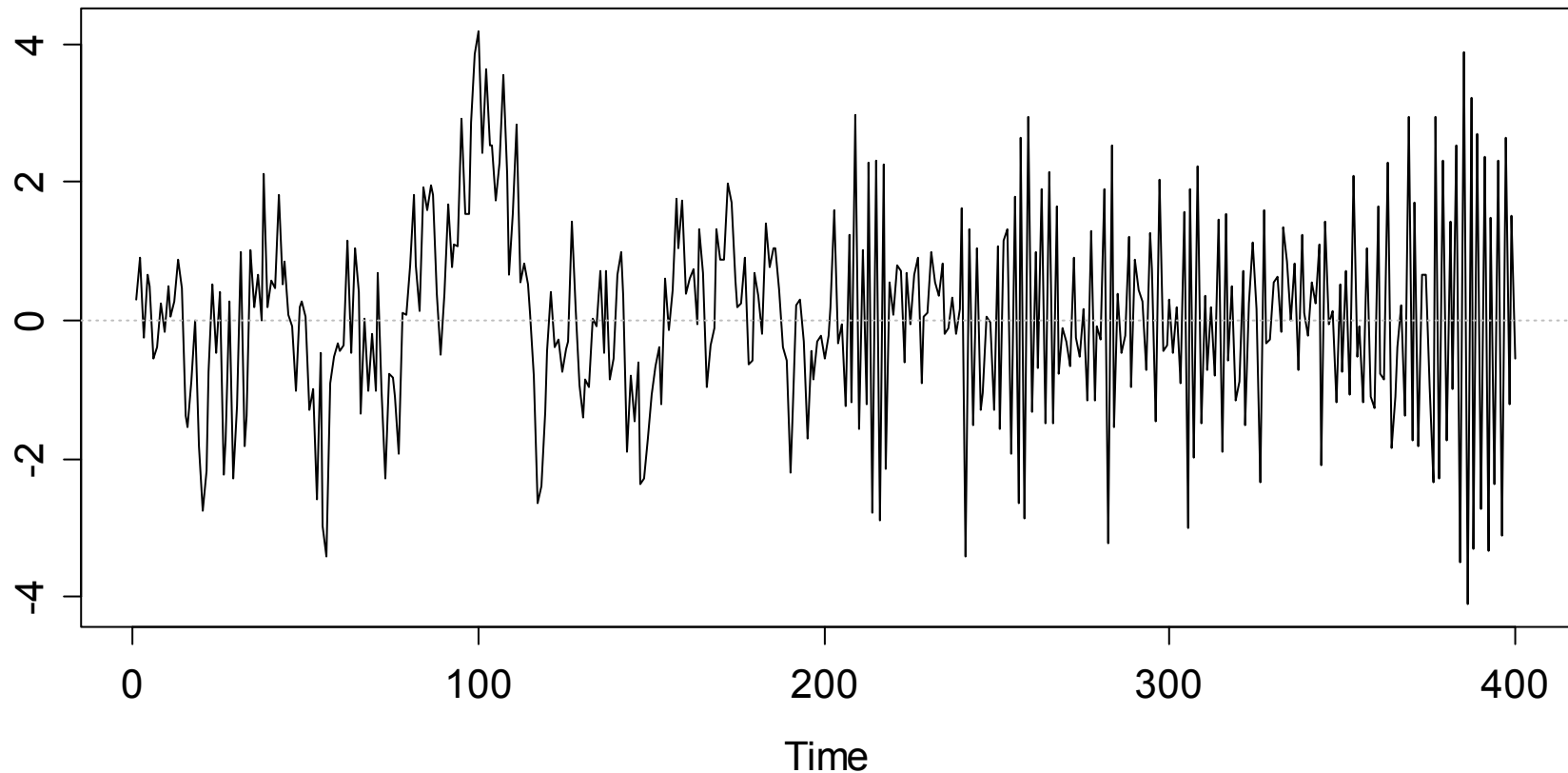


Applied Time Series Analysis

SS 2015 – Mathematical Concepts

Example: Simulated Time Series 4

Simulated Time Series Example



Applied Time Series Analysis

SS 2015 – Time Series in R

Time Series in R

- In **R**, there are *objects*, which are organized in a large number of *classes*. These classes e.g. include *vectors*, *data frames*, *model output*, *functions*, and many more. Not surprisingly, there are also *several classes for time series*.
- We focus on **ts**, the basic class for regularly spaced time series in **R**. This class is comparably simple, as it can only represent time series with *fixed interval records*, and *only uses numeric time stamps*, i.e. enumerates the index set.
- For defining a **ts** object, we have to supply the *data*, but also the *starting time* (as argument *start*), and the *frequency* of measurements as argument *frequency*.

Applied Time Series Analysis

SS 2015 – Time Series in R

Time Series in R: Example

Data: number of days per year with traffic holdups in front of the Gotthard road tunnel north entrance in Switzerland.

2004	2005	2006	2007	2008	2009	2010
88	76	112	109	91	98	139

```
> rawdat <- c(88, 76, 112, 109, 91, 98, 139)
```

```
> ts.dat <- ts(rawdat, start=2004, freq=1)
```

```
> ts.dat
```

```
Time Series: Start = 2004
```

```
End = 2010; Frequency = 1
```

```
[1] 88 76 112 109 91 98 139
```


Applied Time Series Analysis

SS 2015 – Time Series in R

Time Series in R: Example

```
> plot(ts.dat, ylab="# of Days", main="Traffic Holdups")
```



Applied Time Series Analysis

SS 2015 – Time Series in R

Further Topics in R

The scriptum discusses some further topics which are of interest when doing time series analysis in R:

- *Handling of dates and times in R*
 - *Reading/Importing data into R*
- **Please thoroughly read and study these chapters. Examples will be shown/discussed in the exercises.**

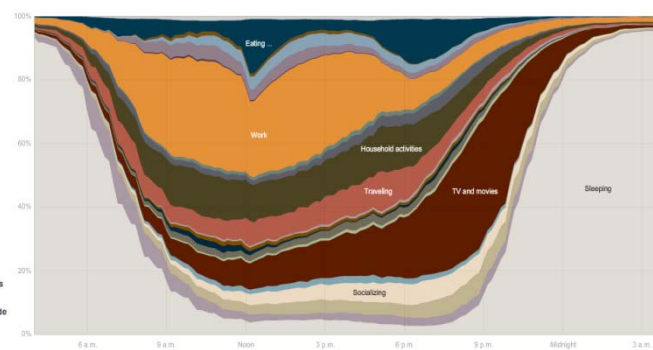
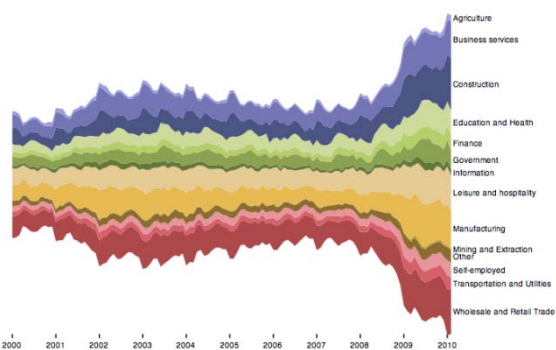
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Descriptive Analysis

As always, when working with data, it is important to first gain an overview. In time series analysis, the following is required:

- *Understanding the context of the data and the data source*
- *Making suitable plots, looking for structure and outliers*
- *Thinking about transformations, e.g. to reduce skewness*
- *Judging stationarity and achieve it by decomposition*
- *For stationary series, the analysis of autocorrelations*

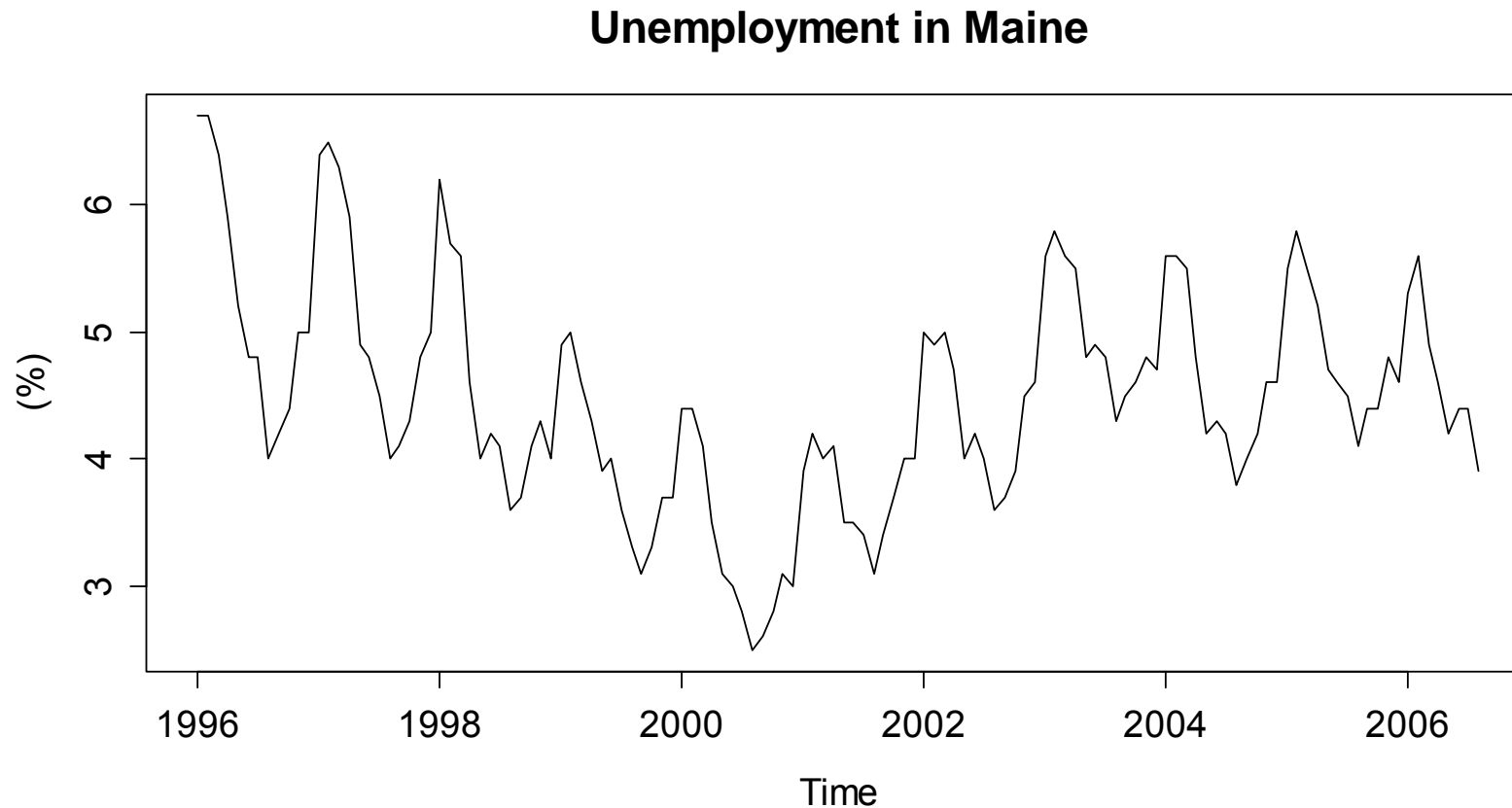


Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Visualization: Time Series Plot

```
> plot(tsd, ylab="(%)", main="Unemployment in Maine")
```



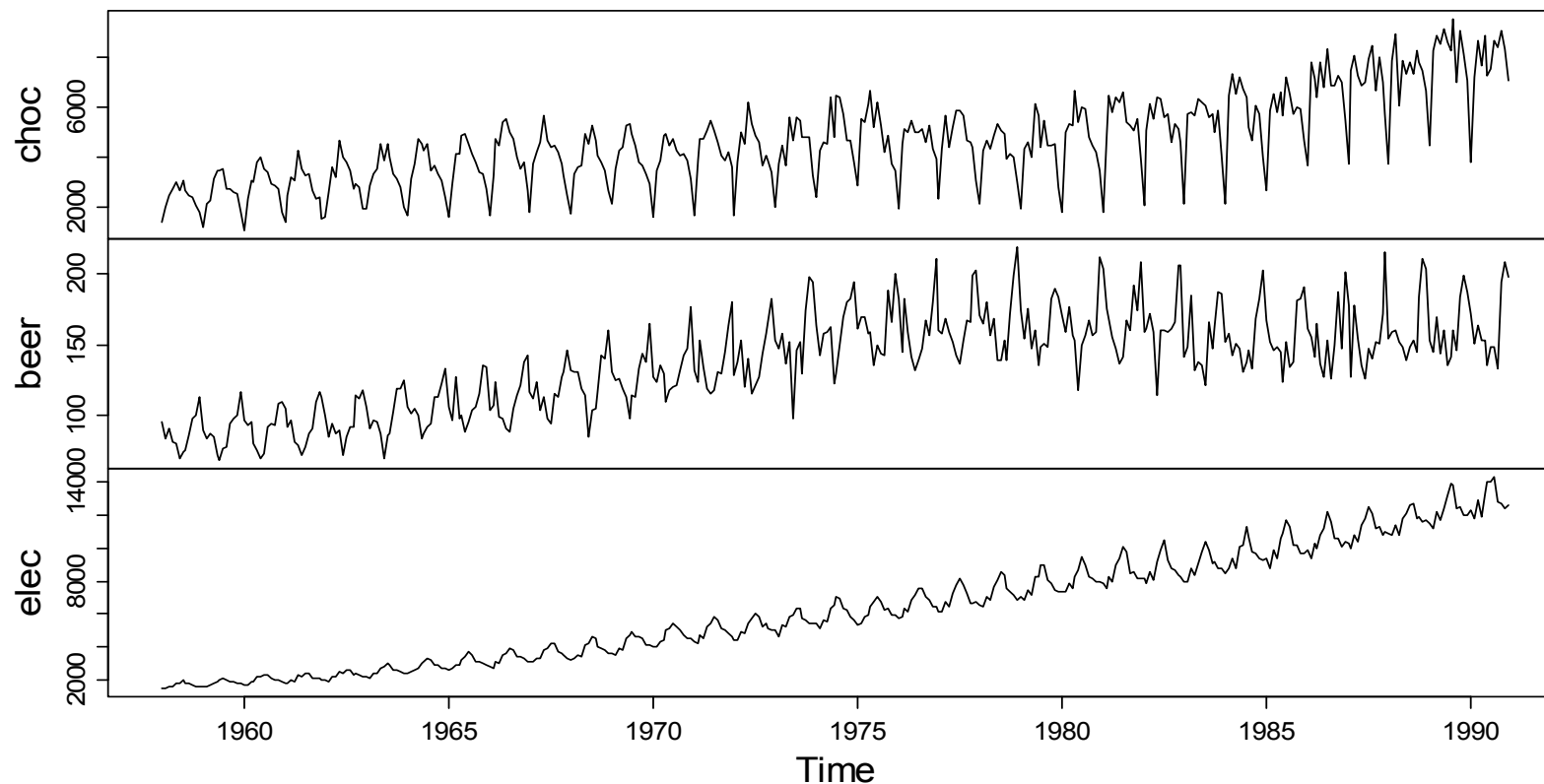
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Multiple Time Series Plots

```
> plot(tsd, main="Chocolate, Beer & Electricity")
```

Chocolate, Beer & Electricity



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Only One or Multiple Frames?

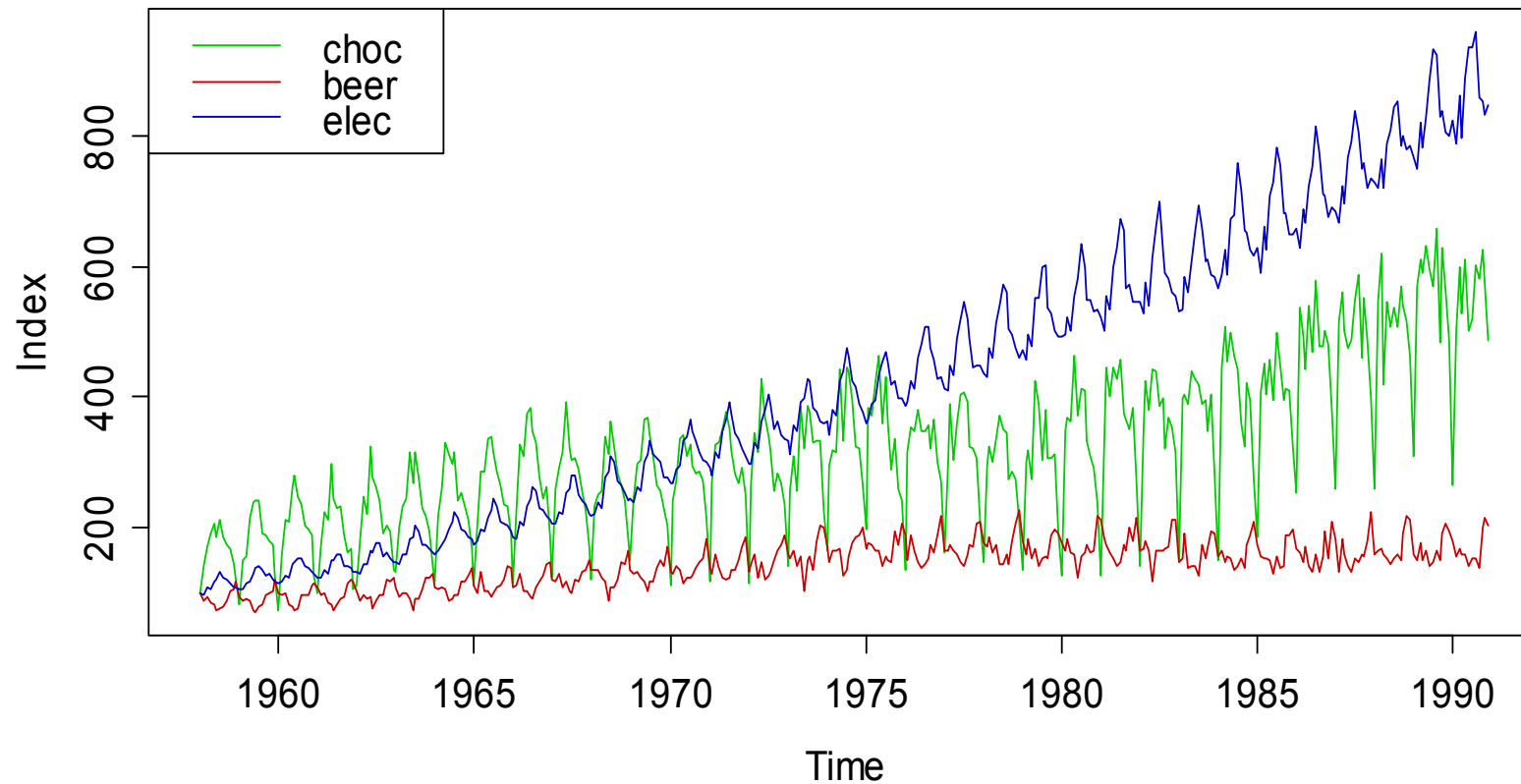
- Due to different scale/units it is often impossible to directly plot multiple time series in one single frame. Also, multiple frames are convenient for visualizing the series.
- If the relative development of multiple series is of interest, then we can (manually) index the series and (manually) plot them into one single frame.
- This clearly shows the magnitudes for trend and seasonality. However, the original units are lost.
- For details on how indexing is done, see the scriptum.

Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Multiple Time Series Plots

Indexed Chocolate, Beer & Electricity



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Transformations

For strictly stationary time series, we have: $X_t \sim F$

We did not specify the distribution F and there is no restriction to it. However, many popular time series models are based on:

- 1) *Gaussian distribution*
- 2) *linear relations between the variables*

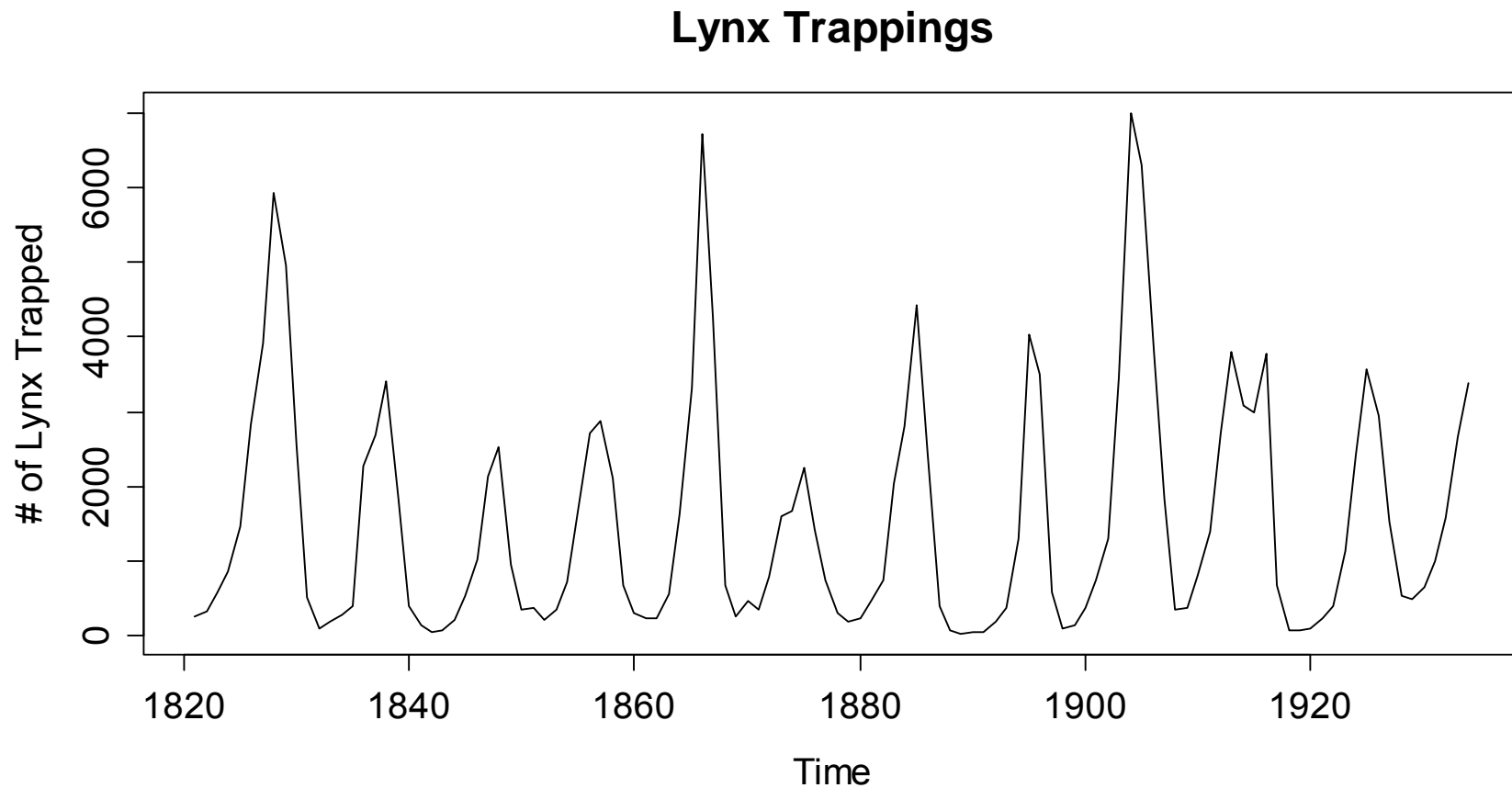
If the data show different behaviour, we can often improve the situation by transforming x_1, \dots, x_n to $g(x_1), \dots, g(x_n)$. The most popular and practically relevant transformation is:

$$g(\cdot) = \log(\cdot)$$

Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Transformations: Lynx Data

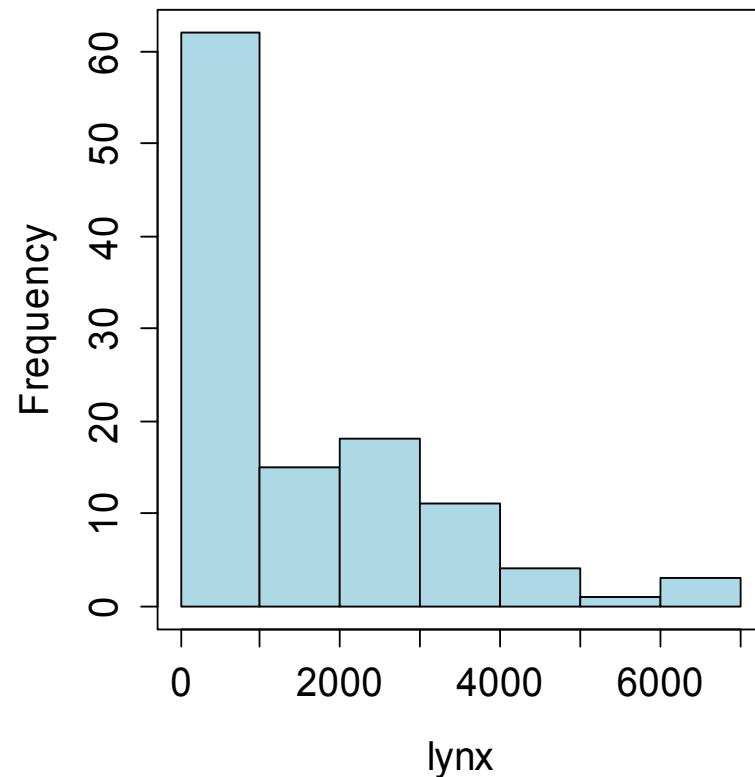


Applied Time Series Analysis

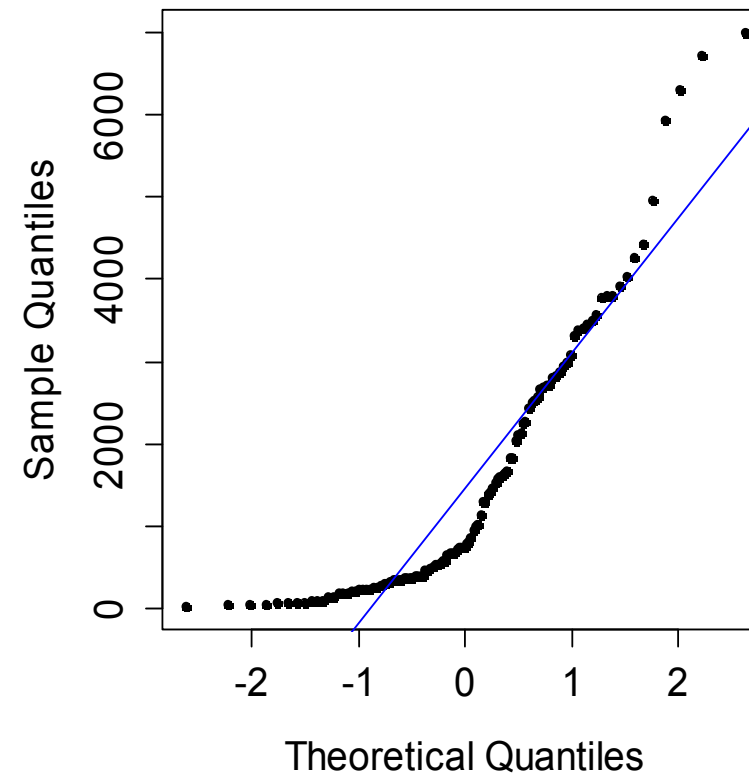
SS 2015 – Descriptive Analysis

Transformations: Lynx Data

Histogram of lynx



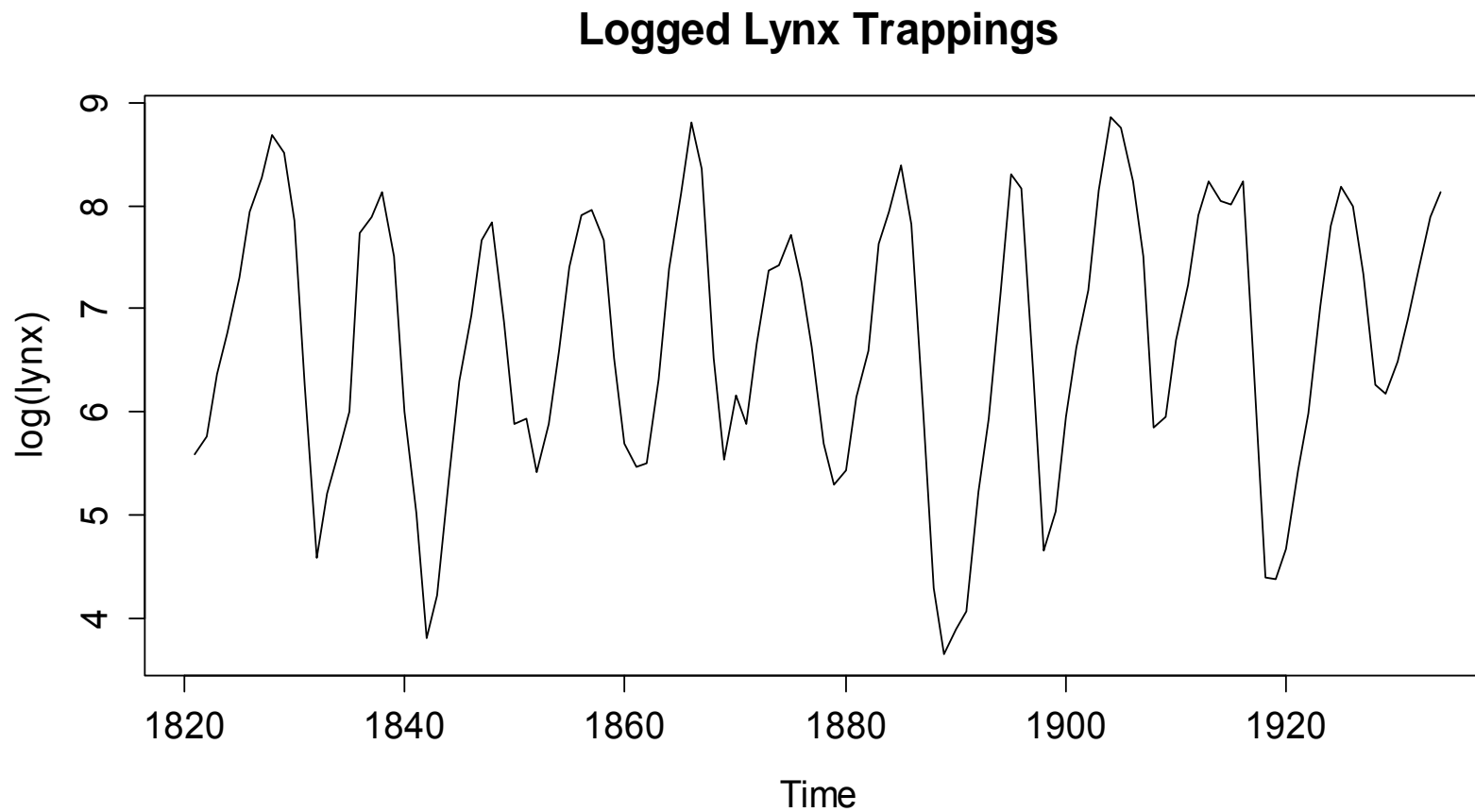
Normal Q-Q Plot



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Transformations: Lynx Data



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

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)*

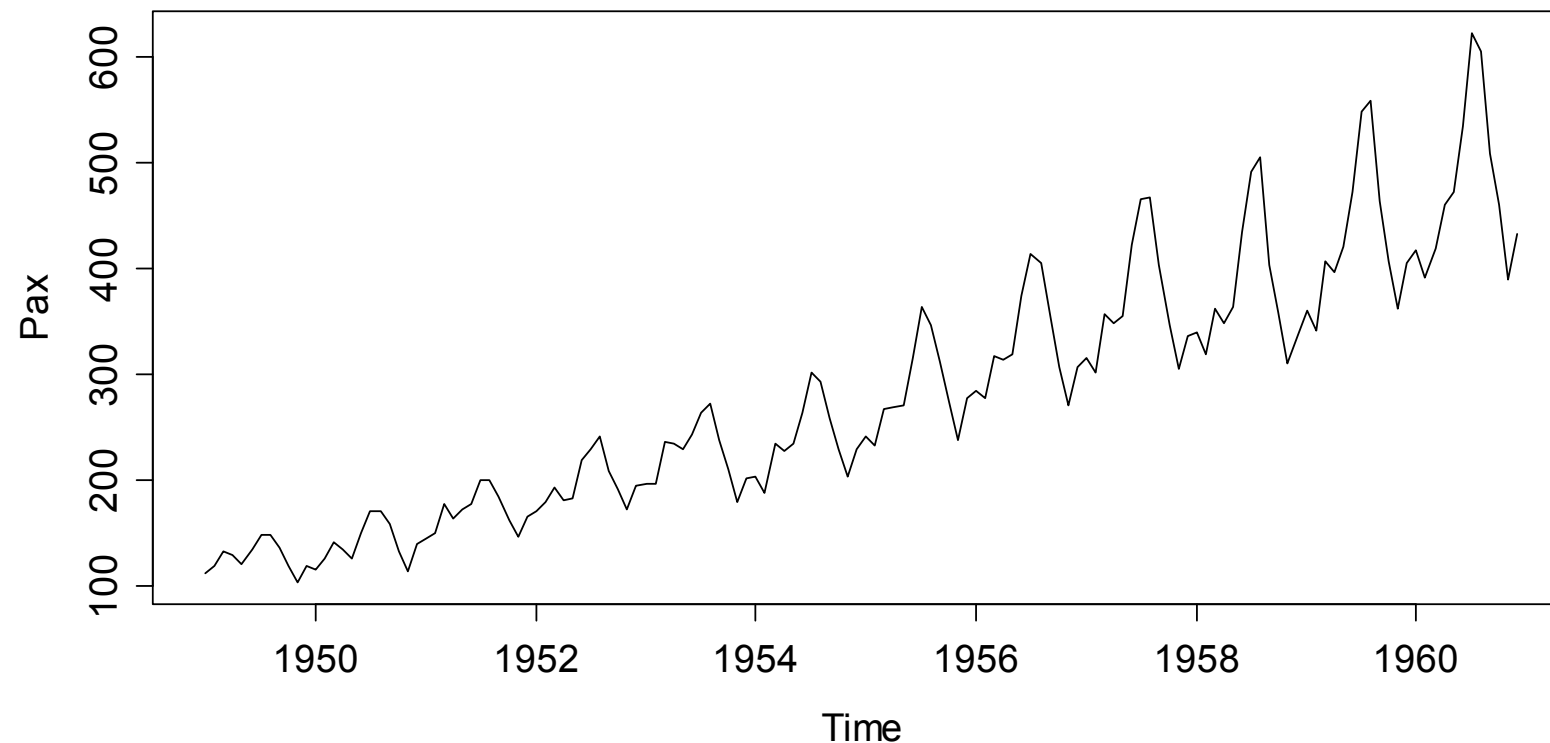
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Multiplicative Decomposition

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

Passenger Bookings



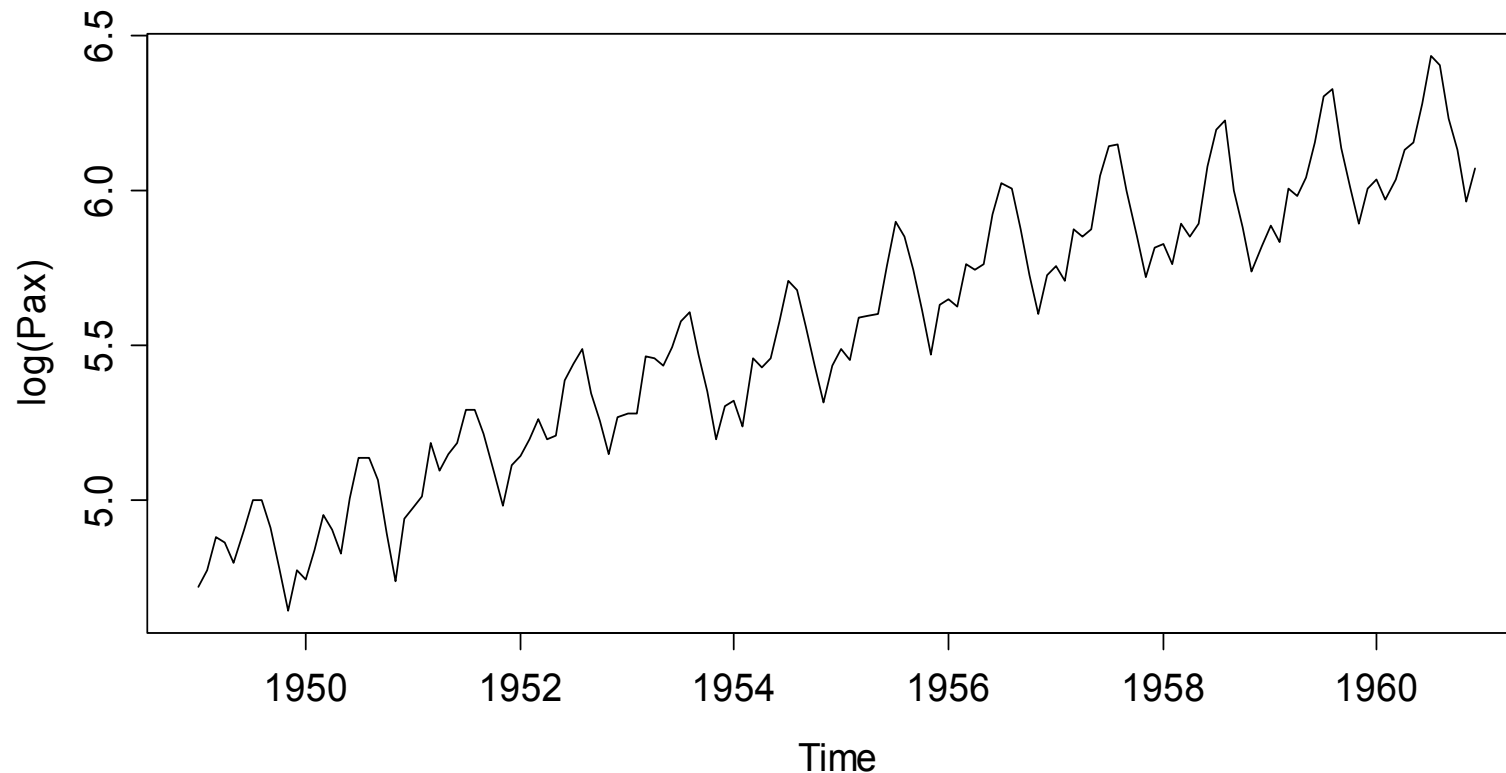
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Multiplicative Decomposition

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

Logged Passenger Bookings



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Differencing: Removing a Trend

→ see blackboard...

Summary:

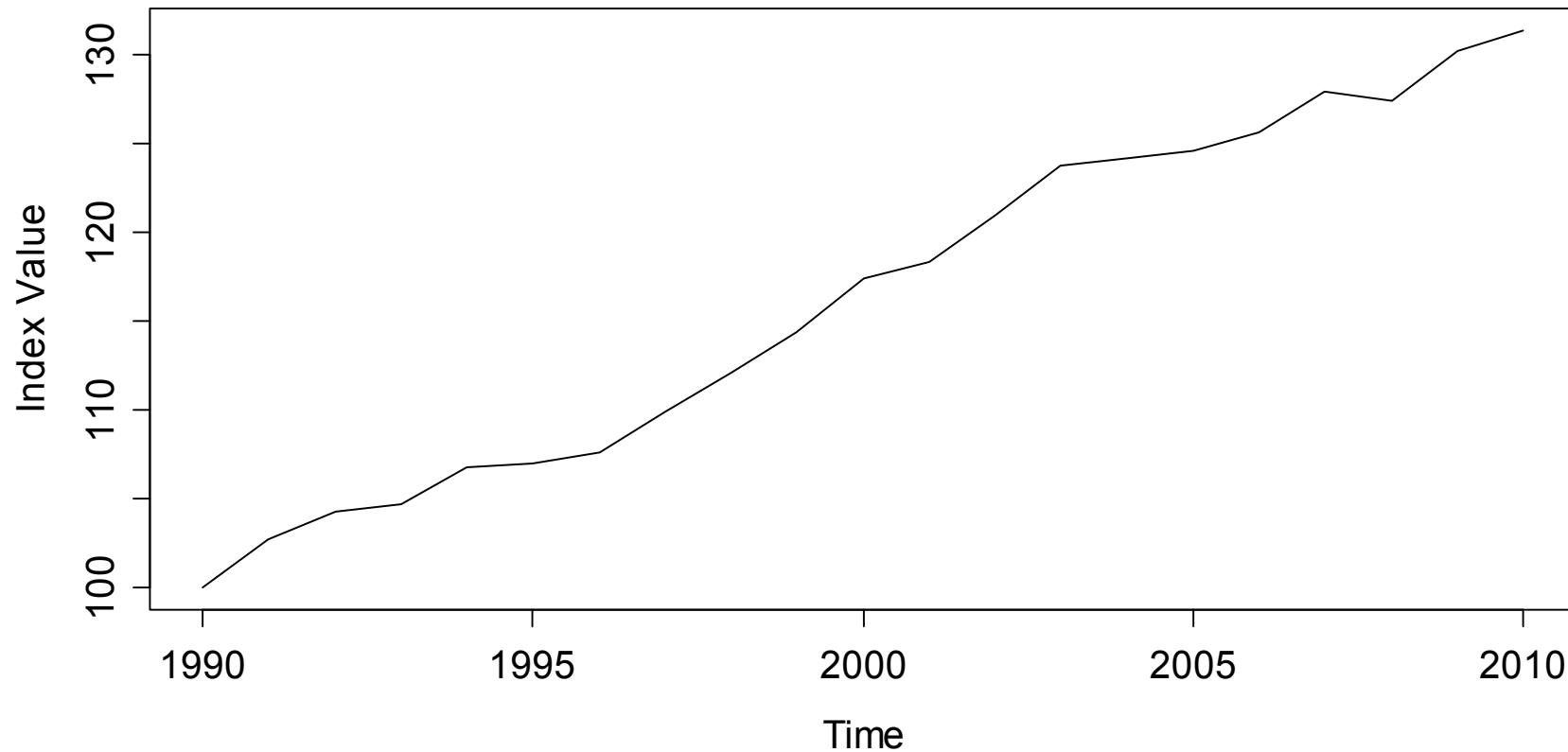
- Differencing means analyzing the observation-to-observation changes in the series, but no longer the original.
- This may (or may not) remove trend/seasonality, but does not yield estimates for m_t and s_t , and not even for R_t .
- Differencing changes the dependency in the series, i.e it artificially creates new correlations.

Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Differencing: Example

Swiss Traffic Index

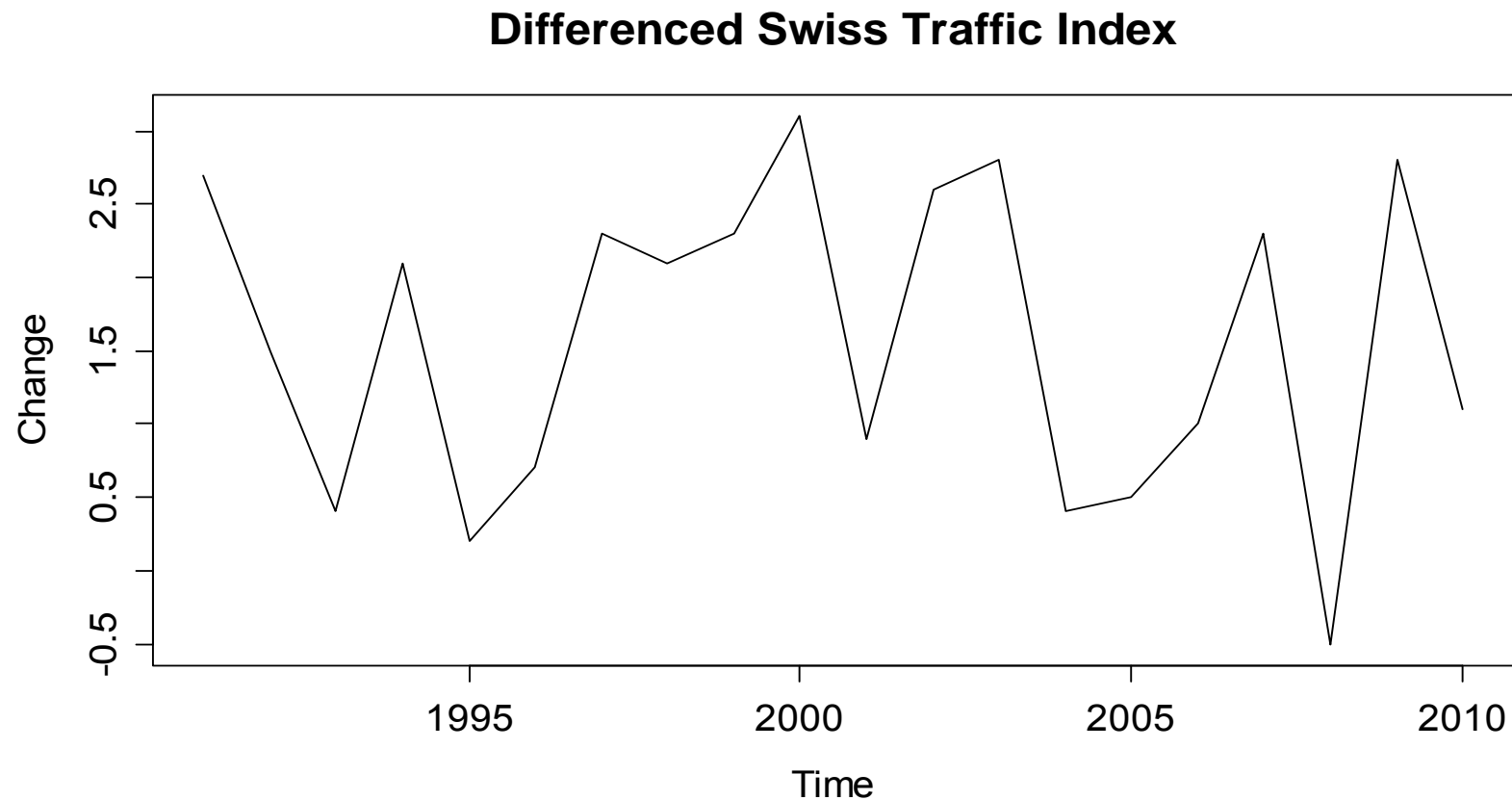


Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Differencing: Example

```
> plot(diff(SwissTraffic), main=...)
```



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Differencing: Further Remarks

- If log-transformed series are difference (i.e. the SMI series), we are considering (an approximation to) the relative changes:

$$Y_t = \log(X_t) - \log(X_{t-1}) = \log\left(\frac{X_t}{X_{t-1}}\right) = \log\left(\frac{X_t - X_{t-1}}{X_{t-1}} + 1\right) \approx \frac{X_t - X_{t-1}}{X_{t-1}}$$

- The backshift operator “go back 1 step” allows for convenient notation with all differencing operations:

Backshift operator: $B(X_t) = X_{t-1}$

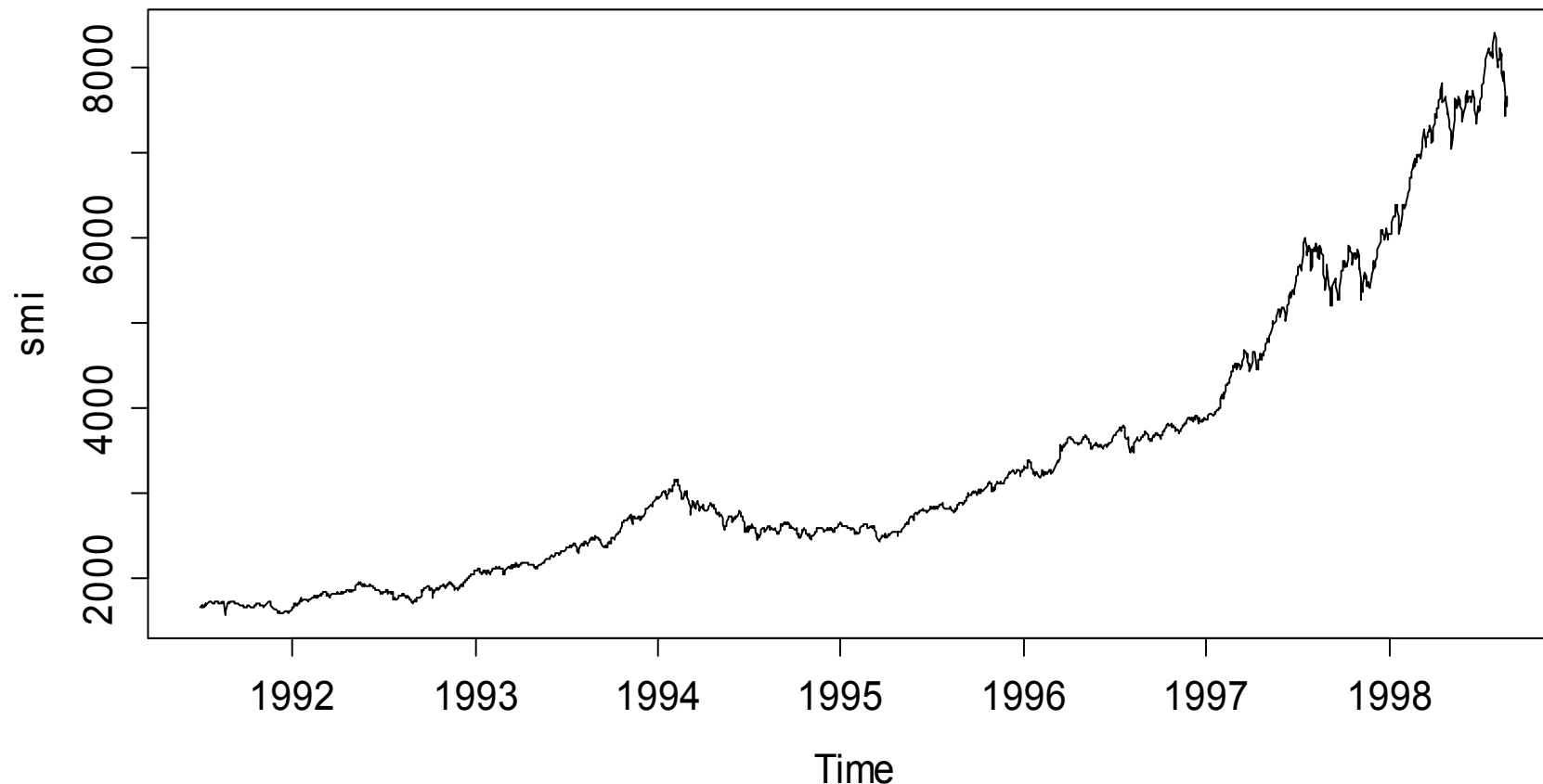
Differencing: $Y_t = (1 - B)X_t = X_t - X_{t-1}$

Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Differencing Series with Transformation

SMI Daily Closing Value

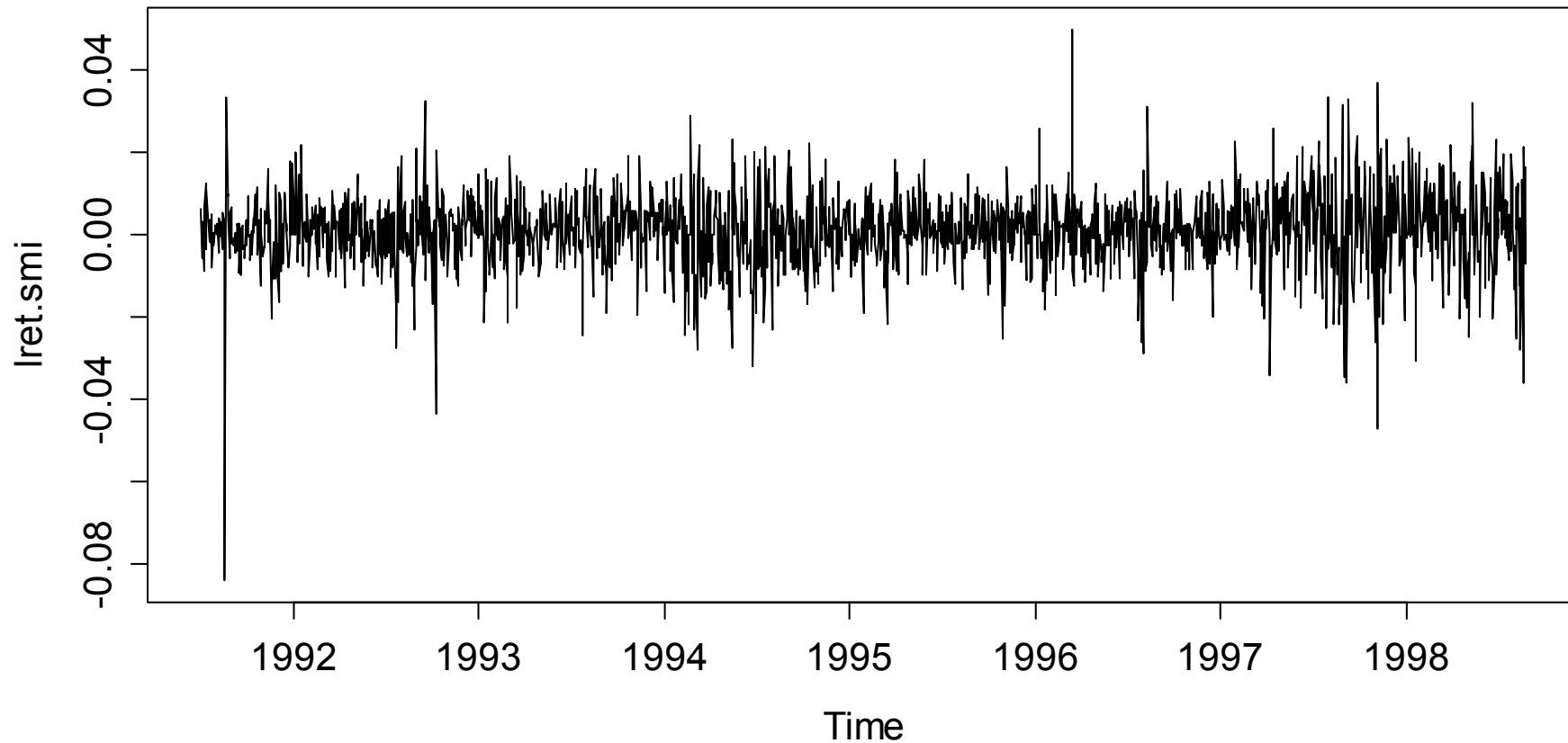


Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Differencing Series with Transformation

SMI Log>Returns



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Higher-Order Differencing

The “normal” differencing from above managed to remove any linear trend from the data. In case of polynomial trend, that is no longer true. But we can take higher-order differences:

$$\begin{aligned}X_t &= \alpha + \beta_1 t + \beta_2 t^2 + R_t, \quad R_t \text{ stationary} \\Y_t &= (1-B)^2 X_t \\&= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\&= R_t - 2R_{t-1} + R_{t-2} + 2\beta_2\end{aligned}$$

A quadratic trend can be removed by taking second-order differences. However, what we obtain is not an estimate of the remainder term R_t , but something that is much more complicated.

Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Removing Seasonal Effects

Time series with seasonal effects can be made stationary through differencing by comparing to the previous periods' value.

$$Y_t = (1 - B^p)X_t = X_t - X_{t-p}$$

- Here, p is the frequency of the series.
- A potential trend which is exactly linear will be removed by the above form of seasonal differencing.
- In practice, trends are rarely linear but slowly varying: $m_t \approx m_{t-1}$
However, here we compare m_t with m_{t-p} , which means that seasonal differencing often fails to remove trends completely.

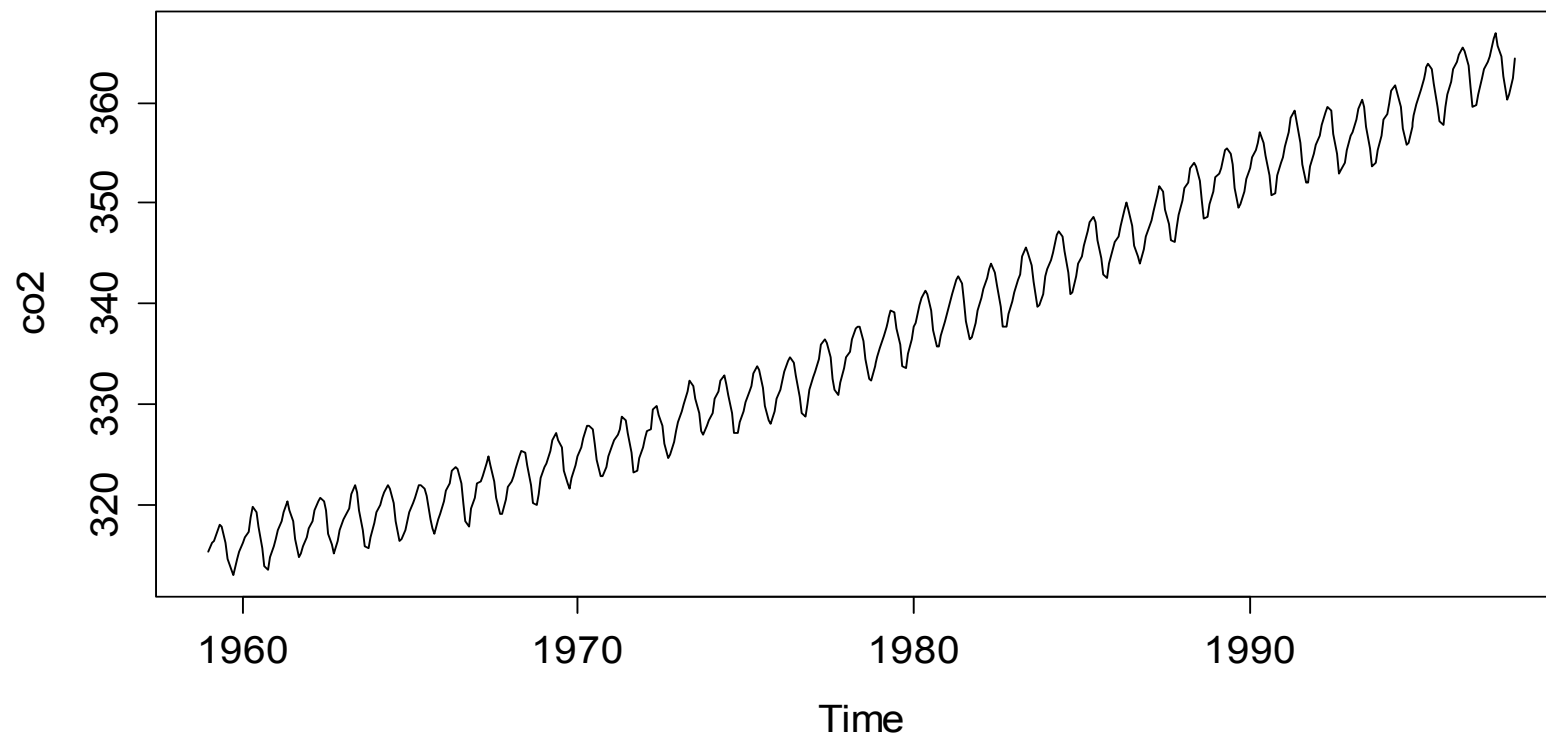
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Seasonal Differencing: Example

```
> data(co2); plot(co2, main=...)
```

Mauna Loa CO2 Concentrations



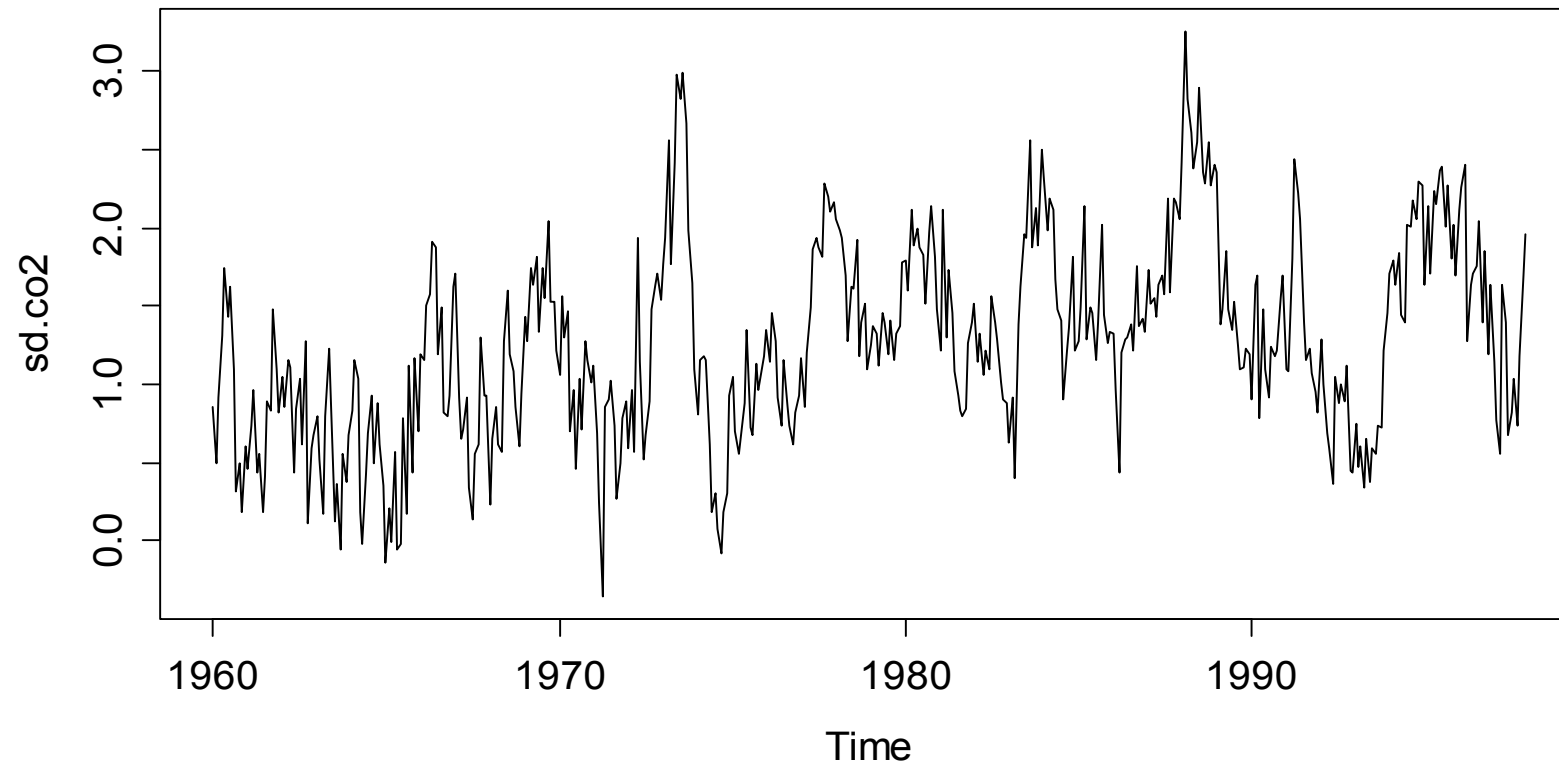
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Seasonal Differencing: Example

```
> sd.co2 <- diff(co2, lag=12)
```

Differenced Mauna Loa Data (p=12)



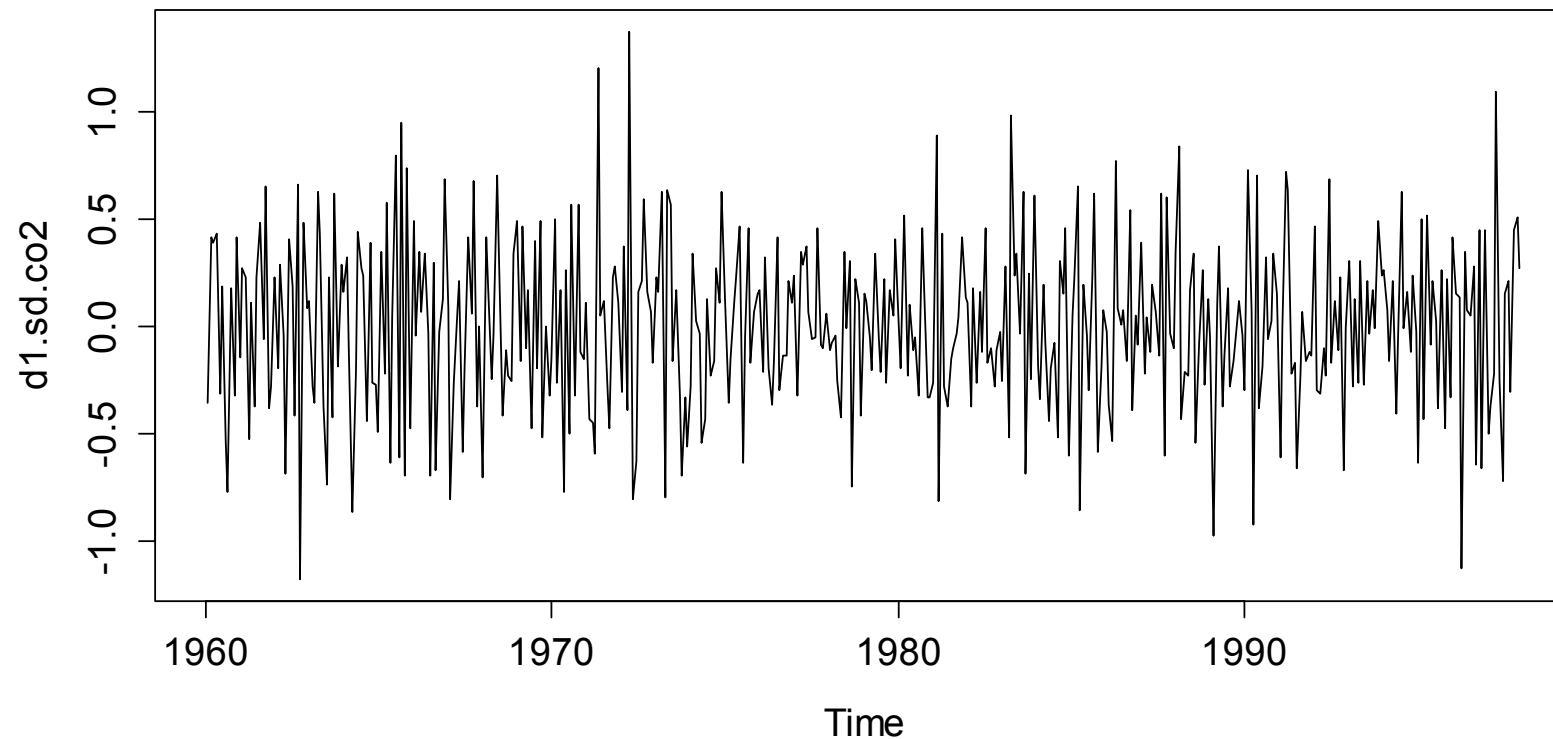
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Seasonal Differencing: Example

This is: $Z_t = (1 - B)Y_t = (1 - B)(1 - B^{12})X_t$

Twice Differenced Mauna Loa Data (p=12, p=1)



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Differencing: Remarks

Some advantages and disadvantages:

- + trend and seasonal effect can be removed
- + procedure is very quick and very simple to implement

- \hat{m}_t , \hat{s}_t and \hat{R}_t are not known, and cannot be visualised
- resulting time series will be shorter than the original
- differencing leads to strong artificial dependencies
- extrapolation of \hat{m}_t , \hat{s}_t is not possible

Applied Time Series Analysis

SS 2015 – Descriptive Analysis

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.

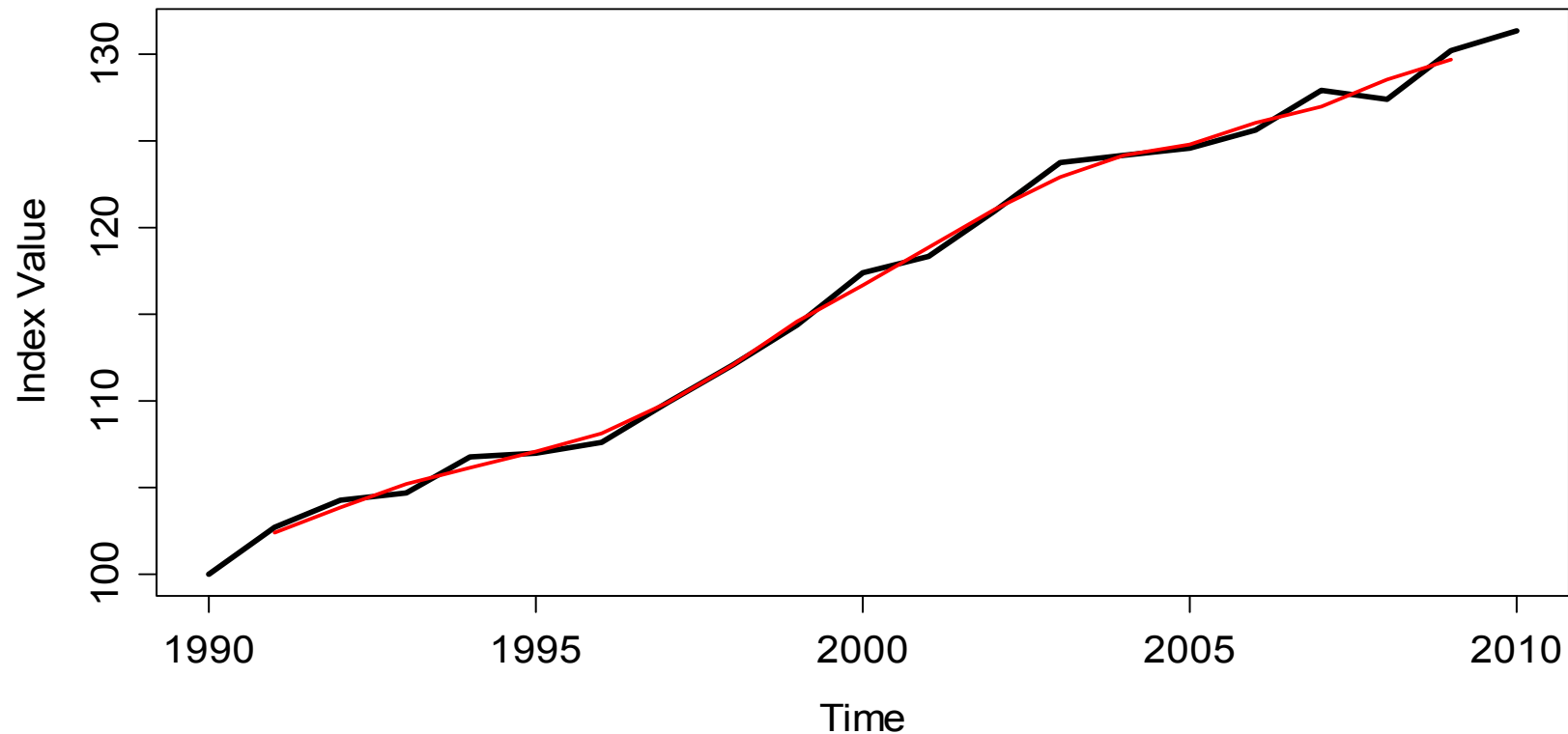
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Trend Estimation with the Running Mean

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

Swiss Traffic Index with Running Mean



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

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:

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

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.

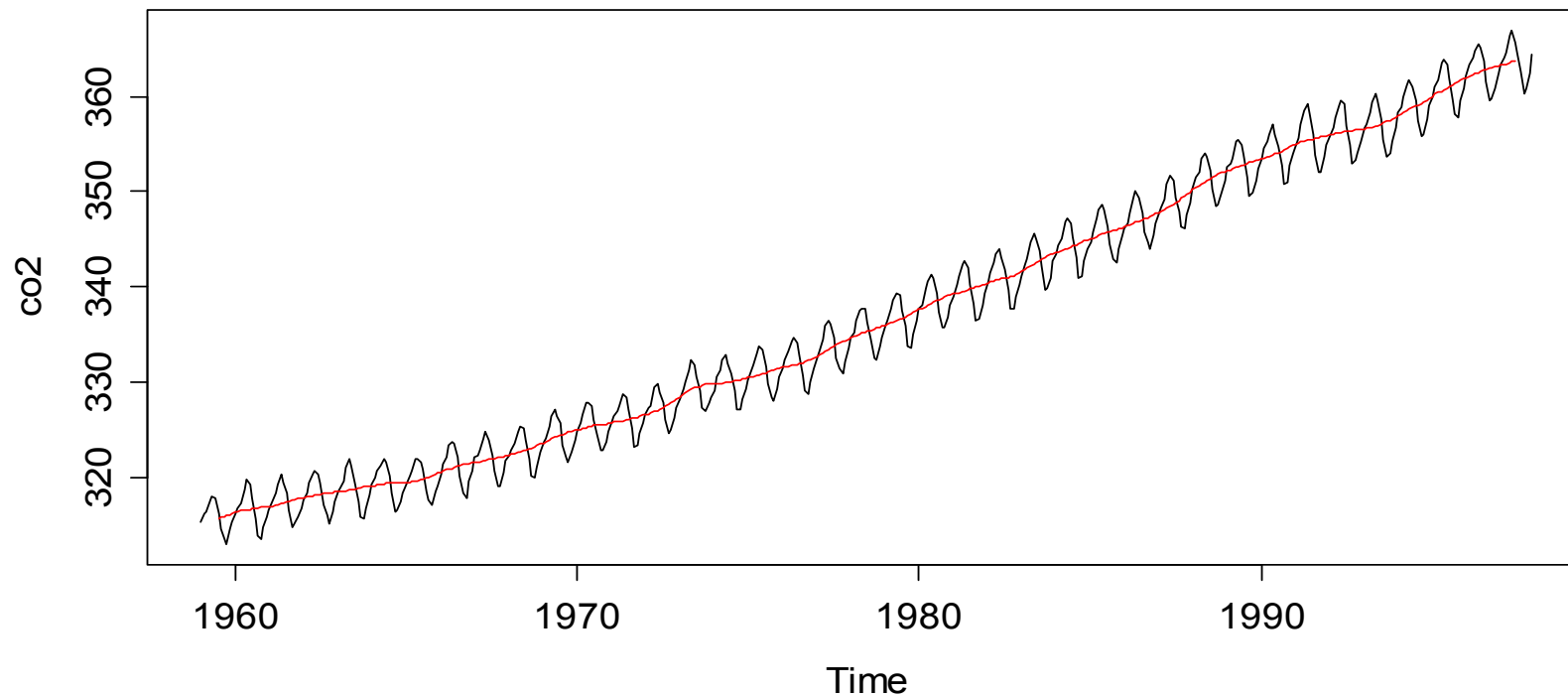
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Trend Estimation for Mauna Loa Data

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

Mauna Loa CO2 Concentrations



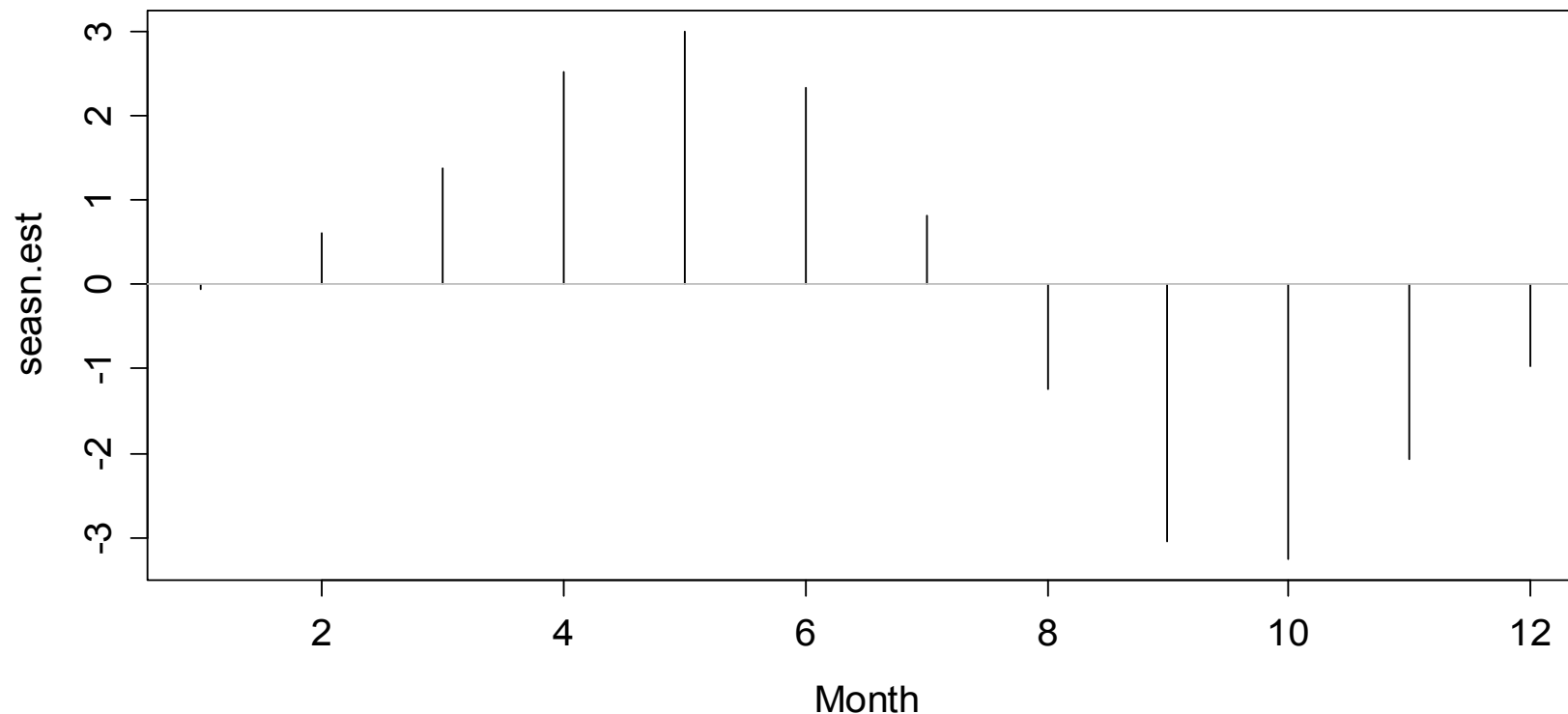
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Estimating the Seasonal Effects

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

Seasonal Effects for Mauna Loa Data



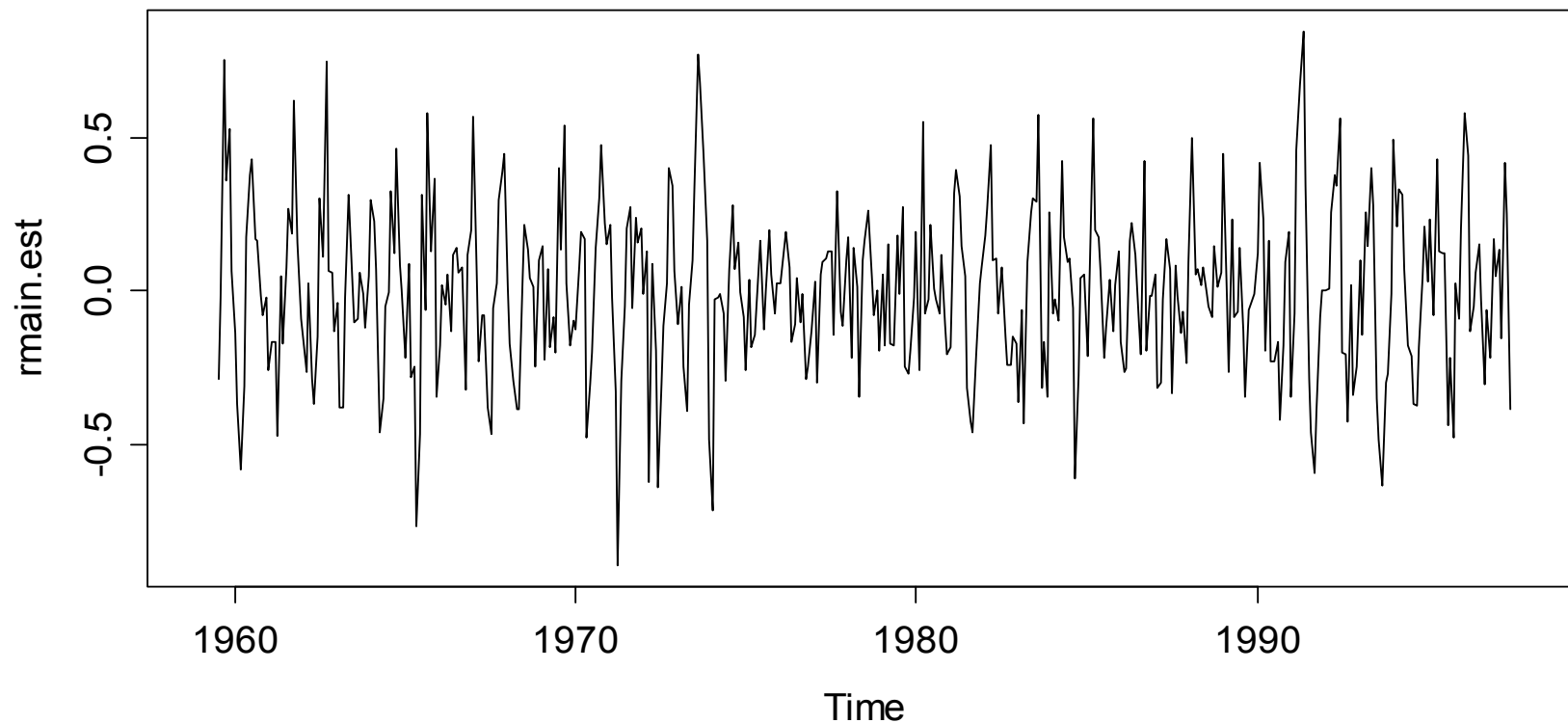
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Estimating the Remainder Term

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

Estimated Stochastic Remainder Term



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Smoothing, Filtering: Part 3

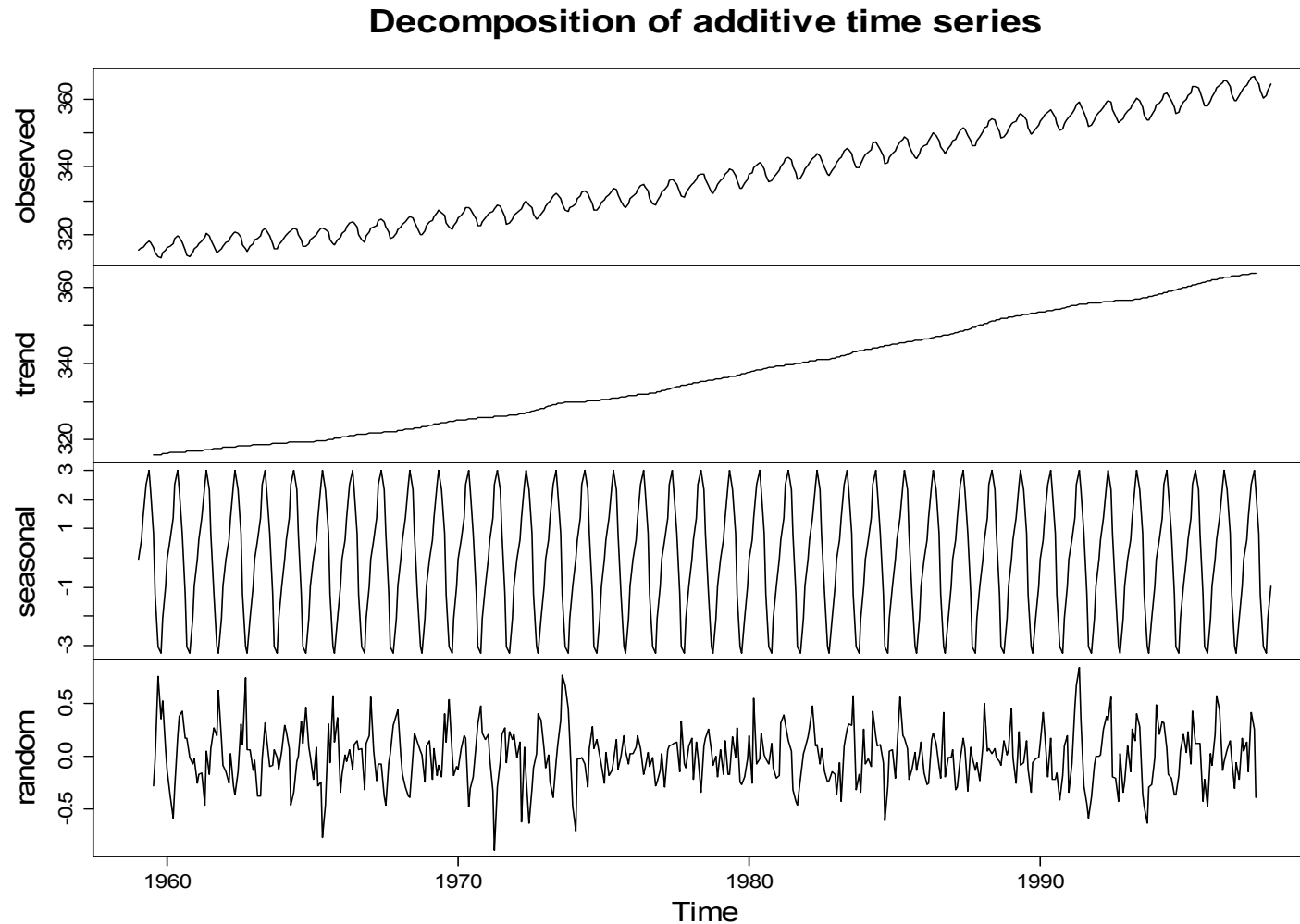
- The smoothing approach is based on estimating the trend first, and then the seasonality.
- The generalization to other periods than $p = 12$, i.e. monthly data is straightforward. Just choose a symmetric window and use uniformly distributed coefficients that sum up to 1.
- The sum over all seasonal 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 2015 – Descriptive Analysis

Estimating the Remainder Term

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



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

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

Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Smoothing, Filtering: STL-Decomposition

The **S**easonal-**T**rend Decomposition Procedure by **L**oess

- 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

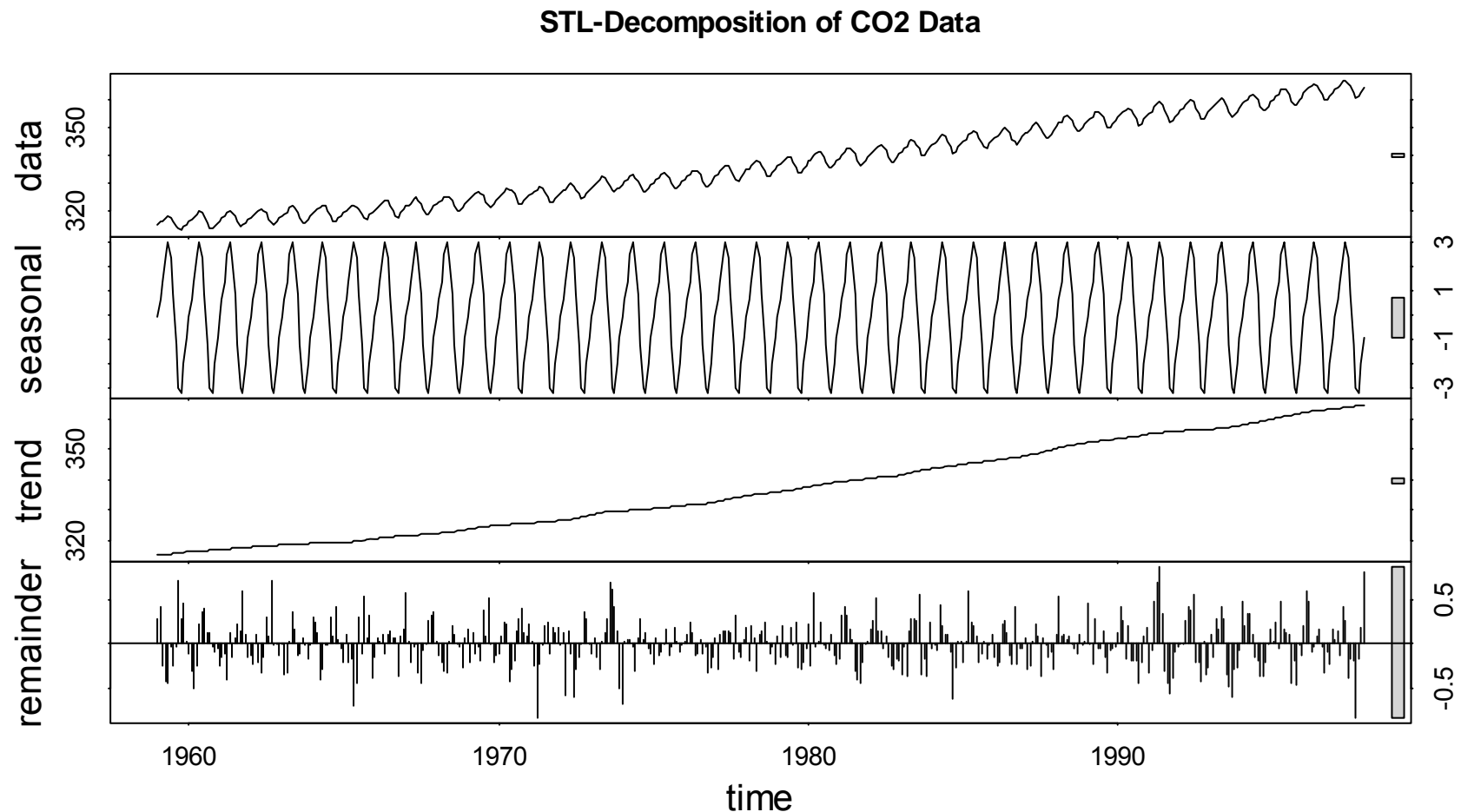
→ **Good method for „having a quick look at the data“**

Applied Time Series Analysis

SS 2015 – Descriptive Analysis

STL-Decomposition for Periodic Series

```
> co2.stl <- stl(co2, s.window="periodic")  
> plot(co2.stl, main="STL-Decomposition of CO2 Data")
```



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Using the `stl()` Function in R

```
stl {stats} R Documentation  
  
Seasonal Decomposition of Time Series by Loess  
  
Description  
Decompose a time series into seasonal, trend and irregular components using loess, acronym STL.  
  
Usage  
  
stl(x, s.window, s.degree = 0,  
    t.window = NULL, t.degree = 1,  
    l.window = nextodd(period), l.degree = t.degree,  
    s.jump = ceiling(s.window/10),  
    t.jump = ceiling(t.window/10),  
    l.jump = ceiling(l.window/10),  
    robust = FALSE,  
    inner = if(robust) 1 else 2,  
    outer = if(robust) 15 else 0,  
    na.action = na.fail)
```

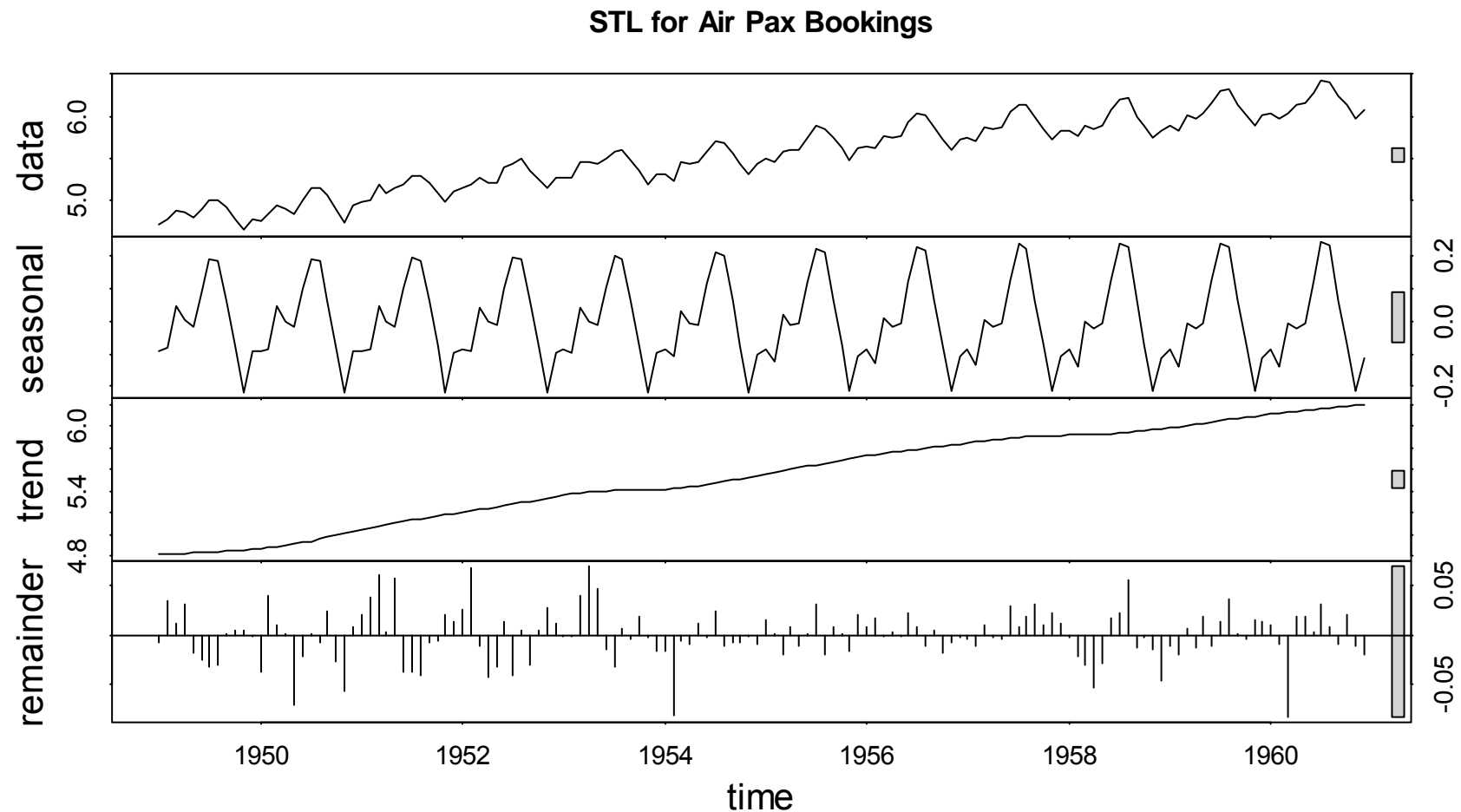
We need to supply argument `x` (i.e. the data) and `s.window` (for seasonal smoothing), either by setting it to `"periodic"` or to a numerical value. We can adjust `t.window` to a numerical value for altering the trend smoothing. Leave the rest alone!

Applied Time Series Analysis

SS 2015 – Descriptive Analysis

STL for Series with Evolving Seasonality

```
> lap.stl <- stl(lap, s.window=13)  
> plot(lap.stl, main="STL for Air Pax Bookings")
```



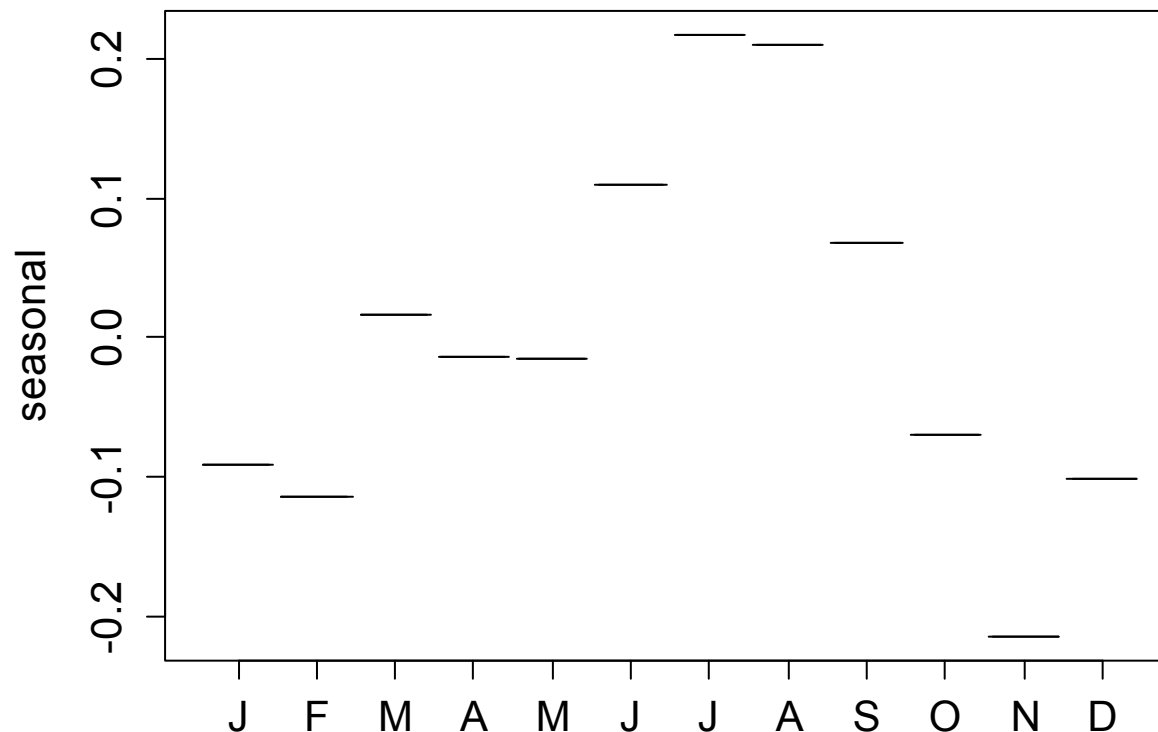
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

STL for Series with Evolving Seasonality

```
> monthplot(stl(lap, s.window="periodic"))
```

Monthplot, s.window="periodic"



Constant Seasonality:

Check the STL plot on the previous slide for assessing whether this is reasonable or not!

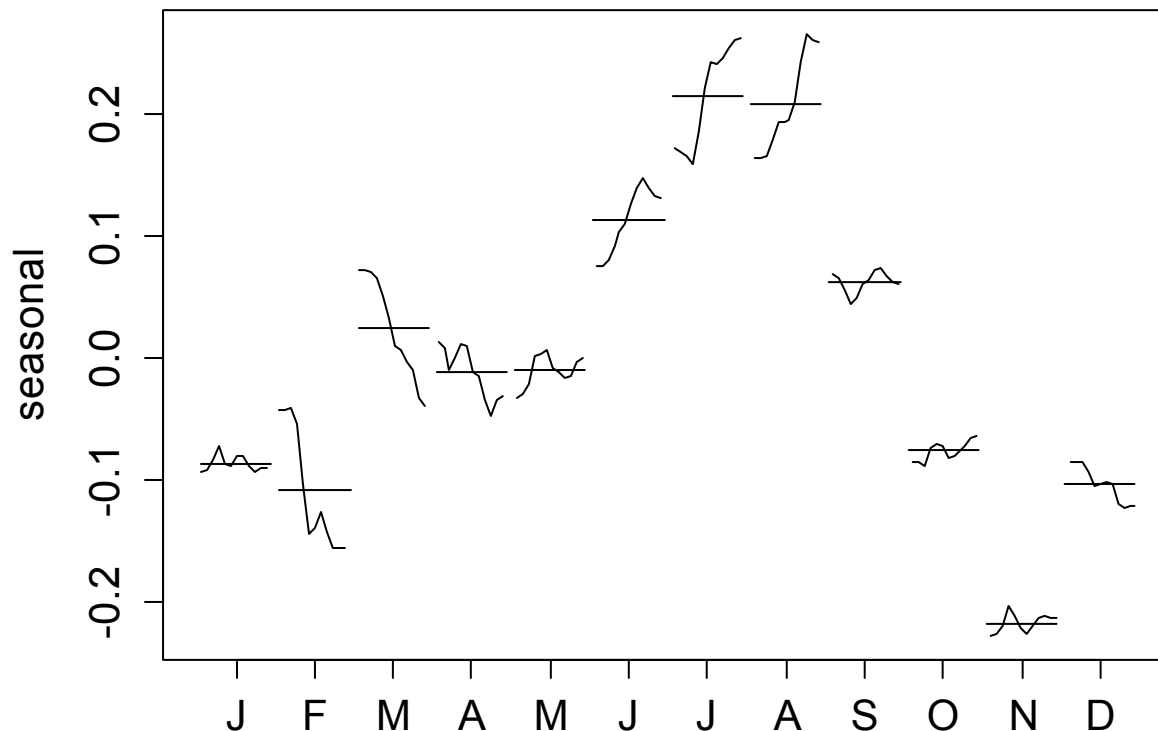
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

STL for Series with Evolving Seasonality

```
> monthplot(stl(lap, s.window=5))
```

Monthplot, s.window=5



Evolving Seasonality:
Too little smoothing in the seasonal effect, the changes are irregular.
As a remedy, increase parameter `s.window`

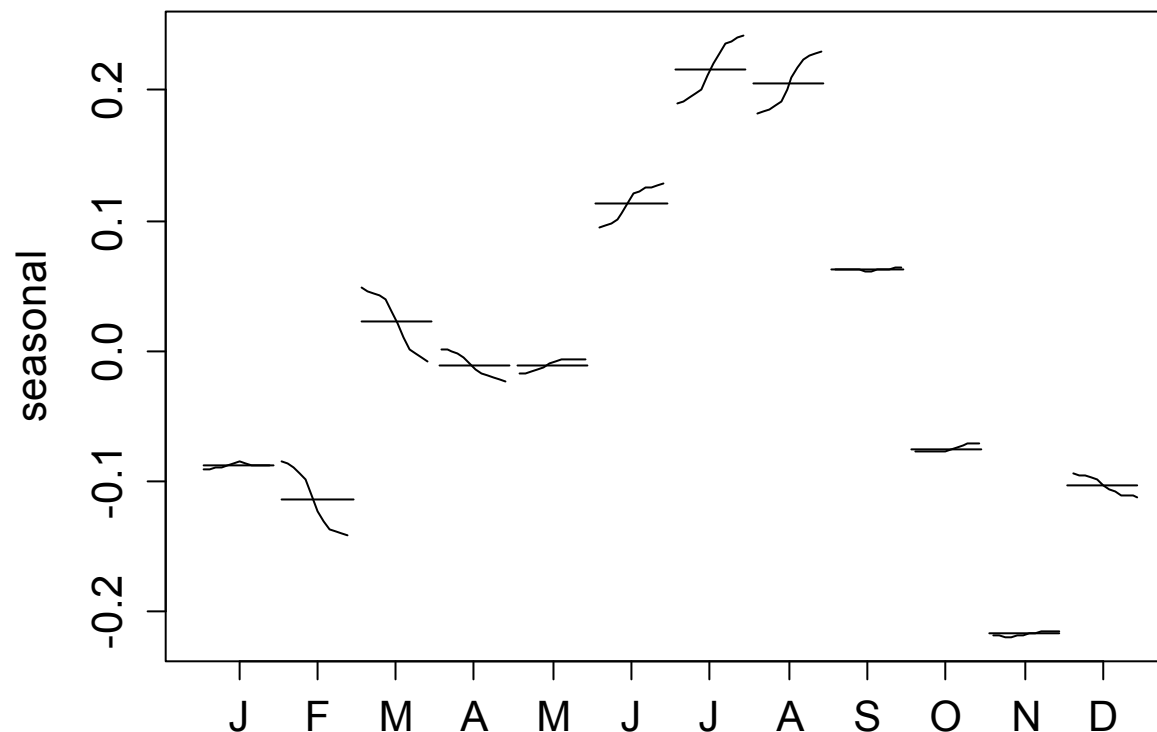
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

STL for Series with Evolving Seasonality

```
> monthplot(stl(lap, s.window=13))
```

Monthplot, s.window=13



Evolving Seasonality:

Adequate amount of smoothing will well chosen `s.window`

Applied Time Series Analysis

SS 2015 – Descriptive Analysis

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

Applied Time Series Analysis

SS 2015 – Descriptive Analysis

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!**

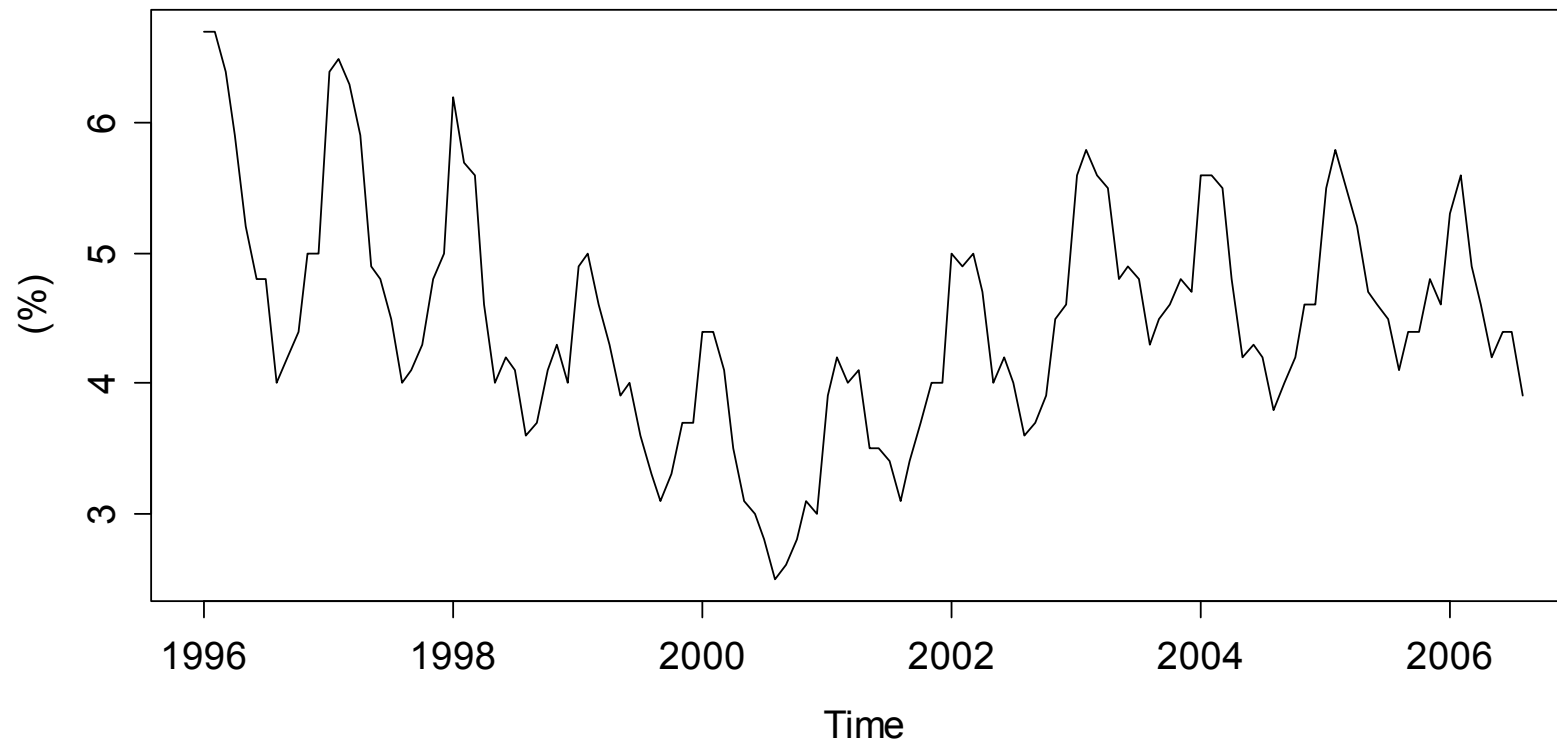
Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Parametric Modelling: Example

Maine unemployment data: Jan/1996 – Aug/2006

Unemployment in Maine



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

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 $t \in \{1, 2, \dots, 128\}$

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

Remark: choice of the polynomial degree is crucial!

Applied Time Series Analysis

SS 2015 – Descriptive Analysis

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.

Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Important Hints for Fitting

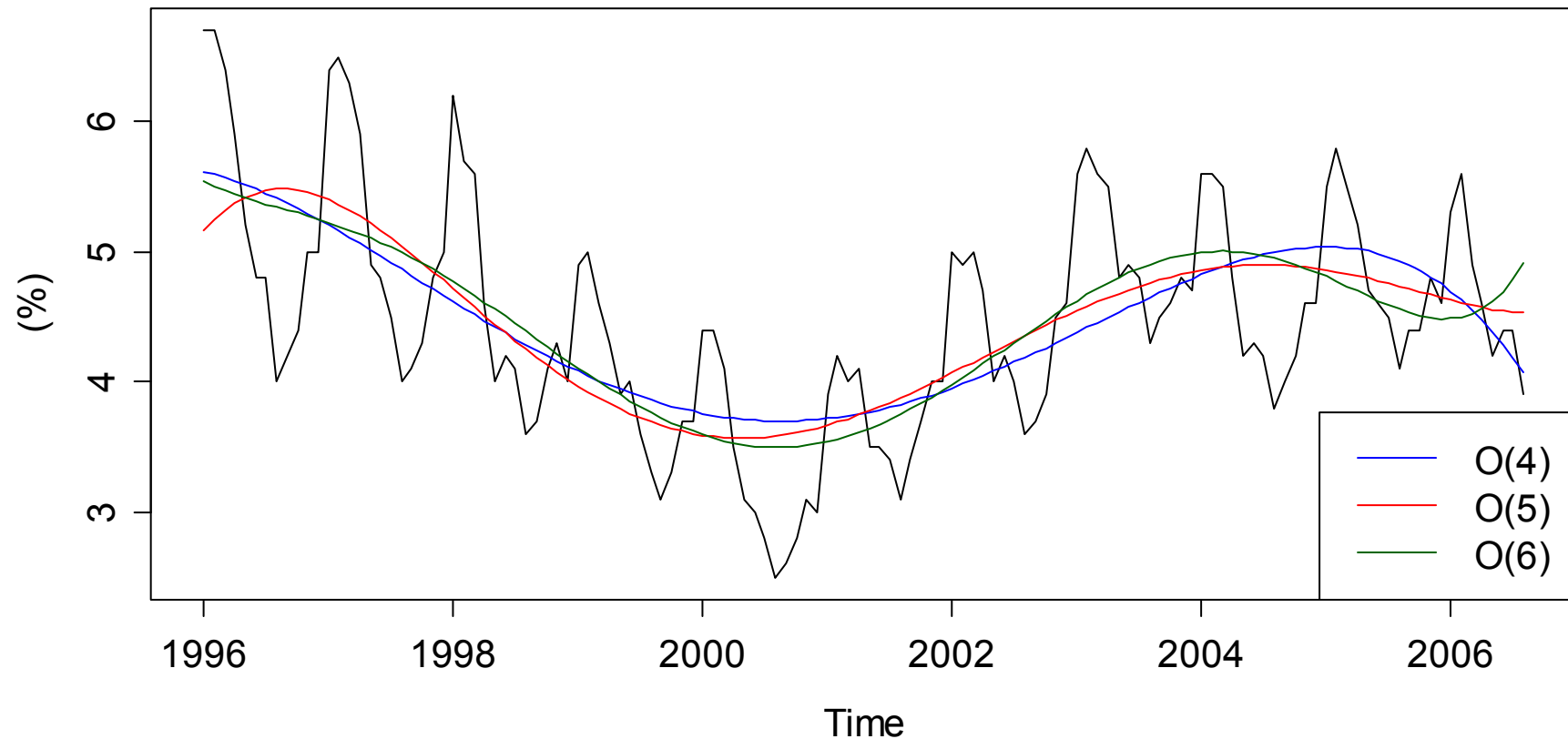
- The main predictor used in polynomial parametric modeling is the time of the observations. It can be obtained by typing `time(maine)`.
- For avoiding numerical and collinearity problems, it is essential to center the time/predictors!
- R sets the first factor level to 0, seasonality is thus expressed 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

SS 2015 – Descriptive Analysis

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

Unemployment in Maine

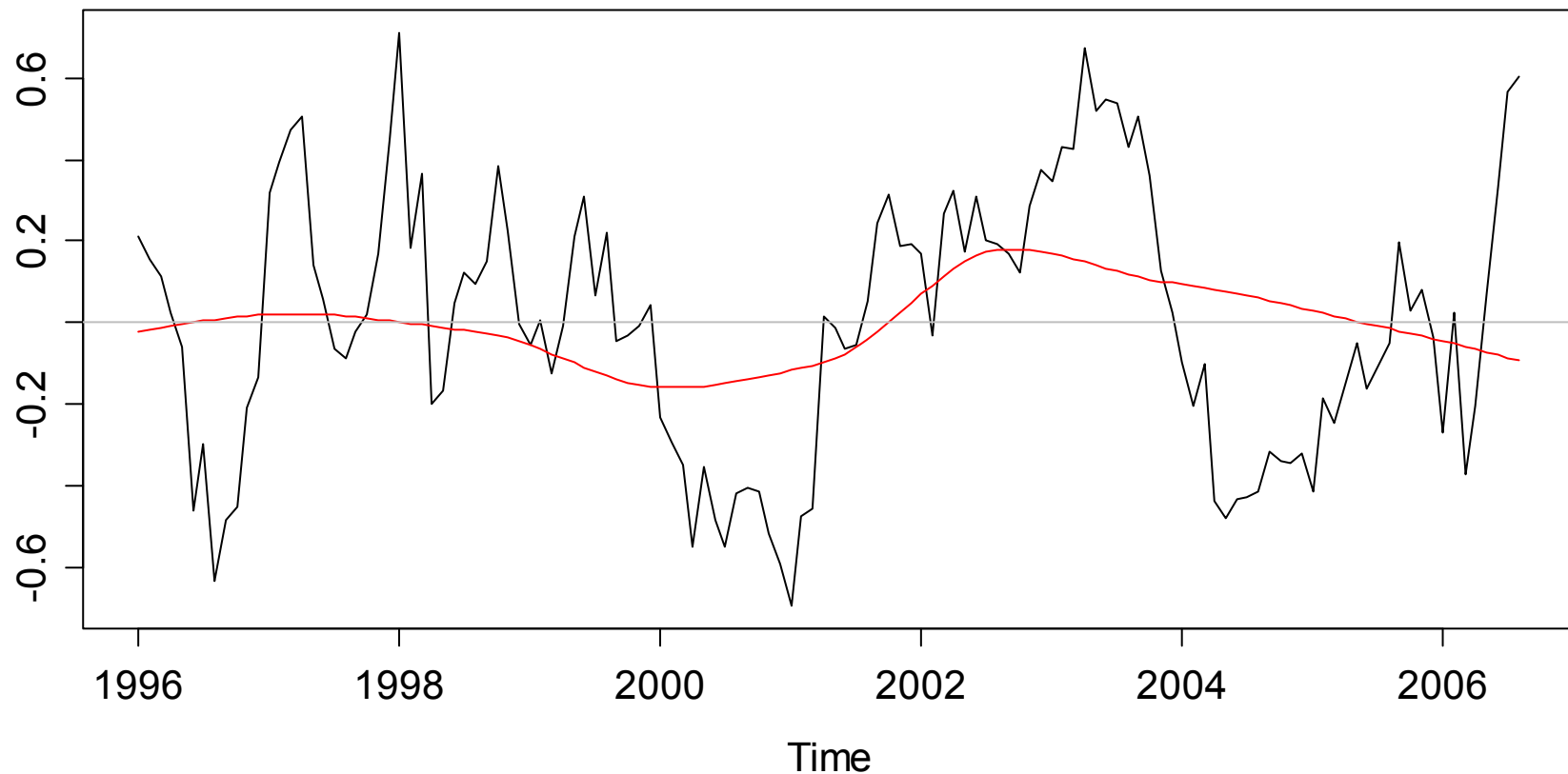


Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Residual Analysis: O(4)

Residuals vs. Time, O(4)

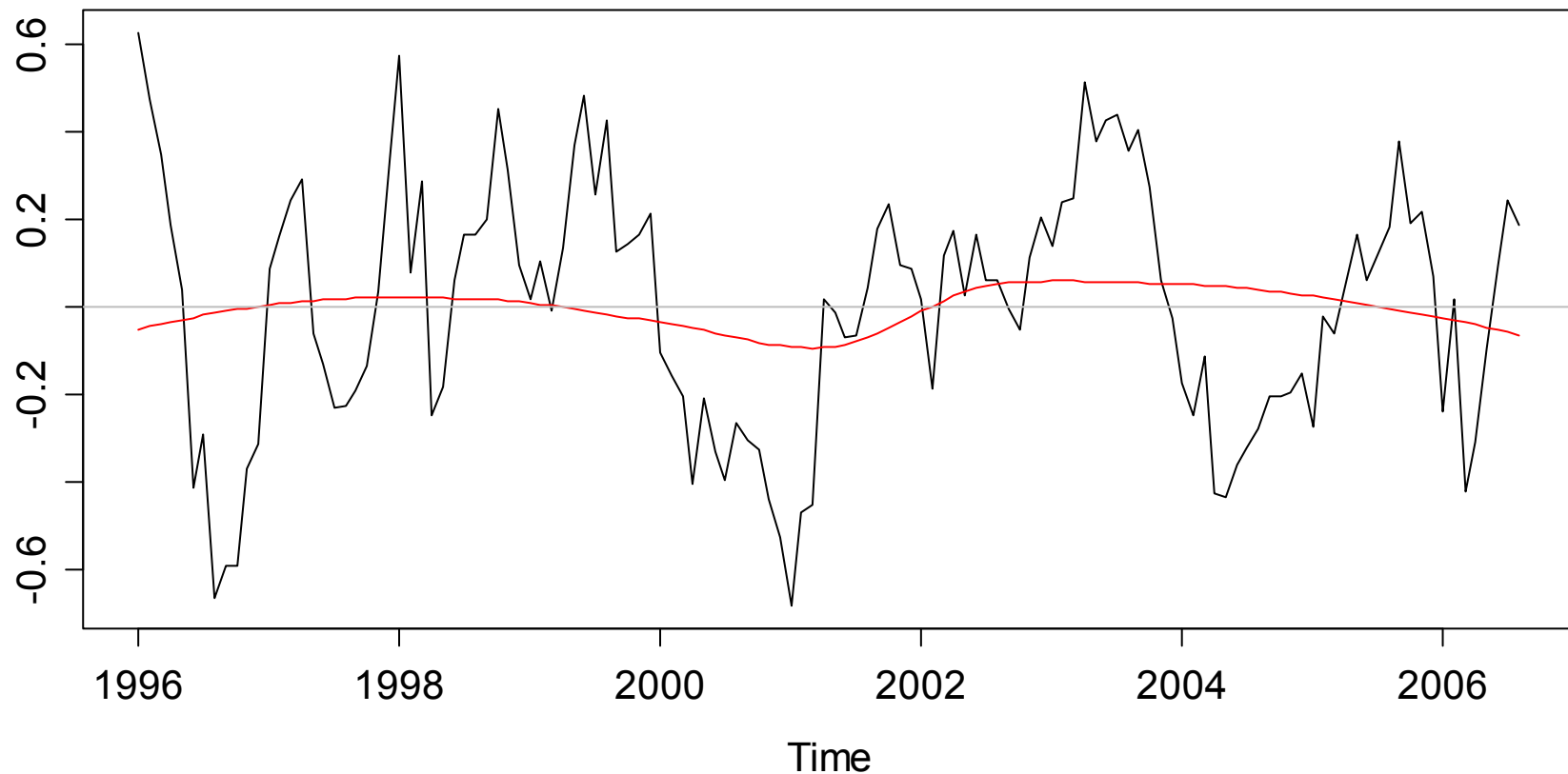


Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Residual Analysis: O(5)

Residuals vs. Time, O(5)

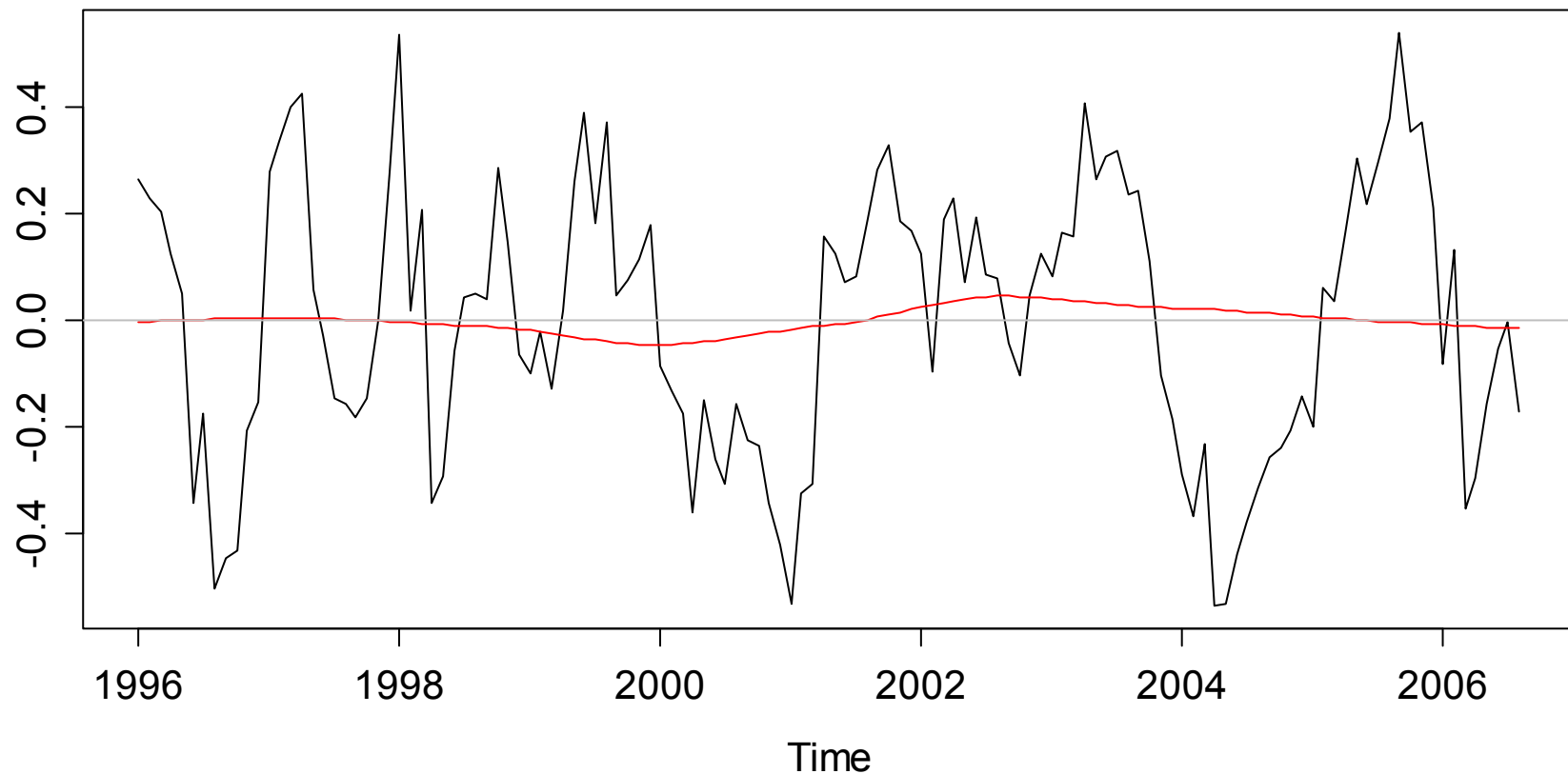


Applied Time Series Analysis

SS 2015 – Descriptive Analysis

Residual Analysis: O(6)

Residuals vs. Time, O(6)



Applied Time Series Analysis

SS 2015 – Descriptive Analysis

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 2015 – Autocorrelation

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

Applied Time Series Analysis

SS 2015 – Autocorrelation

Autocorrelation

The aim of this section is to estimate, explore and understand the dependency structure within a stationary time series.

Def: **Autocorrelation**

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

Autocorrelation is a dimensionless measure for the strength of the linear association between the random variables X_{t+k} and X_t .

There are 2 estimators, i.e. the lagged sample and the plug-in.

→ *see slides & blackboard for a sketch of the two approaches...*

Applied Time Series Analysis

SS 2015 – Autocorrelation

Practical Interpretation of Autocorrelation

We e.g. assume $\rho(k) = 0.7$

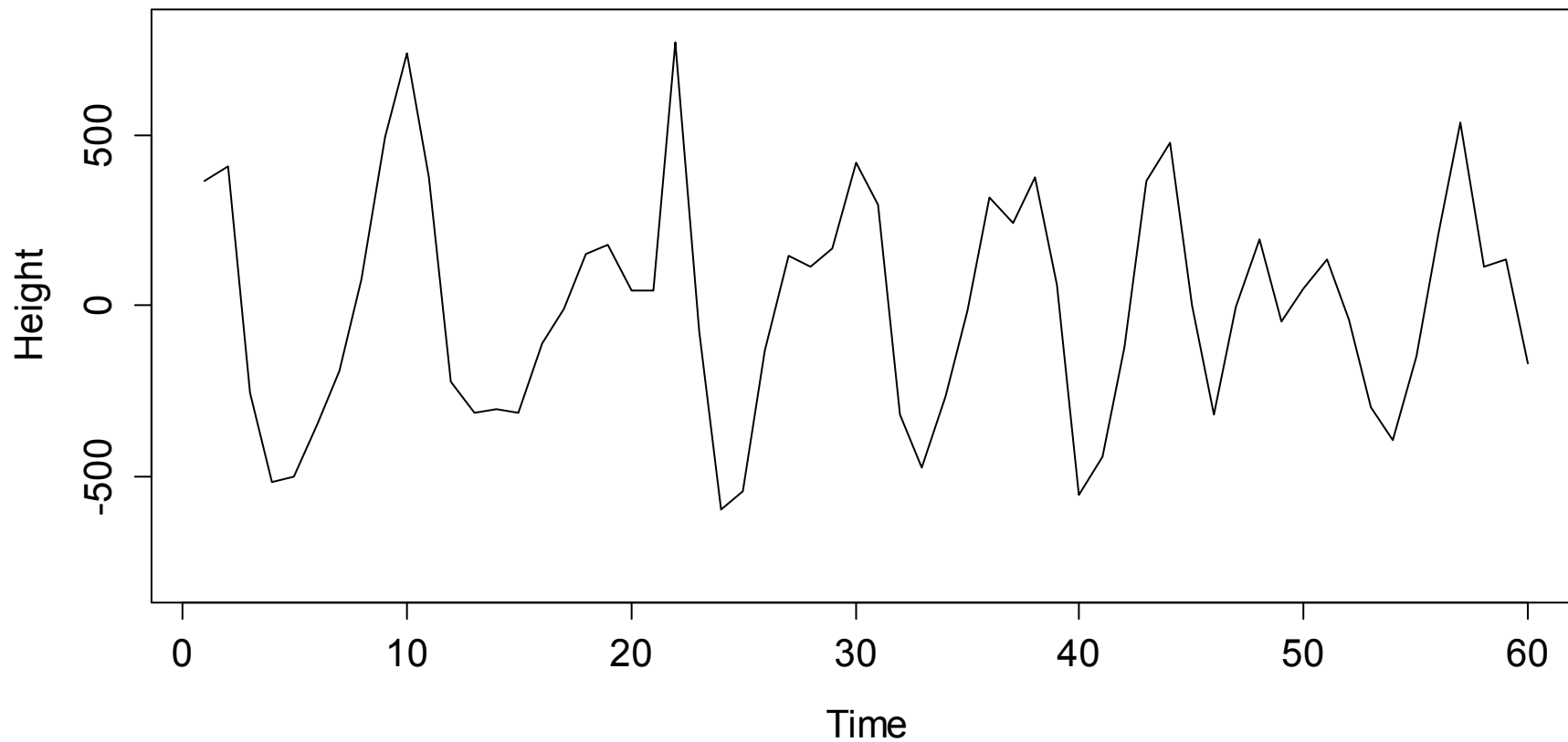
- The square of the autocorrelation, i.e. $\rho(k)^2 = 0.49$, is the percentage of variability explained by the linear association between X_t and its predecessor X_{t-1} .
- Thus, in our example, X_{t-1} accounts for roughly 49% of the variability observed in random variable X_t . Only roughly because the world is not linear.
- From this we can also conclude that any $\rho(k) < 0.4$ is not a strong association, i.e. has a small effect on the next observation only.

Applied Time Series Analysis

SS 2015 – Autocorrelation

Example: Wave Tank Data

Wave Tank Data



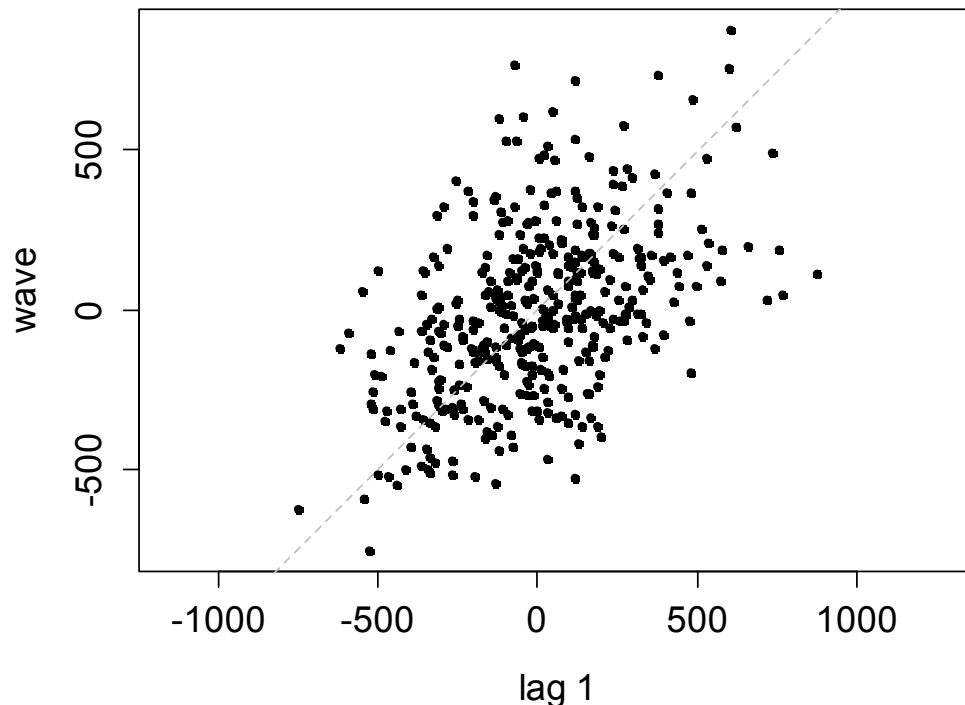
Applied Time Series Analysis

SS 2015 – Autocorrelation

Lagged Scatterplot Approach

Generate a plot of (x_t, x_{t+k}) for all $t = 1, \dots, n - k$ and compute the canonical Pearson correlation coefficient from these data pairs.

Lagged Scatterplot, k=1, cor=0.47



```
> lag.plot(wave, do.lines=FALSE, pch=20)
> title("Lagged Scatter, k=1, cor=0.47")
```

$$\tilde{\rho}(k) = \frac{\sum_{s=1}^{n-k} (x_{s+k} - \bar{x}_{(k)})(x_s - \bar{x}_{(1)})}{\sqrt{\sum_{s=k+1}^n (x_s - \bar{x}_{(k)})^2 \cdot \sum_{t=1}^{n-k} (x_t - \bar{x}_{(1)})^2}}$$

Applied Time Series Analysis

SS 2015 – Autocorrelation

Plug-In Estimation

For obtaining an estimate of $\hat{\rho}(k)$, determine the sample covariance at lag k and divide by the sample variance.

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

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

where

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

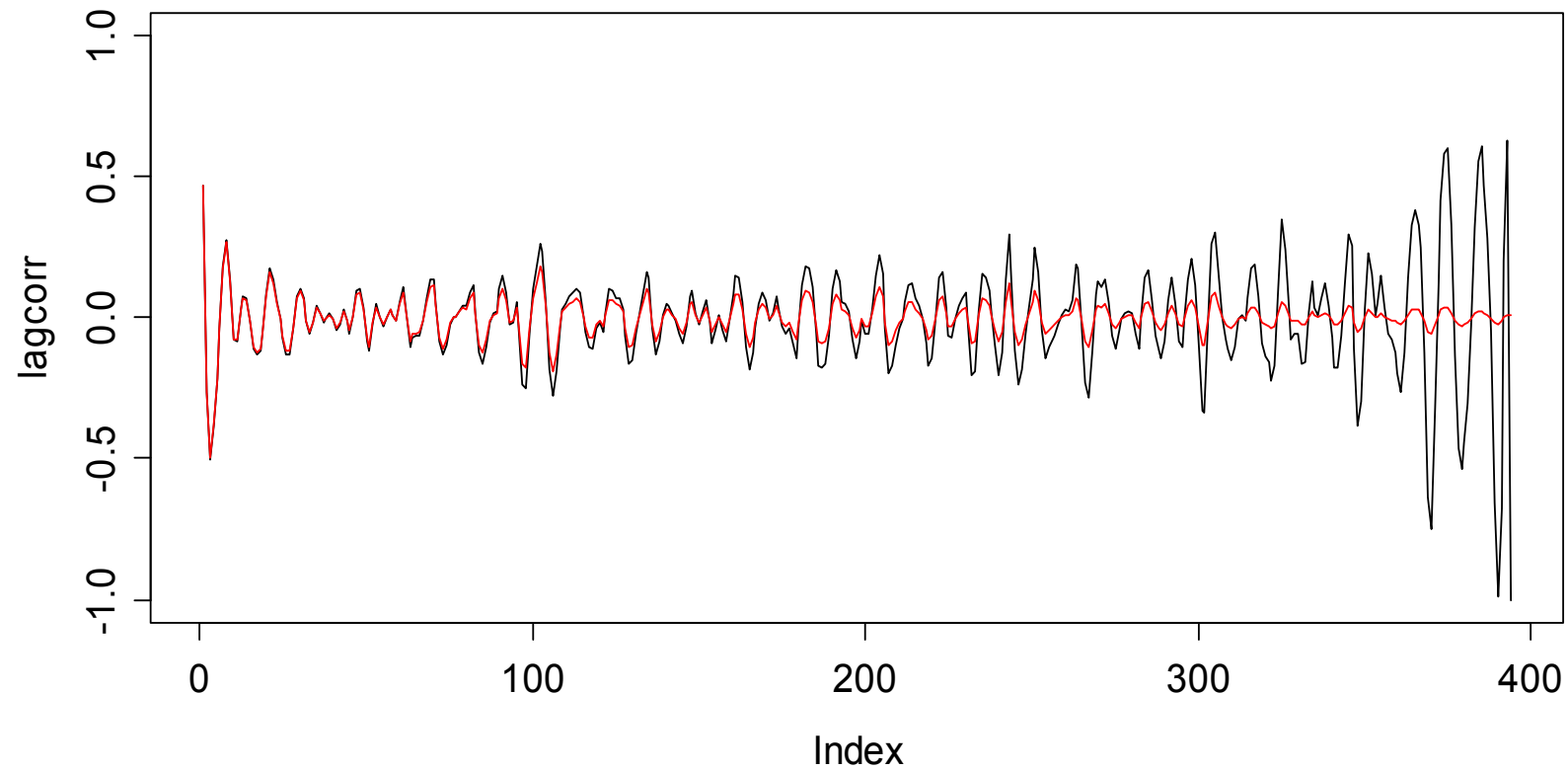
This is the standard approach for computing autocorrelations in time series analysis. It is better than the lagged scatterplot idea.

Applied Time Series Analysis

SS 2015 – Autocorrelation

Comparison Idea 1 vs. Idea 2

ACF Estimation: Lagged Scatterplot vs. Plug-In

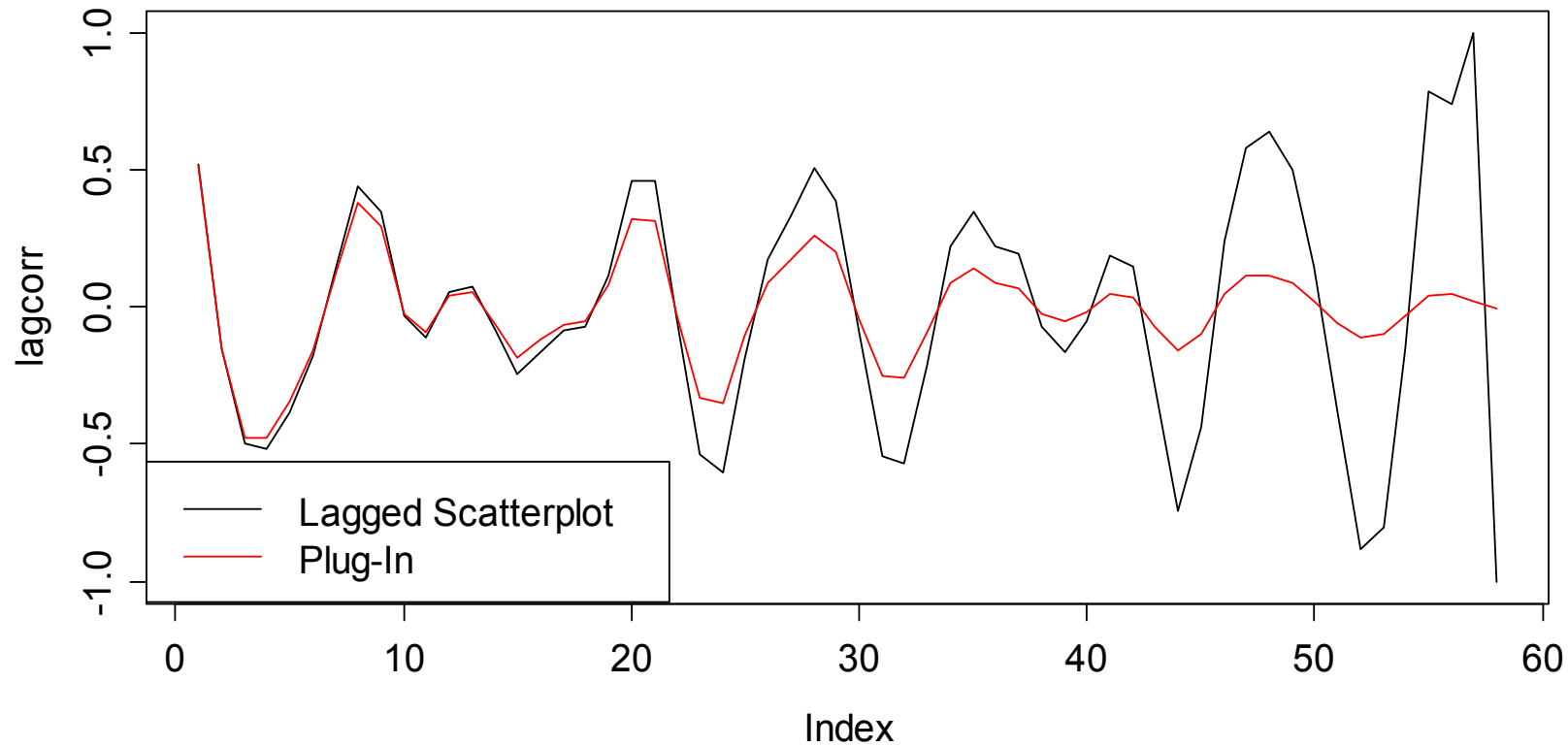


Applied Time Series Analysis

SS 2015 – Autocorrelation

Comparison Idea 1 vs. Idea 2

ACF Estimation: Lagged Scatterplot vs. Plug-In



Applied Time Series Analysis

SS 2015 – Autocorrelation

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!!!**

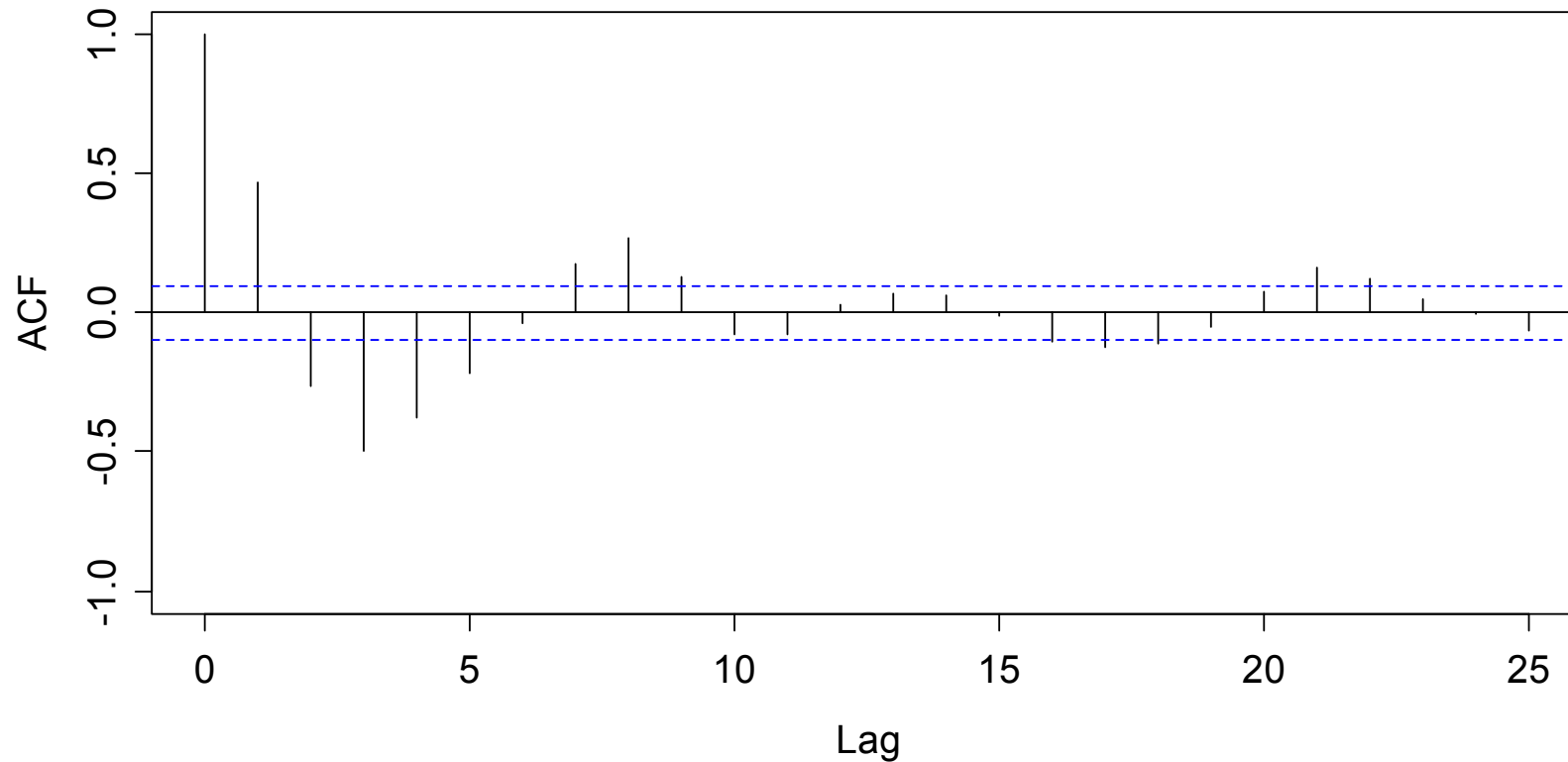
Applied Time Series Analysis

SS 2015 – Autocorrelation

Correlogram

```
> acf(wave, ylim=c(-1,1))
```

Correlogram of Wave Tank Data



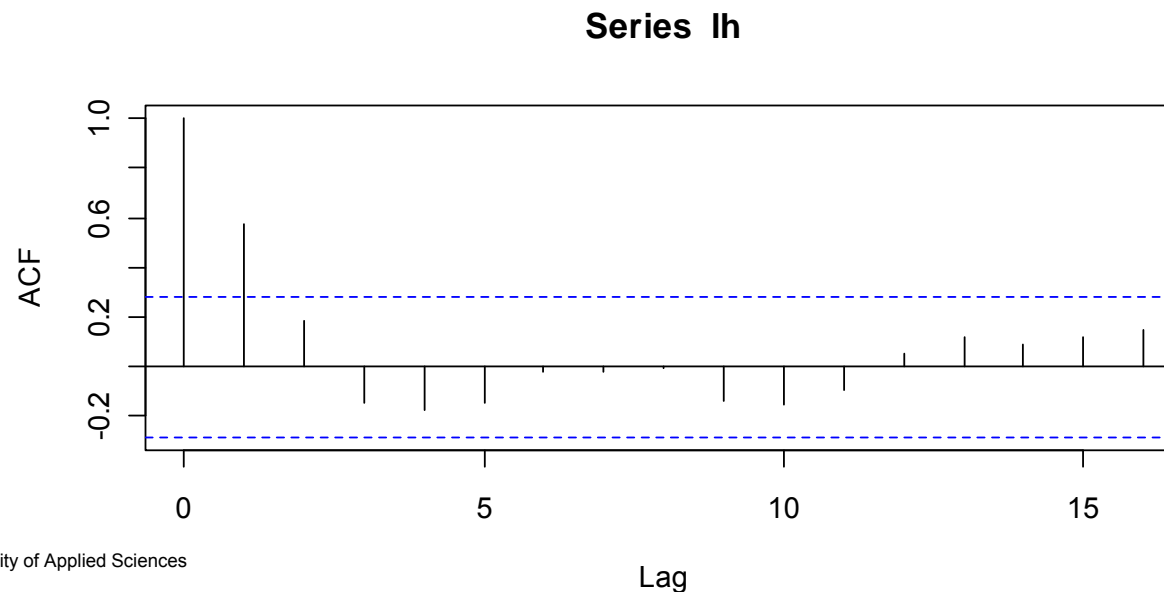
Applied Time Series Analysis

SS 2015 – Autocorrelation

Random Series – Confidence Bands

If a time series is White Noise, i.e. consists of iid random variables X_t , the (theoretical) autocorrelations $\rho(k)$ are all 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.



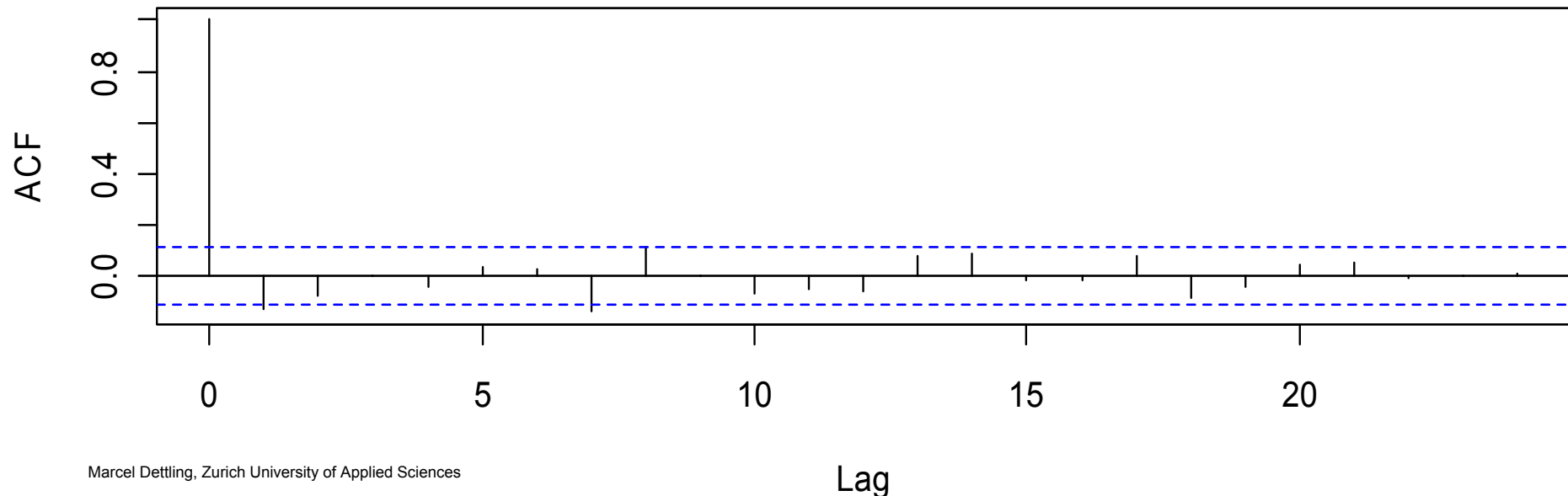
Applied Time Series Analysis

SS 2015 – Autocorrelation

Random Series – Confidence Bands

For long iid series, it can be shown that $\hat{\rho}(k)$ is approximately $N(0, 1/n)$. Thus, under the null hypothesis that a series is iid and hence $\rho(k) = 0$, the 95% acceptance region for the null is given by the interval $\pm 2 / \sqrt{n}$.

i.i.d. Series with n=300



Applied Time Series Analysis

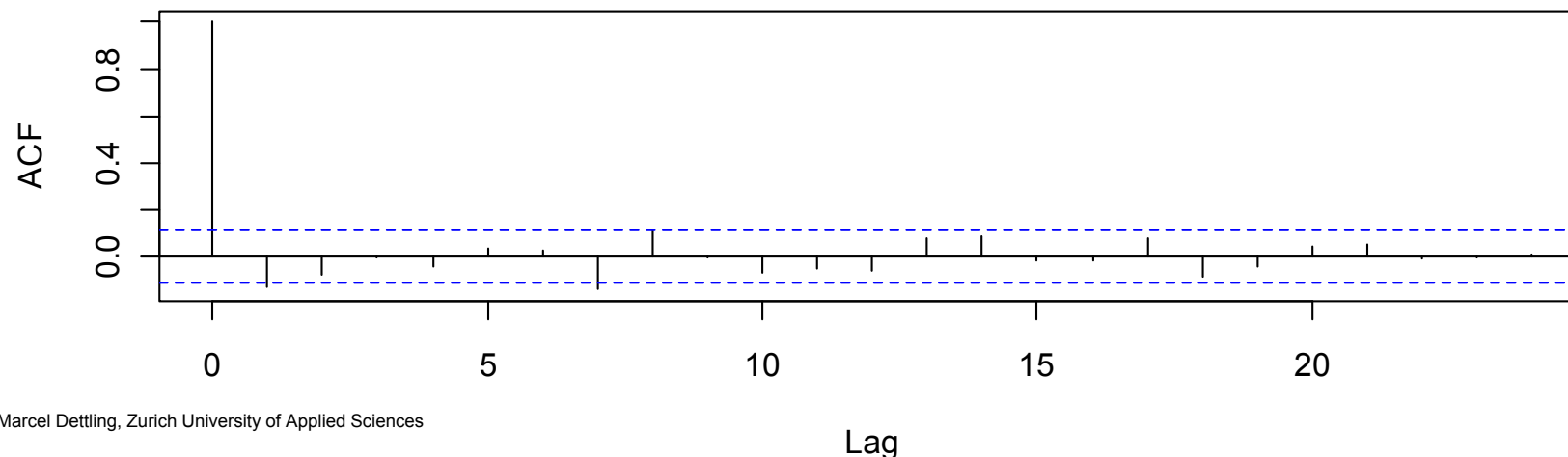
SS 2015 – Autocorrelation

Random Series – Confidence Bands

Thus, even for a (long) i.i.d. time series, we expect that 5% of the estimated autocorrelation coefficients 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.

i.i.d. Series with n=300



Applied Time Series Analysis

SS 2015 – Autocorrelation

Ljung-Box Test

The Ljung-Box approach tests the null hypothesis that a number of autocorrelation coefficients are simultaneously equal to zero. Thus, it tests for significant autocorrelation in a series. The test statistic is:

$$Q(h) = n \cdot (n + 2) \cdot \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n - k} \sim \chi_h^2$$

In R:

```
> Box.test(wave, lag=10, type="Ljung-Box")
```

```
Box-Ljung test
```

```
data: wave
```

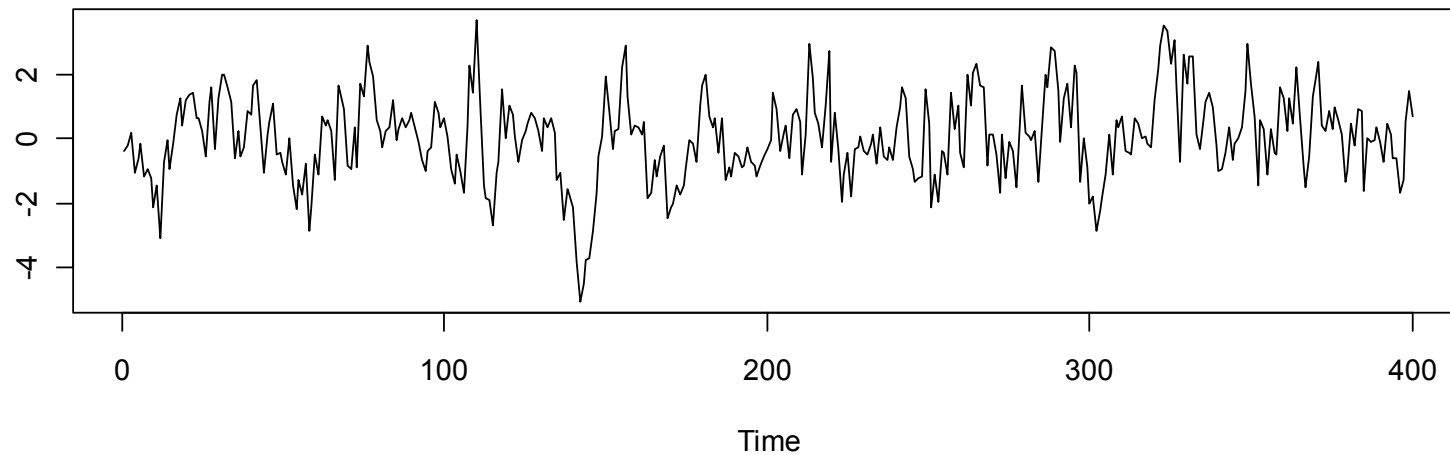
```
X-squared = 344.0155, df = 10, p-value < 2.2e-16
```

Applied Time Series Analysis

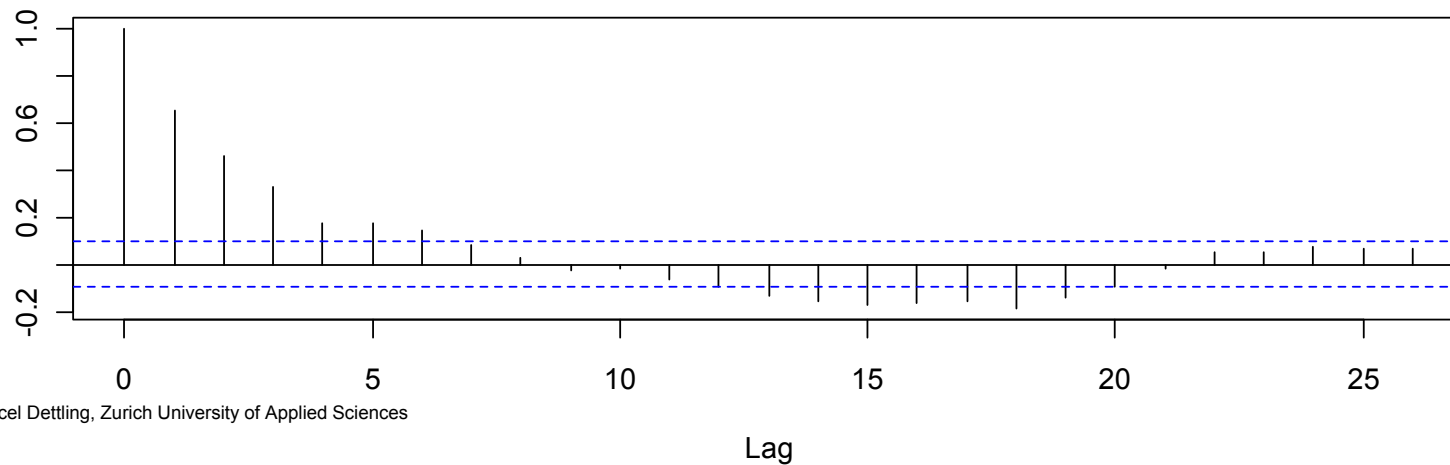
SS 2015 – Autocorrelation

Short Term Positive Correlation

Simulated Short Term Correlation Series



ACF of Simulated Short Term Correlation Series



Applied Time Series Analysis

SS 2015 – Autocorrelation

Short Term Positive 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.

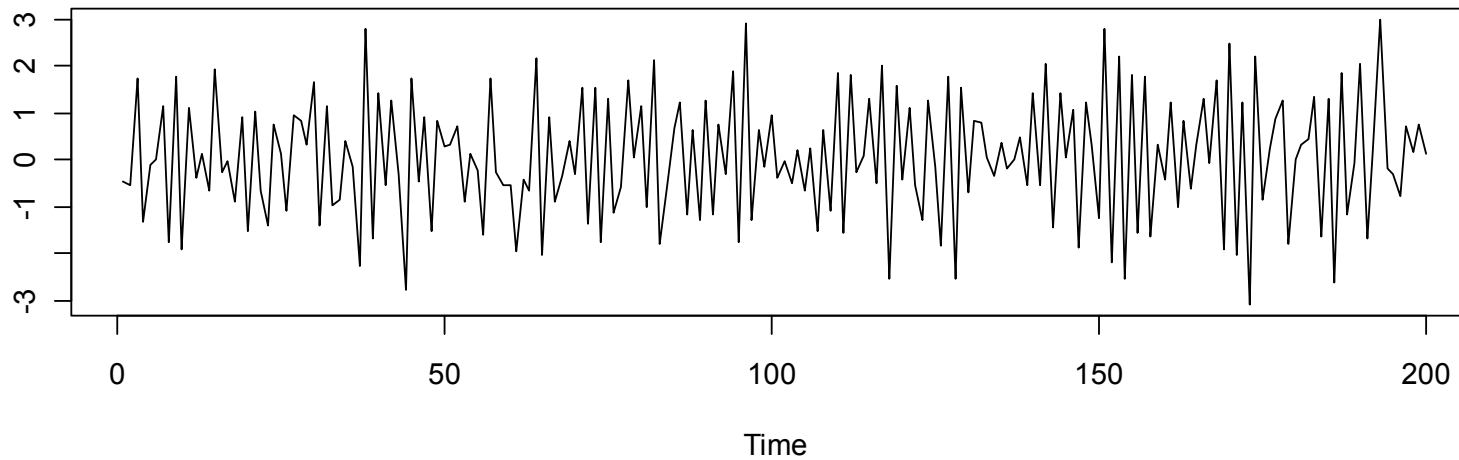
A model called an autoregressive model may be appropriate for series of this type.

Applied Time Series Analysis

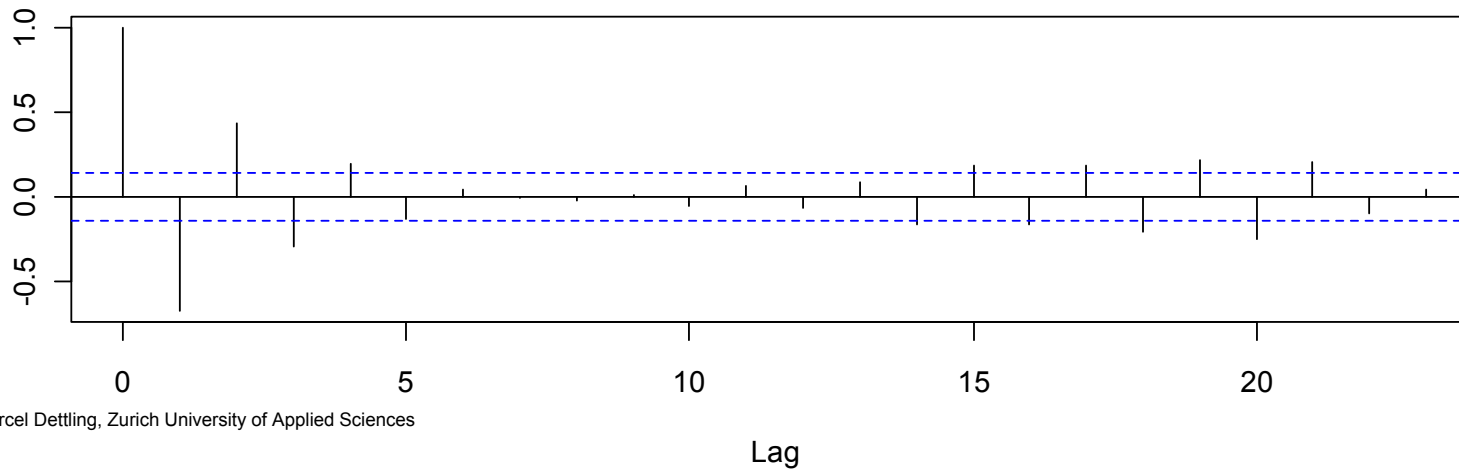
SS 2015 – Autocorrelation

Alternating Time Series

Simulated Alternating Correlation Series



ACF of Simulated Alternating Correlation Series

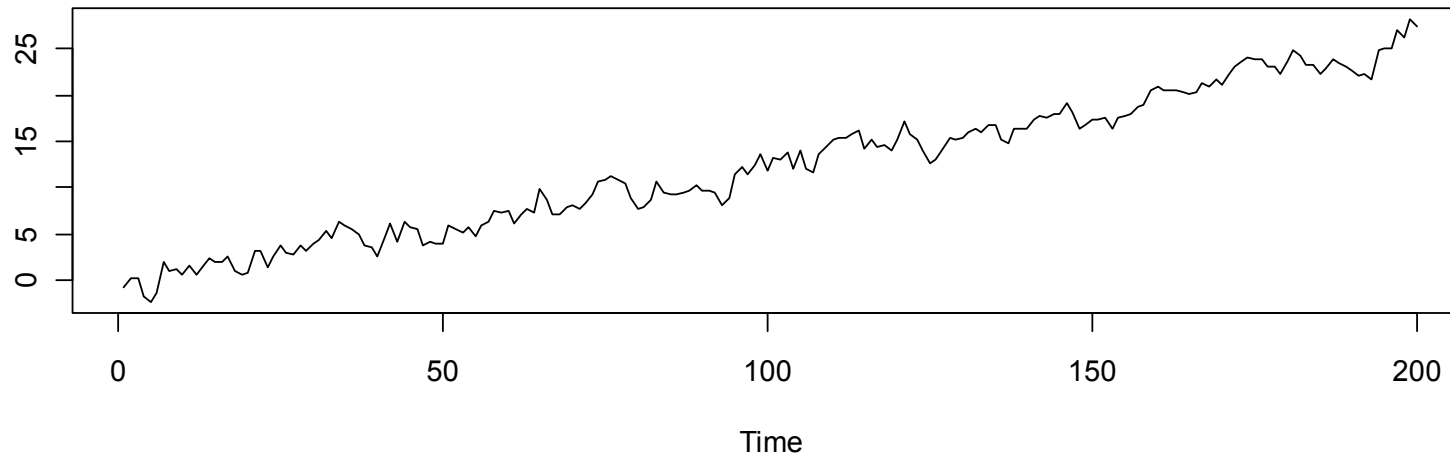


Applied Time Series Analysis

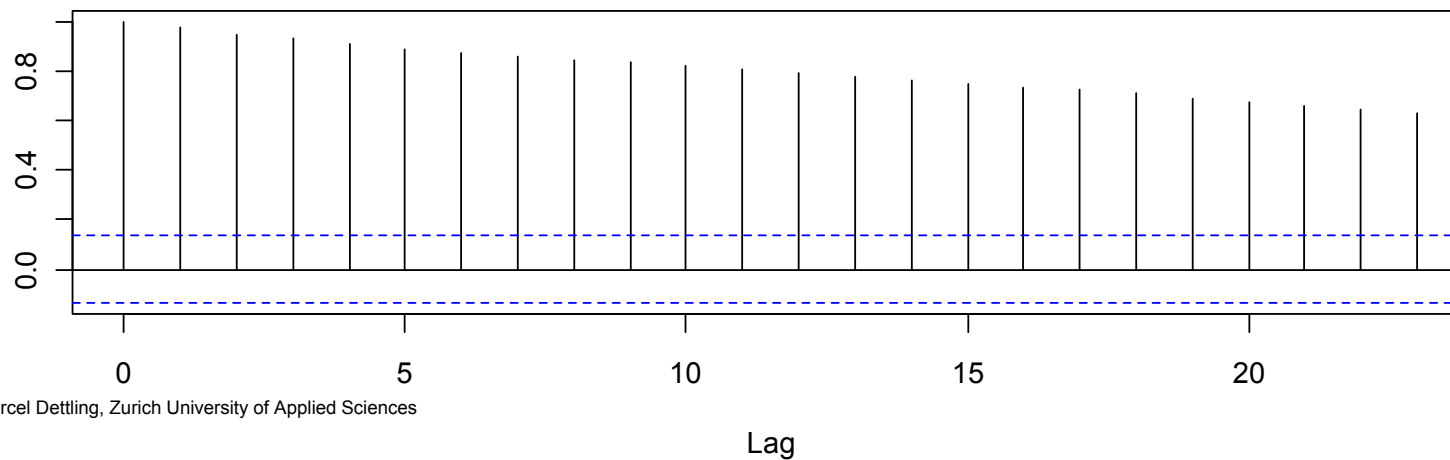
SS 2015 – Autocorrelation

Non-Stationarity in the ACF: Trend

Simulated Series with a Trend



ACF of Simulated Series with a Trend

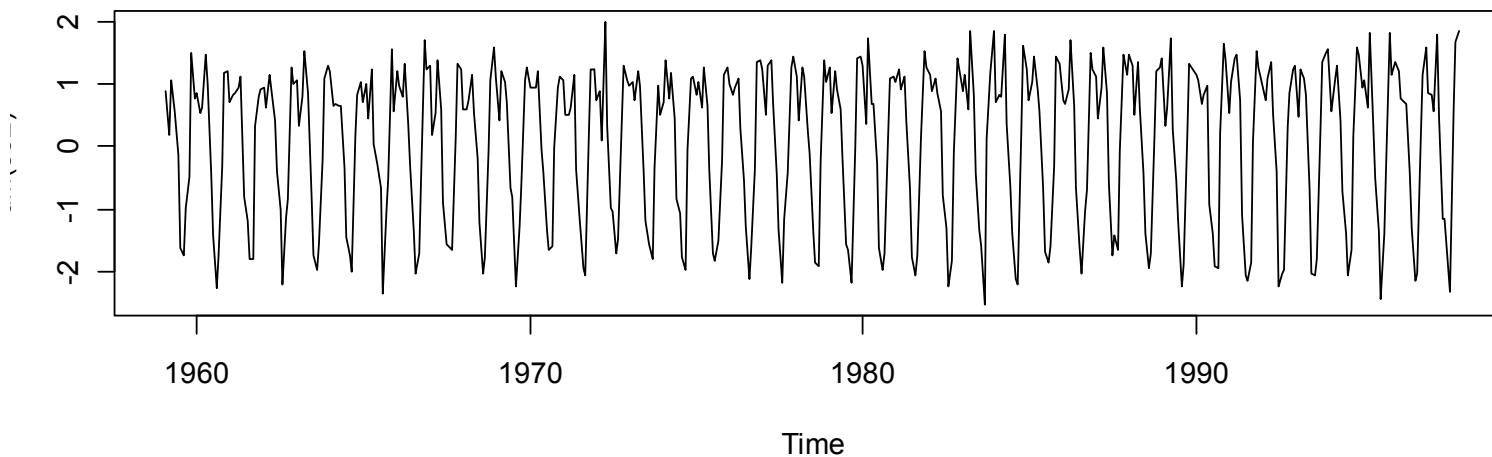


Applied Time Series Analysis

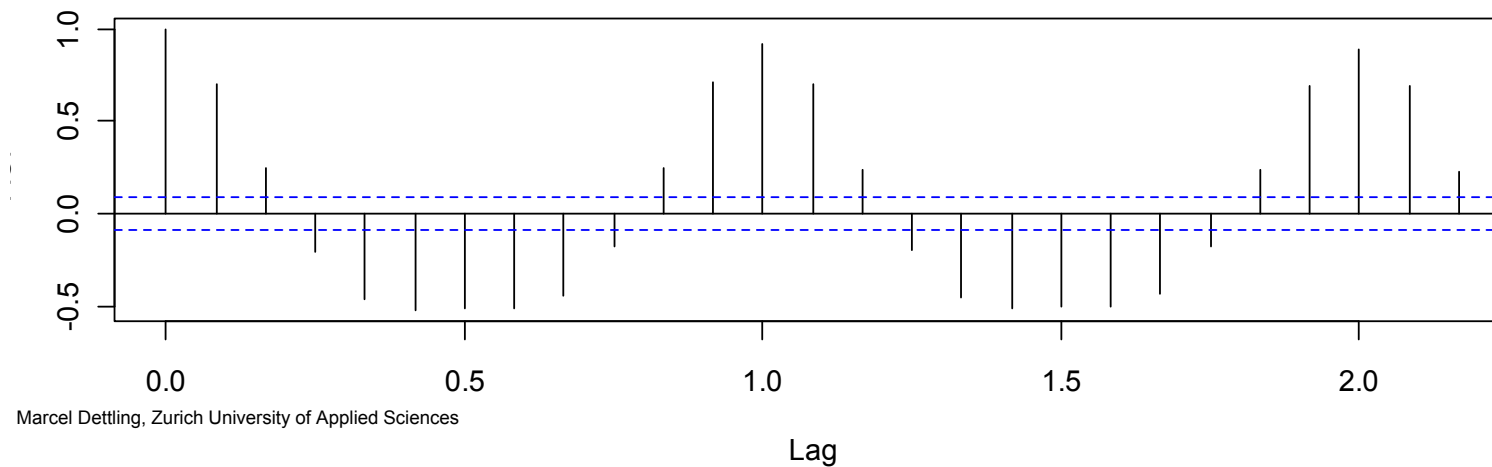
SS 2015 – Autocorrelation

Non-Stationarity in the ACF: Seasonal Pattern

De-Trended Mauna Loa Data



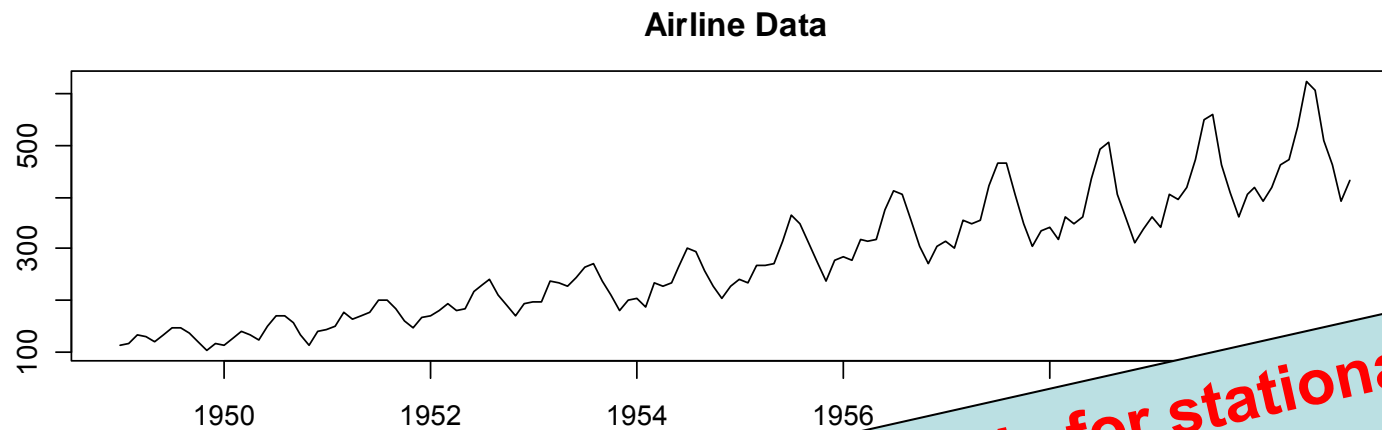
ACF of De-Trended Mauna Loa Data



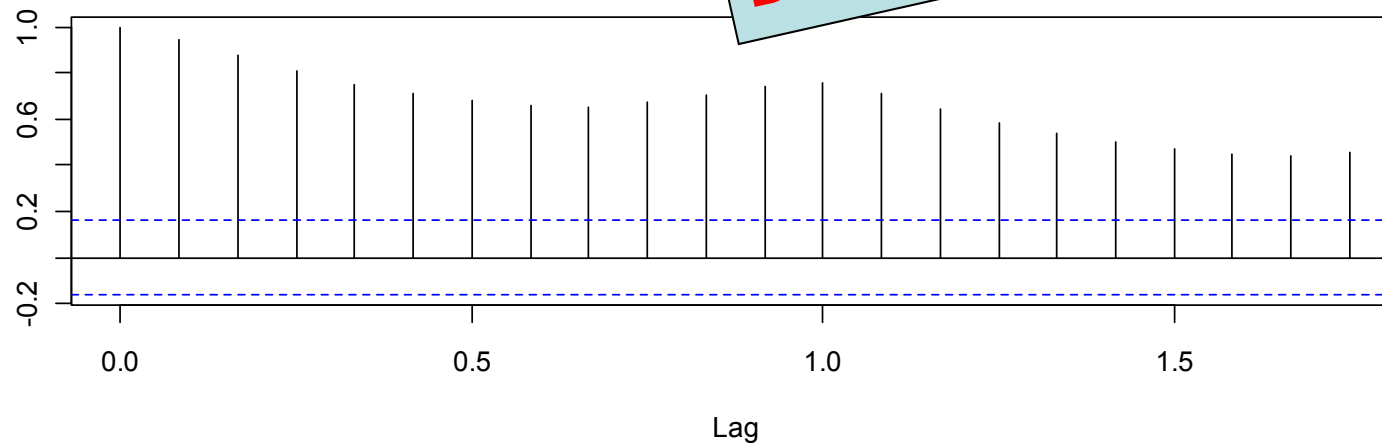
Applied Time Series Analysis

SS 2015 – Autocorrelation

ACF of the Raw Airline Data



ACF of A



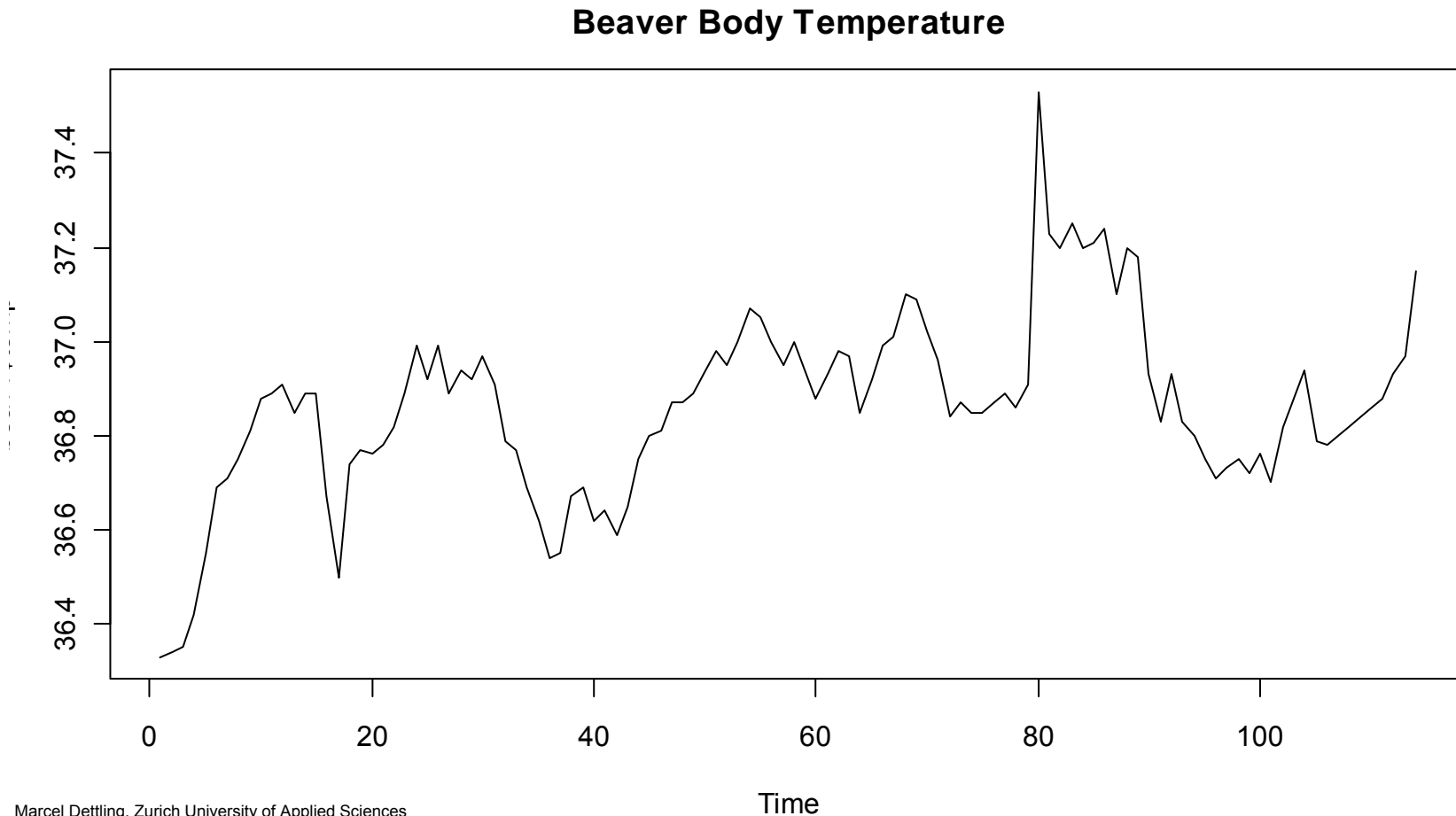
**The ACF is for stationary series only!
Do not use it like this!!!**

Applied Time Series Analysis

SS 2015 – Autocorrelation

Outliers and the ACF

Outliers in the time series strongly affect the ACF estimation!



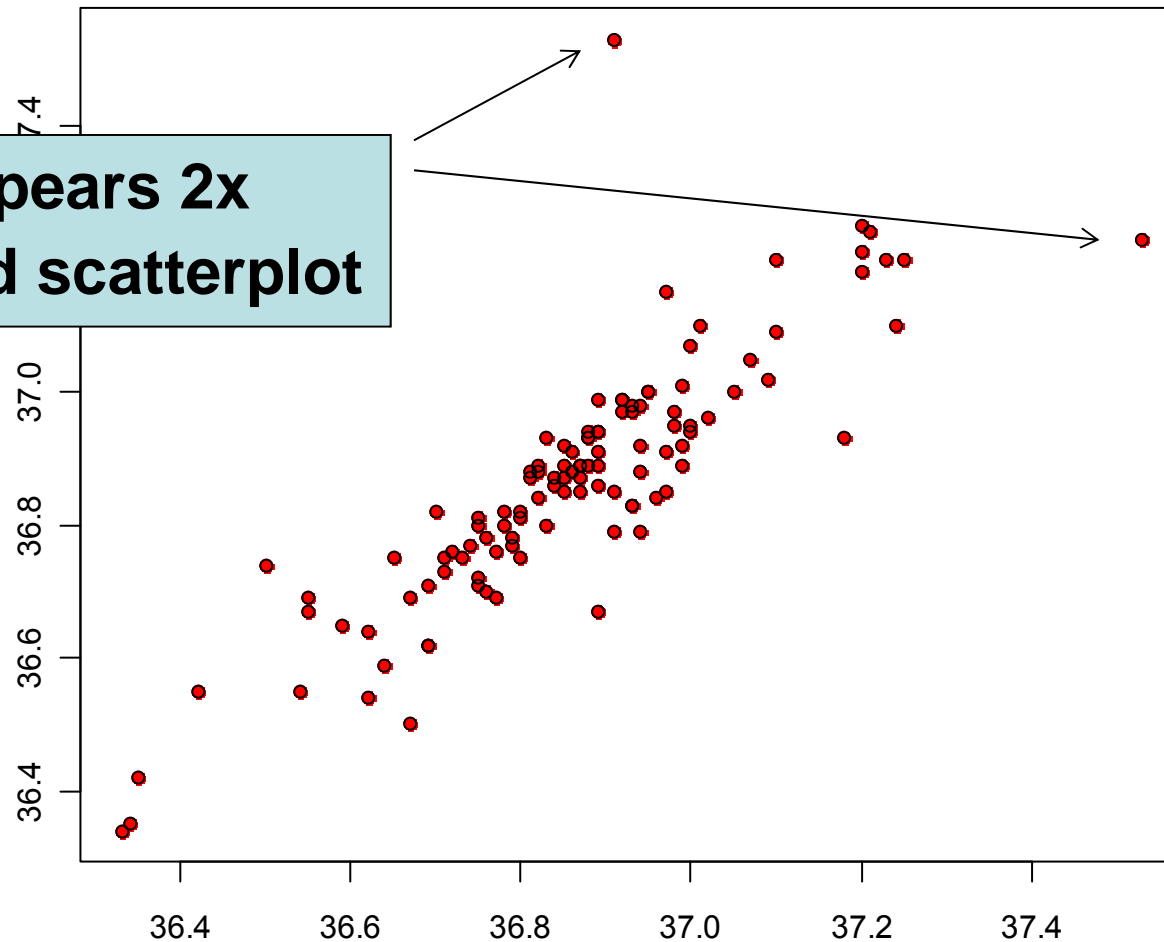
Applied Time Series Analysis

SS 2015 – Autocorrelation

Outliers and the ACF

Lagged Scatterplot with $k=1$ for Beaver Data

**1 Outlier, appears 2x
in the lagged scatterplot**

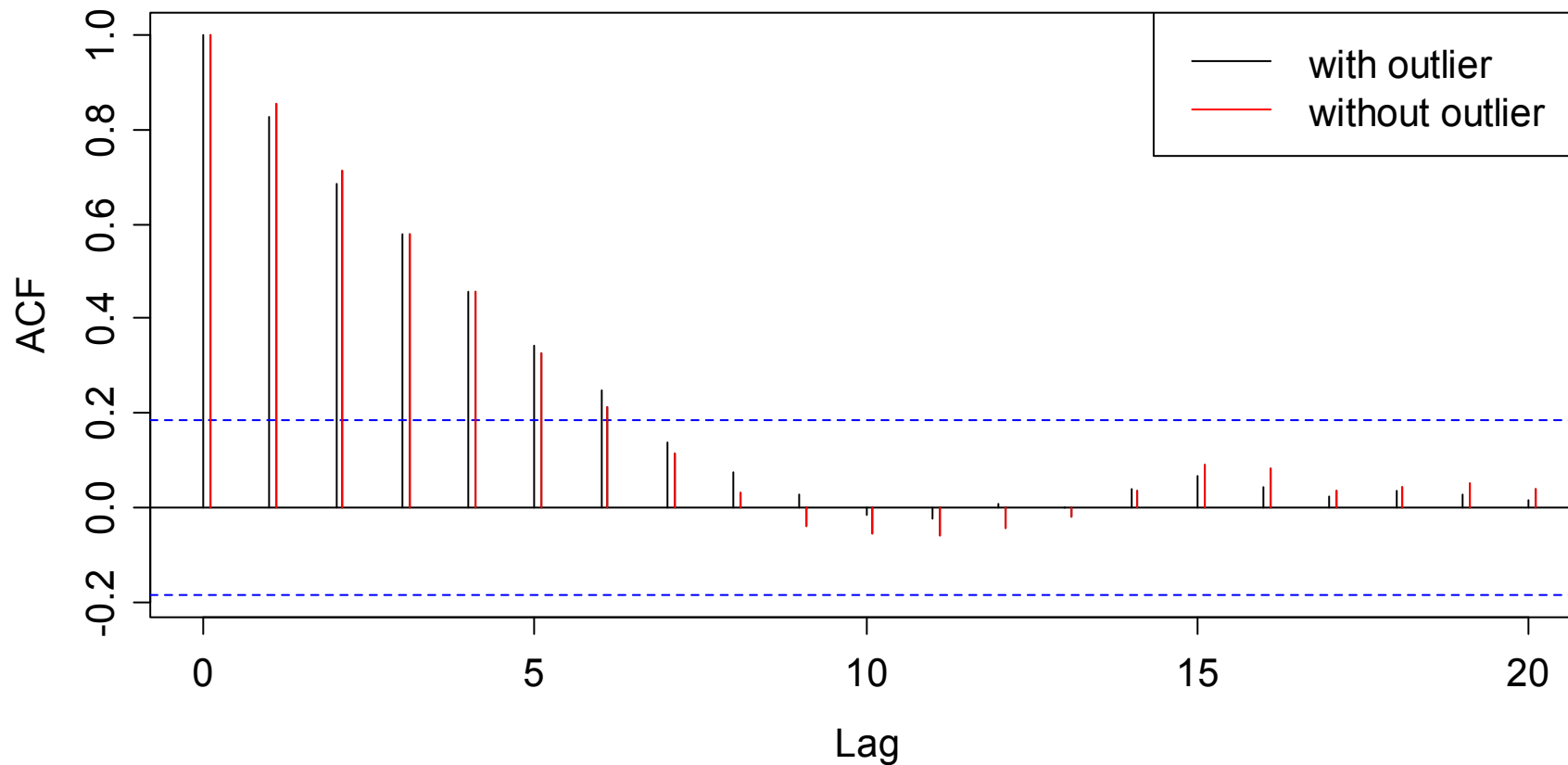


Applied Time Series Analysis

SS 2015 – Autocorrelation

Outliers and the ACF

Correlogram of Beaver Temperature Data



Applied Time Series Analysis

SS 2015 – Autocorrelation

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 bad data point: 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

Applied Time Series Analysis

SS 2015 – Autocorrelation

General Remarks about the ACF

- a) Appearance of the series \Rightarrow Appearance of the ACF
Appearance of the series $\not\Leftarrow$ 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

SS 2015 – Autocorrelation

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:

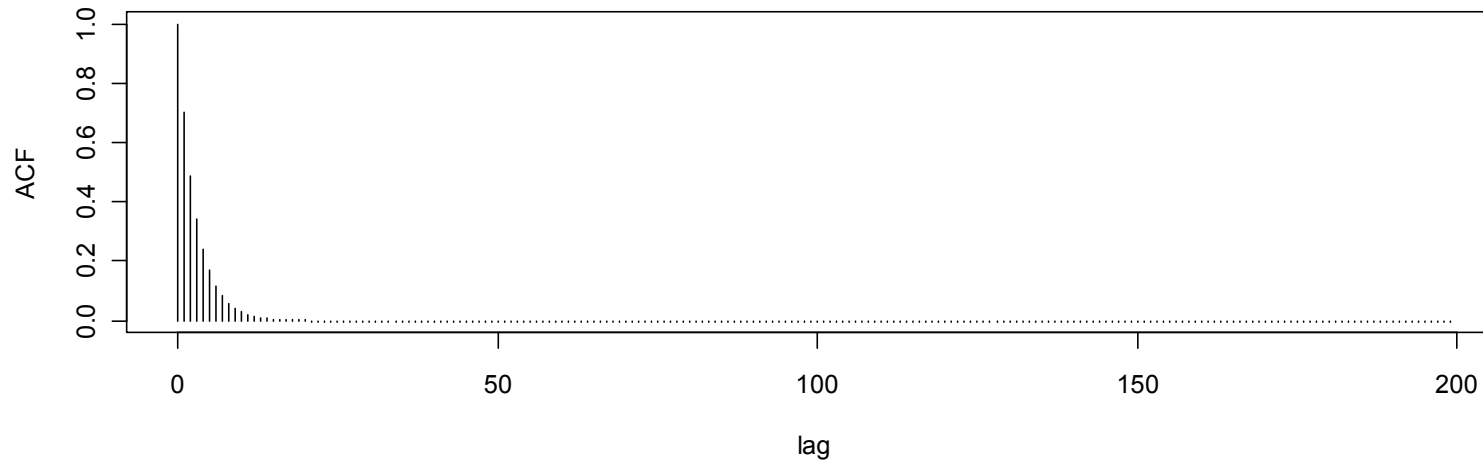
- For an AR(1) time series process, we know the true ACF
- We generate a number of realizations from this process
- We record the ACF estimates and compare to the truth

Applied Time Series Analysis

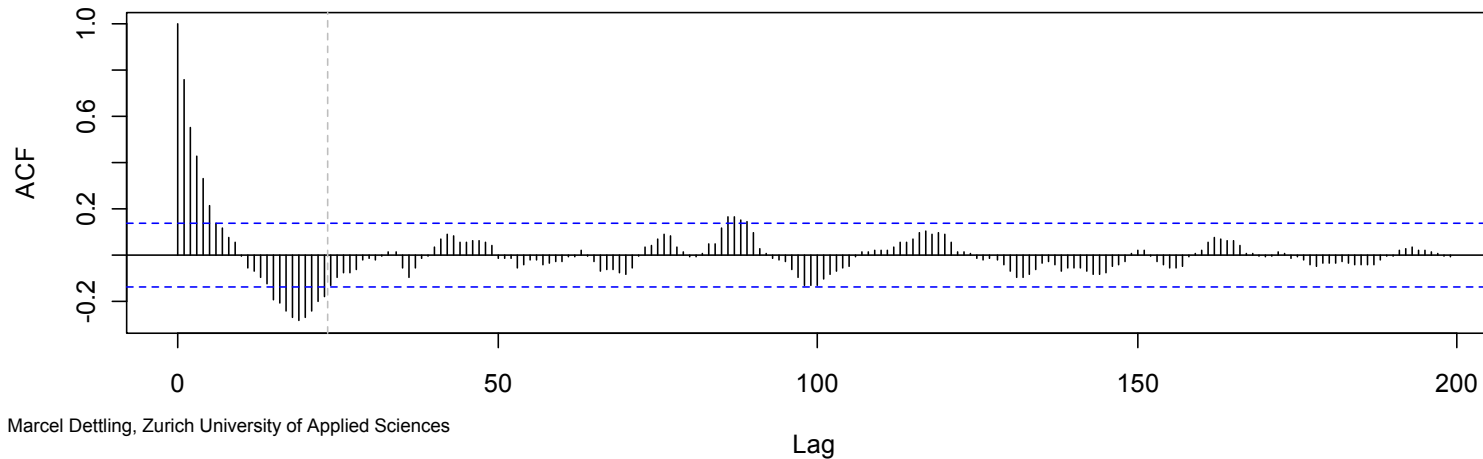
SS 2015 – Autocorrelation

Theoretical vs. Estimated ACF

True ACF of AR(1)-process with $\alpha_1=0.7$



Estimated ACF from an AR(1)-series with $\alpha_1=0.7$



Applied Time Series Analysis

SS 2015 – Autocorrelation

How Well Can We Estimate the ACF?

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

B) Repeat for $i=1, \dots, 1000$

 Simulate a **length n** AR(1)-process

 Estimate the ACF from that realization

End for

C) Boxplot the (bootstrap) sample distribution of ACF-estimates

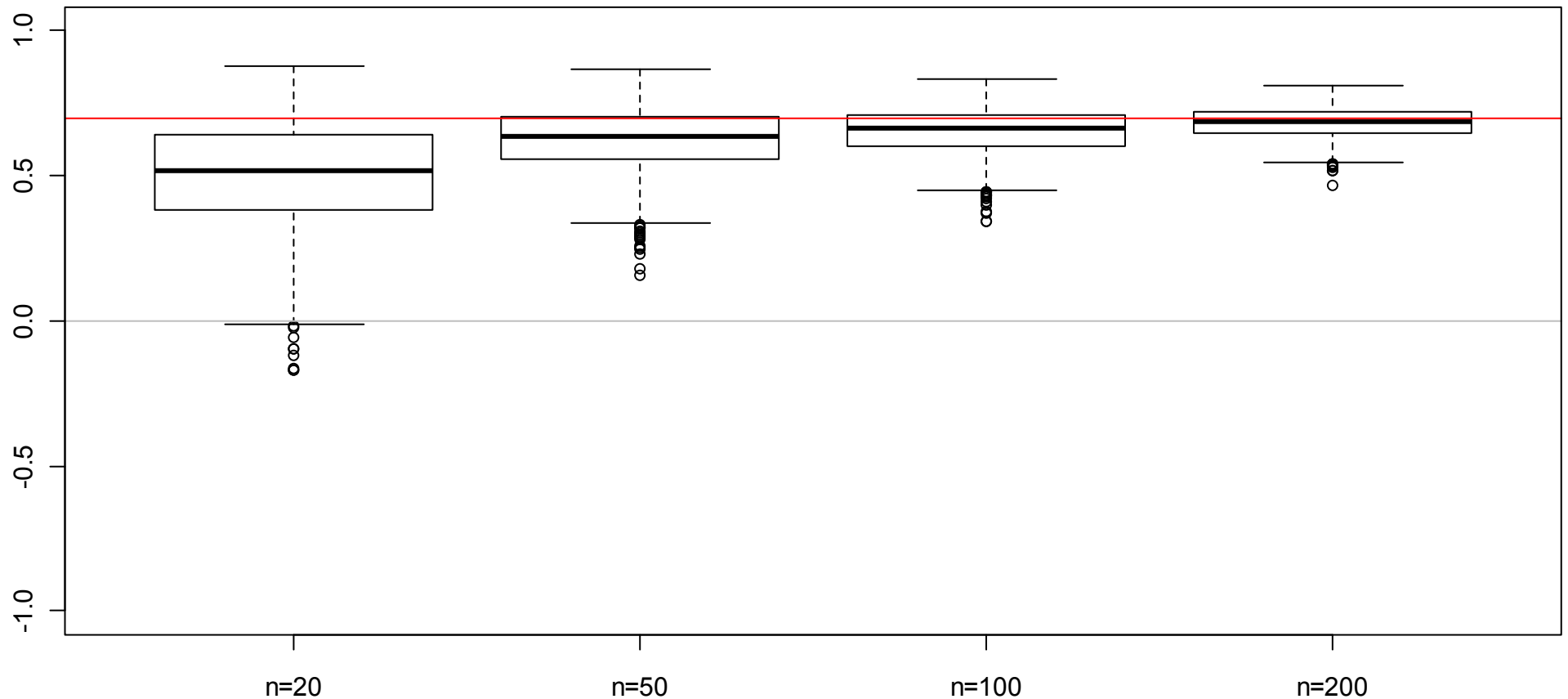
 Do so for different **lags k** and different series **length n**

Applied Time Series Analysis

SS 2015 – Autocorrelation

How Well Can We Estimate the ACF?

Variation in ACF(1) estimation

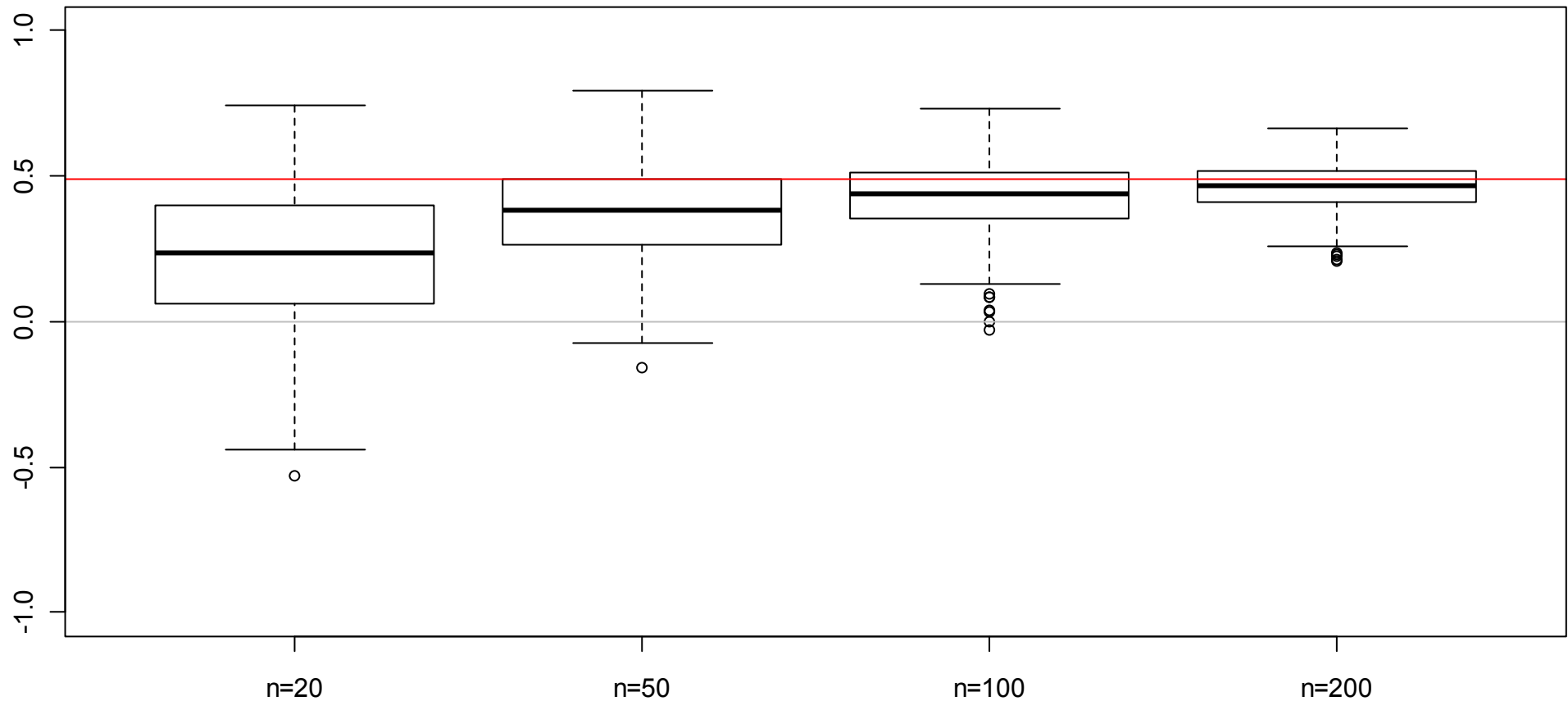


Applied Time Series Analysis

SS 2015 – Autocorrelation

How Well Can We Estimate the ACF?

Variation in ACF(2) estimation

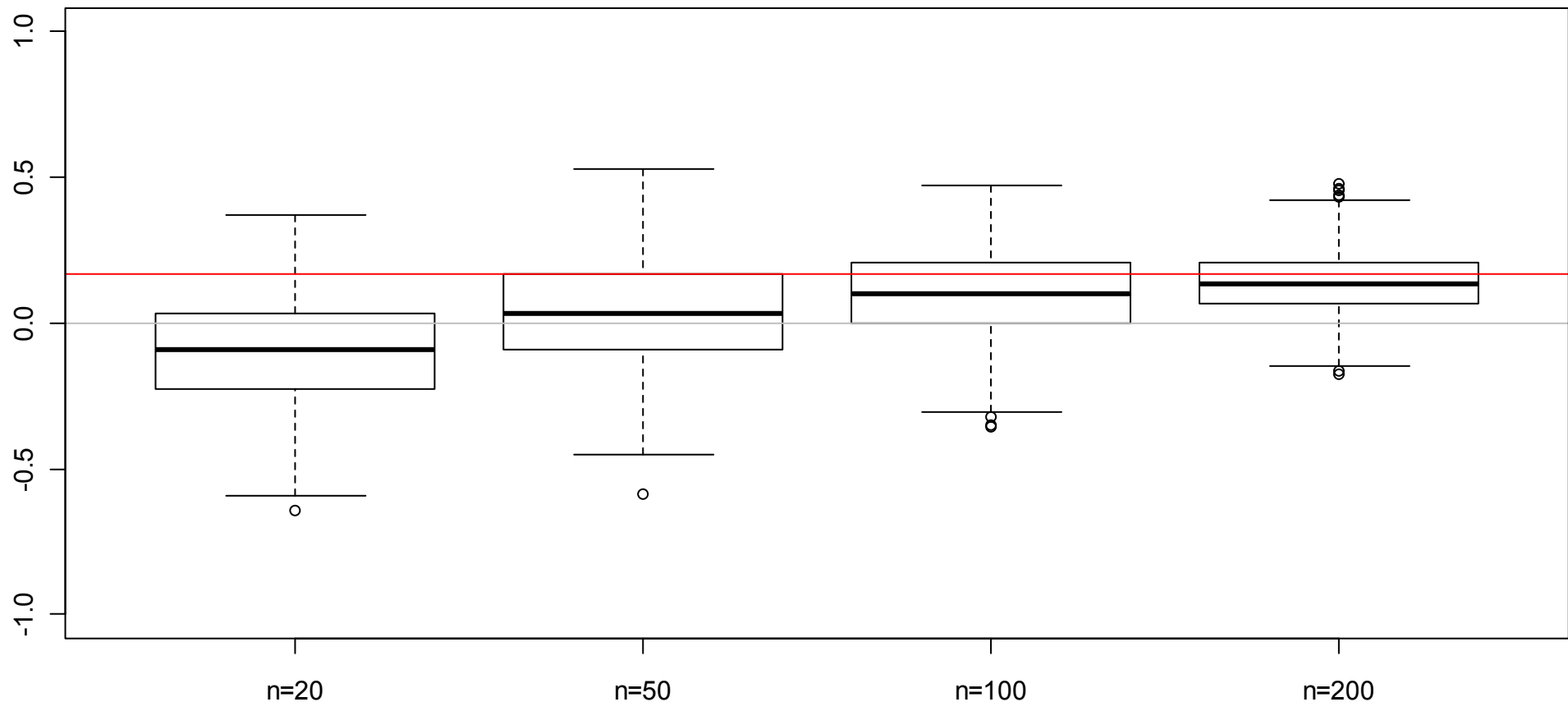


Applied Time Series Analysis

SS 2015 – Autocorrelation

How Well Can We Estimate the ACF?

Variation in ACF(5) estimation

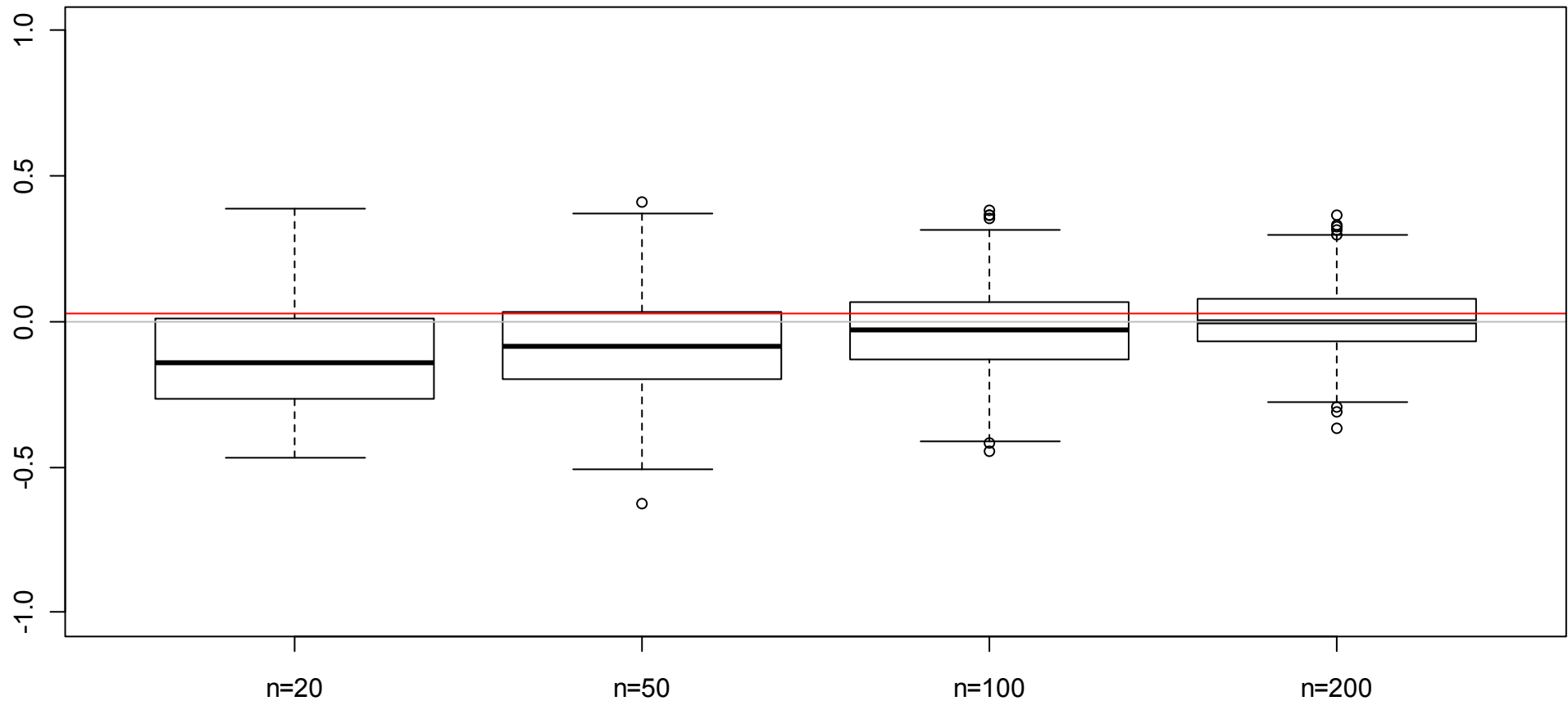


Applied Time Series Analysis

SS 2015 – Autocorrelation

How Well Can We Estimate the ACF?

Variation in ACF(10) estimation



Applied Time Series Analysis

SS 2015 – Autocorrelation

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!

Applied Time Series Analysis

SS 2015 – Autocorrelation

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...**

Applied Time Series Analysis

SS 2015 – Autocorrelation

Partial Autocorrelation Function (PACF)

The k^{th} partial autocorrelation π_k is defined as the correlation between X_{t+k} and X_t , given all the values in between.

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

Interpretation:

- Given a time series X_t , the partial autocorrelation of lag k , is the autocorrelation between X_t and X_{t+k} with the linear dependence of X_{t+1} through to X_{t+k-1} removed.
- One can draw an analogy to regression. The ACF measures the „simple“ dependence between X_t and X_{t+k} , whereas the PACF measures that dependence in a „multiple“ fashion.

Applied Time Series Analysis

SS 2015 – Autocorrelation

Facts About the PACF and Estimation

We have:

- $\pi_1 = \rho_1$
- $\pi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$ for AR(1) models, we have $\pi_2 = 0$,
because $\rho_2 = \rho_1^2$
- For estimating the PACF, we utilize the fact that for any AR(p) model, we have: $\pi_p = \alpha_p$ and $\pi_k = 0$ for all $k > p$.

Thus, for finding $\hat{\pi}_p$, we fit an AR(p) model to the series for various orders p and set $\hat{\pi}_p = \hat{\alpha}_p$

Applied Time Series Analysis

SS 2015 – Autocorrelation

Facts about the PACF

- Estimation of the PACF is implemented in R.
- 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.

Applied Time Series Analysis

SS 2015 – Basics of Modeling

Basics of Modeling

Simulation & Generation

(Time Series) Model → Data

Estimation, Inference & Residual Analysis

Data → (Time Series) Model

We will first discuss the theoretical properties of the most important time series processes and then mainly focus on how to successfully fit models to data.

Applied Time Series Analysis

SS 2015 – Basics of Modeling

A Simple Model: White Noise

A time series (W_1, W_2, \dots, W_n) is a **White Noise series** if the random variables W_1, W_2, \dots are *independent and identically* distributed with *mean zero*.

This implies that all variables W_t have the same variance σ_W^2 , and

$$\text{Cov}(W_i, W_j) = 0 \quad \text{for all } i \neq j.$$

Thus, there are no autocorrelations either: $\rho_k = 0$ for all $k \neq 0$.

If in addition, the variables also follow a *Gaussian distribution*, i.e. $W_t \sim N(0, \sigma_W^2)$, the series is called **Gaussian White Noise**.

The term White Noise is due to the analogy to white light.

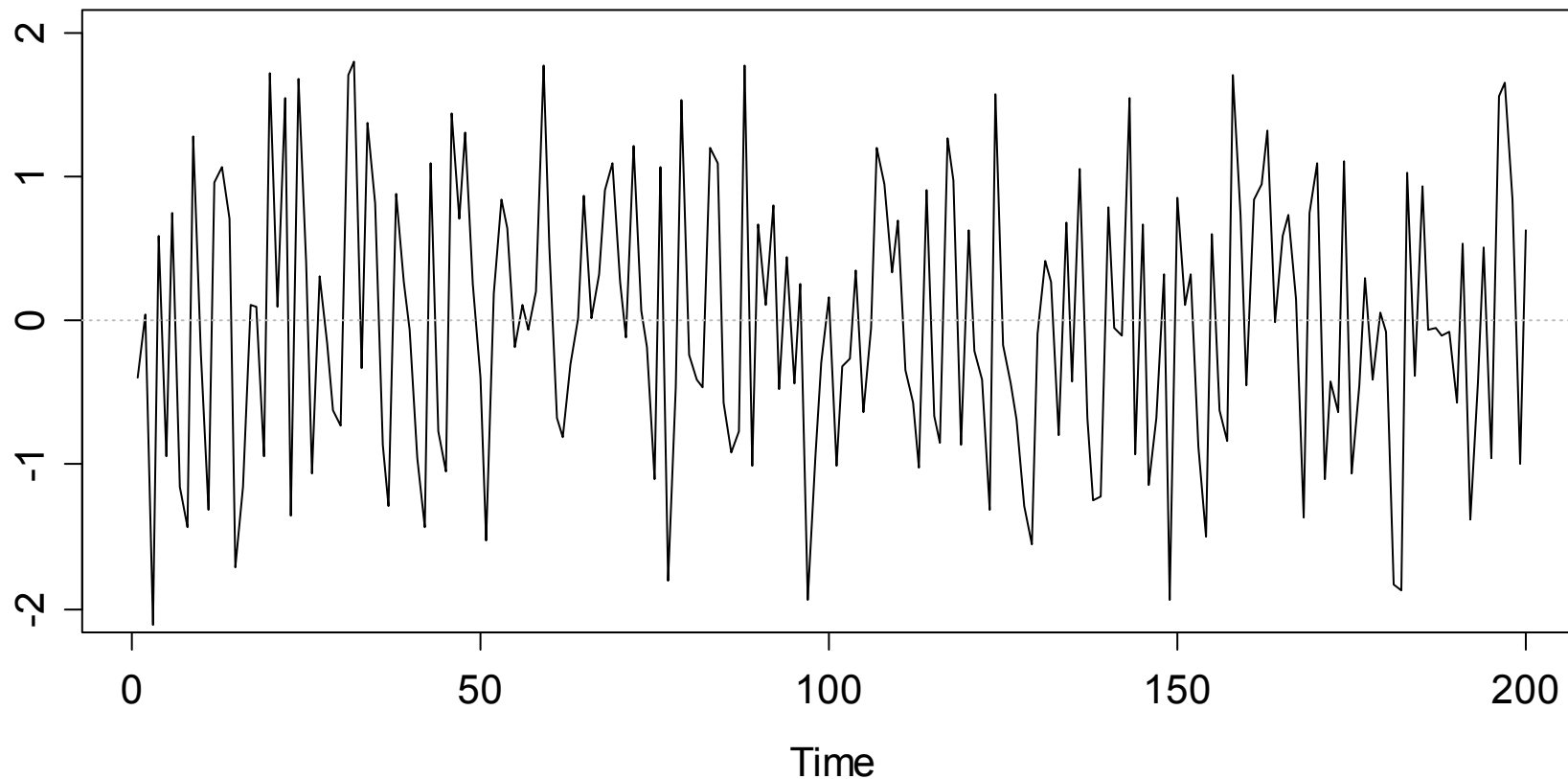
Applied Time Series Analysis

SS 2015 – Basics of Modeling

Example: Gaussian White Noise

```
> plot(ts(rnorm(200, mean=0, sd=1)))
```

Gaussian White Noise

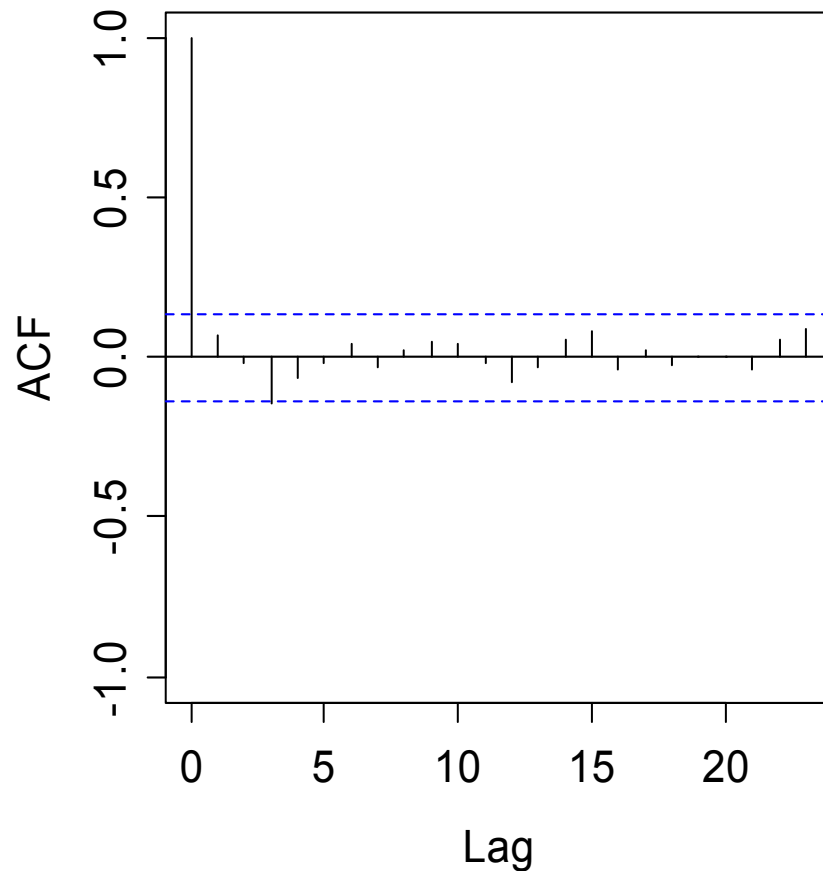


Applied Time Series Analysis

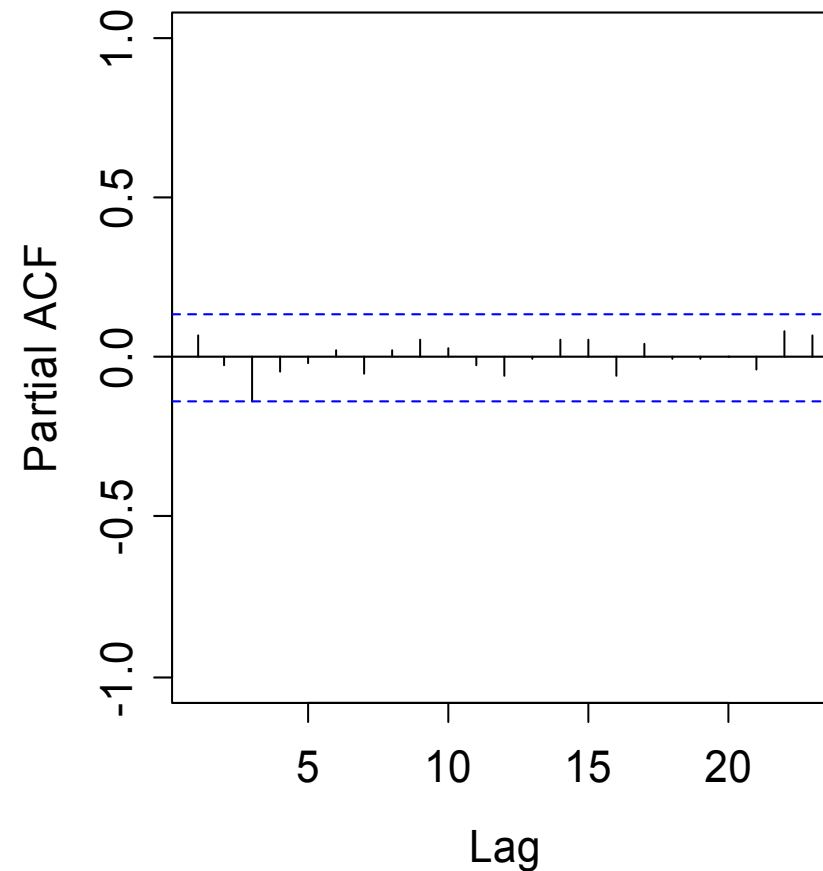
SS 2015 – Basics of Modeling

Example: Gaussian White Noise

ACF



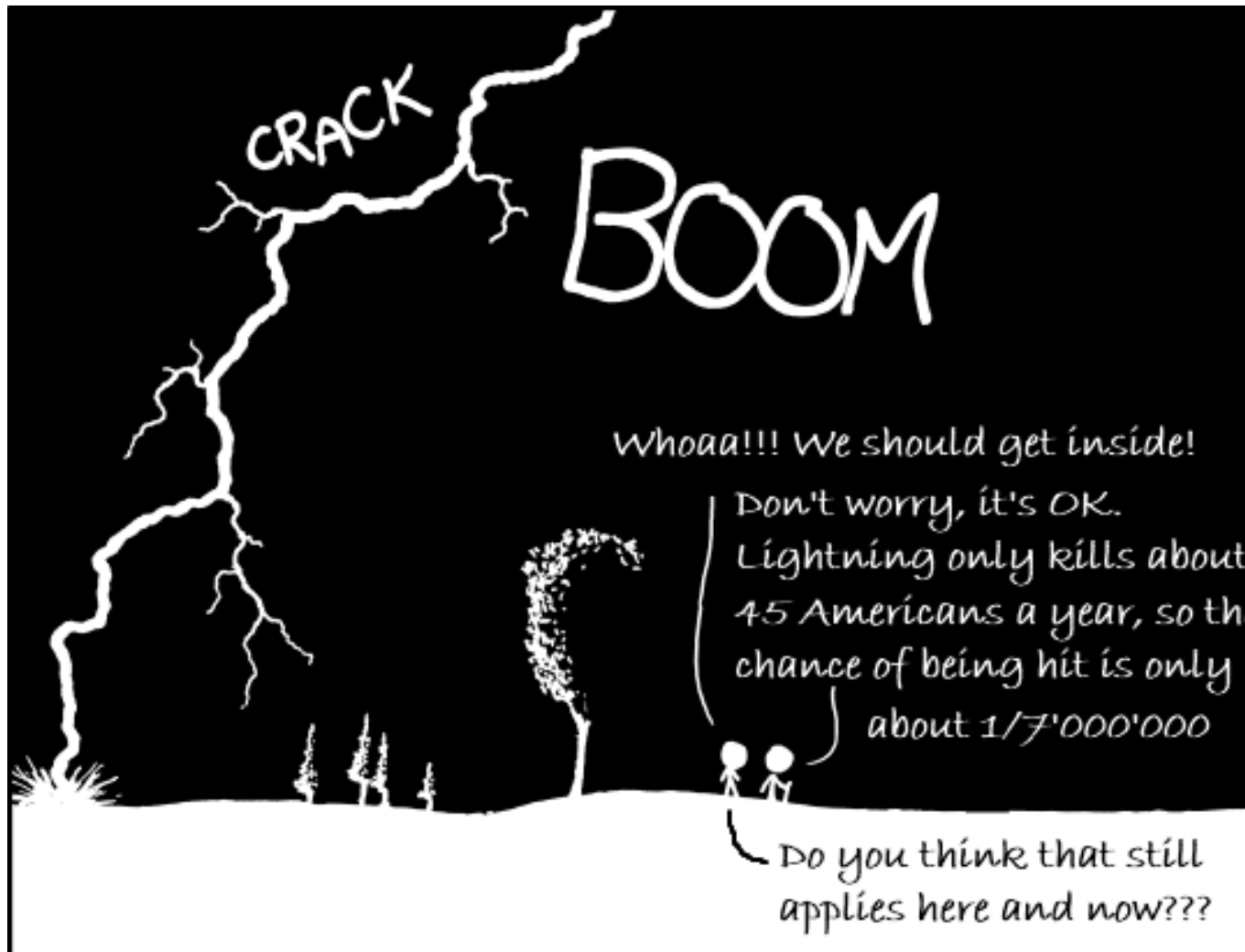
PACF



Applied Time Series Analysis

SS 2015 – Basics of Modeling

Estimating the Conditional Mean



→ see blackboard...

Applied Time Series Analysis

SS 2015 – Basics of Modeling

Time Series Modeling

There is a wealth of time series models

- AR autoregressive model
- MA moving average model
- ARMA combination of AR & MA
- ARIMA non-stationary ARMAs
- SARIMA seasonal ARIMAs
- ...

We start by discussing autoregressive models. They are perhaps the simplest and most intuitive time series models that exist.

Applied Time Series Analysis

SS 2015 – Autoregressive Models

Basic Idea for AR(p)-Models

We have a process where the random variable X_t depends on an auto-regressive linear combination of the preceding X_{t-1}, \dots, X_{t-p} , plus a „completely independent“ term called innovation E_t .

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + E_t$$

Here, p is called the order of the autoregressive model. Hence, we abbreviate by AR(p). An alternative notation is with the backshift operator B :

$$(1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p) X_t = E_t \quad \text{or short, } \Phi(B) X_t = E_t$$

Here, $\Phi(B)$ is called the characteristic polynomial of the AR(p). It determines most of the relevant properties of the process.

Applied Time Series Analysis

SS 2015 – Autoregressive Models

AR(1)-Model

The simplest model is the AR(1)-model

$$X_t = \alpha_1 X_{t-1} + E_t$$

where

E_t is i.i.d with $E[E_t] = 0$ and $Var(E_t) = \sigma_E^2$
We also require that E_t is independent of $X_s, s < t$

Under these conditions, E_t is a **causal White Noise** process, or an **innovation**. Be aware that this is stronger than the iid requirement: not every iid process is an innovation and that property is central to AR(p)-modelling.

Applied Time Series Analysis

SS 2015 – Autoregressive Models

AR(p)-Models and Stationarity

The following is absolutely essential:

AR(p) models must only be fitted to stationary time series. Any potential trends and/or seasonal effects need to be removed first. We will also make sure that the processes are stationary.

Under which circumstances is an AR(p) stationary?

→ **see blackboard...**

Applied Time Series Analysis

SS 2015 – Autoregressive Models

Stationarity of AR(p)-Processes

As we have seen, any stationary AR(p) meets:

1) $E[X_t] = \mu = 0$

2) The condition on $(\alpha_1, \dots, \alpha_p)$:

All (complex) roots of the characteristic polynomial

$$1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = 0$$

lie outside of the unit circle (can be verified with `polyroot()`)

We can always shift a stationary AR(p) process: $Y_t = m + X_t$
The resulting process is still stationary and allows for greater flexibility in modelling. It is a **shifted AR(p) process**.

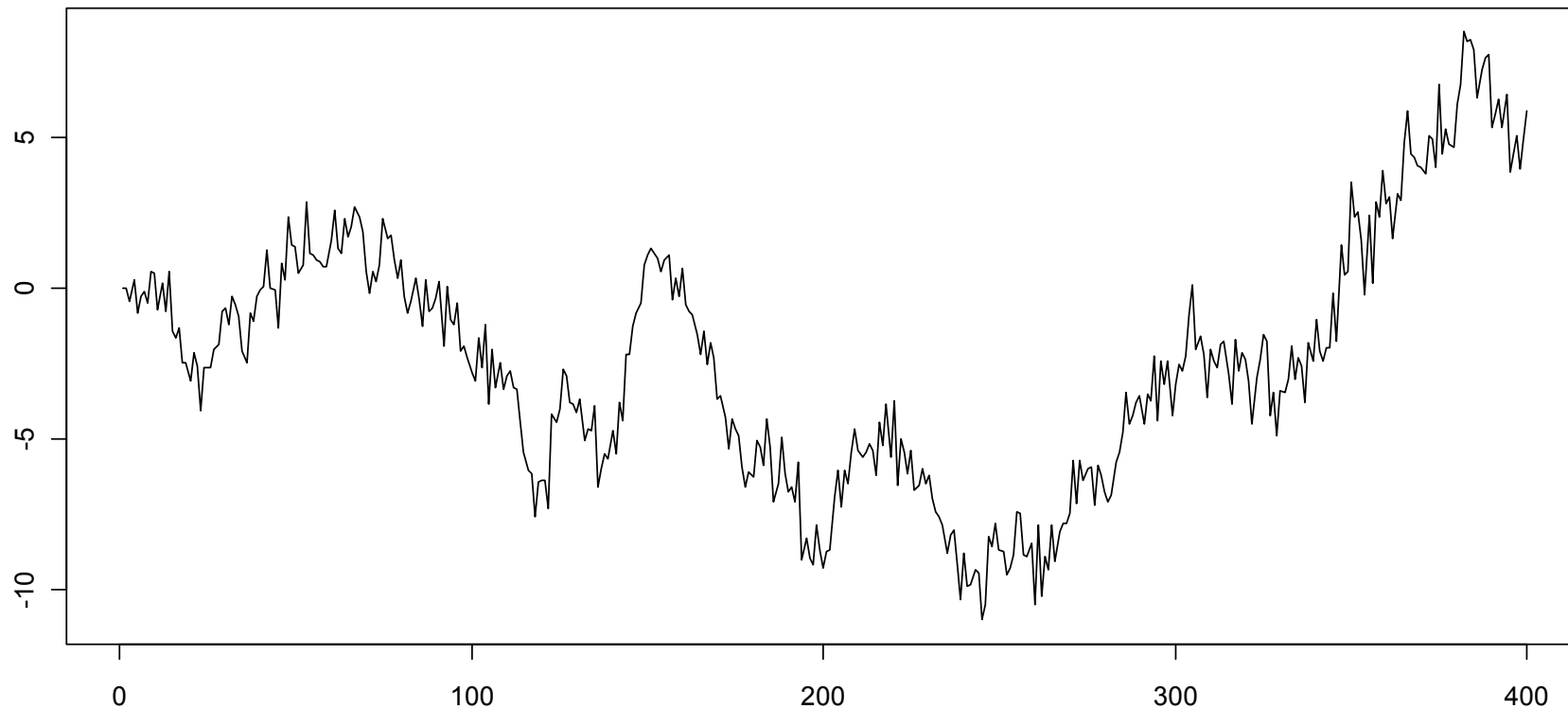
Applied Time Series Analysis

SS 2015 – Autoregressive Models

A Non-Stationary AR(2)-Process

$$X_t = \frac{1}{2} X_{t-1} + \frac{1}{2} X_{t-2} + E_t \text{ is not stationary...}$$

Non-Stationary AR(2)

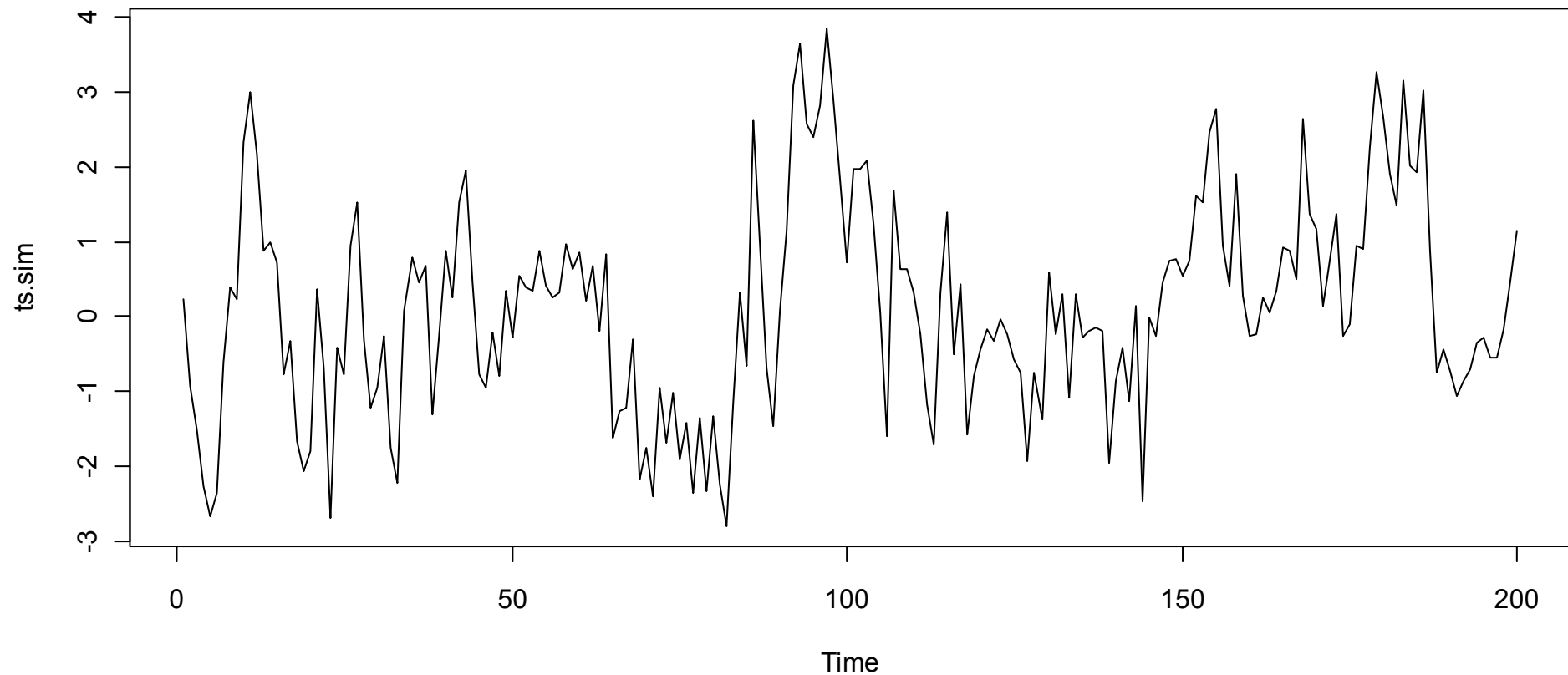


Applied Time Series Analysis

SS 2015 – Autoregressive Models

Simulated AR(1)-Series

Simulated AR(1)-Series: $\alpha_1=0.7$

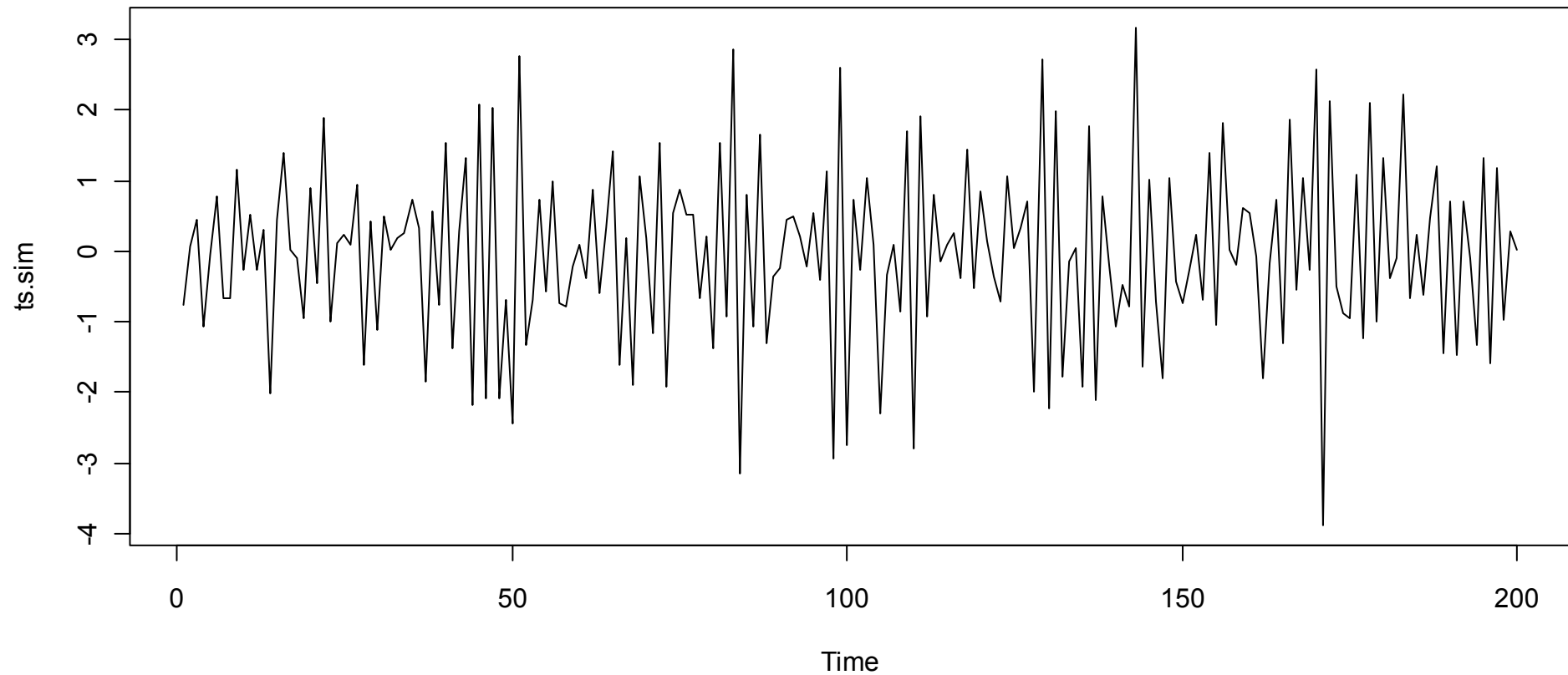


Applied Time Series Analysis

SS 2015 – Autoregressive Models

Simulated AR(1)-Series

Simulated AR(1)-Series: $\alpha_1 = -0.7$

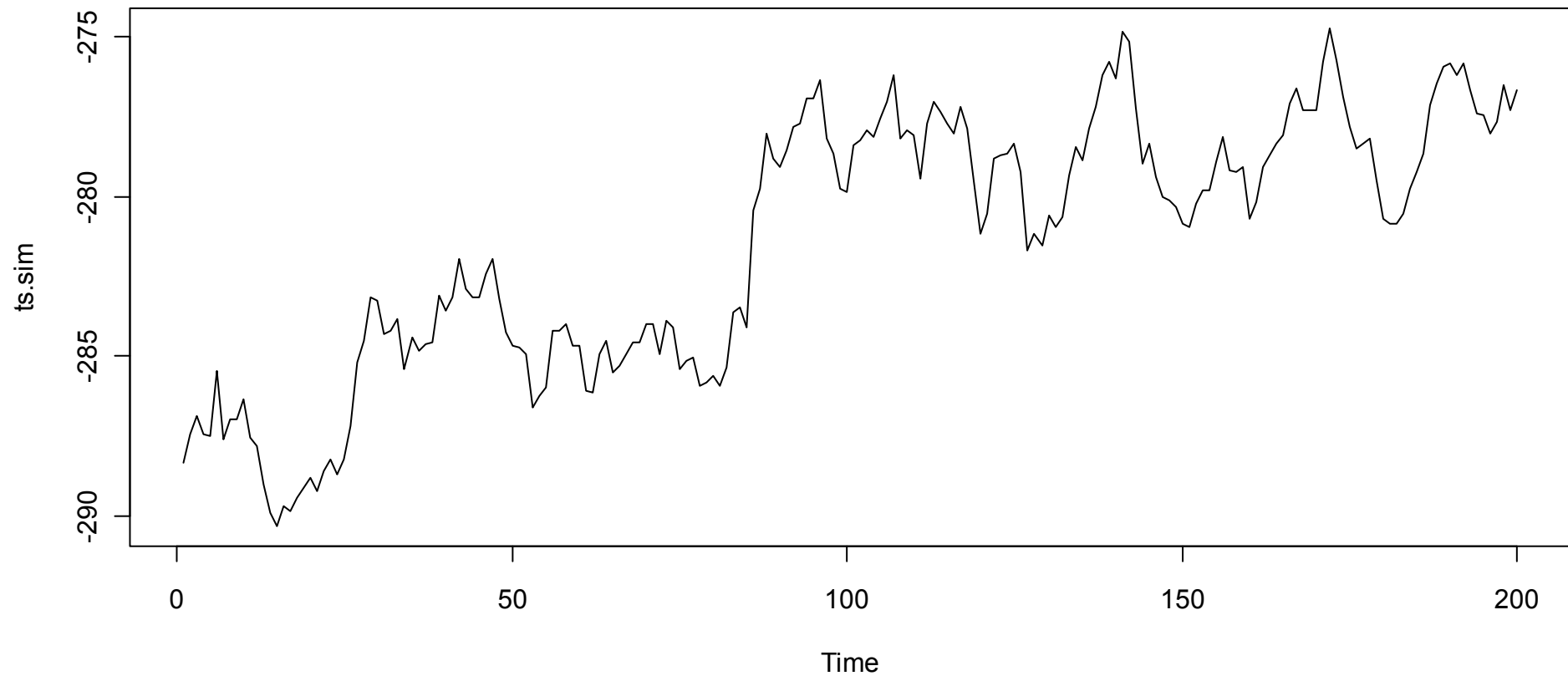


Applied Time Series Analysis

SS 2015 – Autoregressive Models

Simulated AR(1)-Series

Simulated AR(1)-Series: alpha_1=1



Applied Time Series Analysis

SS 2015 – Autoregressive Models

Autocorrelation of AR(p) Processes

On the blackboard...

Yule-Walker Equations

We observe that there exists a linear equation system built up from the AR(p)-coefficients and the ACF-coefficients of up to lag p . These are called Yule-Walker-Equations.

We can use these equations for fitting an AR(p)-model:

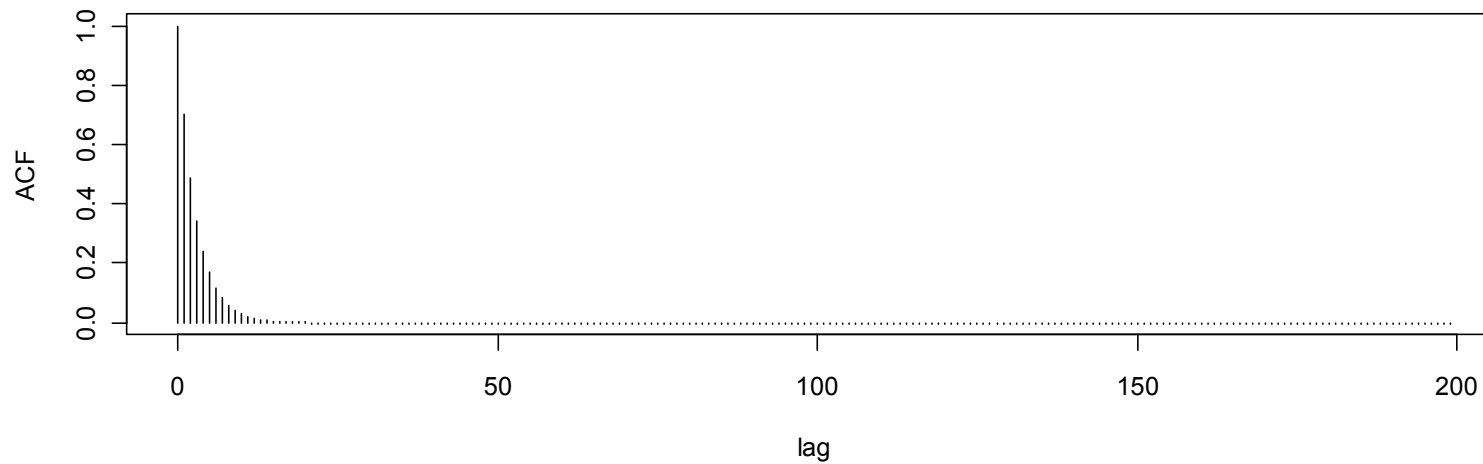
- 1) *Estimate the ACF from a time series*
- 2) *Plug-in the estimates into the Yule-Walker-Equations*
- 3) *The solution are the AR(p)-coefficients*

Applied Time Series Analysis

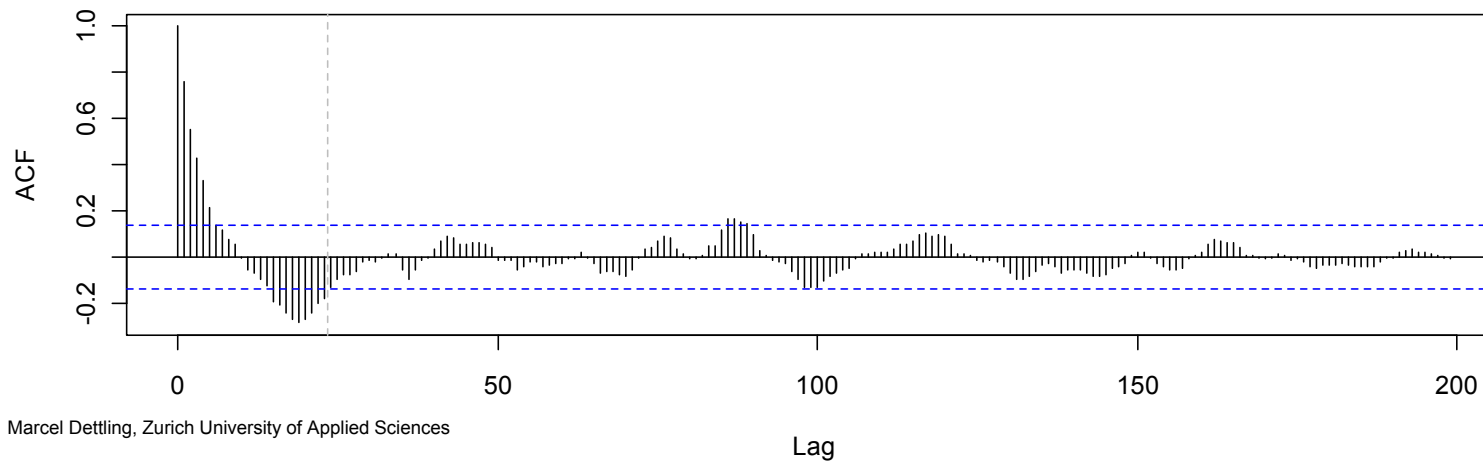
SS 2015 – Autoregressive Models

Theoretical vs. Estimated ACF

True ACF of AR(1)-process with $\alpha_1=0.7$



Estimated ACF from an AR(1)-series with $\alpha_1=0.7$

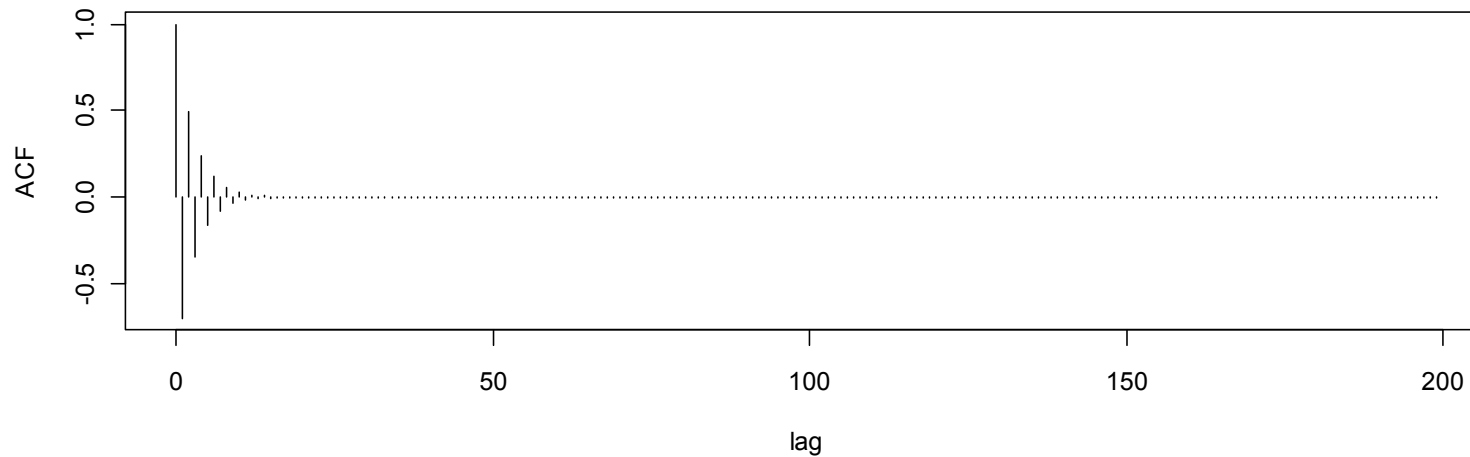


Applied Time Series Analysis

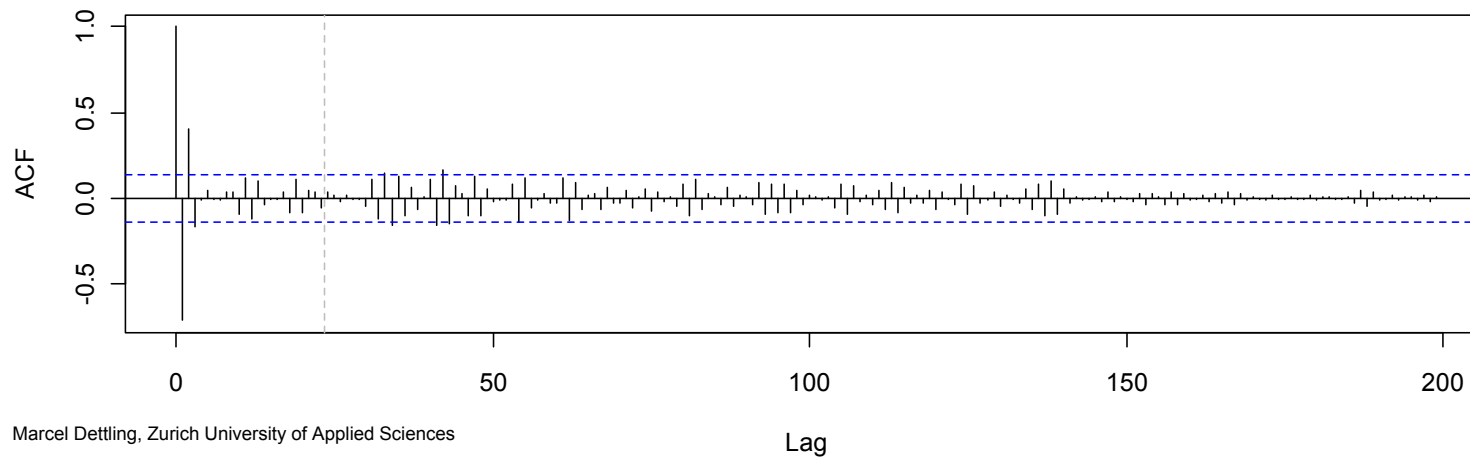
SS 2015 – Autoregressive Models

Theoretical vs. Estimated ACF

True ACF of AR(1)-process with $\alpha_1=-0.7$



Estimated ACF from an AR(1)-series with $\alpha_1=-0.7$



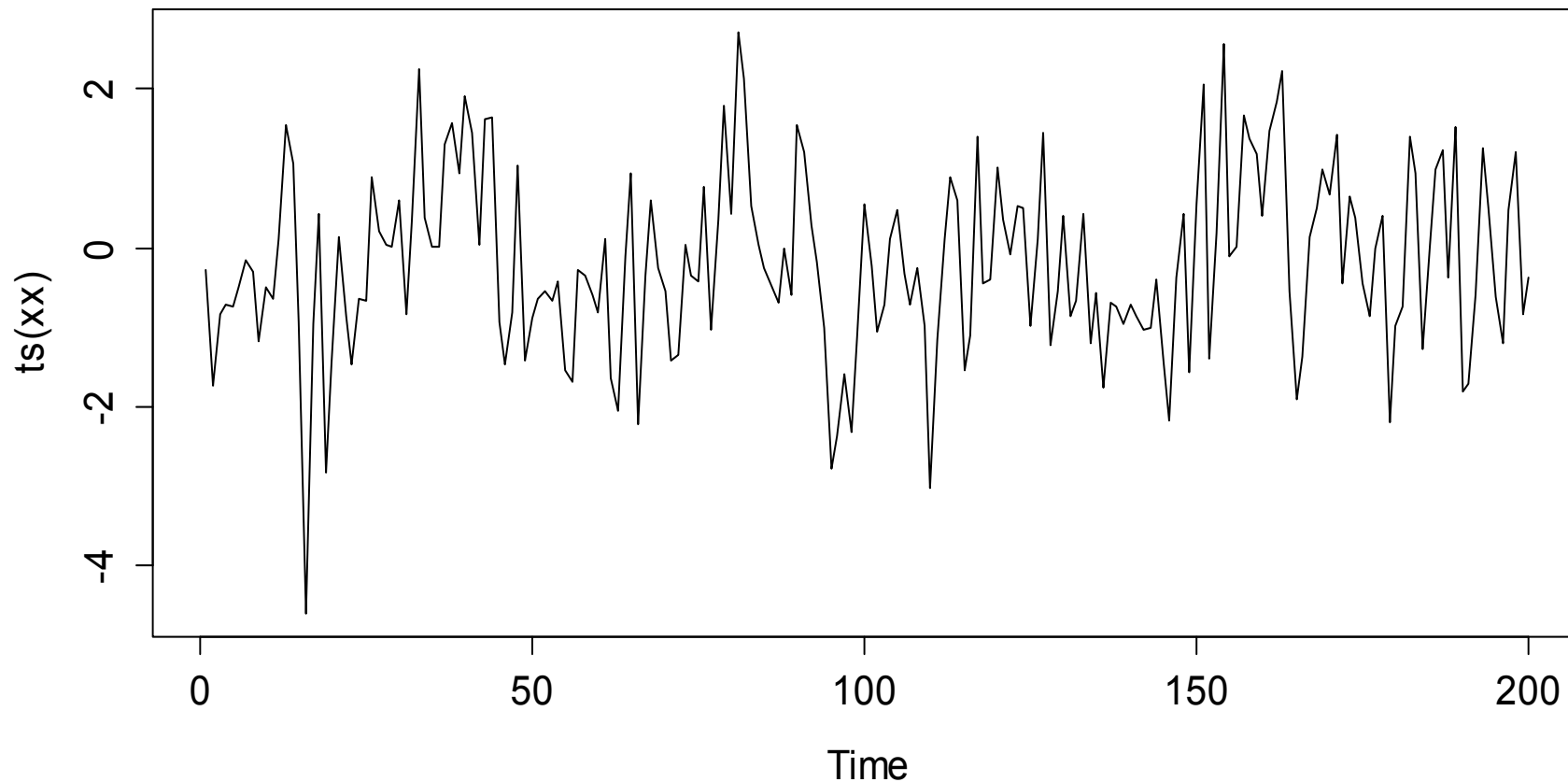
Applied Time Series Analysis

SS 2015 – Autoregressive Models

AR(3): Simulation and Properties

```
> xx <- arima.sim(list(ar=c(0.4, -0.2, 0.3)),
```

AR(3) with $\alpha_1=-0.4$, $\alpha_2=-0.2$, $\alpha_3=0.3$



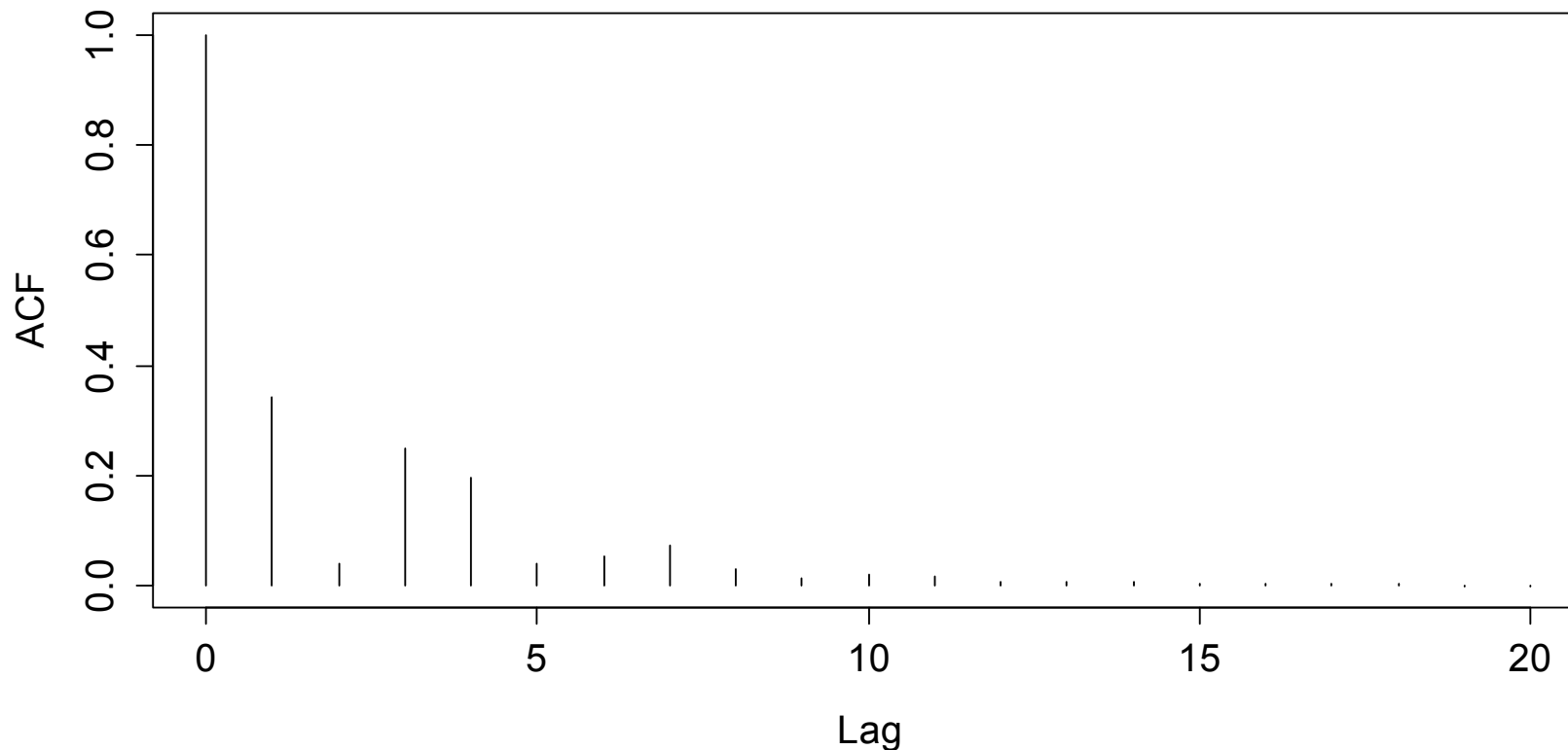
Applied Time Series Analysis

SS 2015 – Autoregressive Models

AR(3): Simulation and Properties

```
> autocorr <- ARMAacf(ar=c(0.4, -0.2, 0.3),...)
> plot(0:20, autocorr, type="h", xlab="Lag")
```

Theoretical Autocorrelation for an AR(3)



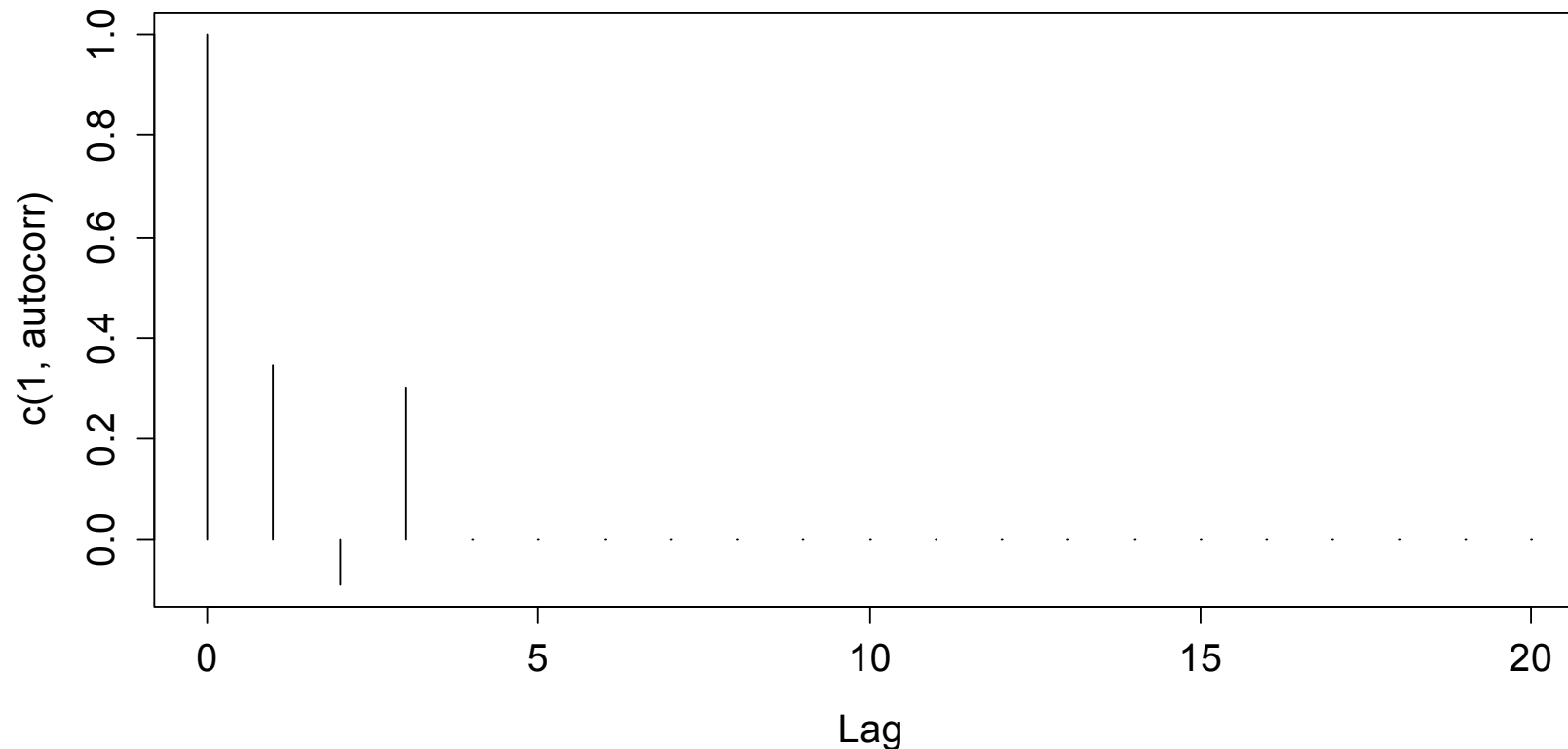
Applied Time Series Analysis

SS 2015 – Autoregressive Models

AR(3): Simulation and Properties

```
> autocorr <- ARMAacf(ar=..., pacf=TRUE, ...)  
> plot(0:20, autocorr, type="h", xlab="Lag")
```

Theoretical Partial Autocorrelation for an AR(3)



Applied Time Series Analysis

SS 2015 – Autoregressive Models

Fitting AR(p)-Models

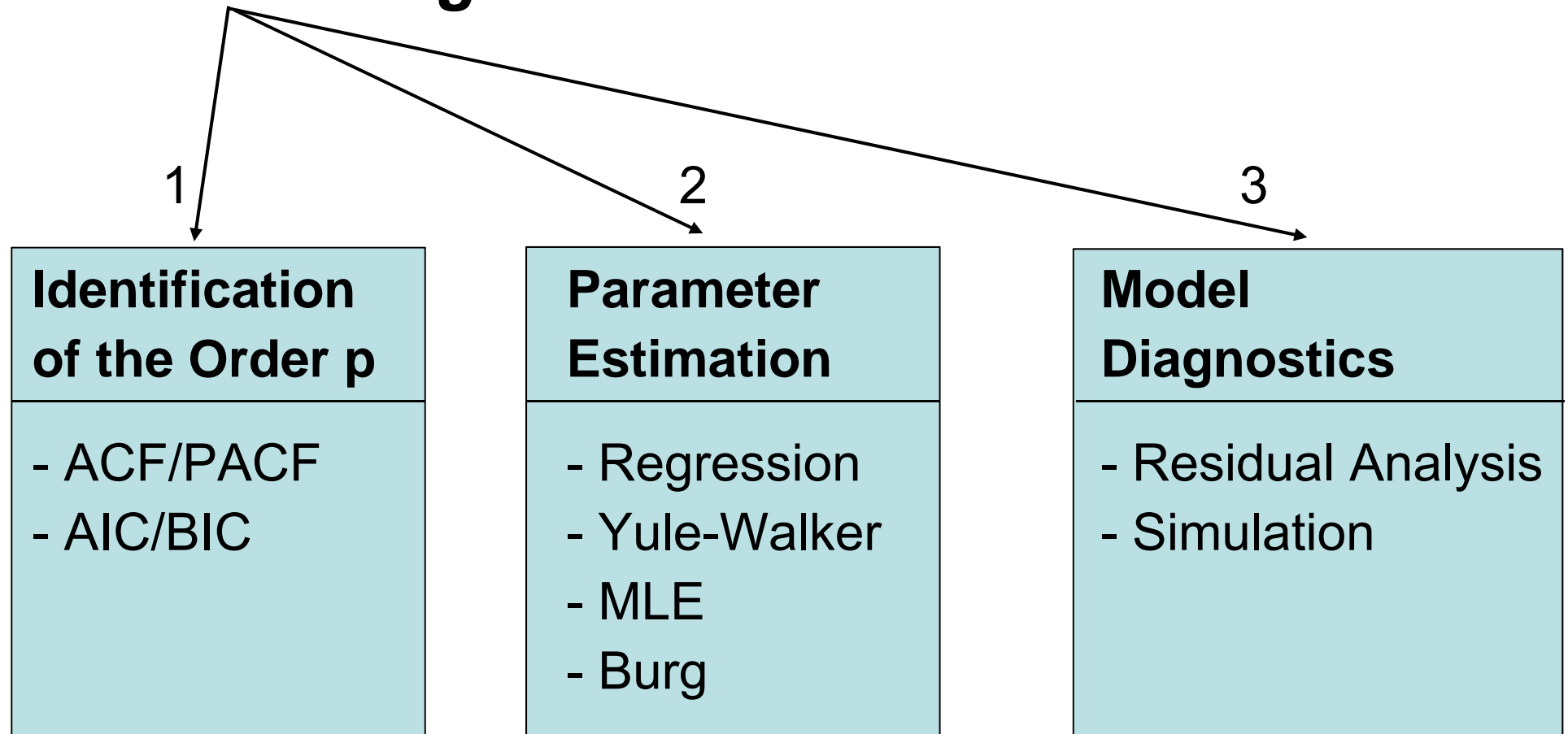
This involves 3 crucial steps:

- 1) **Is an AR(p) suitable, and what is p?**
 - will be based on ACF/PACF-Analysis
- 2) **Estimation of the AR(p)-coefficients**
 - Regression approach
 - Yule-Walker-Equations
 - and more (MLE, Burg-Algorithm)
- 3) **Residual Analysis**
 - to be discussed

Applied Time Series Analysis

SS 2015 – Autoregressive Models

AR-Modelling



Applied Time Series Analysis

SS 2015 – Autoregressive Models

Is an AR(p) suitable, and what is p?

- For all AR(p)-models, the **ACF** decays exponentially quickly, or is an exponentially damped sinusoid.
- For all AR(p)-models, the **PACF** is equal to zero for all lags $k > p$. The behavior before lag p can be anything.

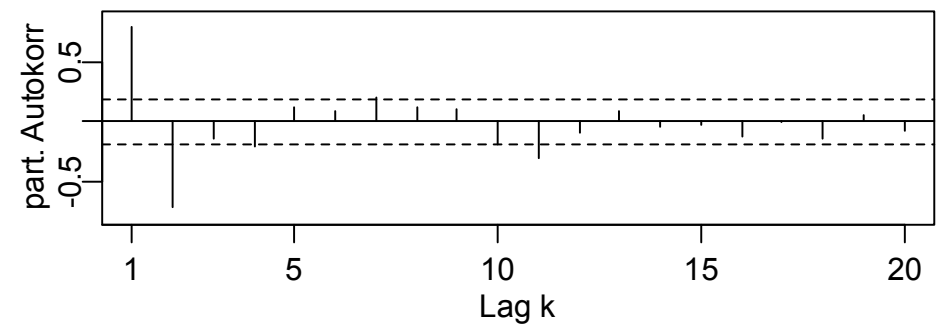
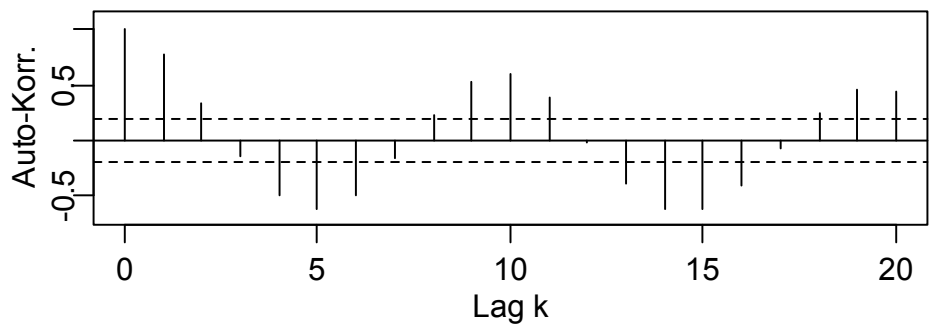
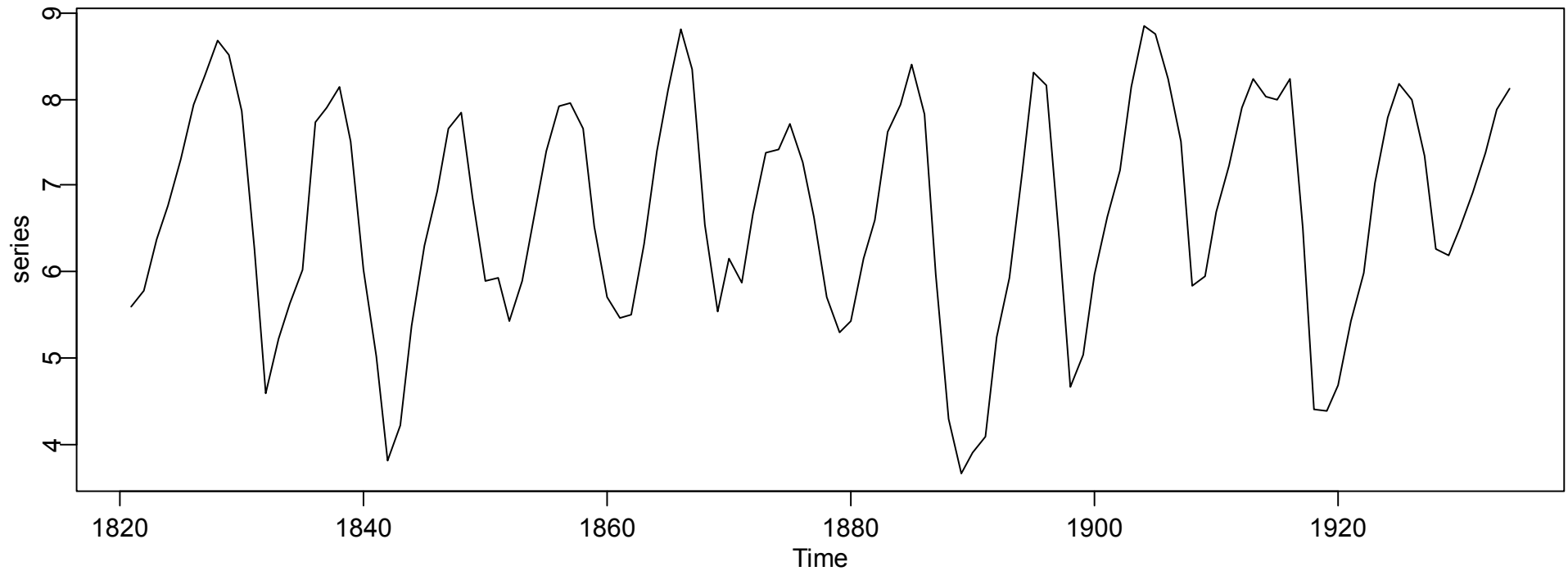
If what we observe is fundamentally different from the above, it is unlikely that the series was generated from an AR(p)-process. We thus need other models, maybe more sophisticated ones.

Remember that the sample ACF has a few peculiarities (bias, variability, compensation issue) and is tricky to interpret!!!

Applied Time Series Analysis

SS 2015 – Autoregressive Models

Model Order for $\log(\text{lynx})$



Applied Time Series Analysis

SS 2015 – Autoregressive Models

Parameter Estimation for AR(p)

Observed time series are rarely centered. Then, it is inappropriate to fit a pure AR(p) process. All R routines by default assume the shifted process $Y_t = m + X_t$. Thus, we face the problem:

$$(Y_t - m) = \alpha_1(Y_{t-1} - m) + \dots + \alpha_p(Y_{t-p} - m) + E_t$$

The goal is to estimate the *global mean* m , the *AR-coefficients* $\alpha_1, \dots, \alpha_p$, and some parameters defining the distribution of the innovation E_t . We usually assume a Gaussian, hence this is σ_E^2 .

We will discuss 4 methods for estimating the parameters:

OLS, Burg's algorithm, Yule-Walker, MLE

Applied Time Series Analysis

SS 2015 – Autoregressive Models

OLS Estimation

If we rethink the previously stated problem:

$$(Y_t - m) = \alpha_1 (Y_{t-1} - m) + \dots + \alpha_p (Y_{t-p} - m) + E_t$$

we recognize a multiple linear regression problem without intercept on the centered observations. What we need to do is:

- 1) Estimate $\hat{m} = \bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$ and determine $x_t = y_t - \hat{m}$
- 2) Run a regression w/o intercept on x_t to obtain $\hat{\alpha}_1, \dots, \hat{\alpha}_p$
- 3) For $\hat{\sigma}_E^2$, take the residual standard error from the output.

This all works without any time series software, but is a bit cumbersome to implement. Dedicated procedures exist...

Applied Time Series Analysis

SS 2015 – Autoregressive Models

OLS Estimation

```
> f.ols <- ar.ols(llynx, aic=F, inter=F, order=2)
```

```
> f.ols
```

```
Coefficients:
```

```
          1          2  
1.3844   -0.7479
```

```
Order selected 2   sigma^2 estimated as 0.2738
```

```
> f.ols$x.mean
```

```
[1] 6.685933
```

```
> sum(na.omit(f.ols$resid)^2)/112
```

```
[1] 0.2737594
```

Applied Time Series Analysis

SS 2015 – Autoregressive Models

Burg's Algorithm

While OLS works, the first p instances are never evaluated as responses. This is cured by Burg's algorithm, which uses the property of time-reversal in stochastic processes. We thus evaluate the RSS of forward and backward prediction errors:

$$\sum_{t=p+1}^n \left\{ \left(X_t - \sum_{k=1}^p \alpha_k X_{t-k} \right)^2 + \left(X_{t-p} - \sum_{k=1}^p \alpha_k X_{t-p+k} \right)^2 \right\}$$

In contrast to OLS, there is no explicit solution and numerical optimization is required. This is done with a recursive method called the Durbin-Levison algorithm (implemented in R).

Applied Time Series Analysis

SS 2015 – Autoregressive Models

Burg's Algorithm

```
> f.burg <- ar.burg(llynx, aic=F, order.max=2)
> f.burg
```

Coefficients:

1	2
1.3831	-0.7461

```
Order selected 2  sigma^2 estimated as  0.2707
```

```
> f.ar.burg$x.mean
[1] 6.685933
```

Note: The innovation variance is estimated from the Durbin-Levinson updates and not from the residuals using the MLE!

Applied Time Series Analysis

SS 2015 – Autoregressive Models

Yule-Walker Equations

The Yule-Walker-Equations yield a LES that connects the true ACF with the true AR-model parameters. We plug-in the estimated ACF coefficients

$$\hat{\rho}(k) = \hat{\alpha}_1 \hat{\rho}(k-1) + \dots + \hat{\alpha}_p \hat{\rho}(k-p) \text{ for } k=1, \dots, p$$

and can solve the LES to obtain the AR-parameter estimates.

- \hat{m} is the arithmetic mean of the time series
- $\hat{\sigma}_E^2$ is obtained from the fitted coefficients via the autocovariance of the series and takes a different value than before!

There is an implementation in R with function `ar.yw()`.

Applied Time Series Analysis

SS 2015 – Autoregressive Models

Yule-Walker Equations

```
> f.ar.yw
```

```
Call: ar.yw.default(x = log(lynx), aic = FALSE,  
order.max = 2)
```

```
Coefficients:
```

```
          1          2  
1.3504 -0.7200
```

```
Order selected 2  sigma^2 estimated as  0.3109
```

While the Yule-Walker method is asymptotically equivalent to OLS and Burg's algorithm, it generally yields a solution with worse Gaussian likelihood on finite samples

Applied Time Series Analysis

SS 2015 – Autoregressive Models

Maximum-Likelihood-Estimation

Idea: Determine the parameters such that, given the observed time series (y_1, \dots, y_n) , the resulting model is the most plausible (i.e. the most likely) one.

This requires the choice of a probability model for the time series. By assuming Gaussian innovations, $E_t \sim N(0, \sigma_E^2)$, any AR(p) process has a multivariate normal distribution:

$$Y = (Y_1, \dots, Y_n) \sim N(m \cdot \underline{1}, V), \text{ with } V \text{ depending on } \underline{\alpha}, \sigma_E^2$$

MLE then provides simultaneous estimates by optimizing:

$$L(\alpha, m, \sigma_E^2) \propto \exp\left(-\frac{1}{2\sigma_E^2} \sum_{t=1}^n (x_t - \hat{x}_t)^2\right)$$

Applied Time Series Analysis

SS 2015 – Autoregressive Models

Maximum-Likelihood Estimation

```
> f.ar.mle
```

```
Call: arima(x = log(lynx), order = c(2, 0, 0))
```

```
Coefficients:
```

	ar1	ar2	intercept
	1.3776	-0.7399	6.6863
s.e.	0.0614	0.0612	0.1349

```
sigma^2=0.2708; log likelihood=-88.58; aic=185.15
```

While MLE by default assumes Gaussian innovations, it still performs reasonably for other distributions as long as they are not extremely skewed or have very precarious outliers.

Applied Time Series Analysis

SS 2015 – Autoregressive Models

Practical Aspects

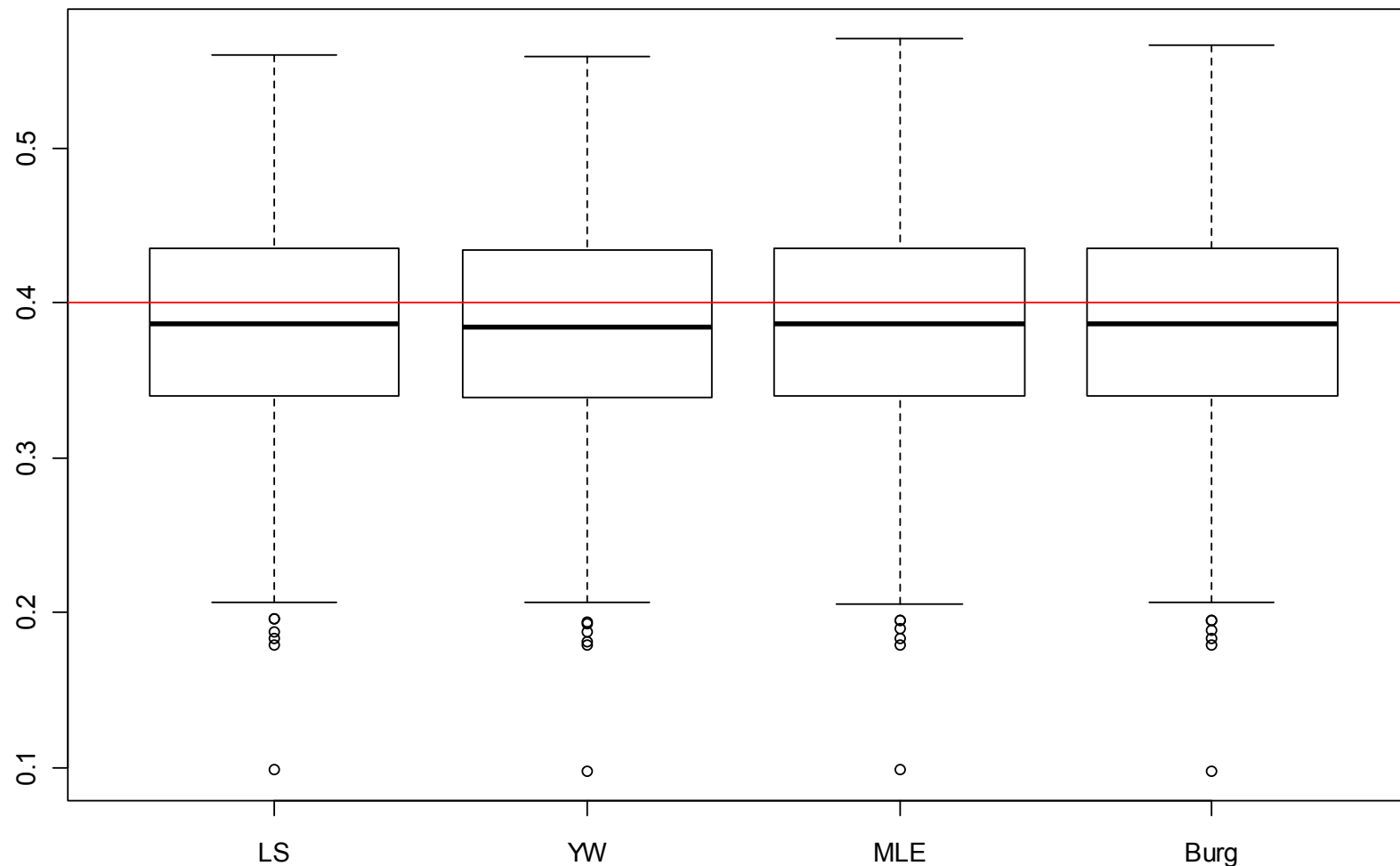
- All 4 estimation methods are asymptotically equivalent.
 - Even on finite samples, the differences are usually small.
 - Under Gaussian distribution, OLS and MLE coincide.
 - OLS/YW: explicit solution; Burg/MLE: numerical solution.
 - Functions `ar.xx()` provide easy AIC estimation of p .
 - Function `arima()` provides standard errors for all parameters.
- > Either work with `ar.burg()` or with `arima()`, depending on whether you want AIC or standard errors. Watch out for warnings if the numerical solution do not converge.

Applied Time Series Analysis

SS 2015 – Autoregressive Models

Comparison: Alpha Estimation vs. Method

Comparison of Methods: $n=200$, $\alpha=0.4$

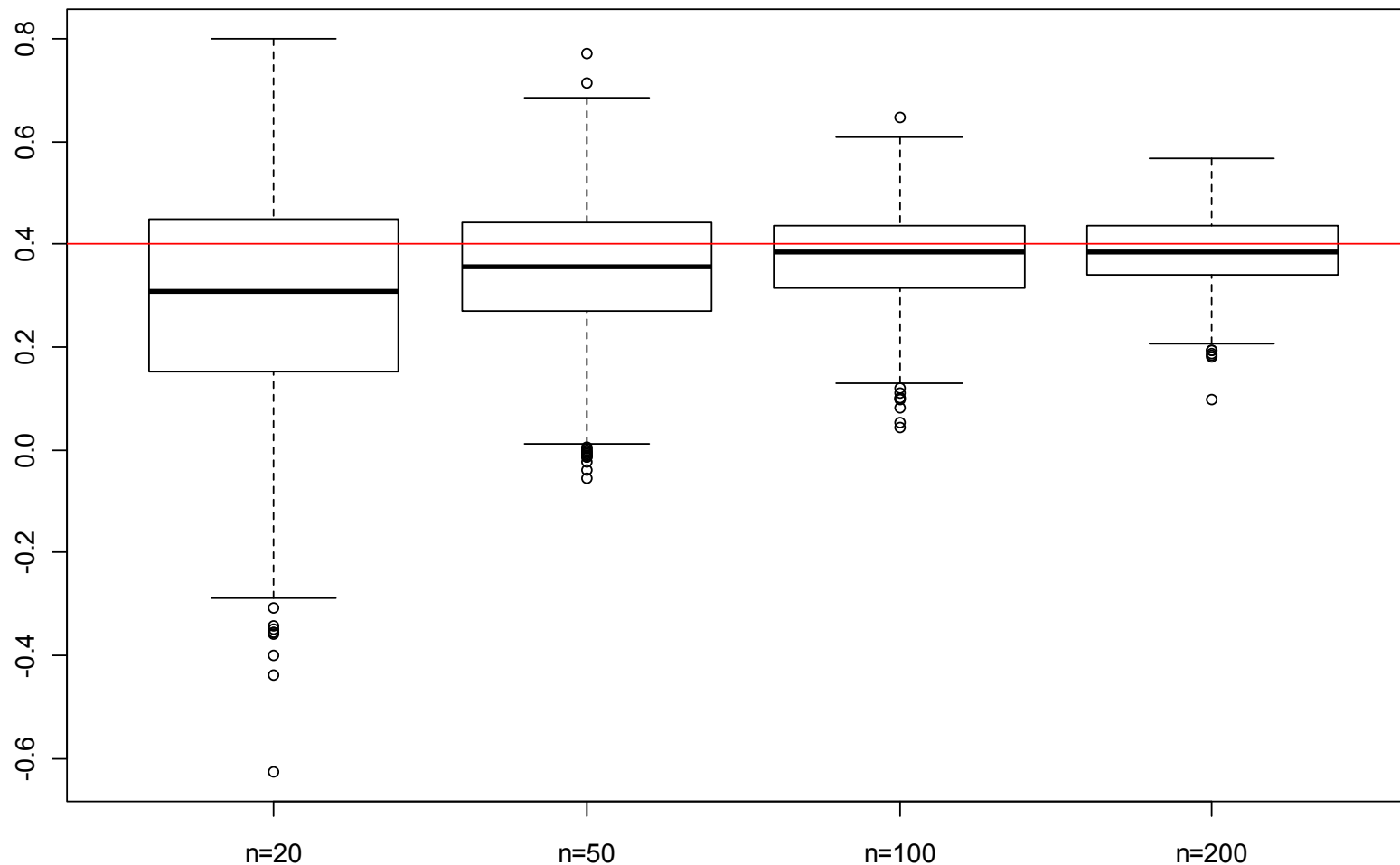


Applied Time Series Analysis

SS 2015 – Autoregressive Models

Comparison: Alpha Estimation vs. n

Comparison for Series Length n: alpha=0.4, method=Burg

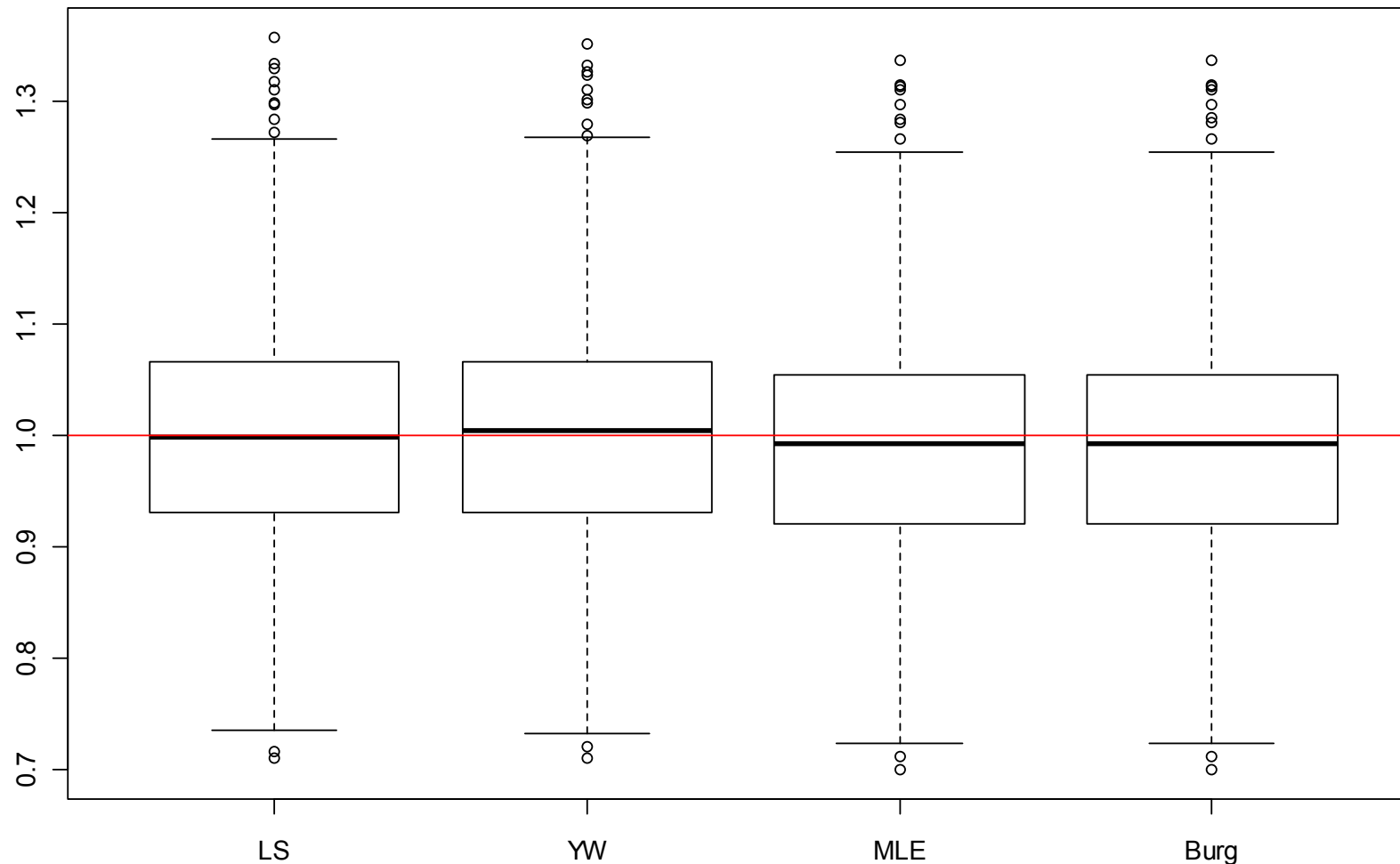


Applied Time Series Analysis

SS 2015 – Autoregressive Models

Comparison: Sigma Estimation vs. Method

Comparison of Methods: $n=200$, $\sigma=1$

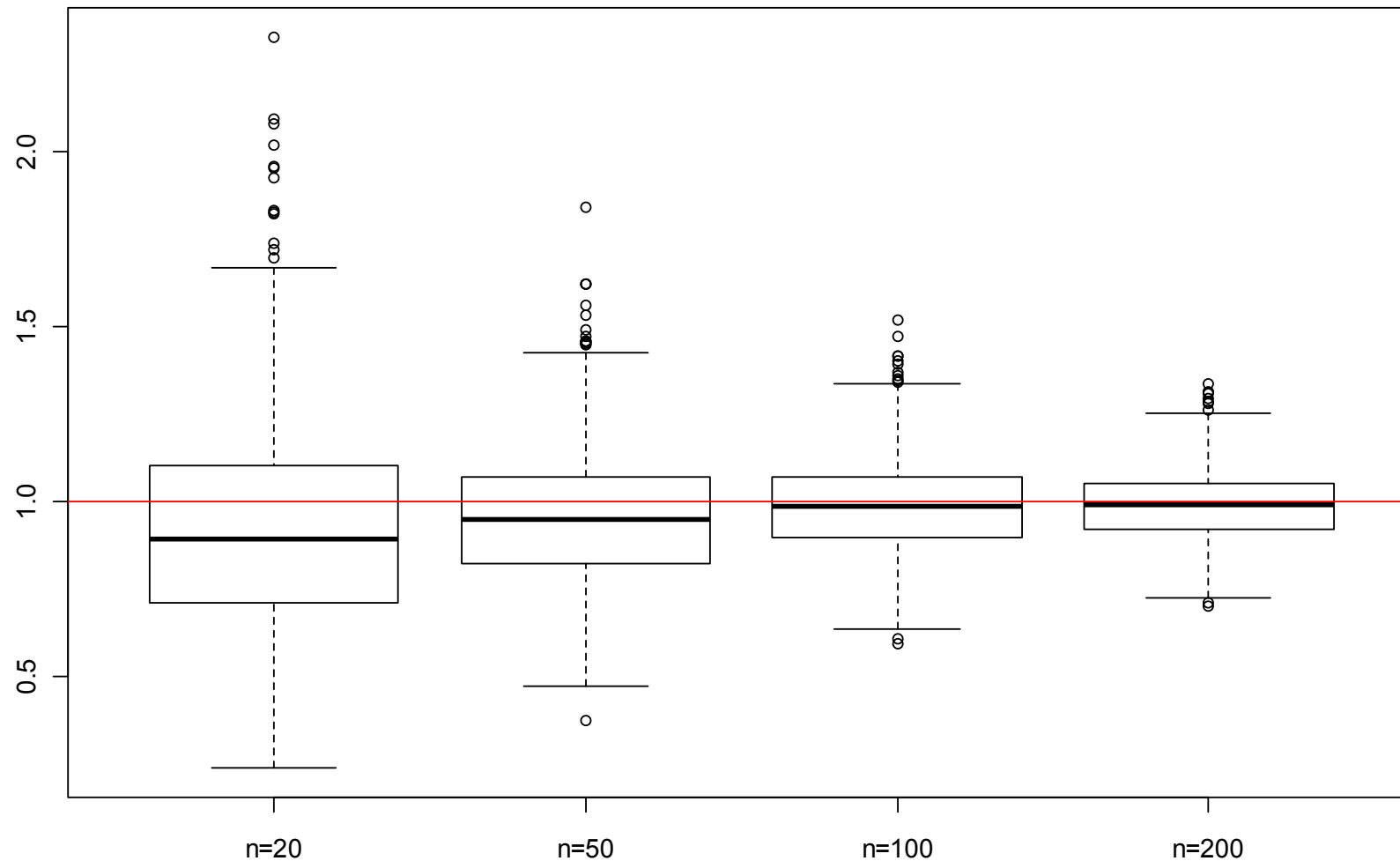


Applied Time Series Analysis

SS 2015 – Autoregressive Models

Comparison: Sigma Estimation vs. n

Comparison for Series Length n: sigma=1, method=Burg



Applied Time Series Analysis

SS 2015 – Autoregressive Models

Model Diagnostics

What we do here is Residual Analysis:

$$\begin{aligned}\text{„residuals“} &= \text{„estimated innovations“} \\ &= \hat{E}_t \\ &= (x_t - \hat{m}) - \left(\hat{\alpha}_1 (x_{t-1} - \hat{m}) - \dots - \hat{\alpha}_p (x_{t-p} - \hat{m}) \right)\end{aligned}$$

Remember the assumptions we made:

$$E_t \text{ i.i.d, } E[E_t] = 0, \text{ Var}(E_t) = \sigma_E^2$$

and probably

$$E_t \sim N(0, \sigma_E^2)$$

Applied Time Series Analysis

SS 2015 – Autoregressive Models

Model Diagnostics

We check the assumptions we made with the following means:

a) Time series plot of \hat{E}_t

b) ACF/PACF plot of \hat{E}_t

c) QQ-plot of \hat{E}_t

→ **The innovation time series \hat{E}_t should look like white noise**

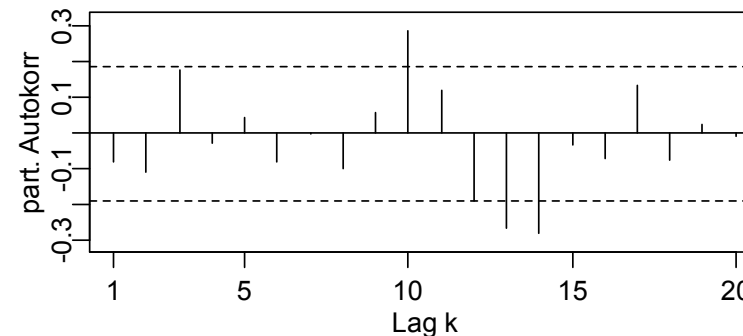
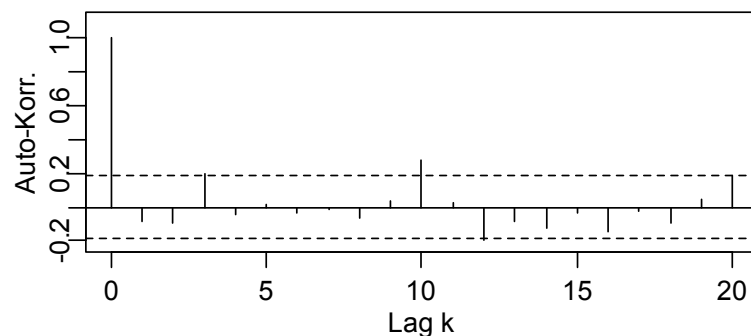
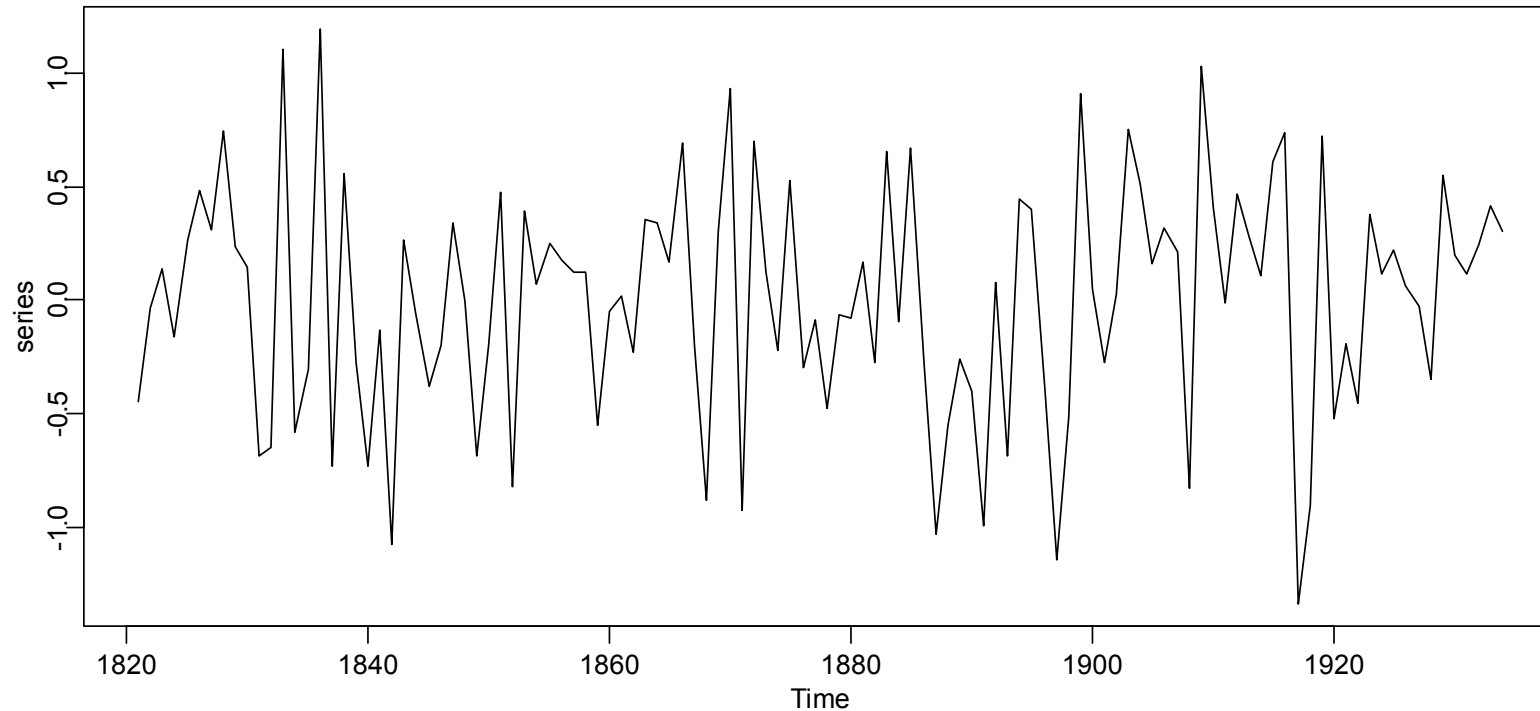
Lynx example:

```
fit <- arima(log(lynx), order=c(2,0,0))  
acf(resid(fit)); pacf(resid(fit))
```

Applied Time Series Analysis

SS 2015 – Autoregressive Models

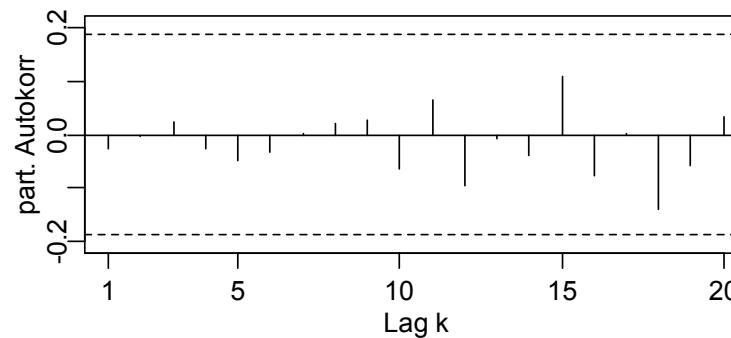
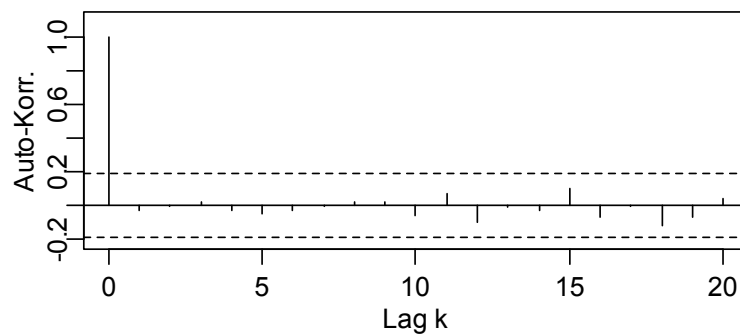
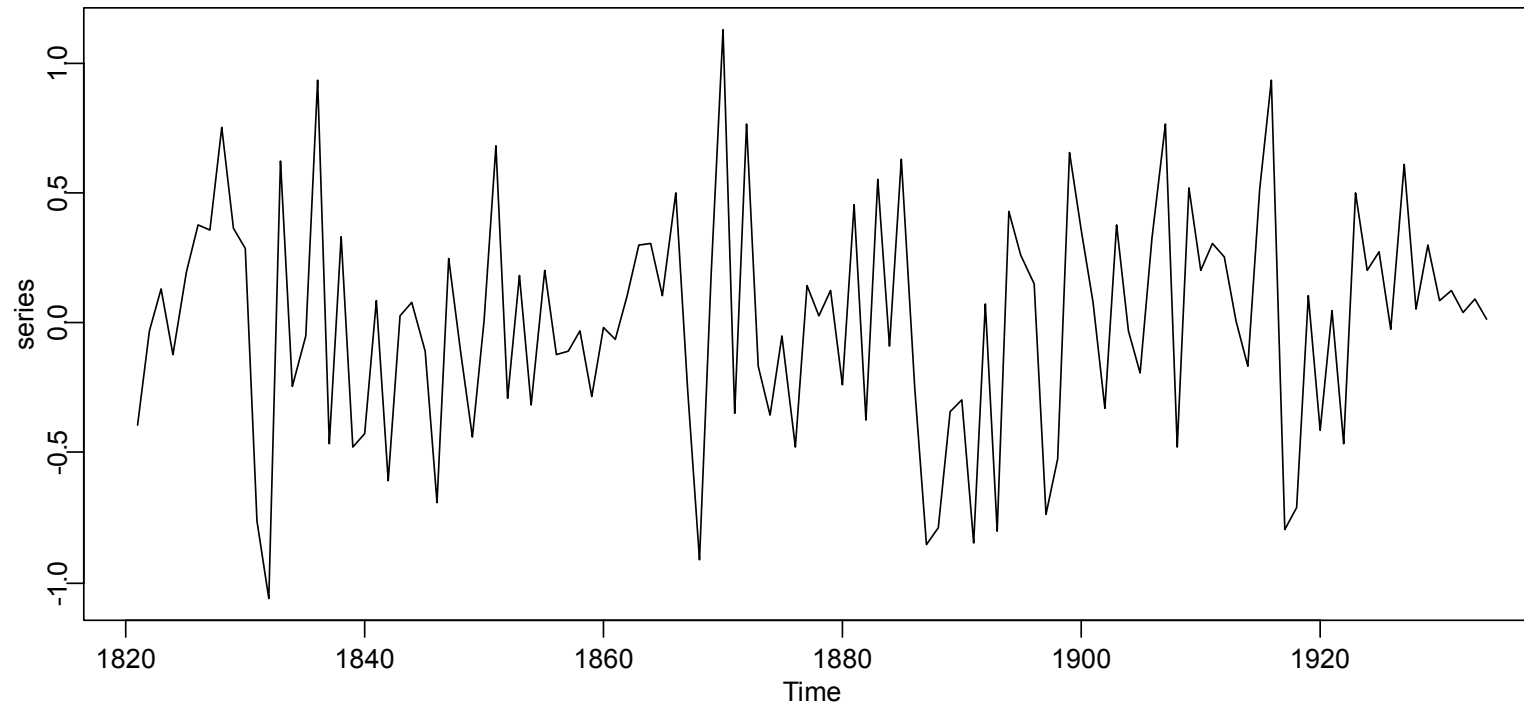
Model Diagnostics: $\log(\text{lynx})$ data, AR(2)



Applied Time Series Analysis

SS 2015 – Autoregressive Models

Model Diagnostics: $\log(\text{lynx})$ data, AR(11)

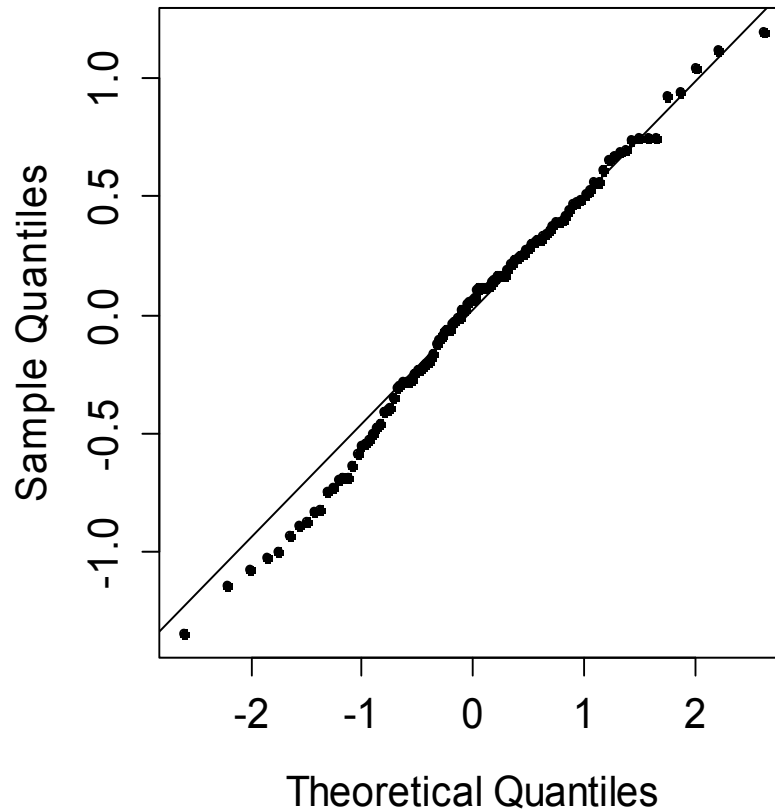


Applied Time Series Analysis

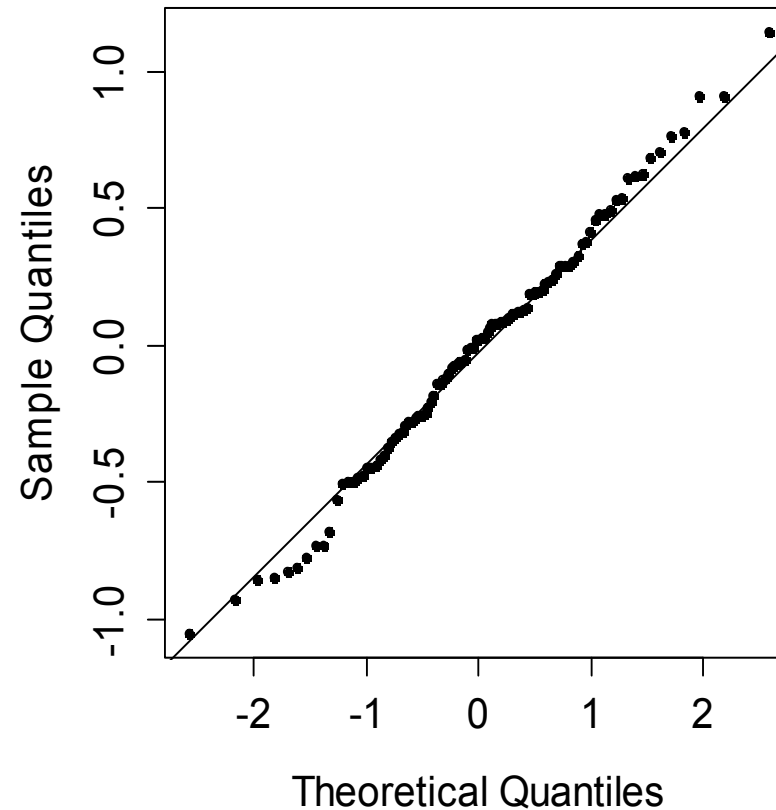
SS 2015 – Autoregressive Models

Model Diagnostics: Normal Plots

Normal QQ-Plot: AR(2)



Normal QQ-Plot: AR(11)



Applied Time Series Analysis

SS 2015 – Autoregressive Models

AIC/BIC

If several alternative models show satisfactory residuals, using the information criteria AIC and/or BIC can help to choose the most suitable one:

$$\text{AIC} = -2\log(L) + 2p$$

$$\text{BIC} = -2\log(L) + \log(n)p$$

where

$L(\alpha, \mu, \sigma^2) = f(x, \alpha, \mu, \sigma^2)$ = „Likelihood Function“

p is the number of parameters and equals p or $p+1$

n is the time series length

Goal: Minimization of AIC and/or BIC

Applied Time Series Analysis

SS 2015 – Autoregressive Models

AIC/BIC

We need (again) a distribution assumption in order to compute the AIC and/or BIC criteria. Mostly, one relies again on i.i.d. normally distributed innovations. Then, the criteria simplify to:

$$\text{AIC} = n \log(\hat{\sigma}_E^2) + 2p$$

$$\text{BIC} = n \log(\hat{\sigma}_E^2) + \log(n)p$$

Remarks:

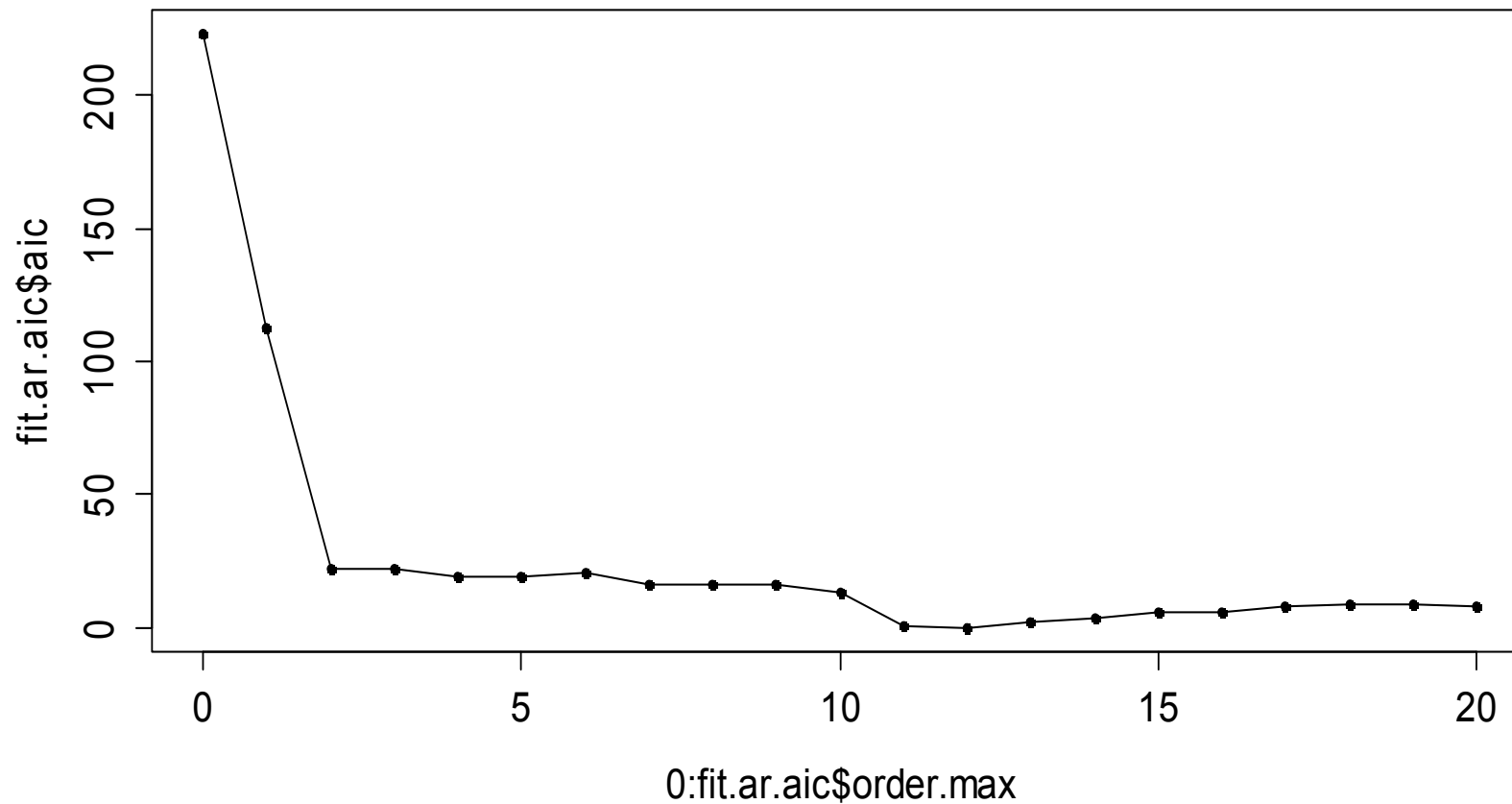
- AIC tends to over-, BIC to underestimate the true p
- Plotting AIC/BIC values against p can give further insight. One then usually chooses the model where the last significant decrease of AIC/BIC was observed

Applied Time Series Analysis

SS 2015 – Autoregressive Models

AIC/BIC

AIC-Values for AR(p)-Models on the Logged Lynx Data



Applied Time Series Analysis

SS 2015 – Autoregressive Models

Diagnostics by Simulation

As a last check before a model is called appropriate, simulating from the estimated coefficients and visually inspecting the resulting series (without any prejudices) to the original can be done.

→ **The simulated series should „look like“ the original. If this is not the case, the model failed to capture (some of) the properties of the original data.**

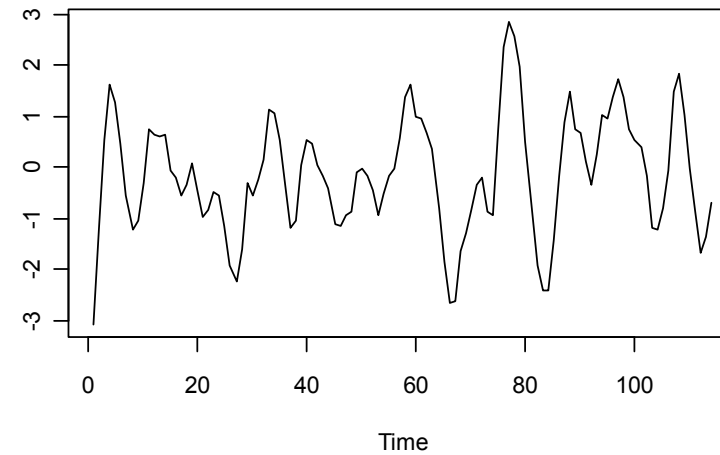
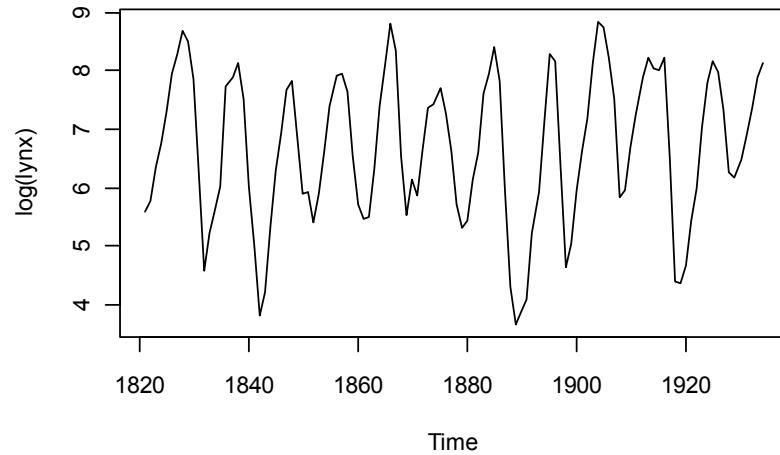
Applied Time Series Analysis

SS 2015 – Autoregressive Models

Diagnostics by Simulation, AR(2)

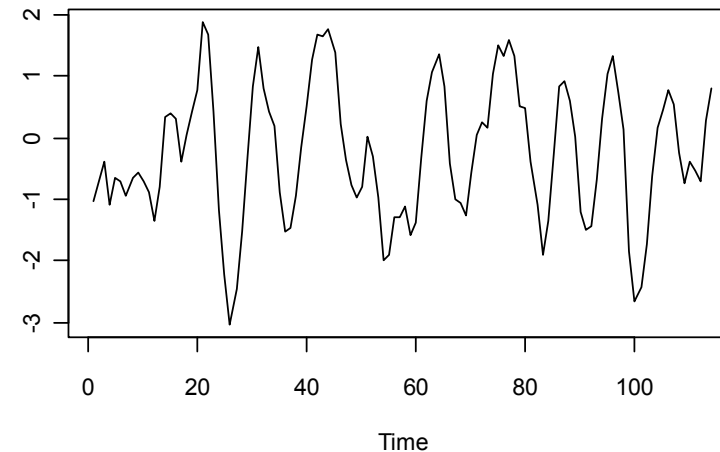
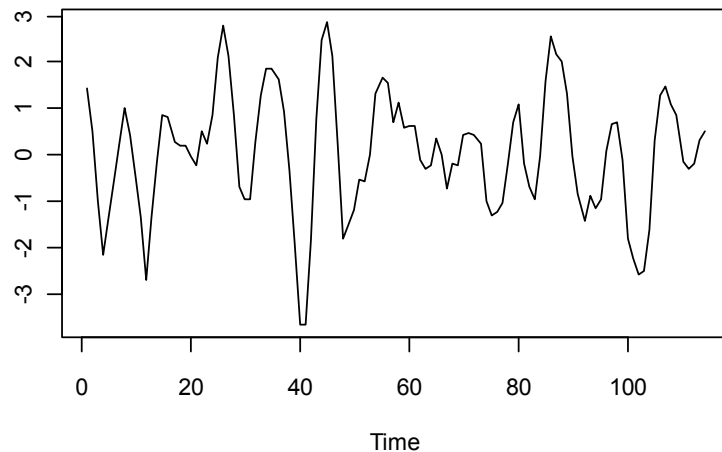
log(lynx)

Simulation 1



Simulation 2

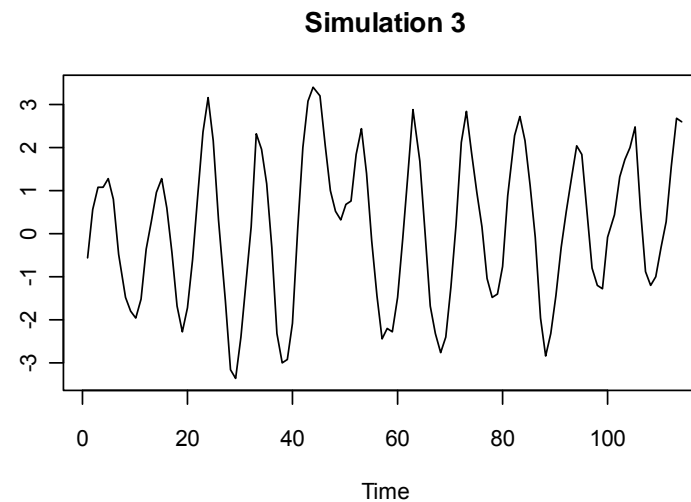
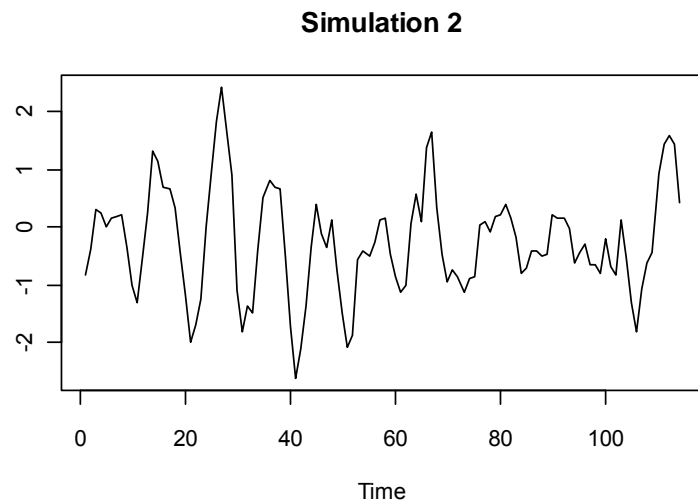
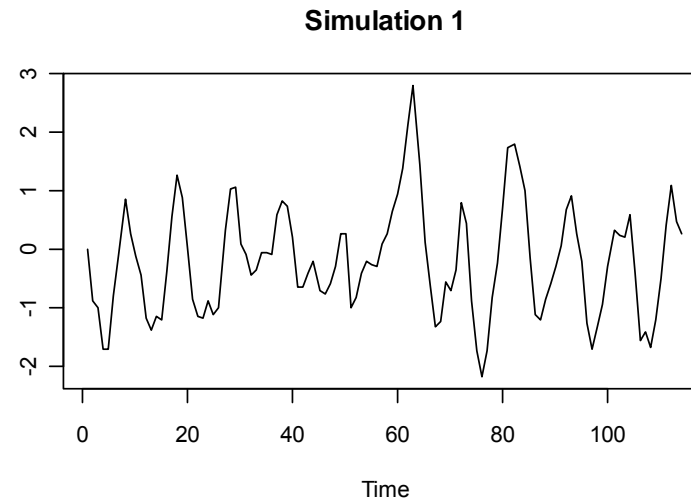
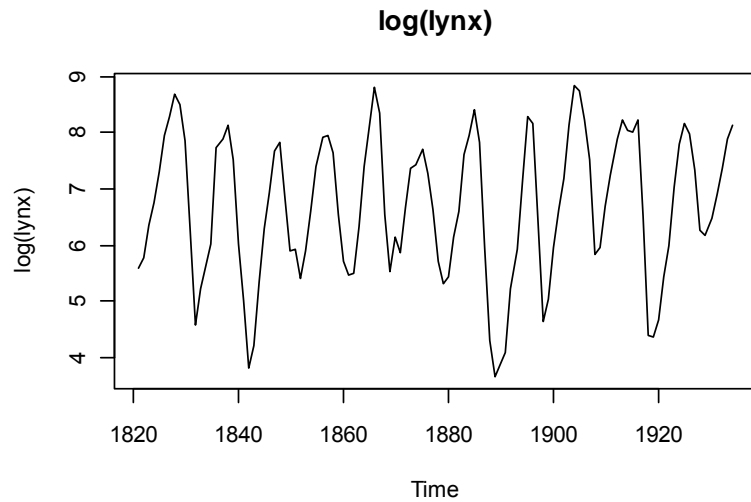
Simulation 3



Applied Time Series Analysis

SS 2015 – Autoregressive Models

Diagnostics by Simulation, AR(1)



Applied Time Series Analysis

SS 2015 – Moving Average Models

Looking Back & Outlook

We did consider shifted **AR(p)-models** $Y_t = m + X_t$ with:

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + E_t$$

where the correlation structure was as follows:

ACF: „exponential decay“

PACF: = 0 for all lags $k > p$

Now, in practice we could well observe a time series whose autocorrelation differs from the above structure.

We will thus discuss **ARMA(p,q) models**, a class that is suitable for modeling a wider spectrum of dependency structures.

Applied Time Series Analysis

SS 2015 – Moving Average Models

Moving Average Models

Whereas for AR(p) models, the current observation of a time series is written as a linear combination of its own past, **MA(q) models** can be seen as an extension of the „pure“ process

$$X_t = E_t, \text{ where } E_t \text{ is a white noise process,}$$

in the sense that past innovation terms E_{t-1}, E_{t-2}, \dots are included, too. We call this a **moving average** model:

$$X_t = E_t + \beta_1 E_{t-1} + \beta_2 E_{t-2} + \dots + \beta_q E_{t-q}$$

This is a time series process that is stationary, but not iid. In many respects, MA(q) models are complementary to AR(p).

Applied Time Series Analysis

SS 2015 – Moving Average Models

Notation for MA(q)-models

The backshift operator, and the characteristic polynomial, allow for convenient notation:

MA(q):
$$X_t = E_t + \beta_1 E_{t-1} + \beta_2 E_{t-2} + \dots + \beta_q E_{t-q}$$

MA(q) with BS:
$$X_t = \left(1 + \beta_1 B + \beta_2 B^2 + \dots + \beta_q B^q\right) E_t$$

MA(q) with BS+CP:
$$X_t = \Theta(B) E_t$$

where

$$\Theta(z) = 1 + \beta_1 z + \beta_2 z^2 + \dots + \beta_q z^q$$

is the characteristic polynomial

Applied Time Series Analysis

SS 2015 – Moving Average Models

Stationarity of MA(1)-Models

We first restrict ourselves to the simple MA(1)-model

$$X_t = E_t + \beta_1 E_{t-1}, \text{ where } E_t \text{ is a White Noise innovation}$$

The series X_t is weakly stationary, no matter what the choice of the parameter β_1 is.

Remember that for proving this, we have to show that:

- the expected value is 0
- the variance is constant and finite
- the autocovariance only depends on the lag k

→ **see the blackboard for the proof**

Applied Time Series Analysis

SS 2015 – Moving Average Models

ACF of the MA(1)-Process

We can deduct the ACF for the MA(1)-process:

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\beta_1}{(1 + \beta_1^2)} < 0.5$$

and

$$\rho(k) = 0 \quad \text{for all } k > 1.$$

Thus, we have a “cut-off” situation, i.e. a similar behavior to the one of the PACF in an AR(1) process. This is why and how AR(1) and MA(1) are complementary.

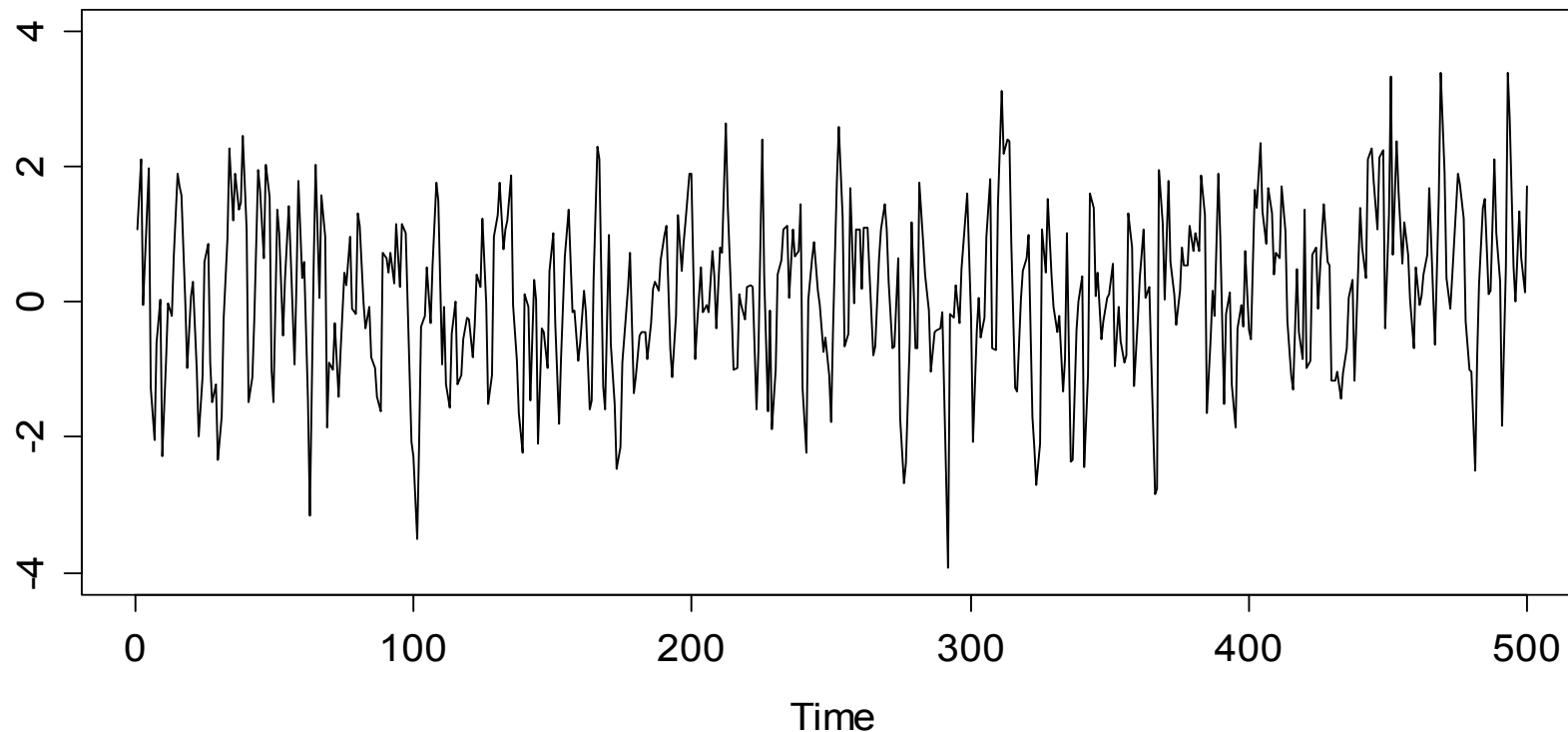
Applied Time Series Analysis

SS 2015 – Moving Average Models

Simulated Process with $\beta_1=0.7$

```
> ts.ma1 <- arima.sim(list(ma=0.7), n=500)
> plot(ts.ma1, ylab="", ylim=c(-4,4))
```

Simulation from a MA(1) Process



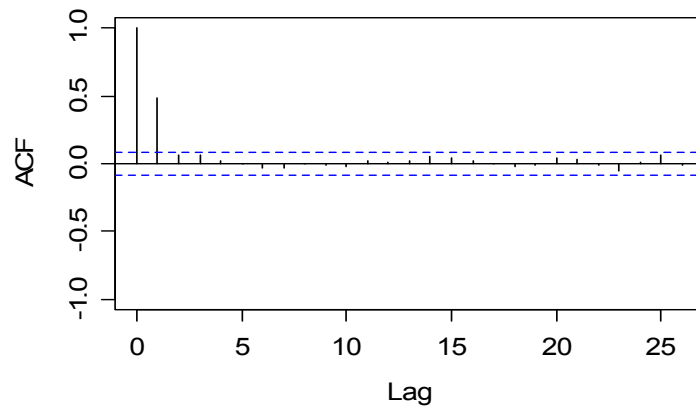
Applied Time Series Analysis

SS 2015 – Moving Average Models

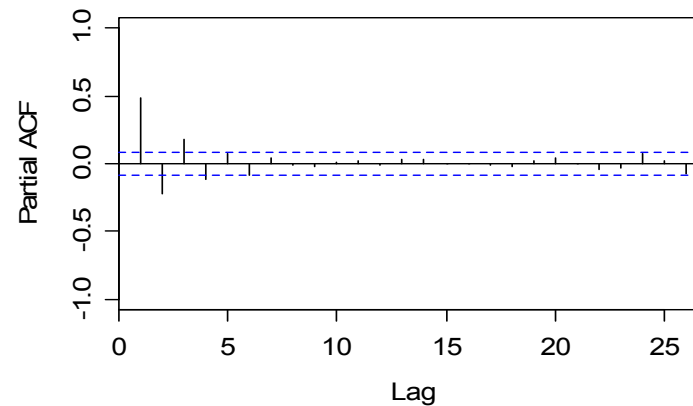
ACF and PACF of MA(1)

```
> acf.true <- ARMAacf(ma=0.7, lag.max=20)  
> pacf.true <- ARMAacf(ma=0.7, pacf=T, lag.m=20)
```

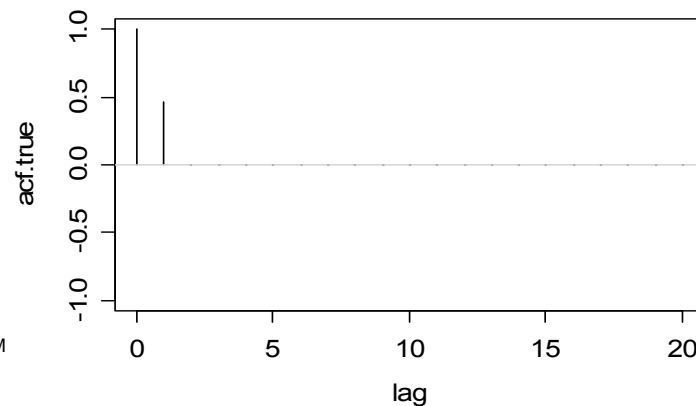
Estimated ACF



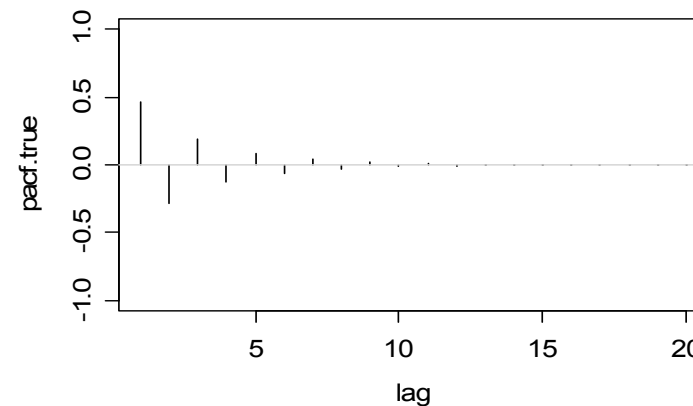
Estimated PACF



True ACF



True PACF



Applied Time Series Analysis

SS 2015 – Moving Average Models

MA(1): Remarks

Without additional assumptions, the ACF of an MA(1) doesn't allow identification of the generating model.

In particular, the two processes

$$X_t = E_t + 0.5 \cdot E_{t-1}$$

$$U_t = E_t + 2 \cdot E_{t-1}$$

have identical ACF:

$$\rho(1) = \frac{\beta_1}{1 + \beta_1^2} = \frac{1 / \beta_1}{1 + (1 / \beta_1^2)}$$

Applied Time Series Analysis

SS 2015 – Moving Average Models

MA(1): Invertibility

- An MA(1)-, or in general an MA(q)-process is said to be invertible if the roots of the characteristic polynomial $\Theta(B)$ lie outside of the unit circle.
- Under this condition, there exists only one MA(q)-process for any given ACF. But please note that any MA(q) is stationary, no matter if it is invertible or not.
- The condition on the characteristic polynomial translates to restrictions on the coefficients. For any MA(1)-model, $|\beta_1| < 1$ is required.
- R function `polyroot()` can be used for finding the roots.

Applied Time Series Analysis

SS 2015 – Moving Average Models

Practical Importance of Invertibility

The condition of invertibility is not only a technical issue, but has important practical meaning. Invertible MA(1)-processes can be written as an AR(∞):

$$\begin{aligned} X_t &= E_t + \beta_1 E_{t-1} \\ &= E_t + \beta_1 (X_{t-1} - \beta_1 E_{t-2}) \\ &= \dots \\ &= E_t + \beta_1 X_{t-1} - \beta_1^2 X_{t-2} + \beta_1^3 X_{t-3} + \dots \\ &= E_t + \sum_{i=1}^{\infty} \psi_i X_{t-i} \end{aligned}$$

Invertibility is practically relevant for model fitting!

Applied Time Series Analysis

SS 2015 – Moving Average Models

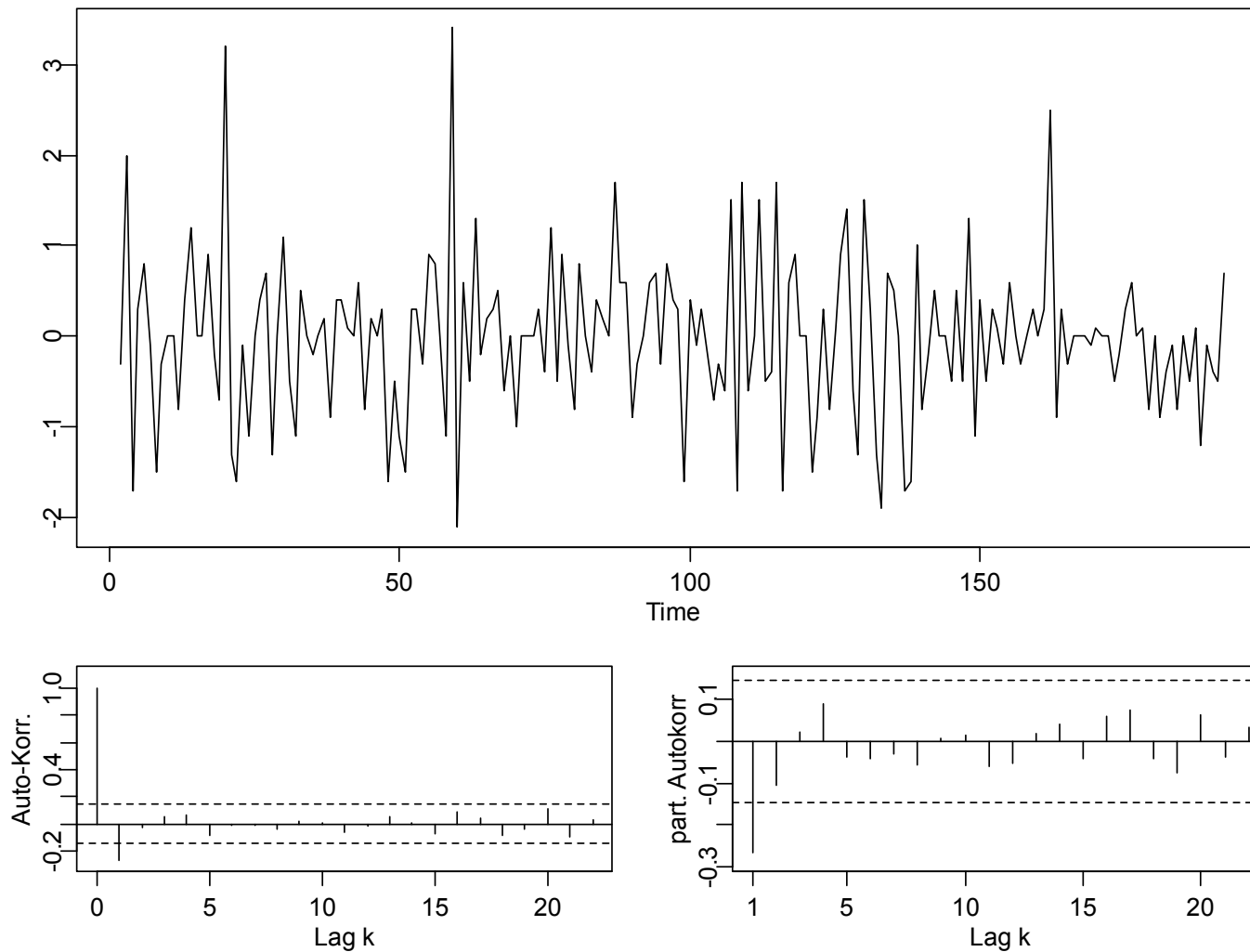
MA(1): Example

- daily return of an AT&T bond from 04/1975 to 12/1975
- the time series has 192 observations
- we are looking at the first-order differences
- an MA(1) model seems to fit the data (→ next slide)
- since we are looking at a differenced series, this is in fact an ARIMA(0,1,1) model (→ will be discussed later...)

Applied Time Series Analysis

SS 2015 – Moving Average Models

MA(1): Example



Applied Time Series Analysis

SS 2015 – Moving Average Models

MA(q)-Models

The MA(q)-model is defined as follows:

$$X_t = E_t + \beta_1 E_{t-1} + \beta_2 E_{t-2} + \dots + \beta_q E_{t-q} ,$$

where E_t are i.i.d. innovations (=a white noise process).

The ACF of this process can be computed from the coefficients:

$$\rho(k) = \frac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^q \beta_i^2} , \quad \text{for all } k=1, \dots, q \text{ with } \beta_0 = 1$$

$$\rho(k) = 0 , \quad \text{for all } k > q$$

Applied Time Series Analysis

SS 2015 – Moving Average Models

ACF/PACF of MA(q)

ACF

- the ACF of an MA(q) has a cut-off at lag $k=q$
- it behaves thus like the PACF of an AR(q)-model

PACF

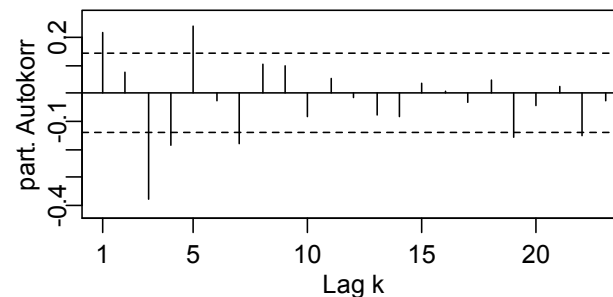
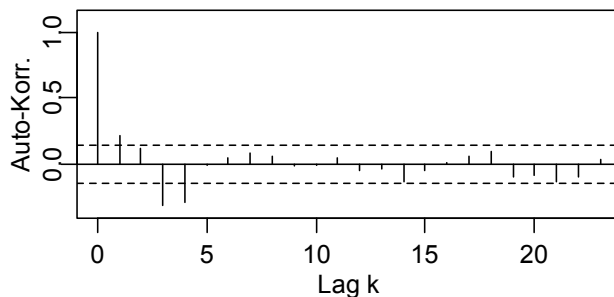
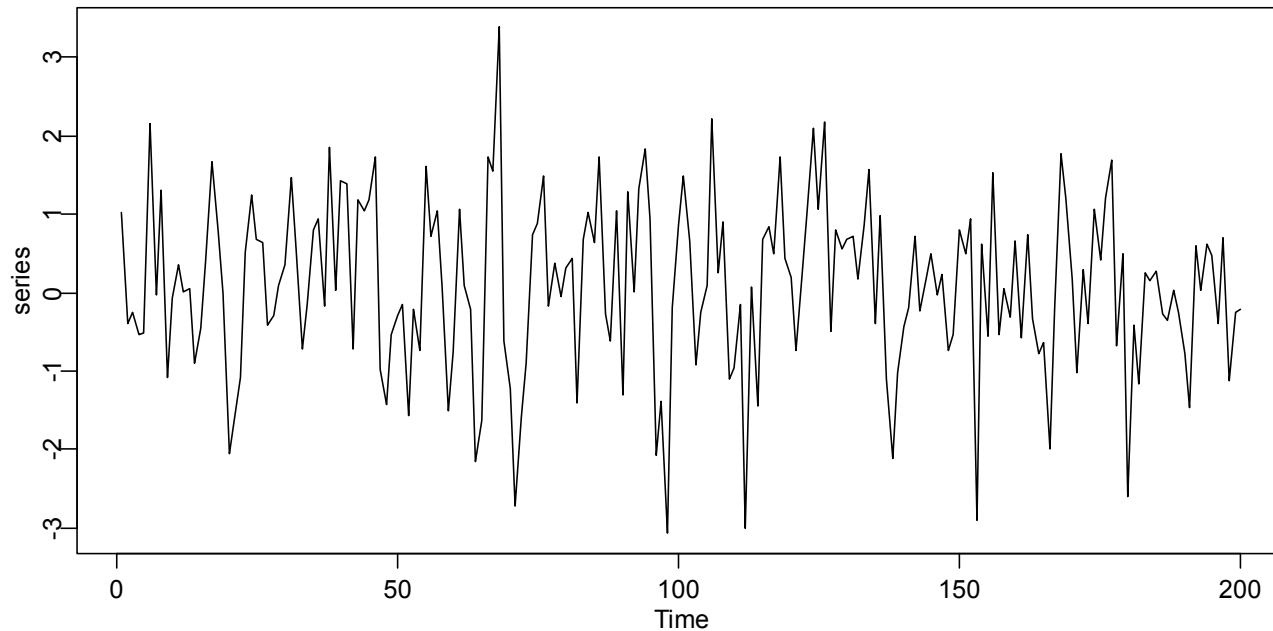
- the PACF is (again) complicated to determine, but:
- the PACF of an MA(q) has an „exponential decay“
- it behaves thus like the ACF of an AR-model

Applied Time Series Analysis

SS 2015 – Moving Average Models

MA(4): Example

$$X_t = E_t + 0.3 \cdot E_{t-1} + 0.3 \cdot E_{t-2} - 0.2 \cdot E_{t-3} - 0.2 \cdot E_{t-4}, \quad E_t \sim N(0,1)$$



Applied Time Series Analysis

SS 2015 – ARMA(p,q)

ARMA(p,q)-Models

An ARMA(p,q)-model combines AR(p) and MA(q):

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + E_t + \beta_1 E_{t-1} + \dots + \beta_q E_{t-q}$$

where E_t are i.i.d. innovations (=a white noise process).

It's easier to write an ARMA(p,q) with the characteristic polynomial:

$$\Phi(B)X_t = \Theta(B)E_t, \text{ where}$$

$$\Phi(z) = 1 - \alpha_1 z - \dots - \alpha_p z^p \quad \text{is the cP of the AR-part, and}$$

$$\Theta(z) = 1 + \beta_1 z + \dots + \beta_q z^q \quad \text{is the cP of the MA-part}$$

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Stationarity/Invertibility of ARMA(p,q)

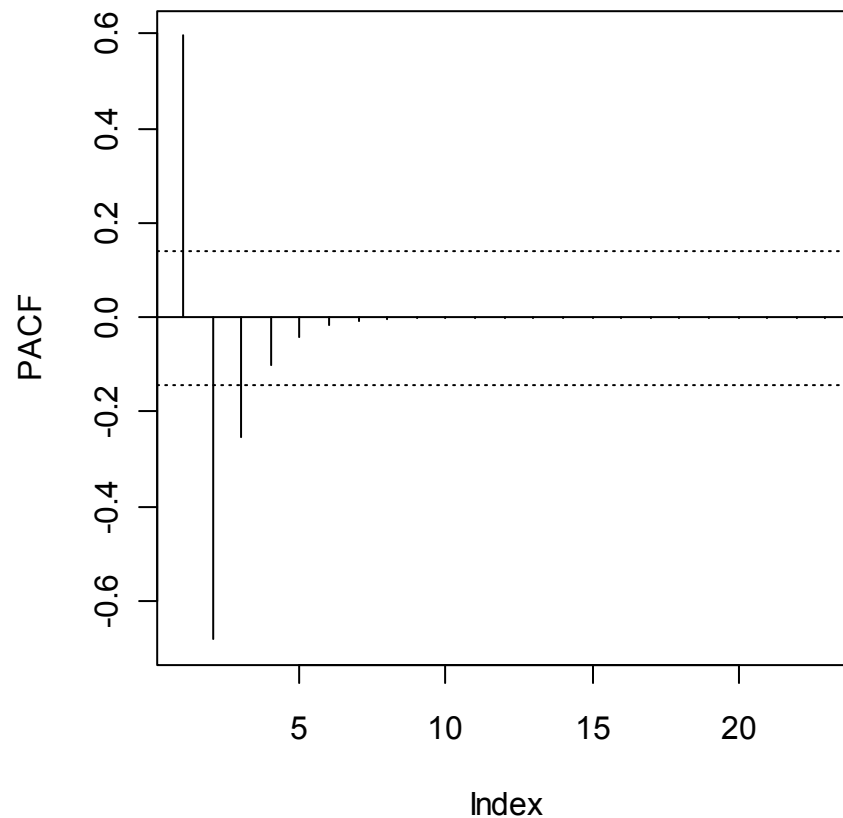
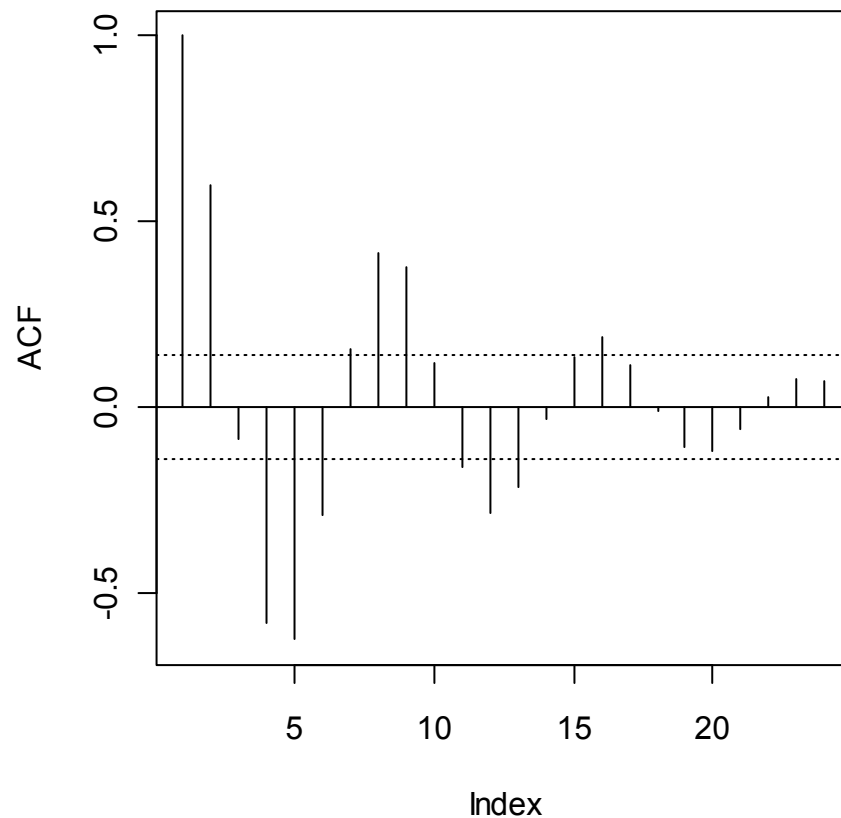
- both properties are determined by the cP
- the AR-cP determines stationarity
- the MA-cP determines invertibility
- condition: roots of the cP outside of the unit circle
- stationarity: model can be written as a MA(∞)
- invertibility: model can be written as an AR(∞)

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

True ACF/PACF of an ARMA(2,1)

$$X_t = 1.2 \cdot X_{t-1} - 0.8 \cdot X_{t-2} + E_t + 0.4 \cdot E_{t-1}, \quad E_t \sim N(0,1)$$

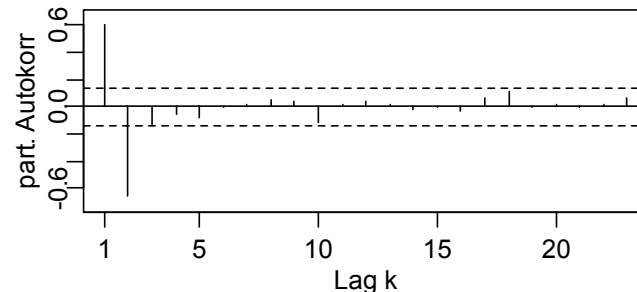
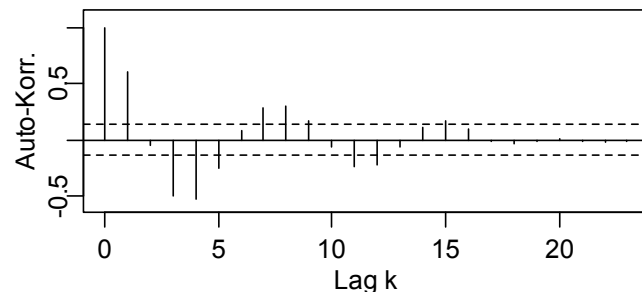
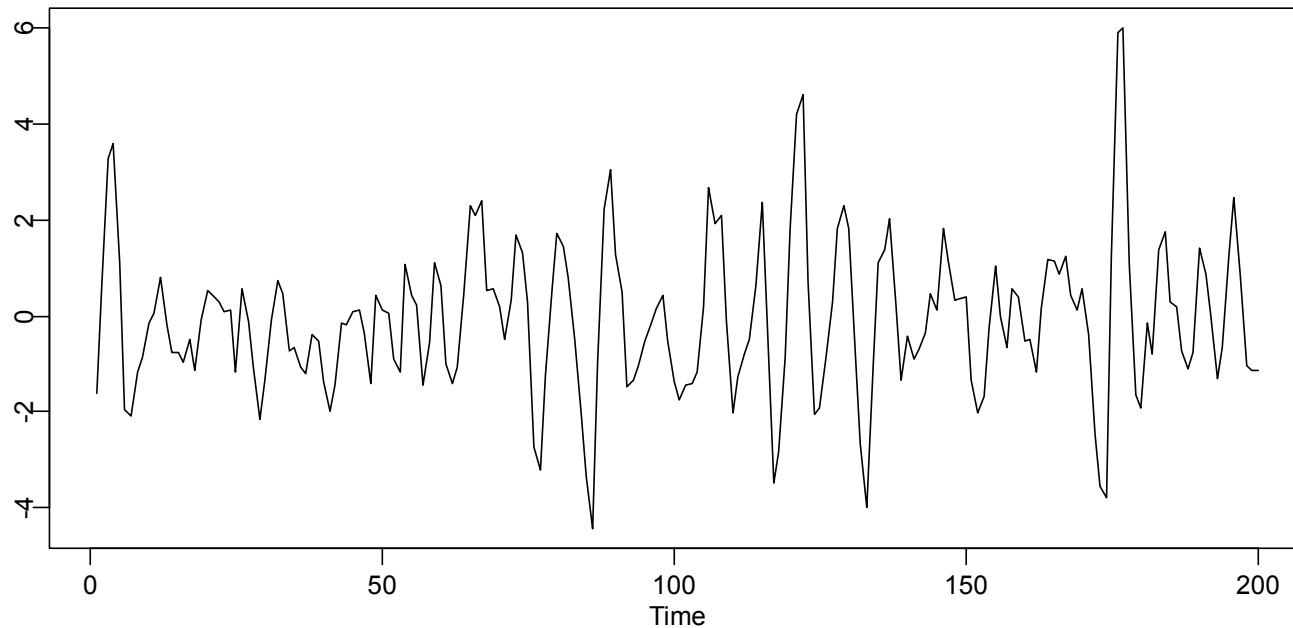


Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Simulated ACF/PACF of an ARMA(2,1)

$$X_t = 1.2 \cdot X_{t-1} - 0.8 \cdot X_{t-2} + E_t + 0.4 \cdot E_{t-1}, \quad E_t \sim N(0,1)$$



Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Properties of ACF/PACF in ARMA(p,q)

	ACF	PACF
AR(p)	exponential decay	cut-off at lag p
MA(q)	cut-off at lag q	exponential decay
ARMA(p,q)	mix decay/cut-off	mix decay/cut-off

→ all linear time series processes can be approximated by an ARMA(p,q) with possibly large p,q. They are thus are very rich class of models.

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Fitting ARMA(p,q)

What needs to be done?

- 1) **Achieve stationarity**
→ transformations, differencing, modeling, ...
- 2) **Choice of the order**
→ determining (p,q)
- 3) **Parameter estimation**
→ Estimation of α , β , m , σ_E^2
- 4) **Residual analysis**
→ if necessary, repeat 1), and/or 2)-4)

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Identification of the Order (p,q)

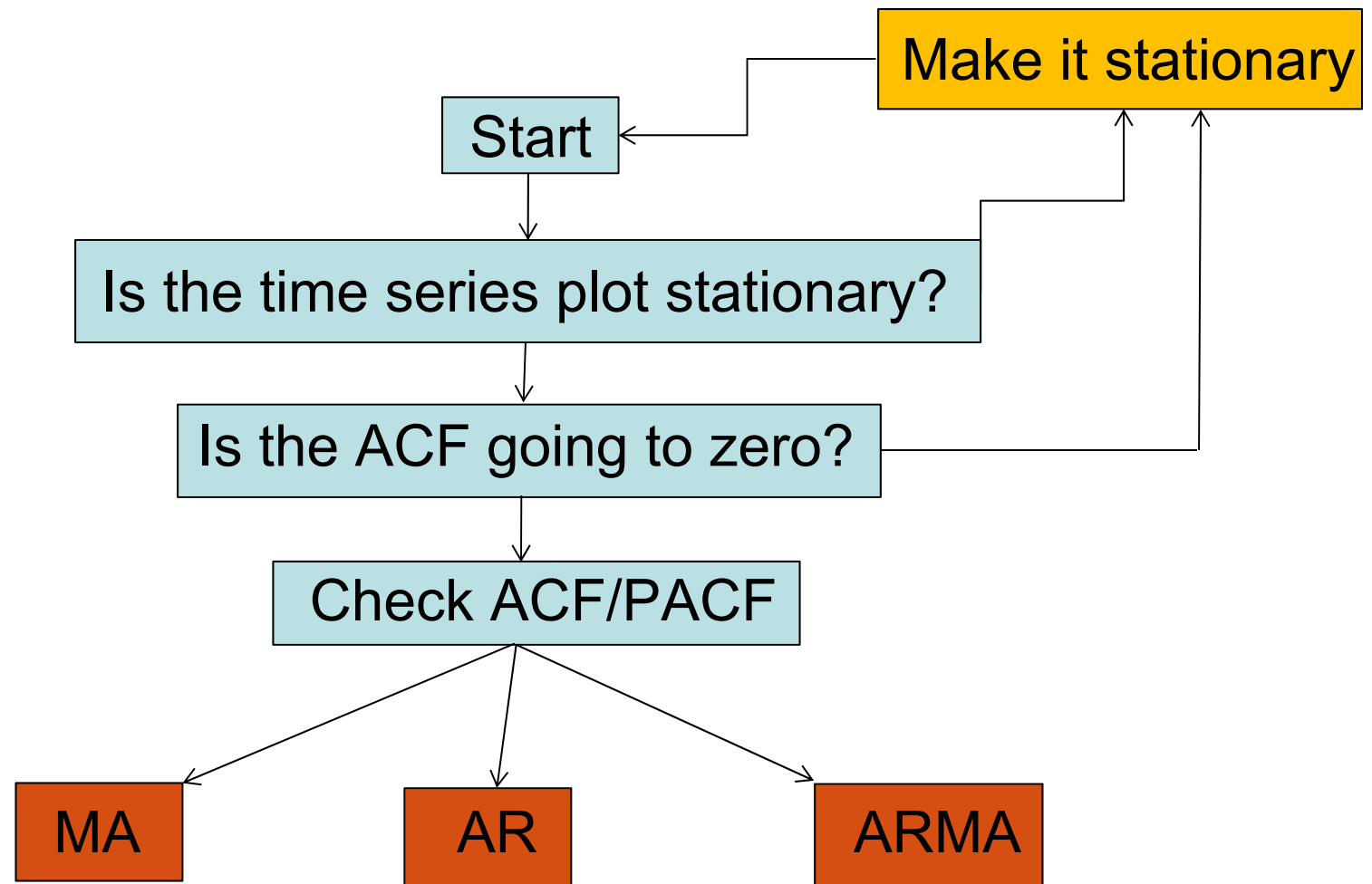
Please note:

- We only have one single realization of the time series with finite length.
 - The plots (etc.) we base the order choice on are not „facts“, but are estimations with uncertainty.
 - This holds especially for the ACF/PACF plots.
 - Every ARMA(p,q) can be written as $AR(\infty)$ or $MA(\infty)$
- **There is usually >1 model that describes the data well.**

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

ARMA(p,q)-Modeling



Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Parameter Estimation

For parameter estimation with AR(p) models, we had 4 choices:

- a) Regression
- b) Yule-Walker
- c) Maximum-Likelihood
- d) Burg's Algorithm

For ARMA(p,q) models, only two options are remaining, and both of them require numerical optimization:

- 1) Conditional Sum of Squares**
- 2) Maximum-Likelihood**

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Conditional Sum of Squares

Idea: This is an iterative approach where the parameters are determined such that the sum of squared errors (between observations and fitted values) are minimal.

$$S(\hat{\beta}_1, \dots, \hat{\beta}_q) = \sum_{t=1}^n \hat{E}_t^2 = \sum_{t=1}^n (X_t - (\hat{\beta}_1 \hat{E}_{t-1} - \dots - \hat{\beta}_q \hat{E}_{t-q}))^2$$

This requires starting values which are chosen as:

$$\hat{E}_0 = 0, \hat{E}_{-1} = 0, \dots, \hat{E}_{1-q} = 0$$

A numerical search is used to find the parameter values that minimize the entire conditional sum of squares. They also serve as starting values for MLE.

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Maximum-Likelihood-Estimation

Idea: Determine the parameters such that, given the observed time series x_1, \dots, x_n , the resulting model is the most plausible (i.e. the most likely) one.

→ **This requires the choice of a probability distribution for the time series $X = (X_1, \dots, X_n)$**

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Maximum-Likelihood-Estimation

If we assume the ARMA(p,q)-model

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + E_t + \beta_1 E_{t-1} + \dots + \beta_q E_{t-q}$$

and i.i.d. normally distributed innovations

$$E_t \sim N(0, \sigma_E^2)$$

the time series vector has a multivariate normal distribution

$$X = (X_1, \dots, X_n) \sim N(m \cdot \underline{1}, V)$$

with covariance matrix V that depends on the model parameters α , β and σ_E^2 .

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Maximum-Likelihood-Estimation

We then maximize the density of the multivariate normal distribution with respect to the parameters

$$\alpha, \beta, m \text{ and } \sigma_E^2.$$

The observed x-values are hereby regarded as fixed values.

→ **This is a highly complex non-linear optimization problem that requires sophisticated algorithms and starting values which are usually provided by CSS (at least that's the default in R's `arima()`).**

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Maximum-Likelihood-Estimation

```
> r.Pmle <- arima(d.Psqrt,order=c(2,0,0),include.mean=T)
> r.Pmle
```

```
Call: arima(x=d.Psqrt, order=c(2,0,0), include.mean=T)
```

Coefficients:

	ar1	ar2	intercept
	0.275	0.395	3.554
s.e.	0.107	0.109	0.267

```
sigma^2 = 0.6: log likelihood = -82.9, aic = 173.8
```


Applied Time Series Analysis

SS 2015 – ARMA(p,q)

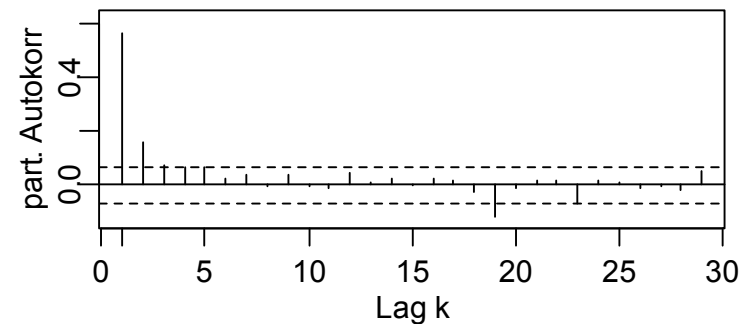
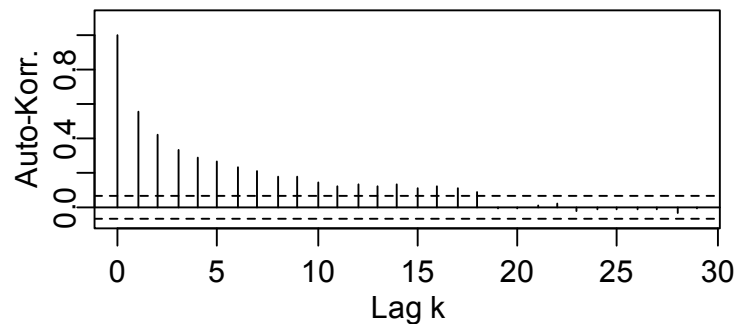
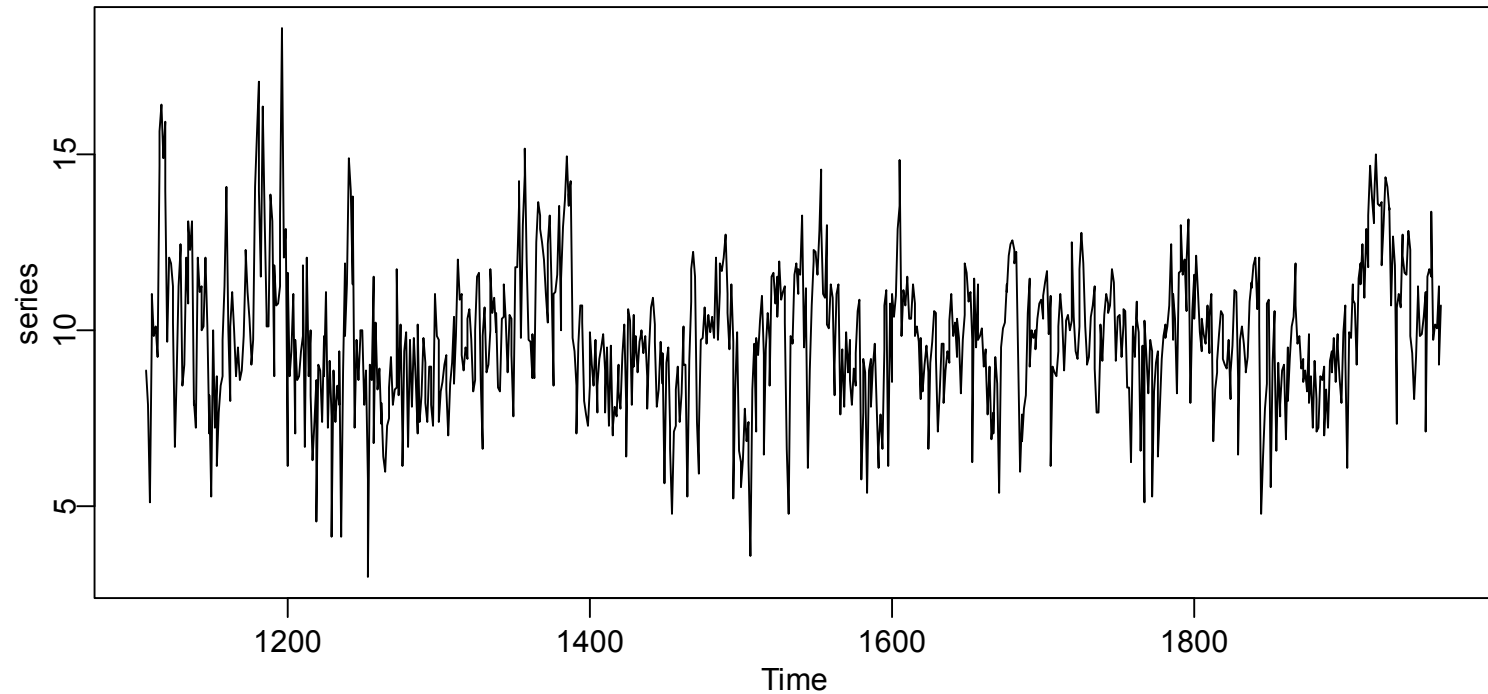
MLE: Remarks

- The MLE approach would work for any distribution. However, for innovation distributions other than Gaussian, the joint distribution might be „difficult“.
- For „reasonable“ deviations from the normality assumption, MLE still yields „good“ results.
- Besides the parameter estimates, we also obtain an estimate of their standard error
- Other software packages such as for example SAS don't rely on MLE, but use CSS, which is in spirit similar to Burg's algorithm.

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

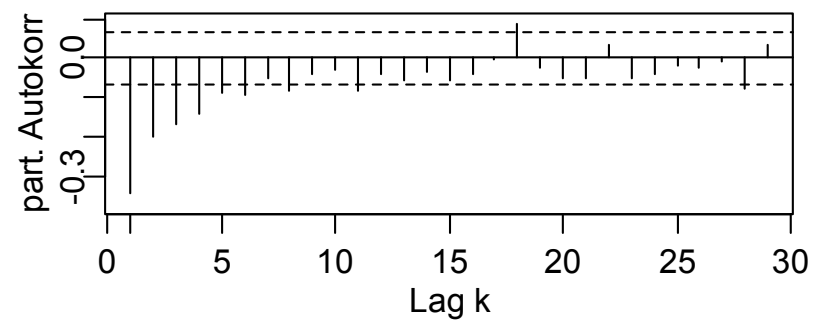
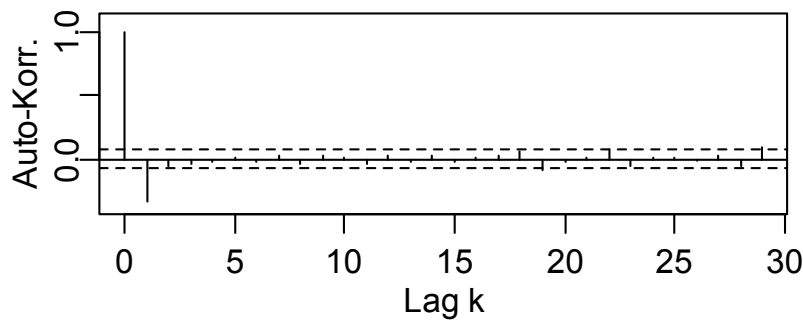
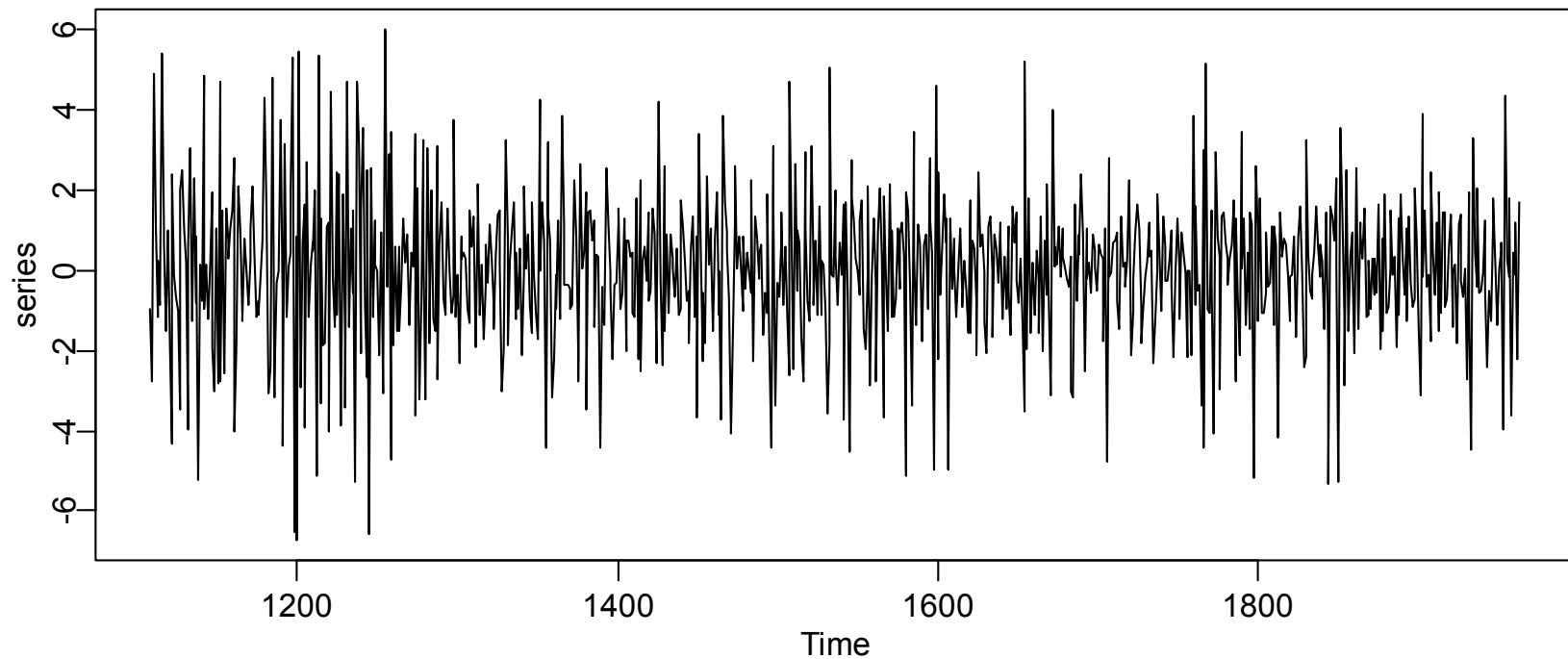
Douglas Fir: Original Data



Applied Time Series Analysis

SS 2015 – ARMA(p,q)

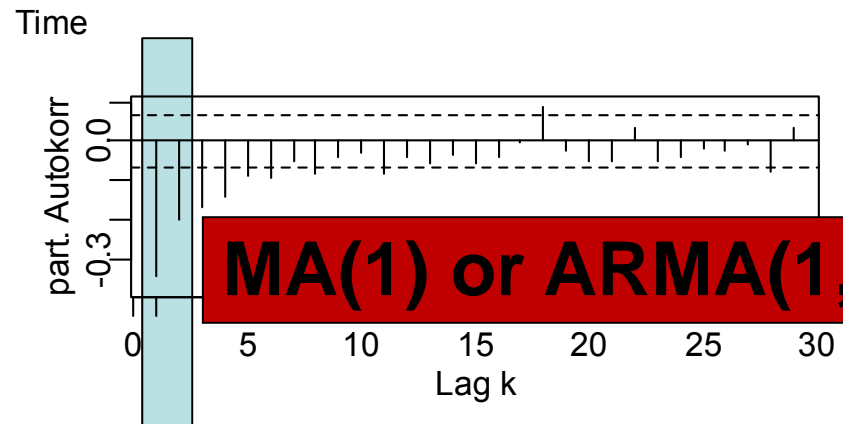
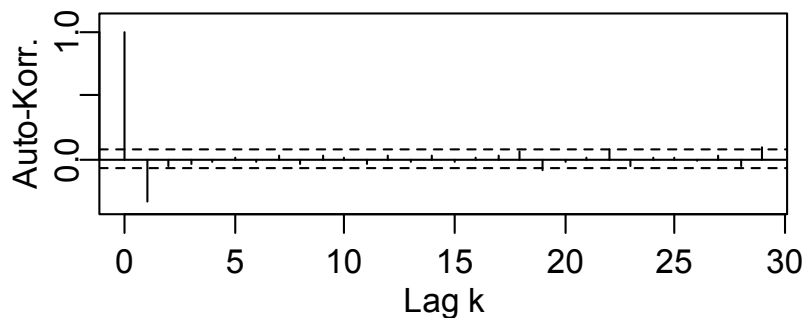
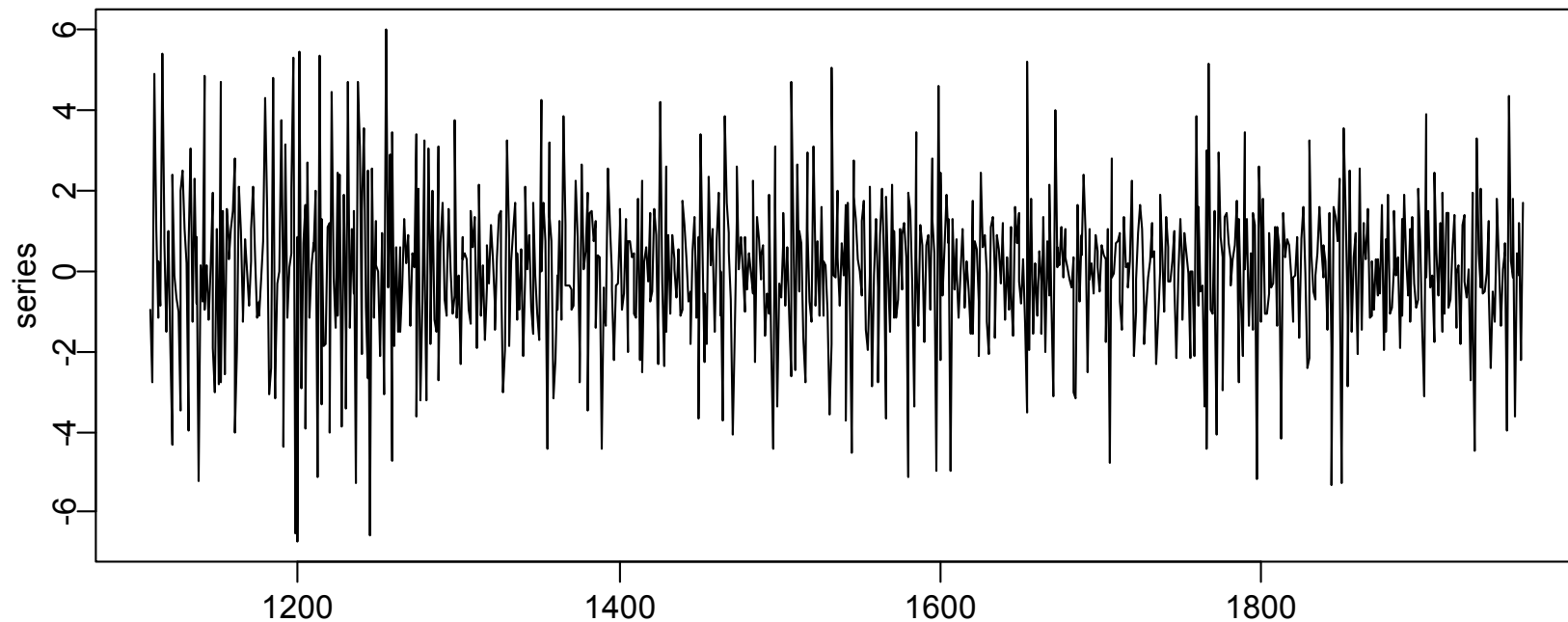
Douglas Fir: Differenced Series



Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Douglas Fir: Differenced Series

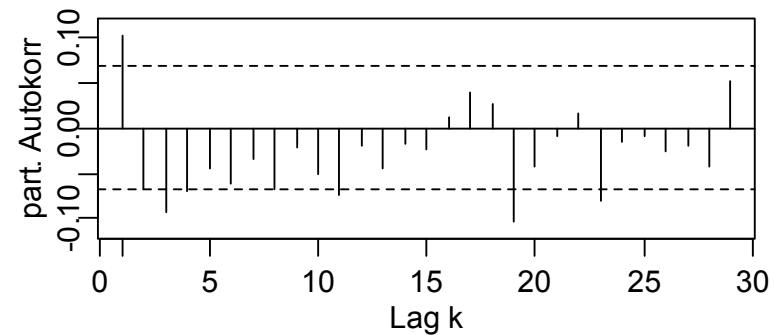
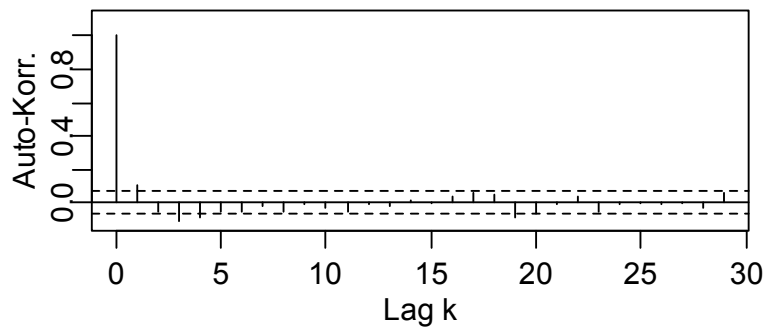
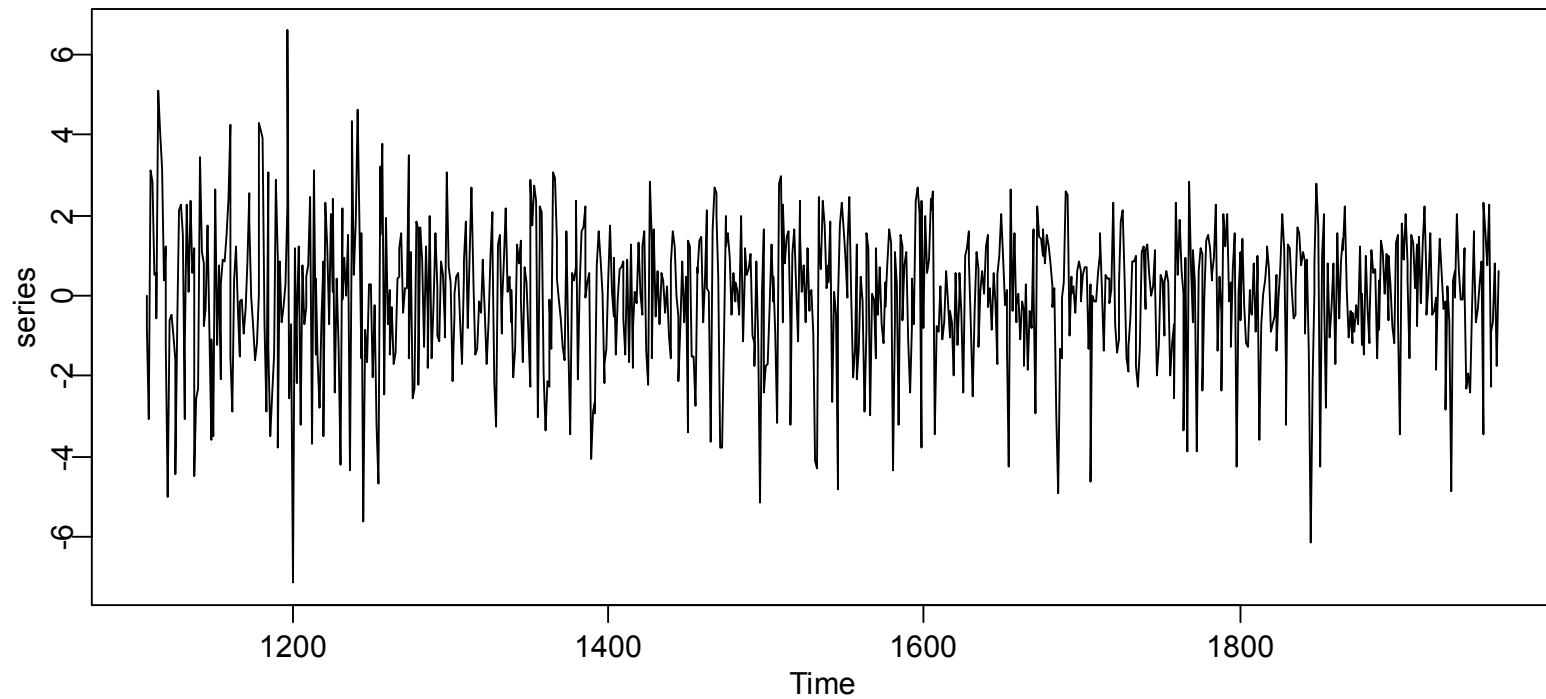


MA(1) or ARMA(1,1)???

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

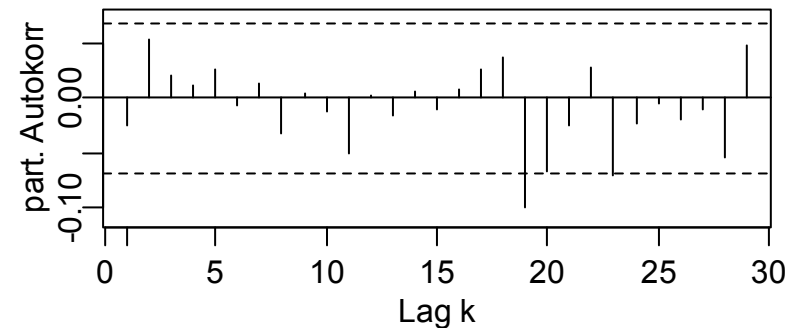
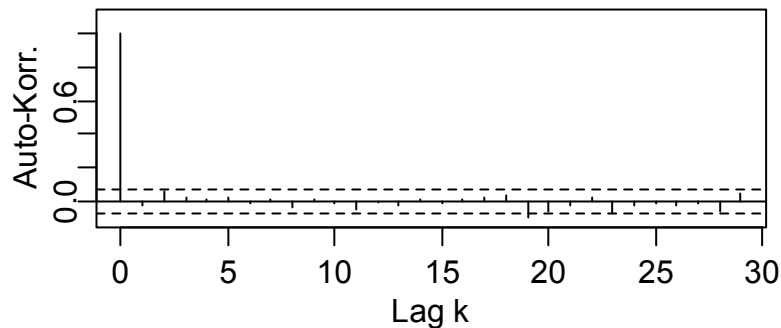
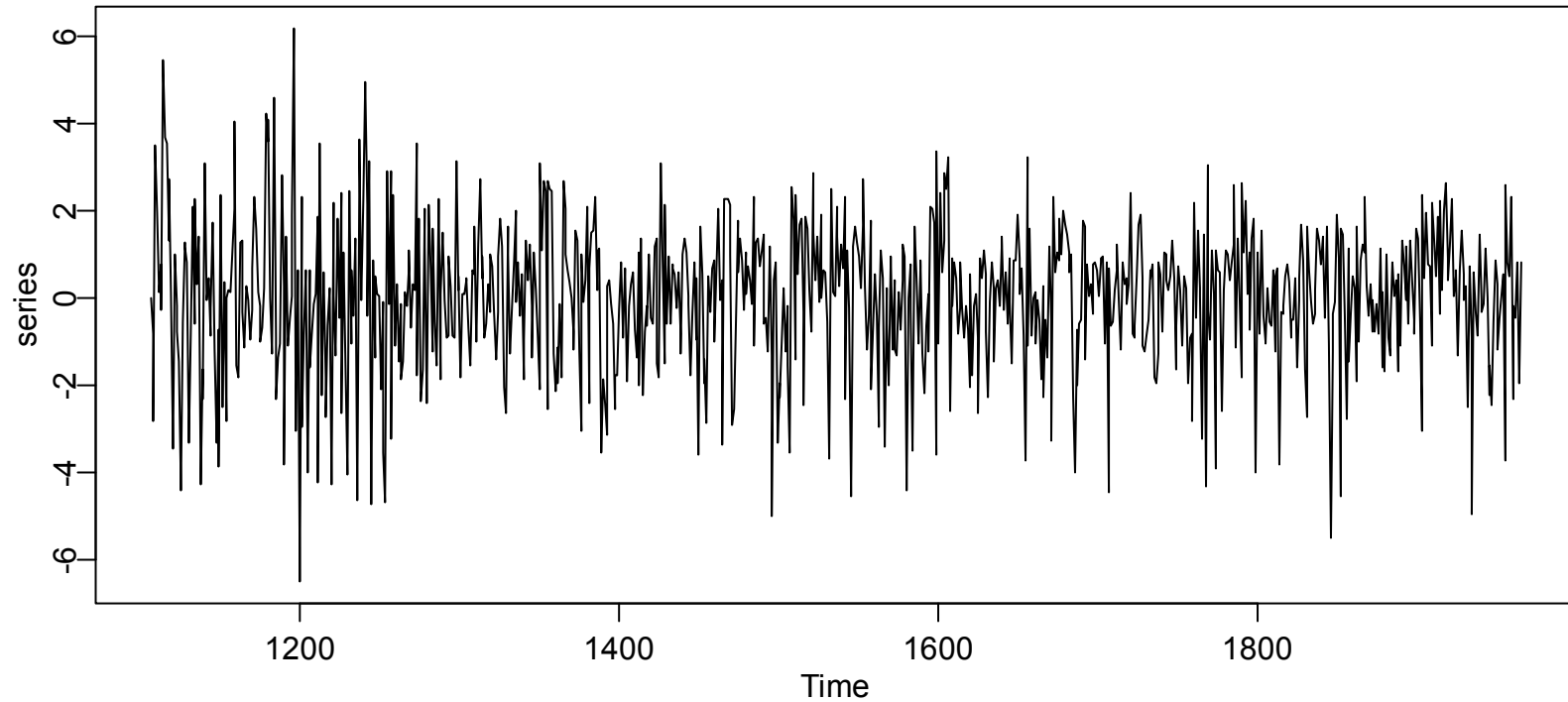
Residuals of MA(1)



Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Residuals of ARMA(1,1)



Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Another Example: Fitting ARMA(p,q)

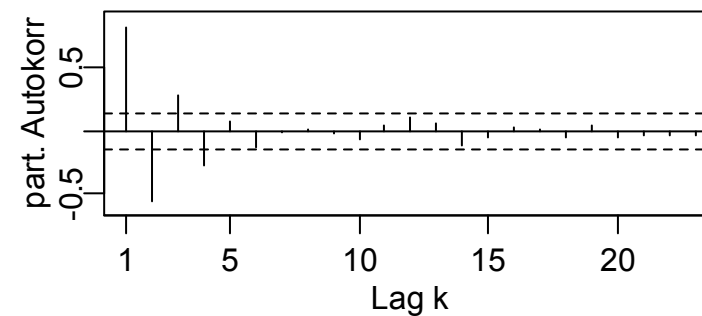
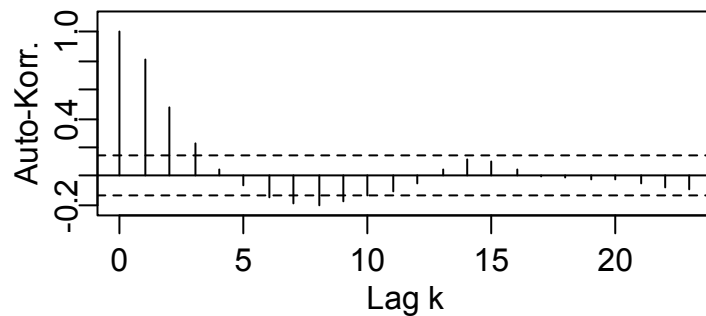
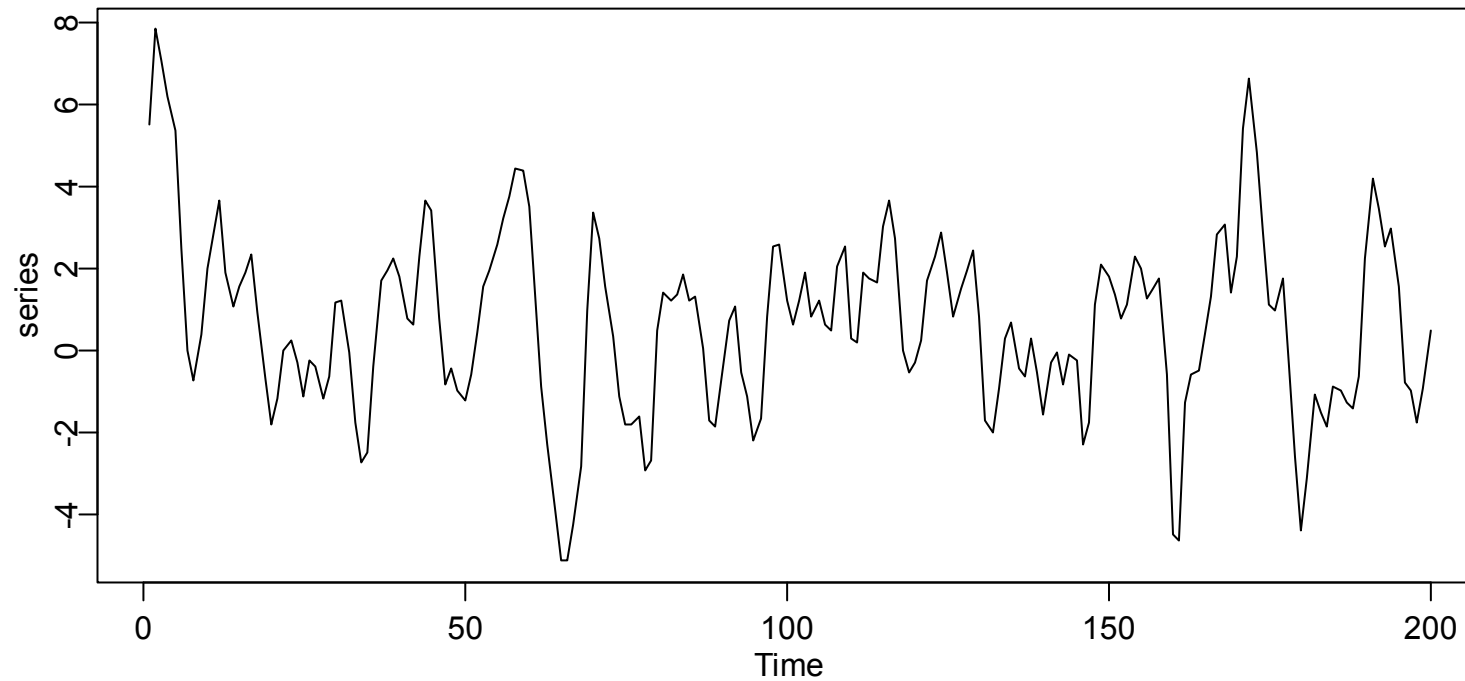
What needs to be done?

- 1) Achieve stationarity
→ transformations, differencing, modeling, ...
- 2) Choice of the order
→ determining (p,q), plus integration order d for ARIMA
- 3) Parameter estimation
→ ML-estimation of $\alpha, \beta, \mu, \sigma_E^2$
- 4) Residual analysis
→ if necessary, repeat 1), and/or 2)-4)

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

The Series, ACF and PACF



Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Model 1: AR(4)

```
> fit1
```

```
Call: arima(x = my.ts, order = c(4, 0, 0))
```

```
Coefficients:
```

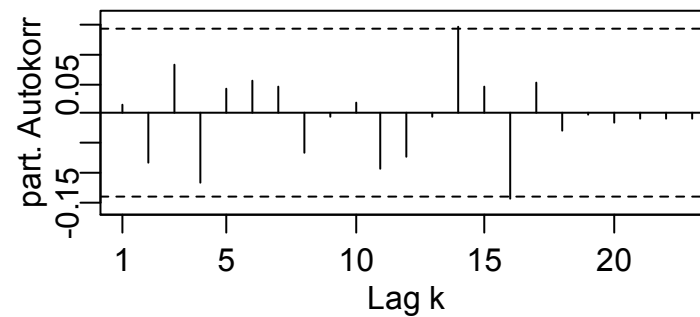
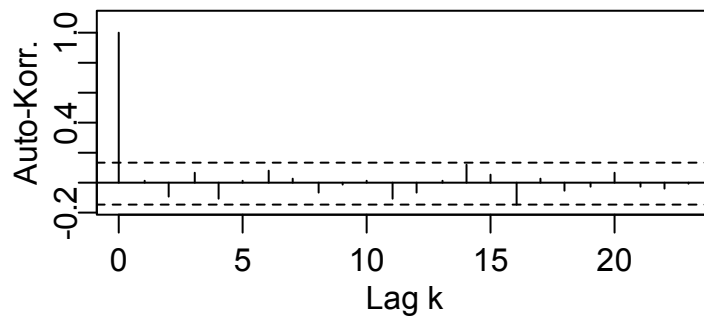
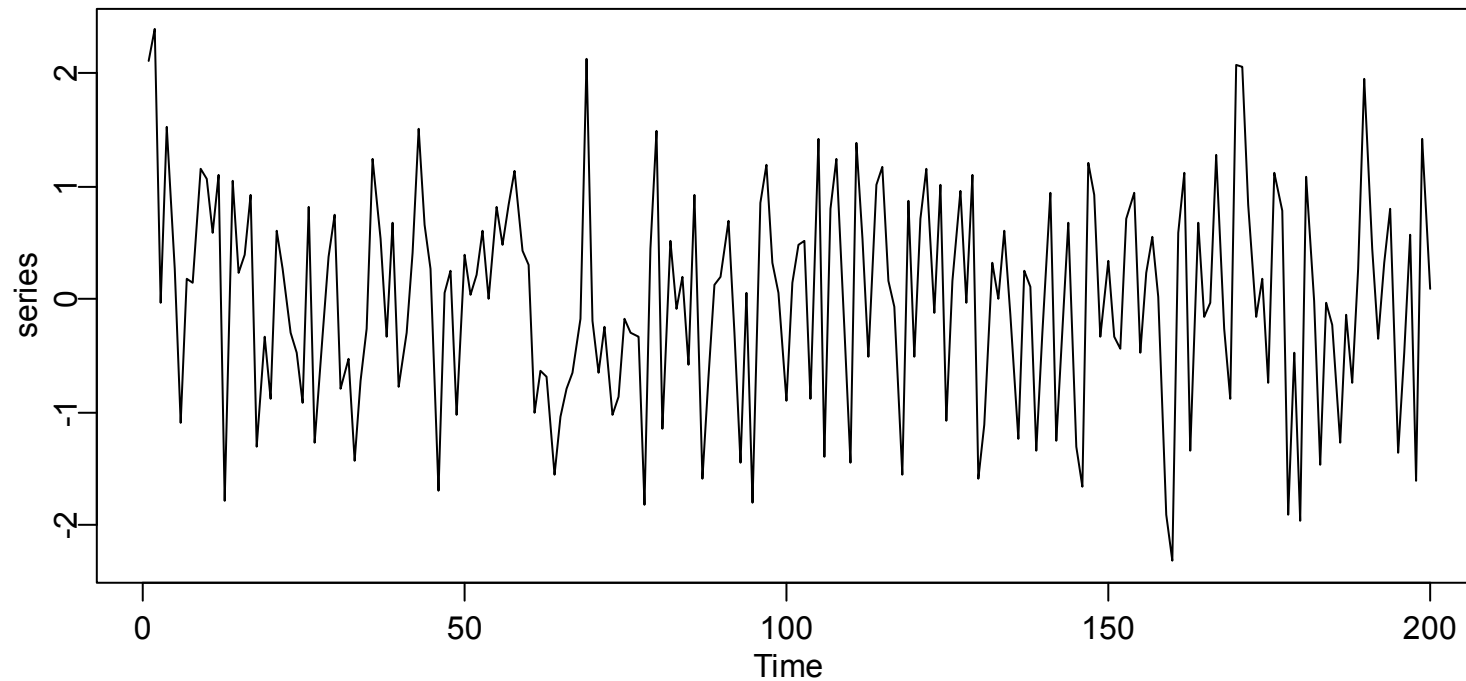
	ar1	ar2	ar3	ar4	intercept
	1.5430	-1.2310	0.7284	-0.3000	0.6197
s.e.	0.0676	0.1189	0.1189	0.0697	0.2573

```
sigma^2=0.8923, log likelihood=-273.67, aic=559.33
```

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Residuals of Model 1: AR(4)



Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Model 2: MA(3)

```
> fit2
```

```
Call: arima(x = my.ts, order = c(0, 0, 3))
```

```
Coefficients:
```

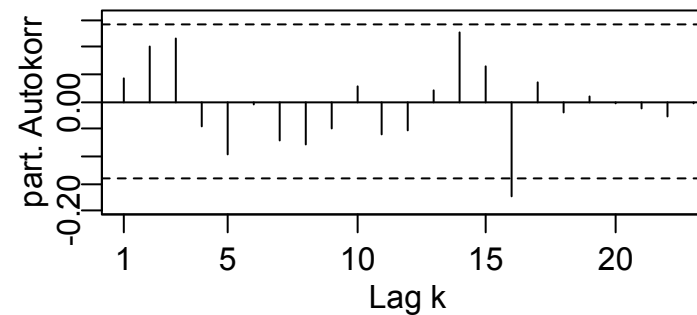
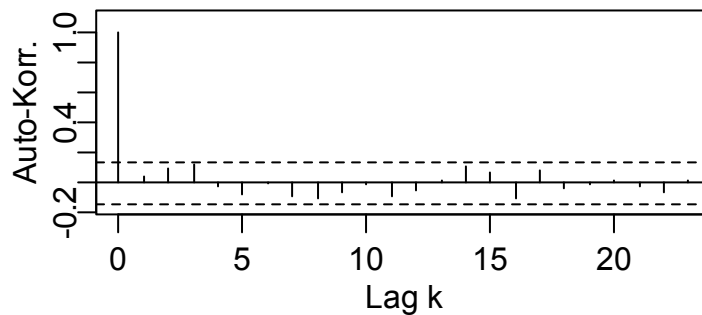
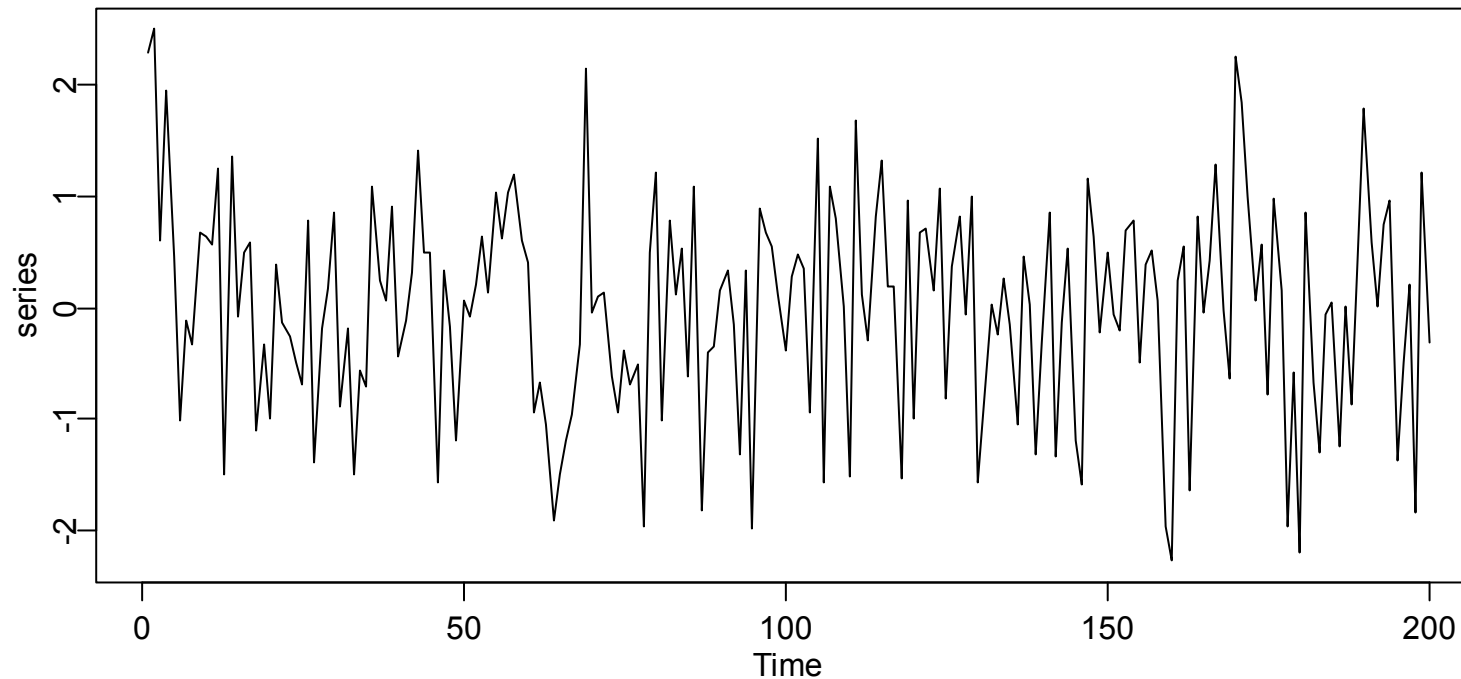
	ma1	ma2	ma3	intercept
	1.5711	1.0056	0.3057	0.6359
s.e.	0.0662	0.0966	0.0615	0.2604

```
sigma^2=0.9098, log likelihood=-275.64, aic=561.29
```

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Residuals of Model 2: MA(3)



Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Model 3: ARMA(1,1)

```
> fit3
```

```
Call: arima(x = my.ts, order = c(1, 0, 1))
```

```
Coefficients:
```

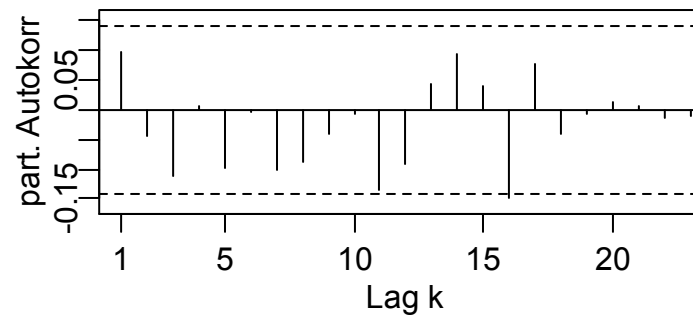
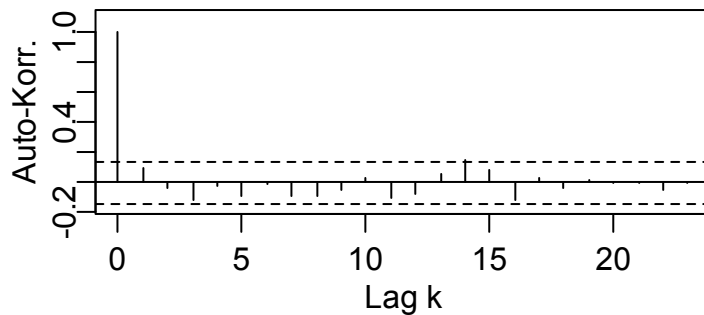
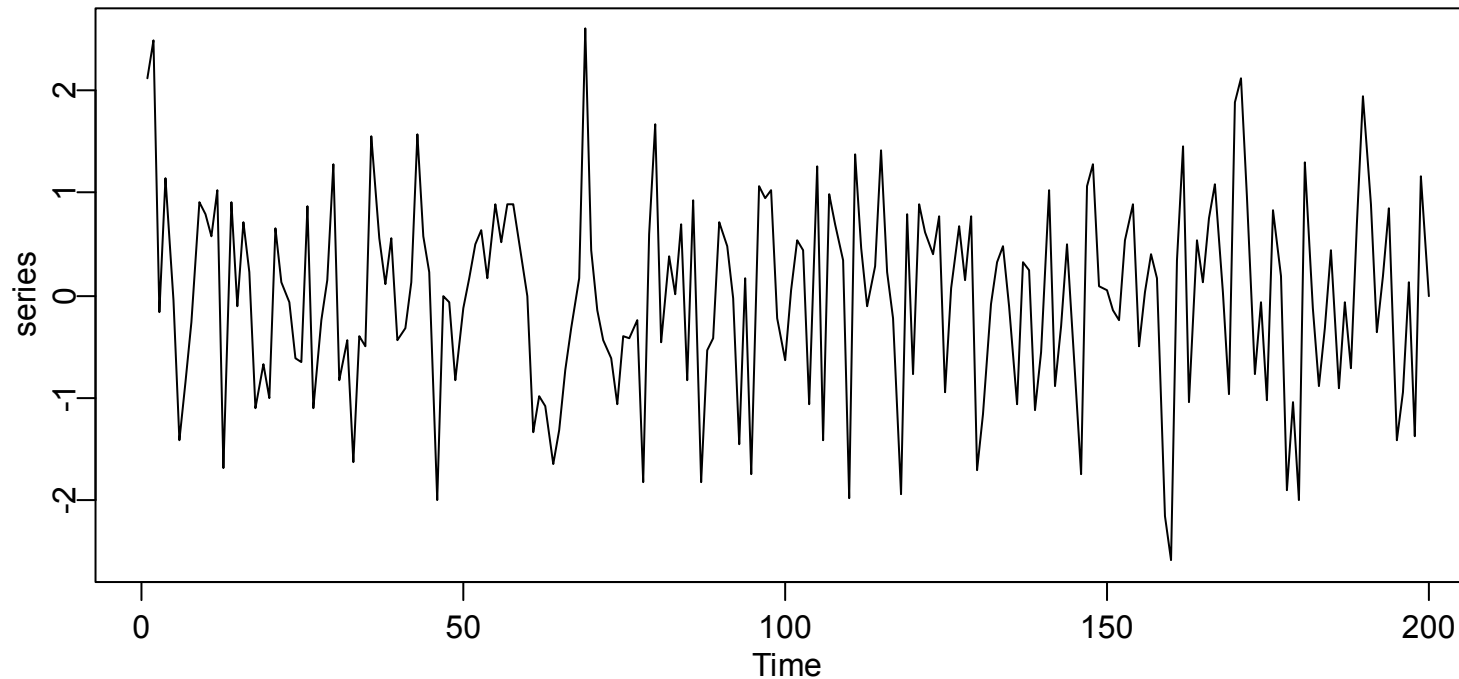
	ar1	ma1	intercept
	0.6965	0.7981	0.6674
s.e.	0.0521	0.0400	0.3945

```
sigma^2=0.9107, log likelihood=-275.72, aic=559.43
```

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Residuals of Model 3: ARMA(1,1)



Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Model 4: ARMA(2,1)

```
> fit4
```

```
Call: arima(x = my.ts, order = c(2, 0, 1))
```

```
Coefficients:
```

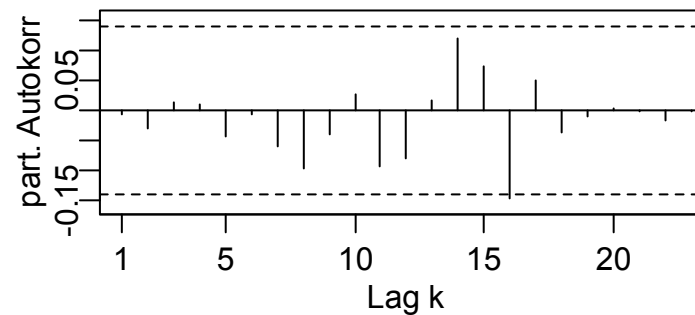
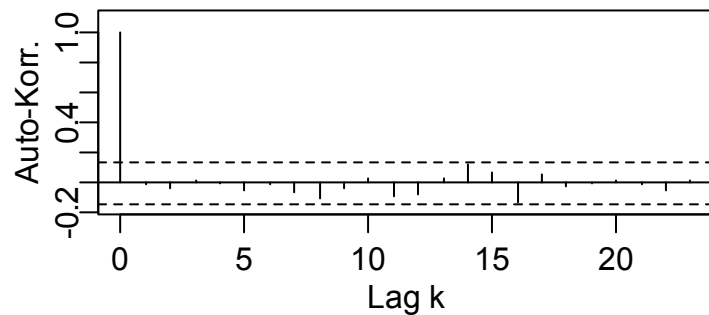
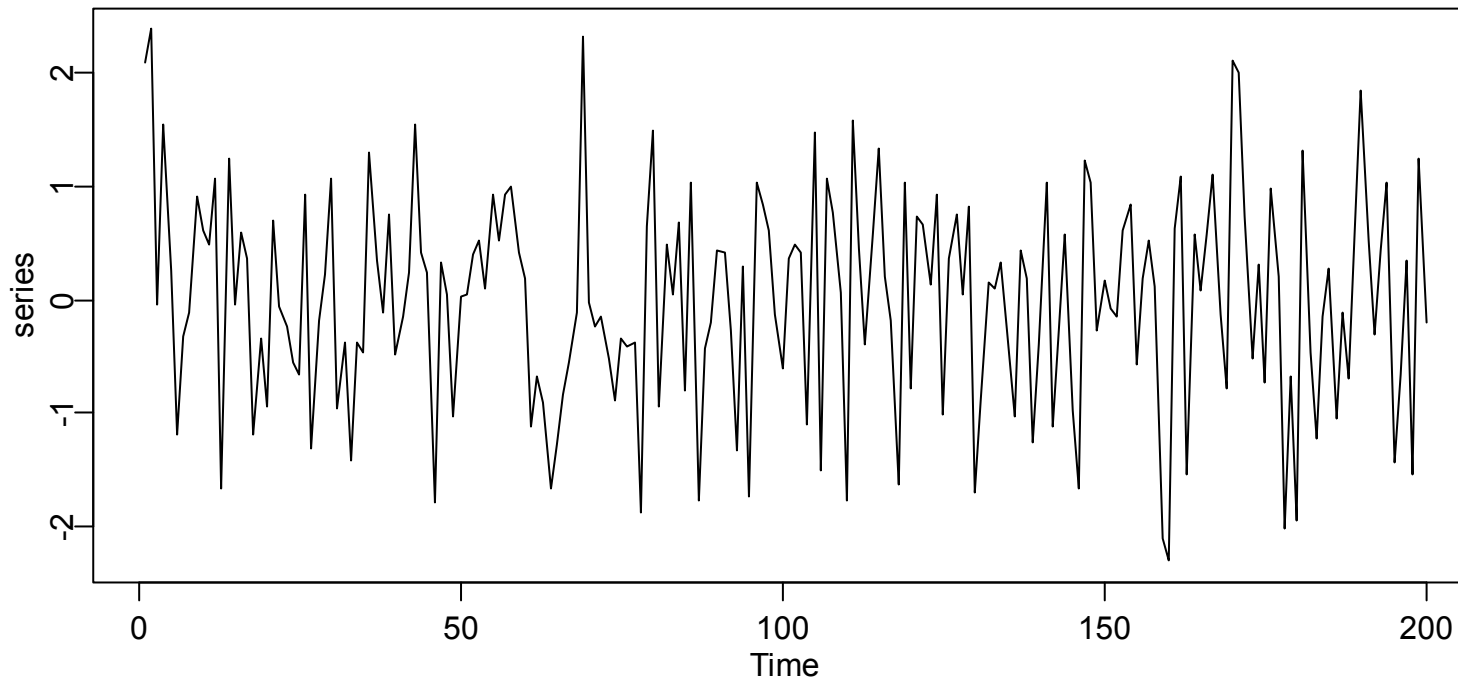
	ar1	ar2	ma1	intercept
	0.8915	-0.2411	0.7061	0.6420
s.e.	0.0855	0.0856	0.0625	0.3208

```
sigma^2=0.8772, log likelihood=-272.01, aic=554.02
```

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Residuals of Model 4: ARMA(2,1)



Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Model 5: ARMA(4,1)

```
> fit5
```

```
Call: arima(x = my.ts, order = c(4, 0, 1))
```

```
Coefficients:
```

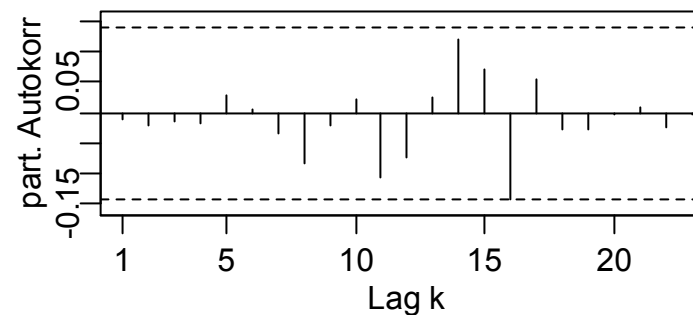
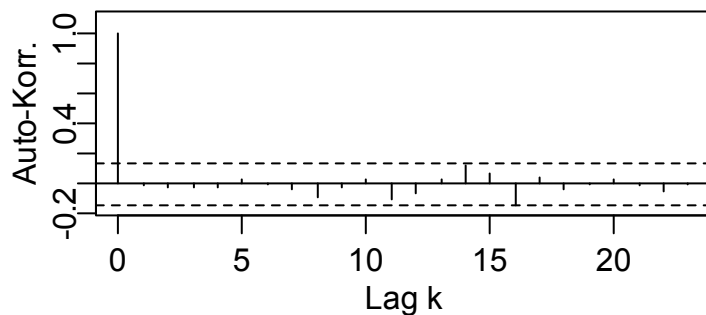
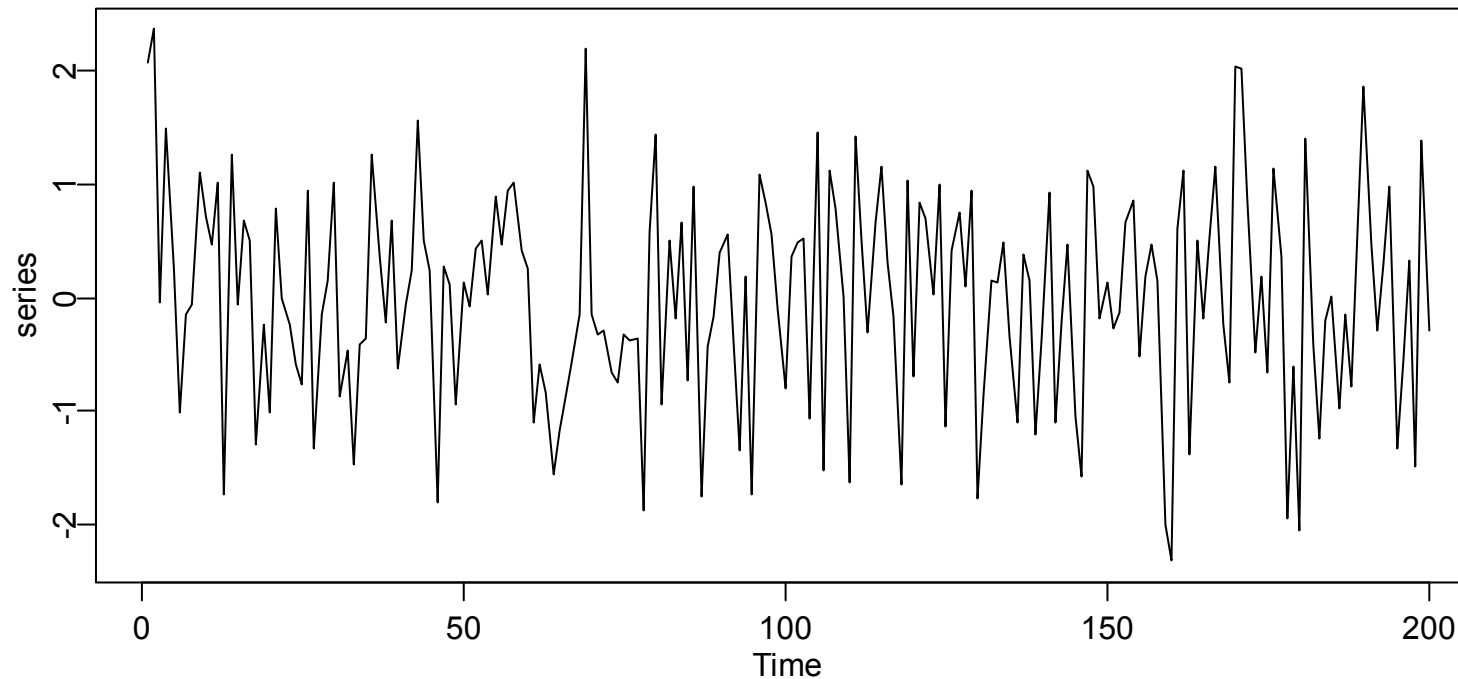
	ar1	ar2	ar3	ar4	ma1	intercept
	1.0253	-0.4693	0.2190	-0.1280	0.5733	0.6312
s.e.	0.1725	0.2658	0.2124	0.1062	0.1653	0.2930

```
sigma^2=0.8708, log likelihood=-271.3, aic = 556.59
```

Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Residuals of Model 5: ARMA(4,1)



Applied Time Series Analysis

SS 2015 – ARMA(p,q)

Summary of the Order Choice Problem

- Regarding ACF/PACF, all 5 models are plausible
→ ARMA(2,1) would be my favorite
- The residuals look fine (i.e. independent) for all 5 models
→ no further evidence for a particular model
- Regarding AIC, the ARMA models do better
→ ARMA(2,1) would be my favorite
- Significance of the coefficients
→ excludes the ARMA(4,1) as the last contender

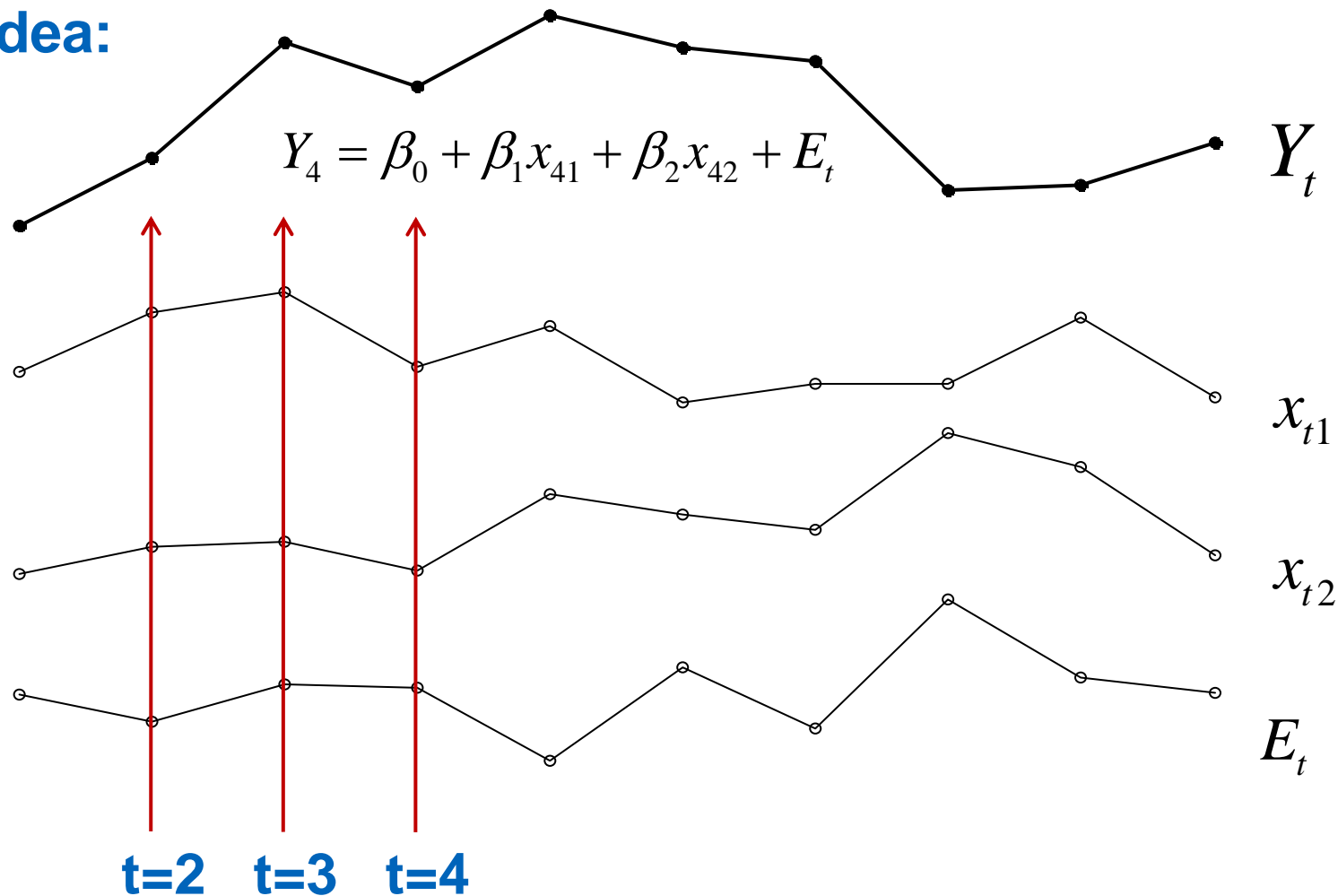
Best choice: ARMA (2,1)

Applied Time Series Analysis

SS 2015 – Time Series Regression

Time Series Regression

Idea:



Applied Time Series Analysis

SS 2015 – Time Series Regression

Time Series Regression

- We speak of time series regression if response and predictors are time series, i.e. if they were observed in a sequence.
- In principle, it is perfectly fine to apply the usual OLS setup

$$Y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_q x_{tp} + E_t$$

Be careful: this assumes that the errors are uncorrelated.

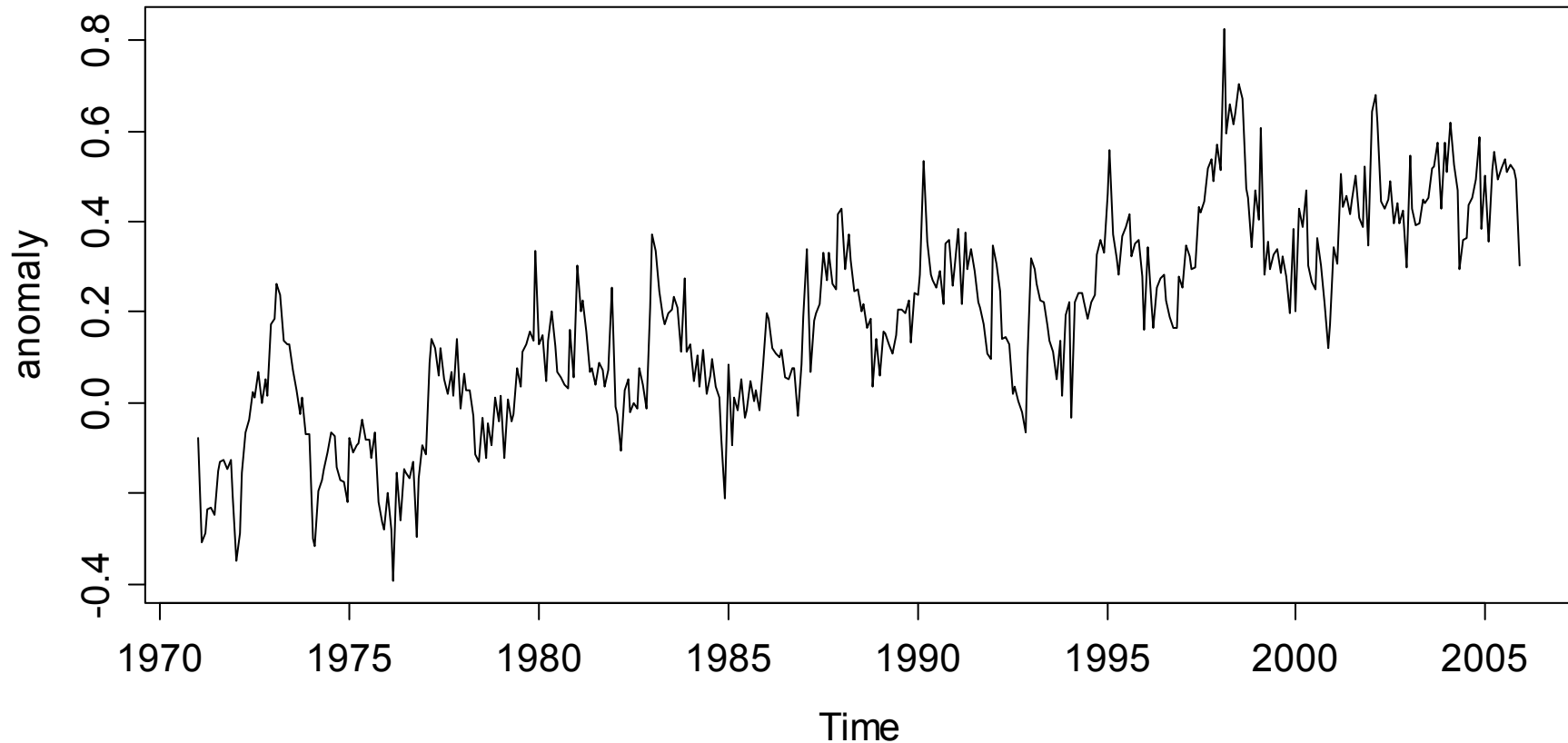
- With correlated errors, the estimates $\hat{\beta}_j$ are still unbiased, but more efficient estimators than OLS exist. The standard errors are wrong, often underestimated, causing spurious significance.
- The **Generalized Least Squares** procedure solves the issue!

Applied Time Series Analysis

SS 2015 – Time Series Regression

Example 1: Global Temperature

Global Temperature Anomalies



Applied Time Series Analysis

SS 2015 – Time Series Regression

Example 1: Global Temperature

Temperature = Trend + Seasonality + Remainder

$$Y_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot 1_{[month="Feb"]} + \dots + \beta_{12} \cdot 1_{[month="Dec"]} + E_t,$$

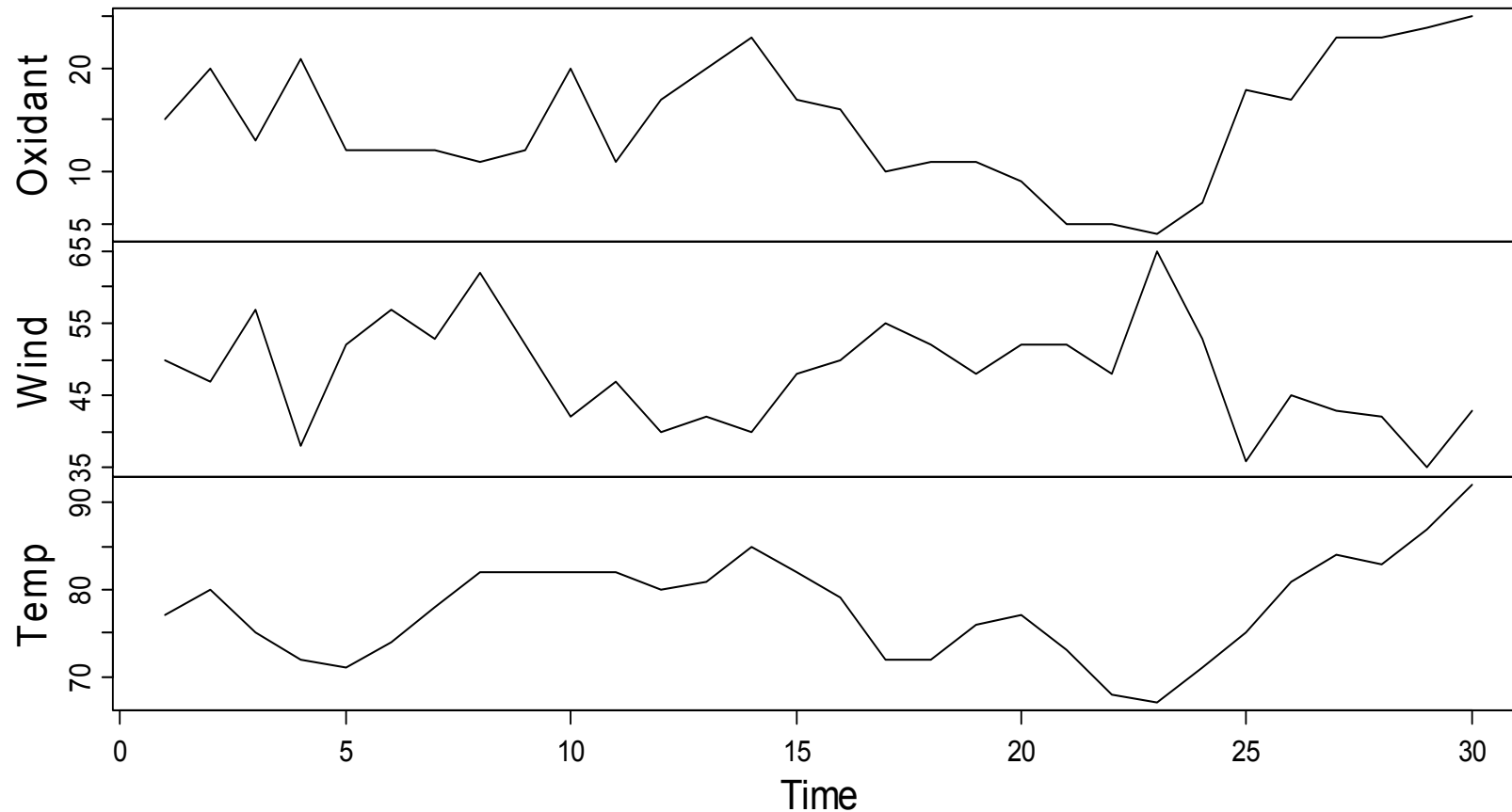
- Recordings from 1971 to 2005, $n = 420$
- The remainder term is usually a stationary time series, thus it would not be surprising if the regression model features correlated errors.
- The applied question which is of importance here is whether there is a significant trend, and a significant seasonal variation

Applied Time Series Analysis

SS 2015 – Time Series Regression

Example 2: Air Pollution

Air Pollution Data



Applied Time Series Analysis

SS 2015 – Time Series Regression

Example 2: Air Pollution

Oxidant = Wind + Temperature + Error

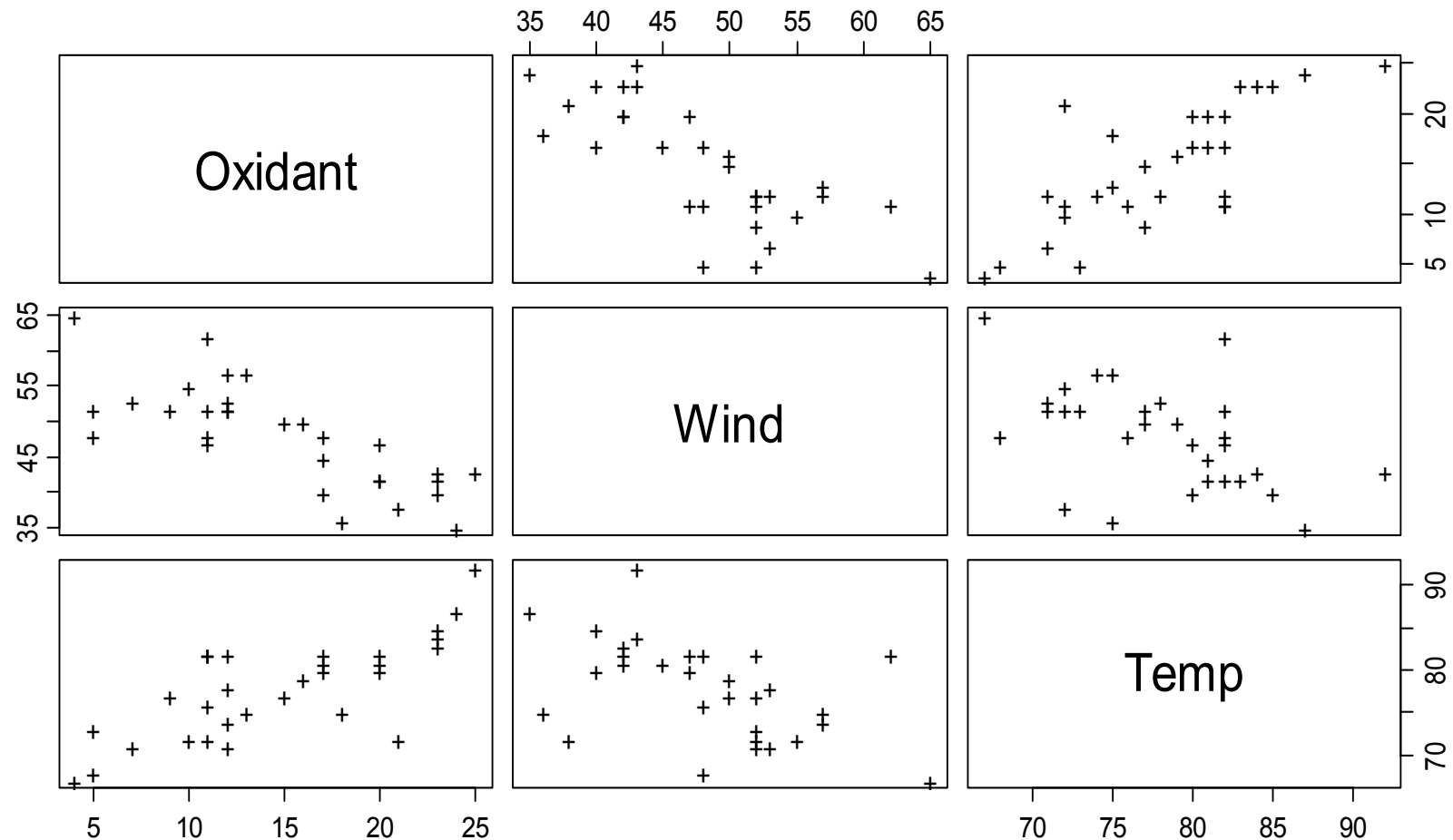
$$Y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + E_t$$

- Recordings from 30 consecutive days, $n = 30$
- The data are from the Los Angeles basin, USA
- The pollutant level is influenced by both wind and temperature, plus some more, unobserved variables.
- It is well conceivable that there is "day-to-day memory" in the pollutant levels, i.e. there are correlated errors.

Applied Time Series Analysis

SS 2015 – Time Series Regression

Example 2: Air Pollution

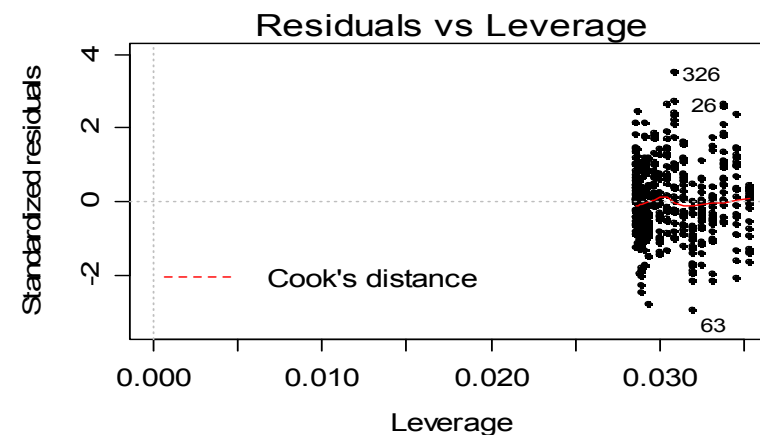
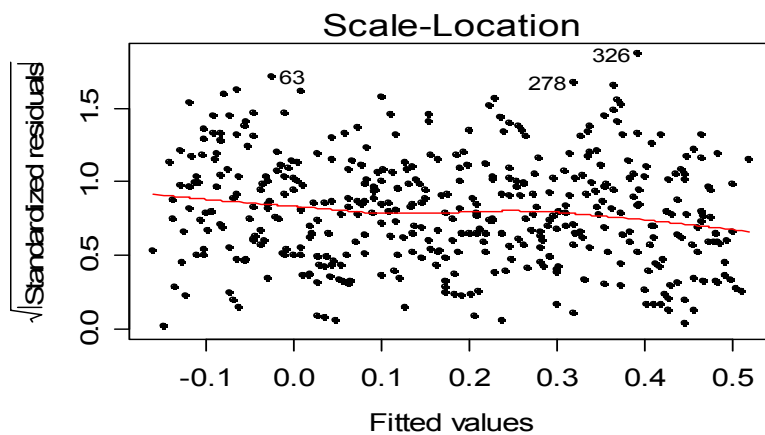
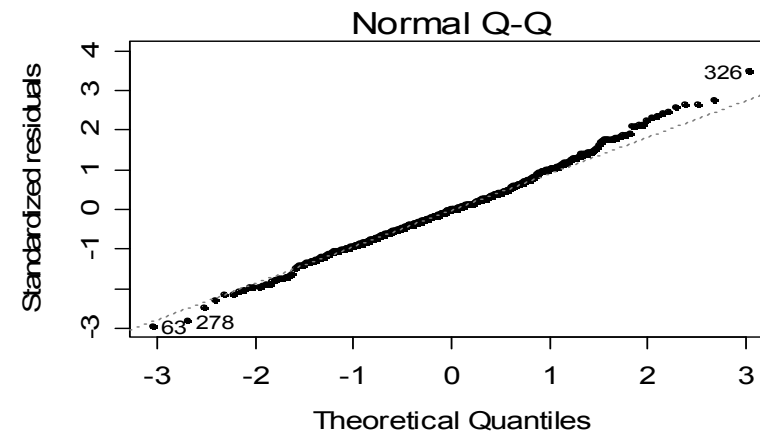
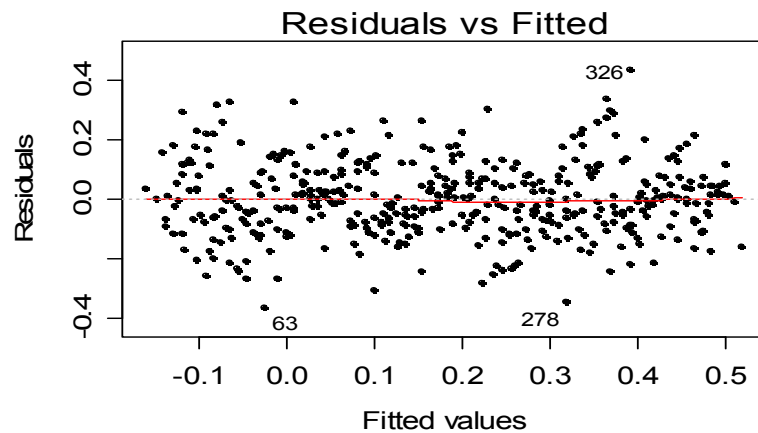


Applied Time Series Analysis

SS 2015 – Time Series Regression

Finding Correlated Errors

1) Start by fitting an OLS regression and analyze residuals



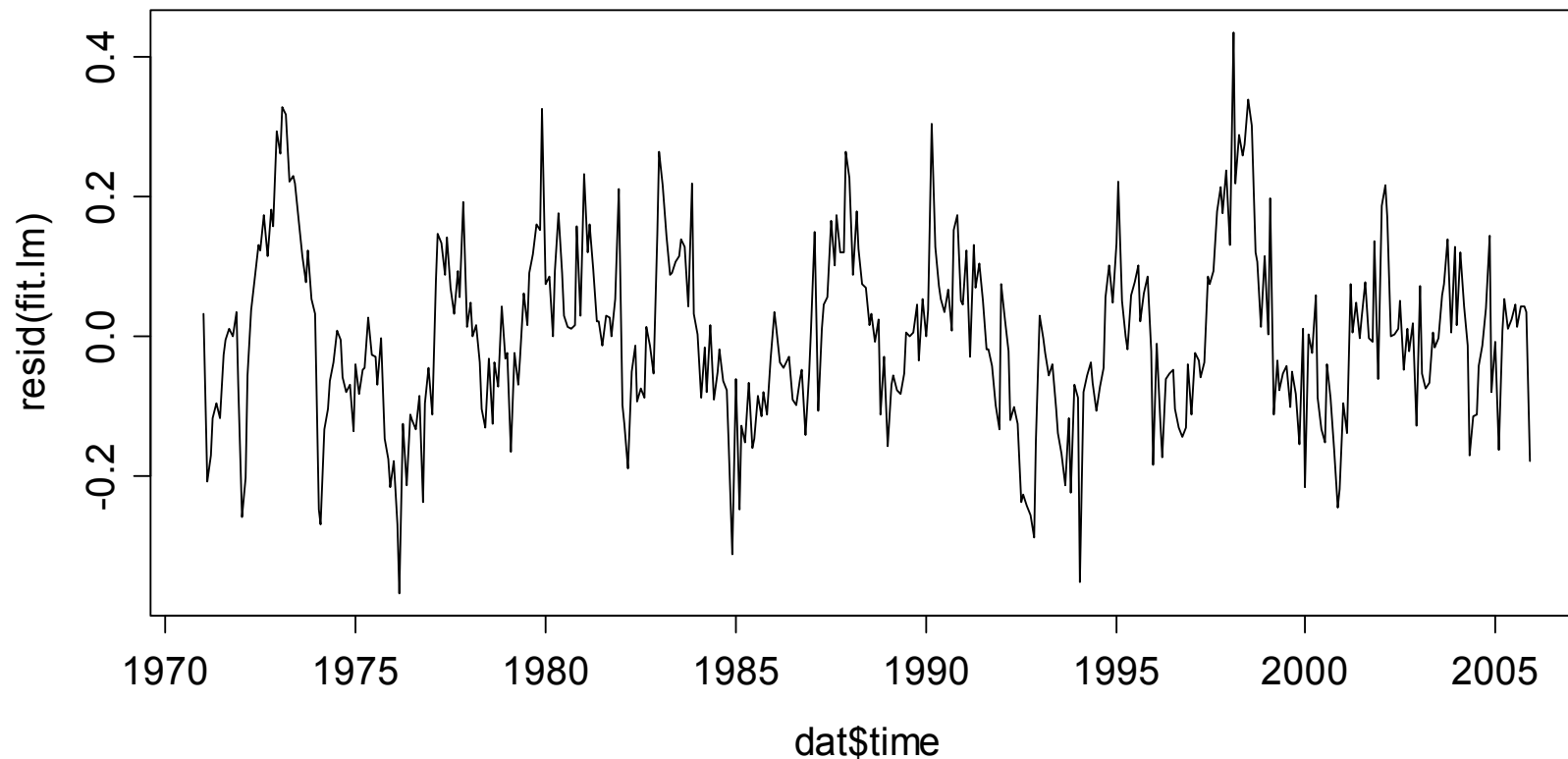
Applied Time Series Analysis

SS 2015 – Time Series Regression

Finding Correlated Errors

2) Continue with a time series plot of OLS residuals

Residuals of the lm() Function



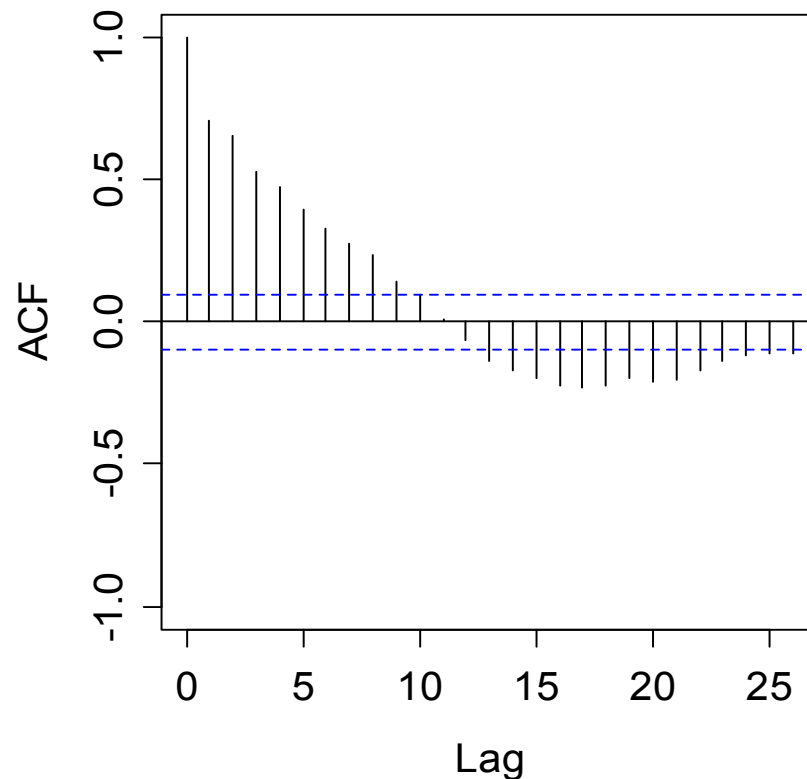
Applied Time Series Analysis

SS 2015 – Time Series Regression

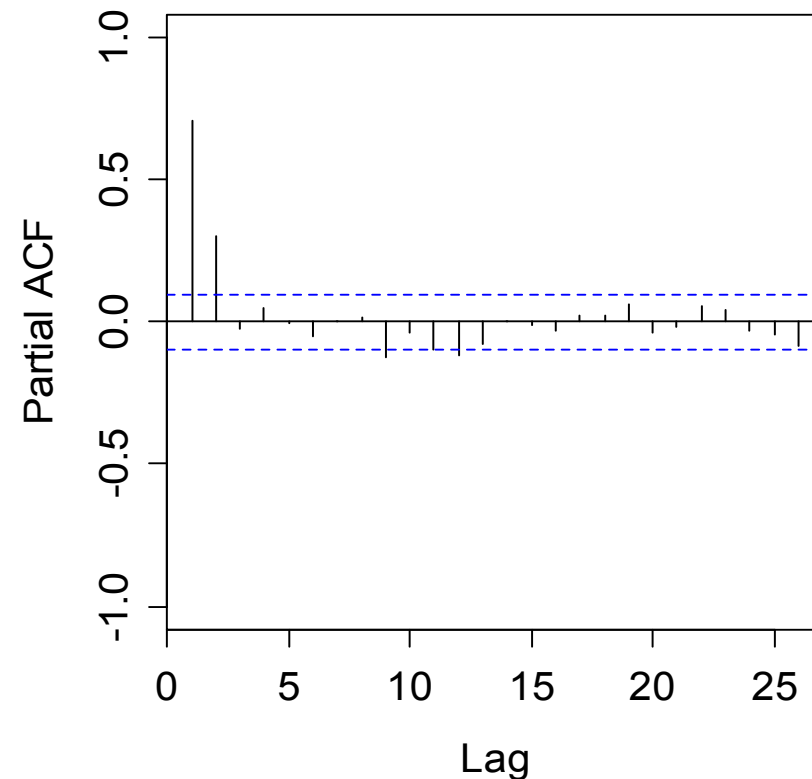
Finding Correlated Errors

3) Also analyze ACF and PACF of OLS residuals

ACF of Residuals



PACF of Residuals



Applied Time Series Analysis

SS 2015 – Time Series Regression

Model for Correlated Errors

→ It seems as if an AR(2) model provides an adequate model for the correlation structure observed in the residuals of the OLS regression model.

```
> fit.ar2 <- ar.burg(resid(fit.lm)); fit.ar2
```

```
Call: ar.burg.default(x = resid(fit.lm))
```

```
Coefficients:
```

```
      1      2  
0.4945 0.3036
```

```
Order selected 2  sigma^2 estimated as 0.00693
```

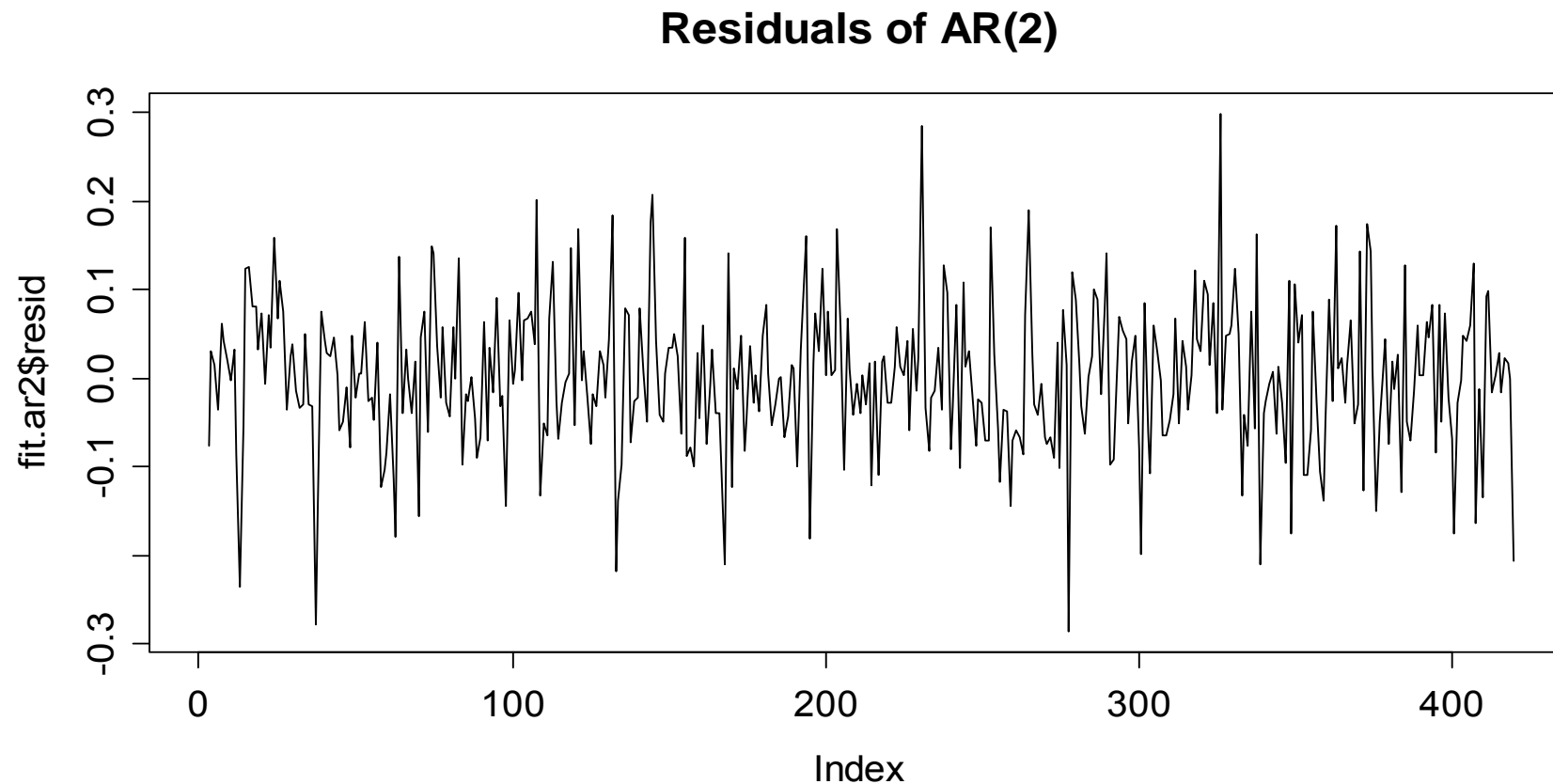
→ Residuals of this AR(2) model must look like white noise!

Applied Time Series Analysis

SS 2015 – Time Series Regression

Does the Model Fit?

5) Visualize a time series plot of the AR(2) residuals

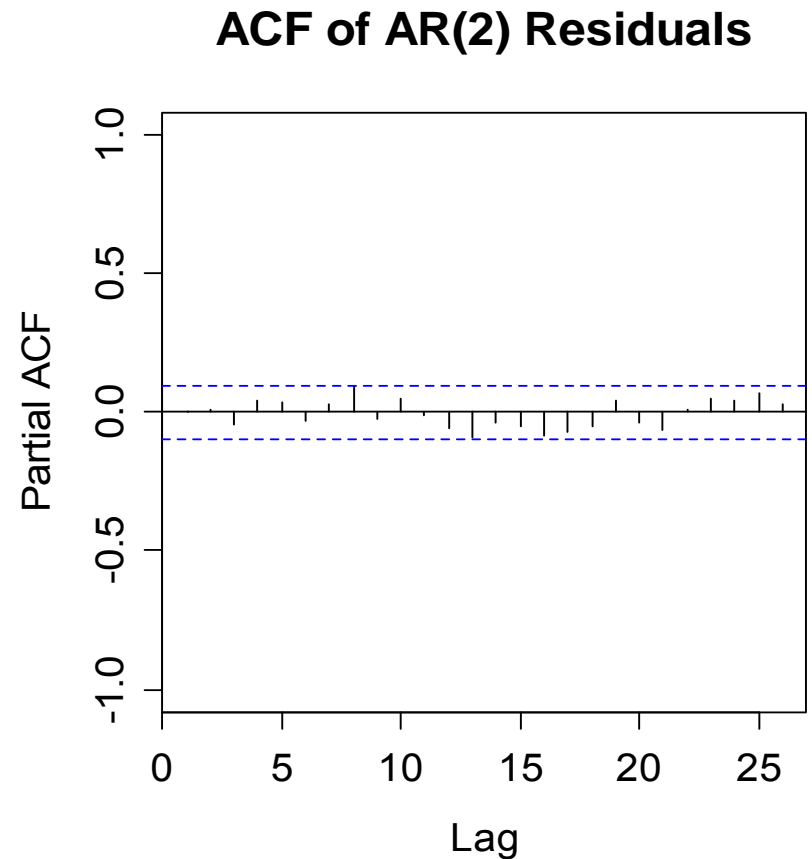
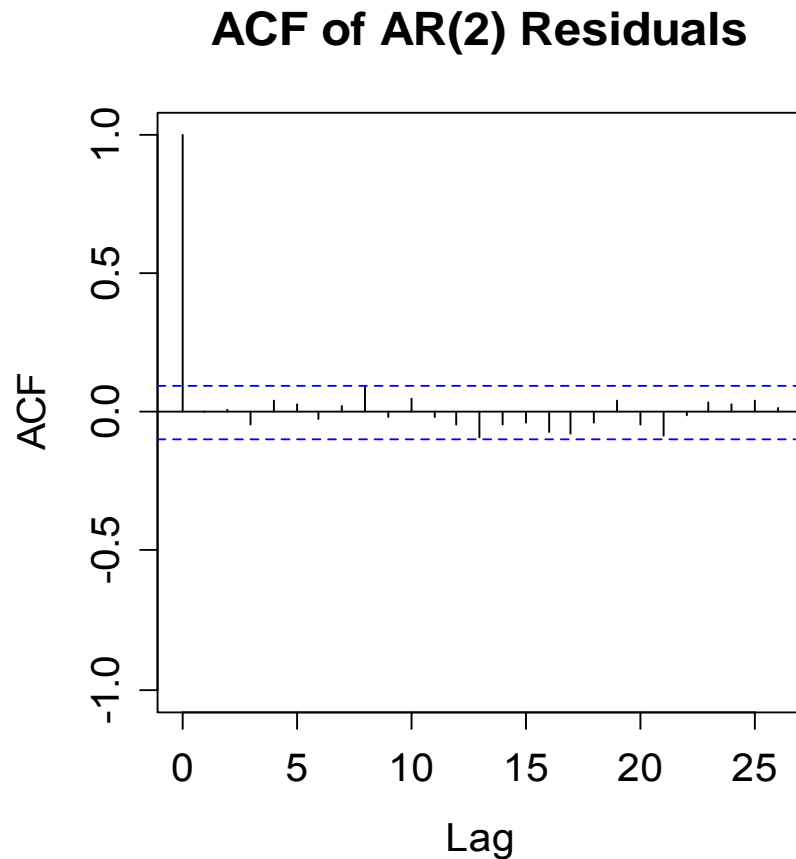


Applied Time Series Analysis

SS 2015 – Time Series Regression

Does the Model Fit?

5) ACF and PACF plots of AR(2) residuals



Applied Time Series Analysis

SS 2015 – Time Series Regression

Global Temperature: Conclusions

- The residuals from OLS regression are visibly correlated.
- An AR(2) model seems appropriate for this dependency.
- The AR(2) yields a good fit, because its residuals have White Noise properties. We have thus understood the dependency of the regression model errors.

→ We need to account for the correlated errors, else the coefficient estimates will be unbiased but inefficient, and the standard errors are wrong, preventing successful inference for trend and seasonality

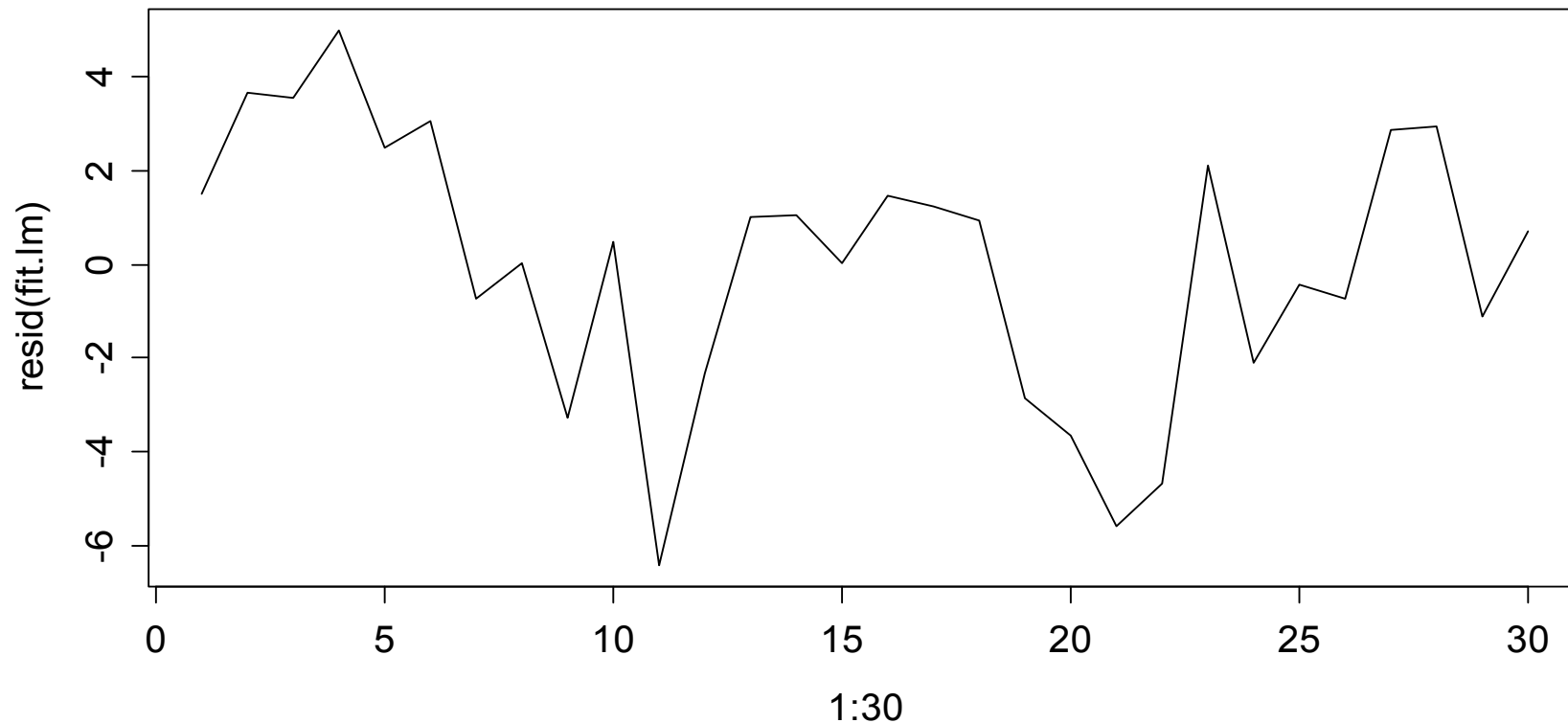
Applied Time Series Analysis

SS 2015 – Time Series Regression

Air Pollution: OLS Residuals

Time series plot: dependence present or not?

Residuals of the lm() Function



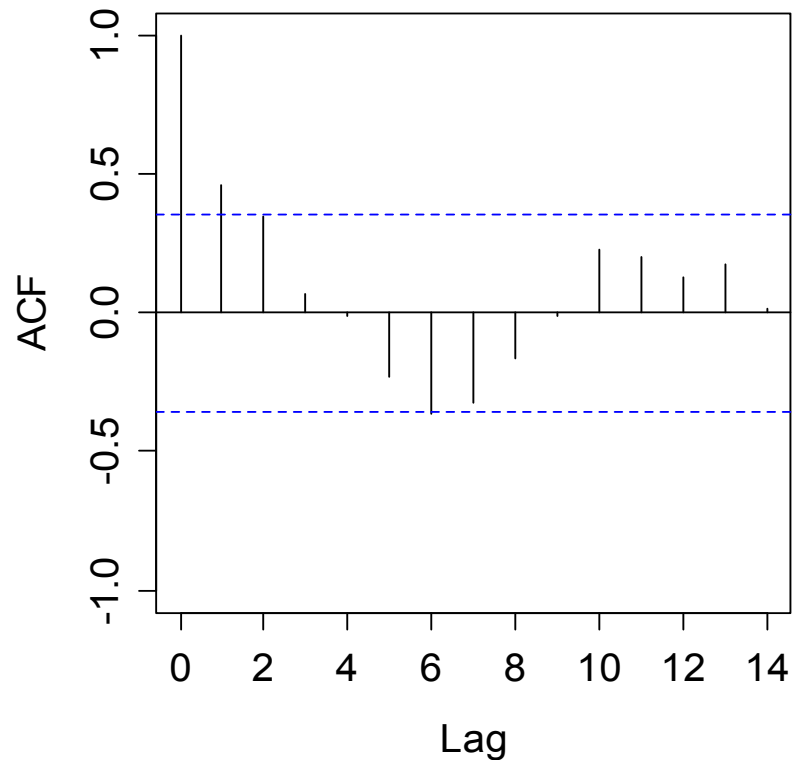
Applied Time Series Analysis

SS 2015 – Time Series Regression

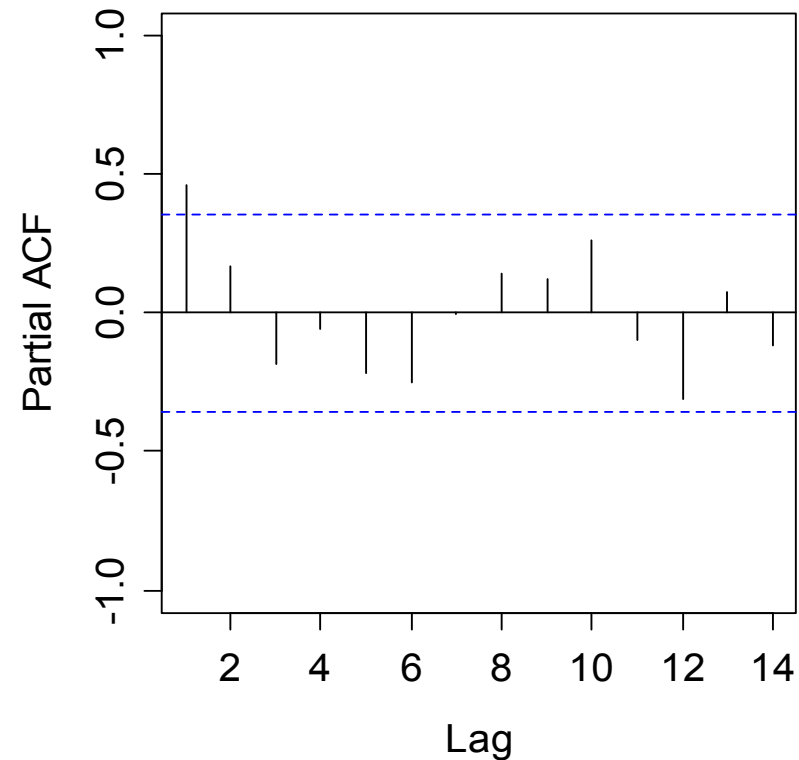
Air Pollution: OLS Residuals

ACF and PACF suggest: *there is AR(1) dependence*

ACF of Residuals



PACF of Residuals



Applied Time Series Analysis

SS 2015 – Time Series Regression

Pollutant Example

```
> summary(erg.poll,corr=F)
```

```
Call: lm(formula = Oxidant ~ Wind + Temp, data = pollute)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-5.20334	11.11810	-0.468	0.644	
Wind	-0.42706	0.08645	-4.940	3.58e-05	***
Temp	0.52035	0.10813	4.812	5.05e-05	***

Residual standard error: 2.95 on 27 degrees of freedom

Multiple R-squared: 0.7773, Adjusted R-squared: 0.7608

F-statistic: 47.12 on 2 and 27 DF, p-value: 1.563e-09

Applied Time Series Analysis

SS 2015 – Time Series Regression

Pollutant Example

```
> summary(erg.poll,corr=F)
```

```
Call: lm(formula = Oxidant ~ Wind + Temp, data = pollute)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-5.20334	11.11810	-0.468	0.644	
Wind	-0.42706	0.08645	-4.940	3.58e-05	***
Temp	0.52035	0.10813	4.812	5.05e-05	***

Residual standard error: 2.95 on 27 degrees of freedom

Multiple R-squared: 0.7773, Adjusted R-squared: 0.7608

F-statistic: 47.12 on 2 and 27 DF, p-value: 1.563e-09

Applied Time Series Analysis

SS 2015 – Time Series Regression

Durbin-Watson Test

- The Durbin-Watson approach is a test for autocorrelated errors in regression modeling based on the test statistic:

$$D = \frac{\sum_{t=2}^N (r_t - r_{t-1})^2}{\sum_{t=1}^N r_t^2}$$

- This is implemented in R: `dwtest()` in `library(lmtest)`. A p-value for the null of no autocorrelation is computed.
- This test does not detect all autocorrelation structures. If the null is not rejected, the residuals may still be autocorrelated.

→ **Never forget to check ACF/PACF of the residuals!**

Applied Time Series Analysis

SS 2015 – Time Series Regression

Durbin-Watson Test

Example 1: Global Temperature

```
> library(lmtest)
> dwtest(fit.lm)
data:  fit.lm
DW = 0.5785, p-value < 2.2e-16
alt. hypothesis: true autocorrelation is greater than 0
```

Example 2: Air Pollution

```
> dwtest(fit.lm)
data:  fit.lm
DW = 1.0619, p-value = 0.001675
alt. hypothesis: true autocorrelation is greater than 0
```

Applied Time Series Analysis

SS 2015 – Time Series Regression

Generalized Least Squares

→ See the blackboard for full explanation

- OLS regression assumes a diagonal error covariance matrix, but there is a generalization to $Var(E) = \sigma^2 \Sigma$.

- If we find $\Sigma = SS^T$, the regression model can be rewritten as:

$$y = X\beta + E$$

$$S^{-1}y = S^{-1}X\beta + S^{-1}E$$

$$y^* = X^*\beta + E^* \quad \text{with } Var(E^*) = \sigma^2 I$$

- One obtains the generalized least square estimates:

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y \quad \text{with } Var(\hat{\beta}) = (X^T \Sigma^{-1} X)^{-1} \sigma^2$$

Applied Time Series Analysis

SS 2015 – Time Series Regression

Generalized Least Squares

For using the GLS approach, i.e. for correcting the dependent errors, we need an estimate of the error covariance matrix Σ .

The two major options for obtaining it are:

- 1) **Cochrane-Orcutt (for AR(p) correlation structure only)**
iterative approach: i) β , ii) α , iii) β
 - 2) **GLS (Generalized Least Squares, for ARMA(p,q))**
simultaneous estimation of β and α
- **Full explanation of the two different approaches is provided on the blackboard!**

Applied Time Series Analysis

SS 2015 – Time Series Regression

GLS: Syntax

Package `nlme` has function `gls()`. It does only work if the correlation structure of the errors is provided. This has to be determined from the residuals of an OLS regression first.

```
> library(nlme)
> corStruct <- corARMA(form=~time, p=2)
> fit.gls <- gls(temp~time+season, data=dat,
                 correlation=corStruct)
```

The output contains the **regression coefficients** and their **standard errors**, as well as the **AR-coefficients** plus some further information about the model (Log-Likeli, AIC, ...).

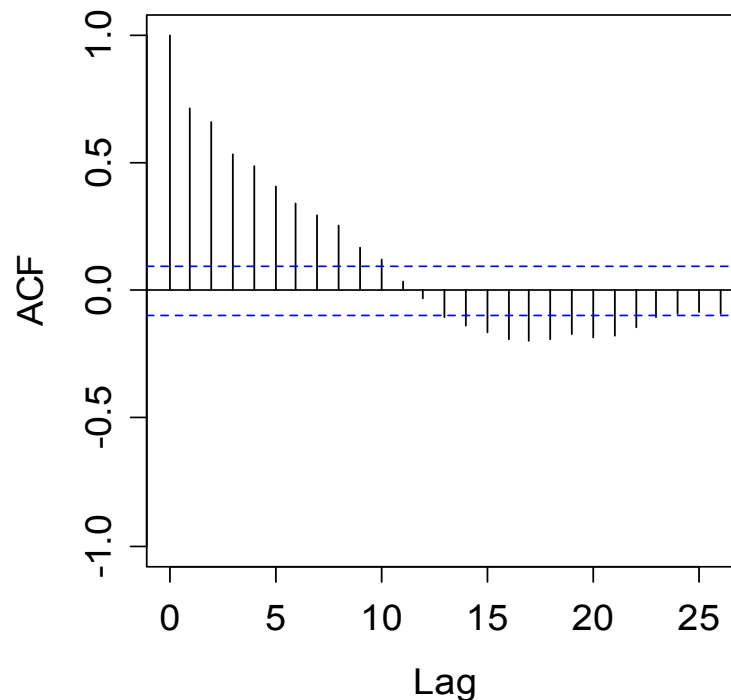
Applied Time Series Analysis

SS 2015 – Time Series Regression

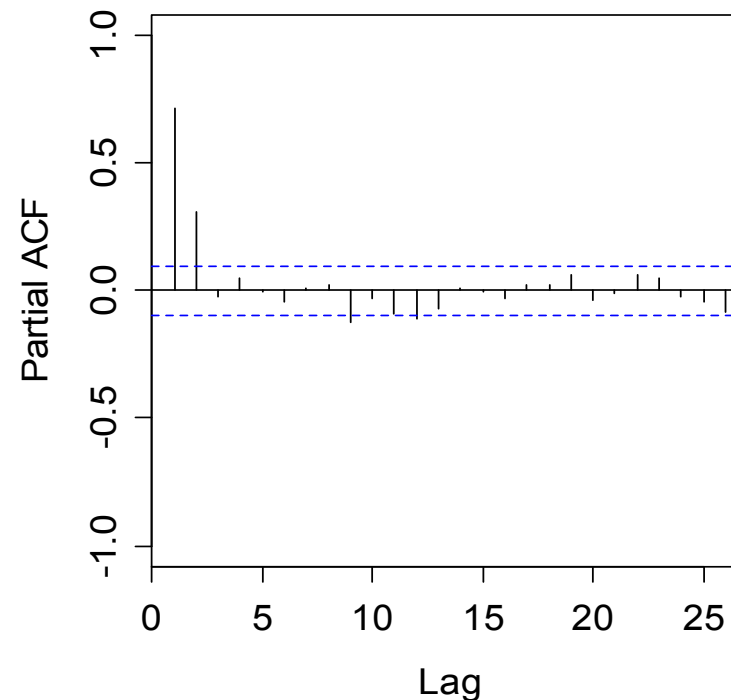
GLS: Residual Analysis

The residuals from a GLS must look like coming from a time series process with the respective structure:

ACF of GLS-Residuals



PACF of GLS-Residuals



Applied Time Series Analysis

SS 2015 – Time Series Regression

GLS/OLS: Comparison of Results

→ The trend in the global temperature is significant!

```
> coef(fit.lm)["time"]
      time
0.01822374
> confint(fit.lm, "time")
           2.5 %      97.5 %
time 0.01702668 0.0194208
```

OLS

```
> coef(fit.gls)["time"]
      time
0.02017553
> confint(fit.gls, "time")
           2.5 %      97.5 %
time 0.01562994 0.02472112
```

GLS

Applied Time Series Analysis

SS 2015 – Time Series Regression

GLS/OLS: Comparison of Results

→ The seasonal effect is not significant!

```
> drop1(fit.lm, test="F")
```

OLS

```
temp ~ time + season
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(F)	
<none>			6.4654	-1727.0			
time	1	14.2274	20.6928	-1240.4	895.6210	<2e-16	***
season	11	0.1744	6.6398	-1737.8	0.9982	0.4472	

```
> anova(fit.gls)
```

GLS

```
Denom. DF: 407
```

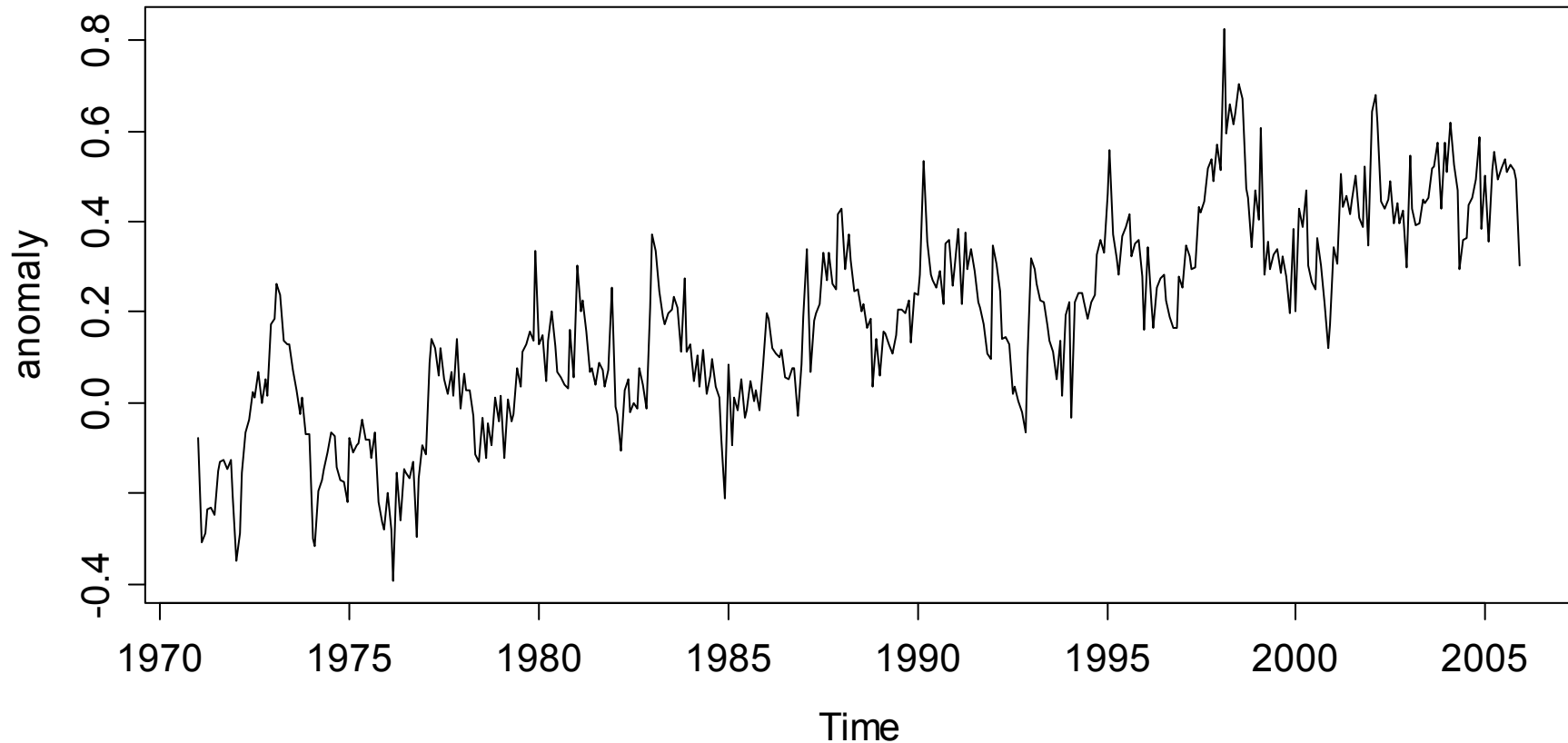
	numDF	F-value	p-value
(Intercept)	1	78.40801	<.0001
time	1	76.48005	<.0001
season	11	0.64371	0.7912

Applied Time Series Analysis

SS 2015 – Time Series Regression

Example 1: Global Temperature

Global Temperature Anomalies



Applied Time Series Analysis

SS 2015 – Time Series Regression

Air Pollution: Results

Both predictors are significant with both approaches...

```
> confint(fit.lm, c("Wind", "Temp"))  
                2.5 %      97.5 %  
Wind -0.6044311 -0.2496841  
Temp  0.2984794  0.7422260
```

OLS

```
> confint(fit.gls, c("Wind", "Temp"))  
                2.5 %      97.5 %  
Wind -0.5447329 -0.2701709  
Temp  0.2420436  0.7382426
```

GLS

→ But still, it is important to use GLS with correlated errors!

Applied Time Series Analysis

SS 2015 – Time Series Regression

Simulation Study: Model

We want to study the effect of correlated errors on the quality of estimates when using the least squares approach:

$$x_t = t / 50$$

$$y_t = x_t + 2x_t^2 + E_t$$

where E_t is from an AR(1)-process with $\alpha = -0.65$ and $\sigma = 0.1$.

We generate 100 realizations from this model and estimate the regression coefficient and its standard error by:

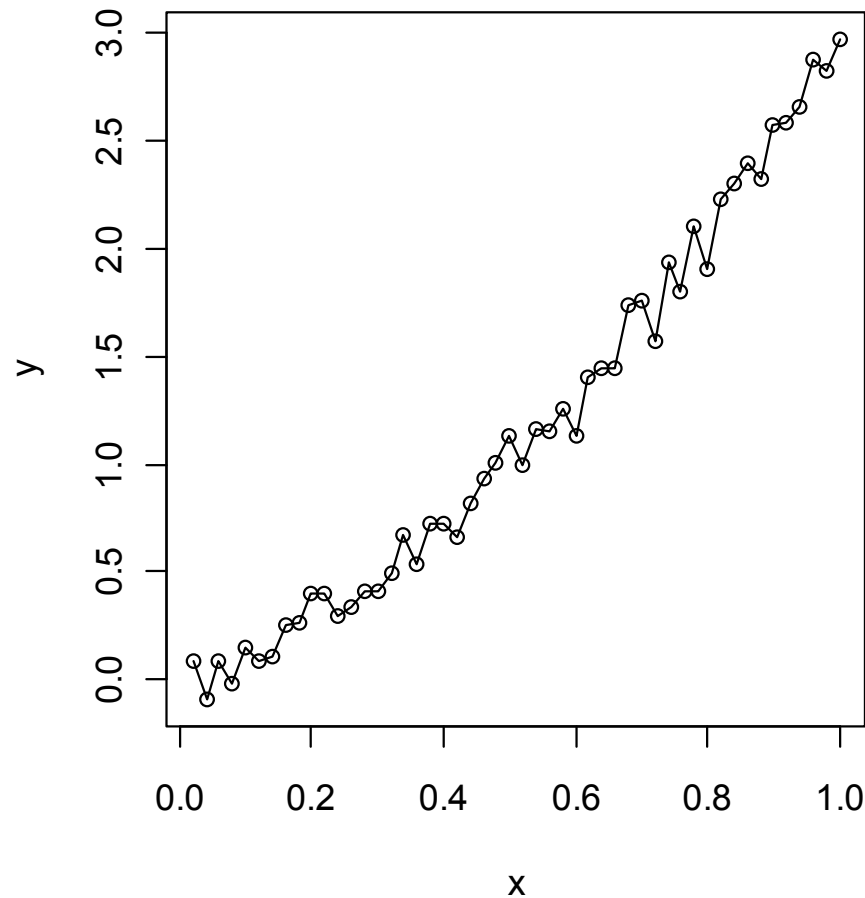
- 1) LS
- 2) GLS

Applied Time Series Analysis

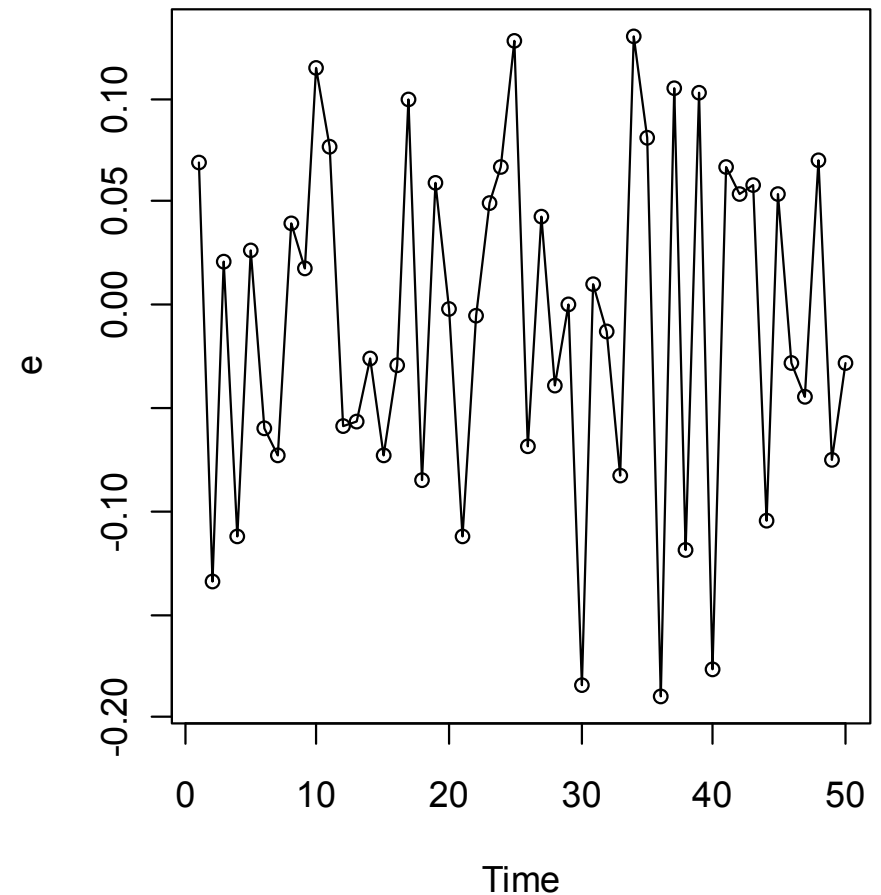
SS 2015 – Time Series Regression

Simulation Study: Series

Series Y_t



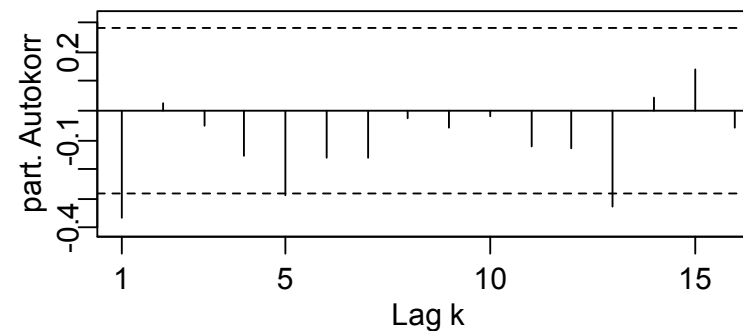
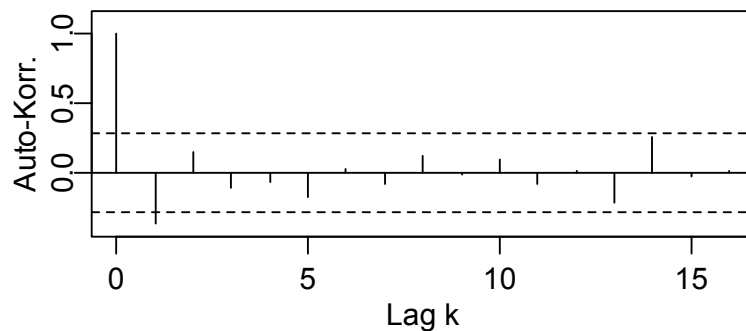
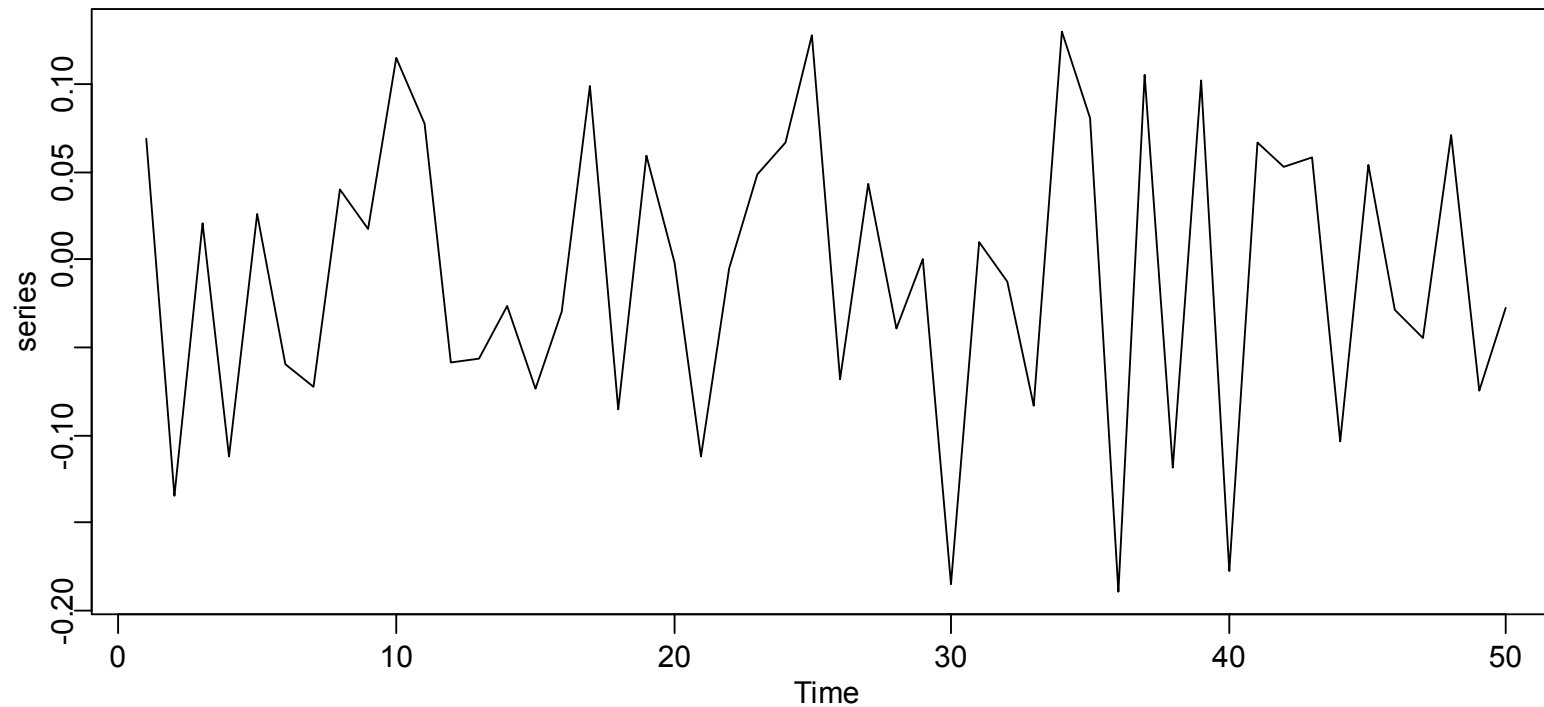
Series E_t



Applied Time Series Analysis

SS 2015 – Time Series Regression

Simulation Study: ACF of the Error Term

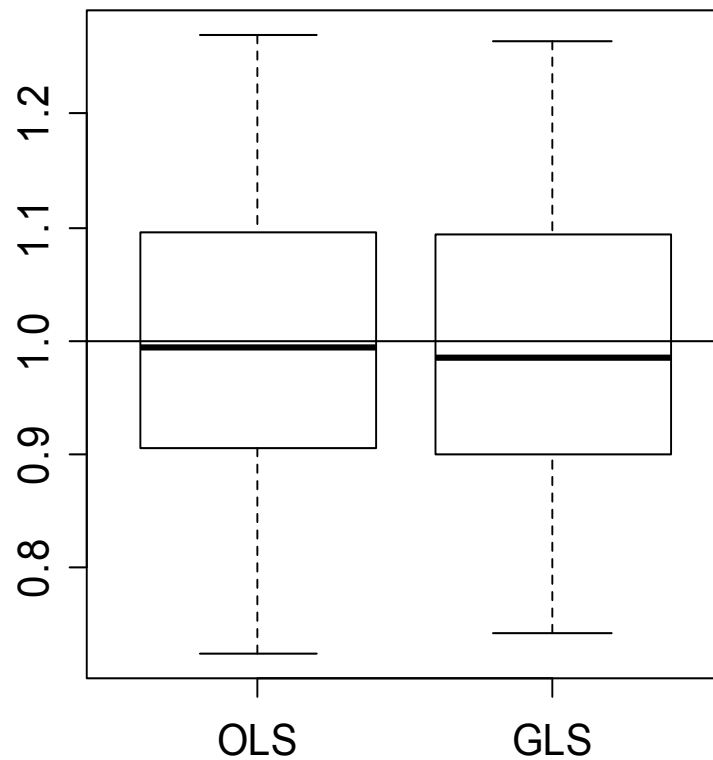


Applied Time Series Analysis

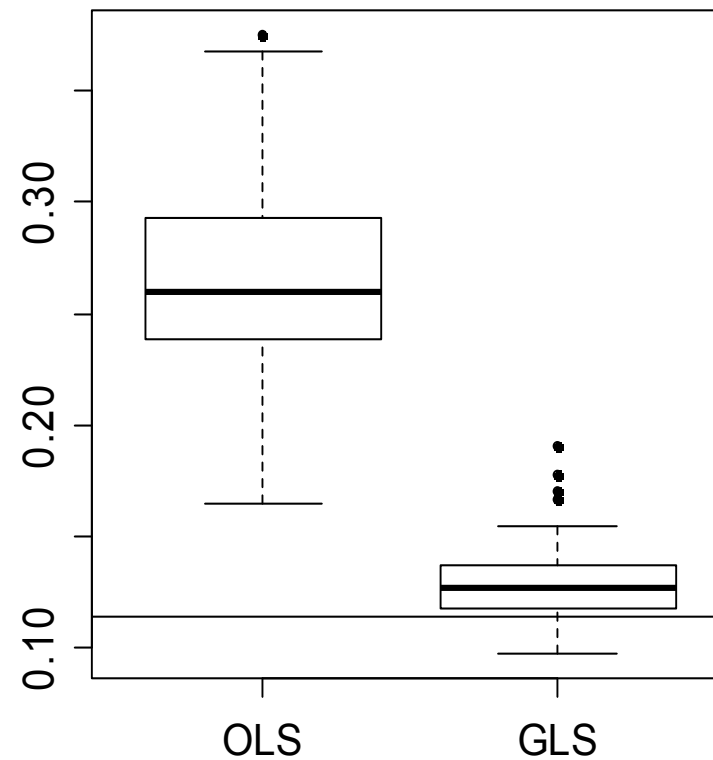
SS 2015 – Time Series Regression

Simulation Study: Results

Coefficient



Standard Error



Applied Time Series Analysis

SS 2015 – Time Series Regression

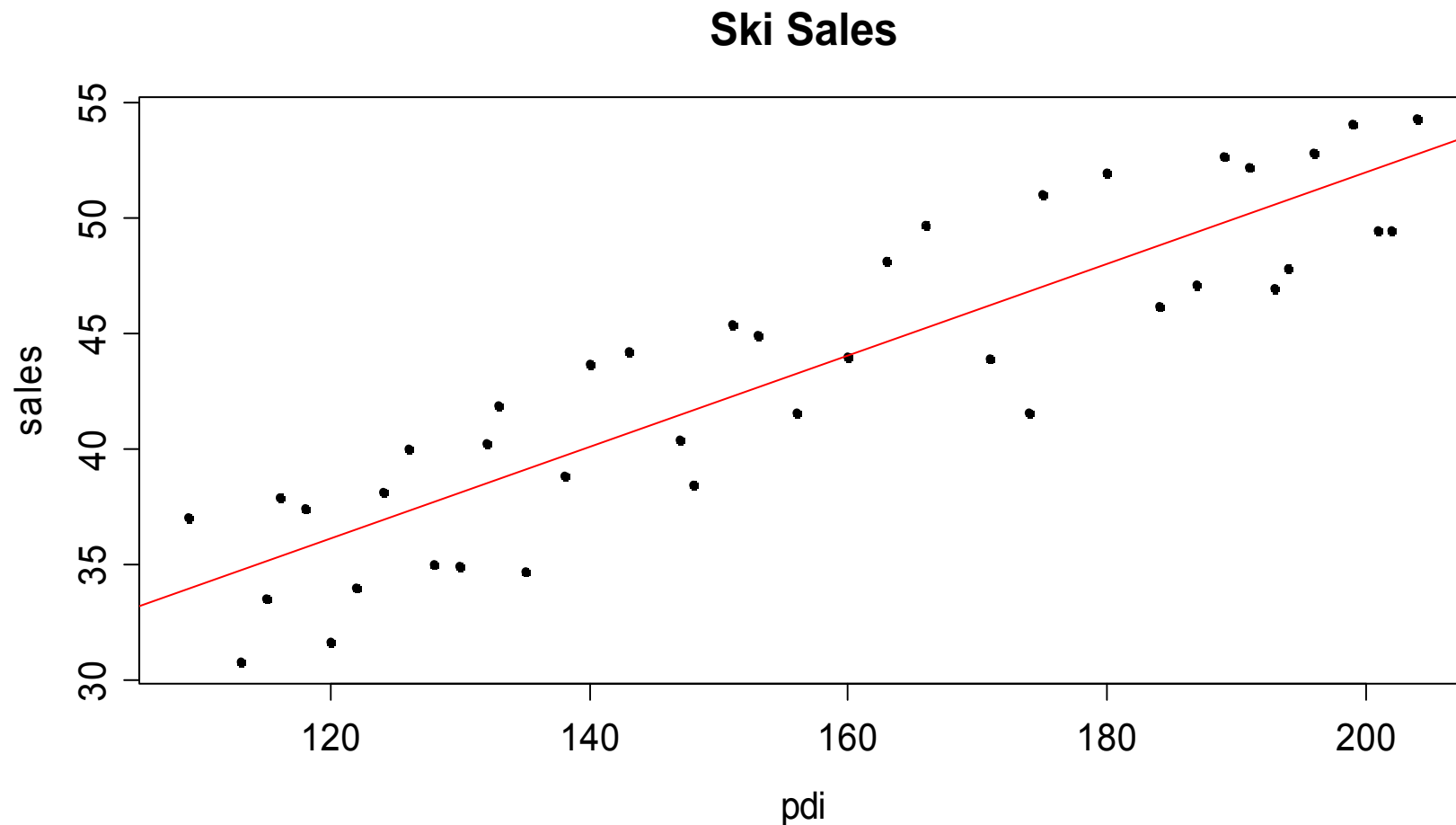
Missing Input Variables

- (Auto-)correlated errors are often caused by the non-presence of crucial input variables.
 - In this case, it is much better to identify the not-yet-present variables and include them in the analysis.
 - However, this isn't always possible.
- **regression with correlated errors can be seen as a sort of emergency kit for the case where the non-present variables cannot be added.**

Applied Time Series Analysis

SS 2015 – Time Series Regression

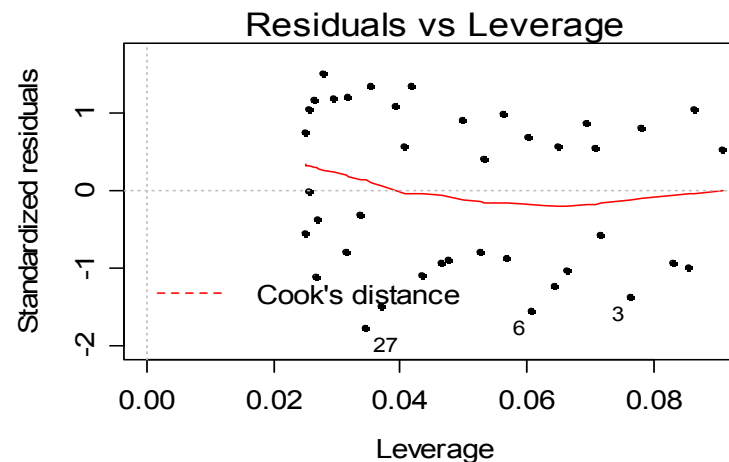
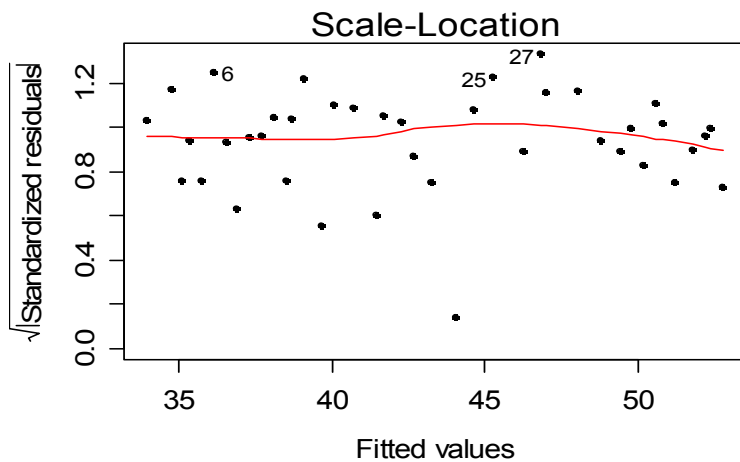
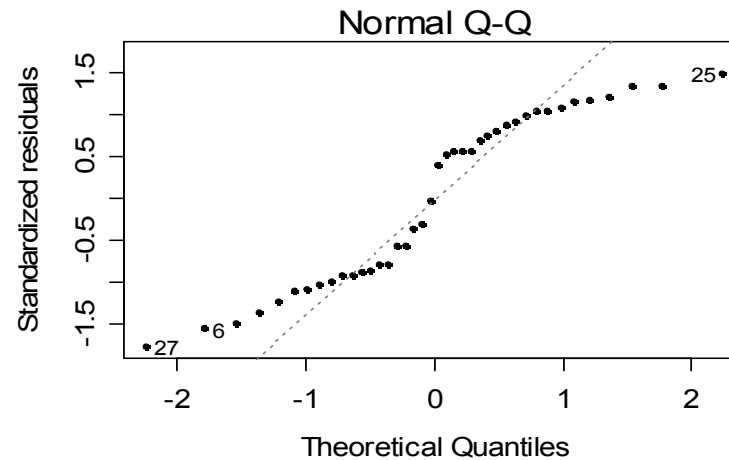
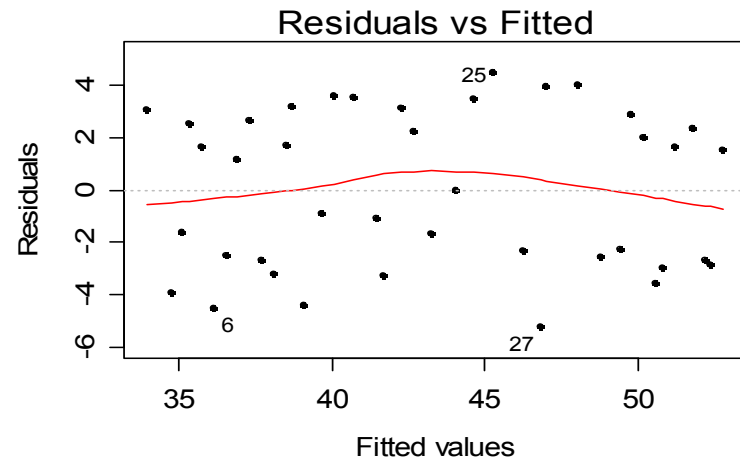
Example: Ski Sales



Applied Time Series Analysis

SS 2015 – Time Series Regression

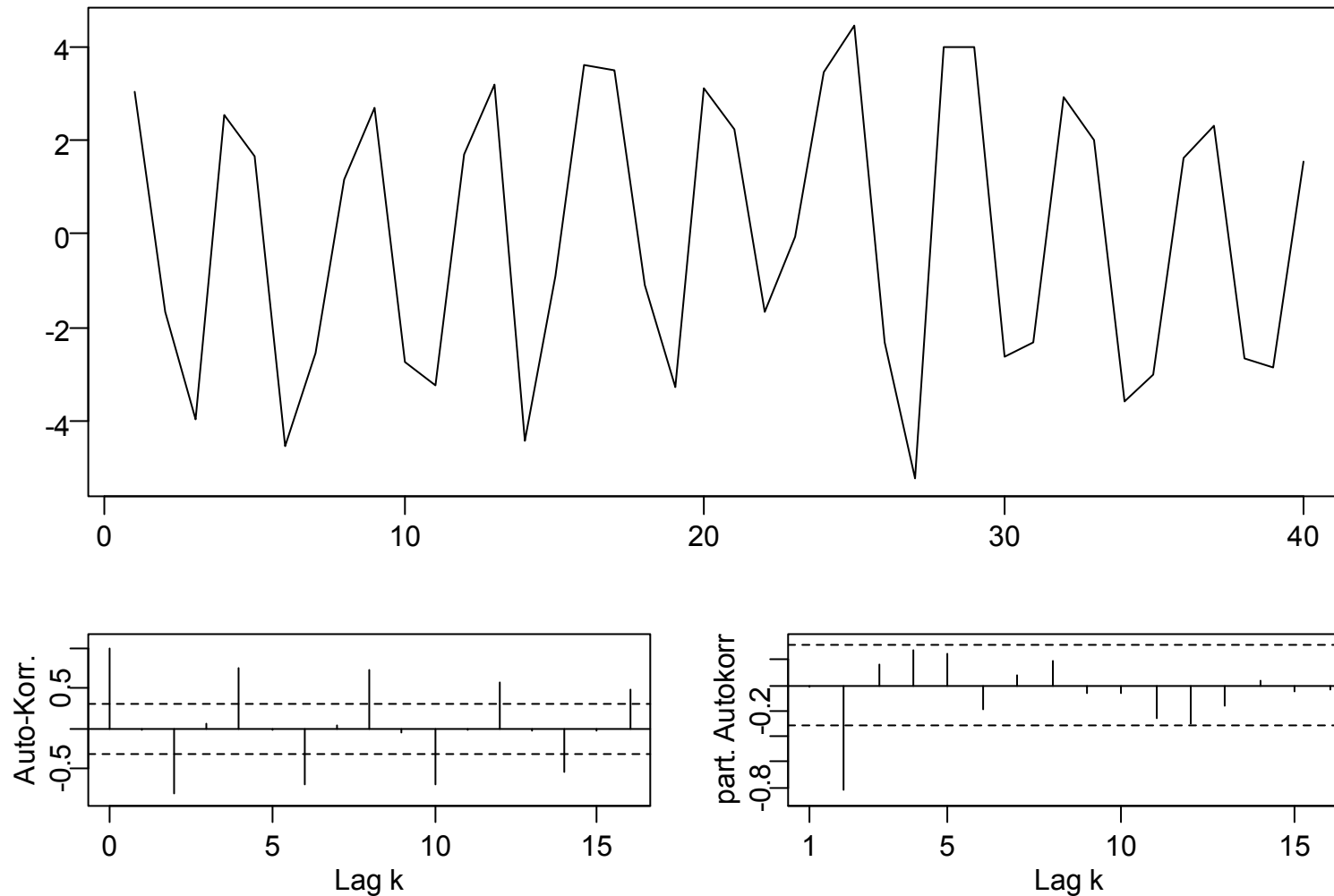
Ski Sales: Residual Diagnostics



Applied Time Series Analysis

SS 2015 – Time Series Regression

Ski Sales: ACF/PACF of Residuals

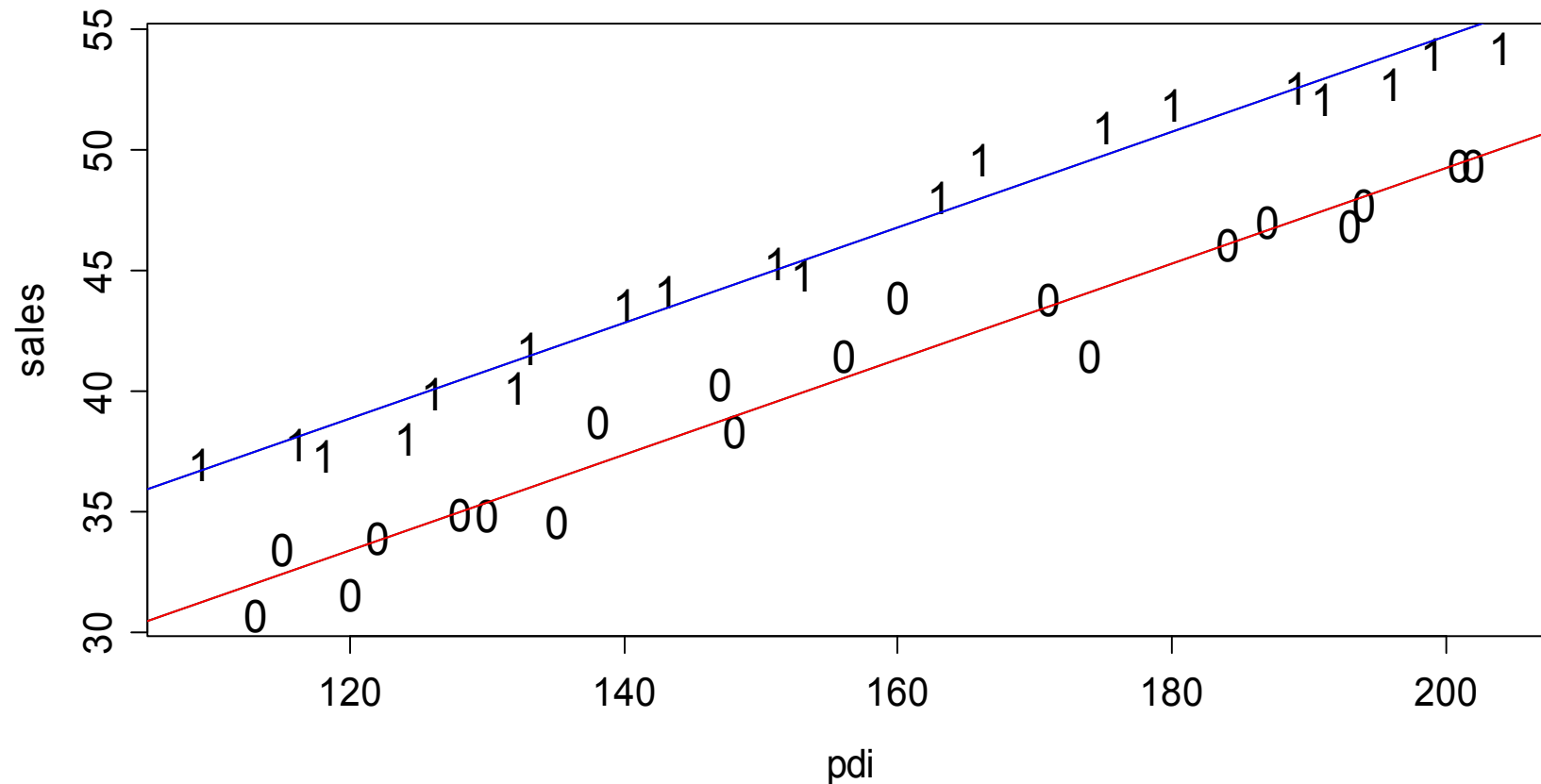


Applied Time Series Analysis

SS 2015 – Time Series Regression

Ski Sales: Model with Seasonal Factor

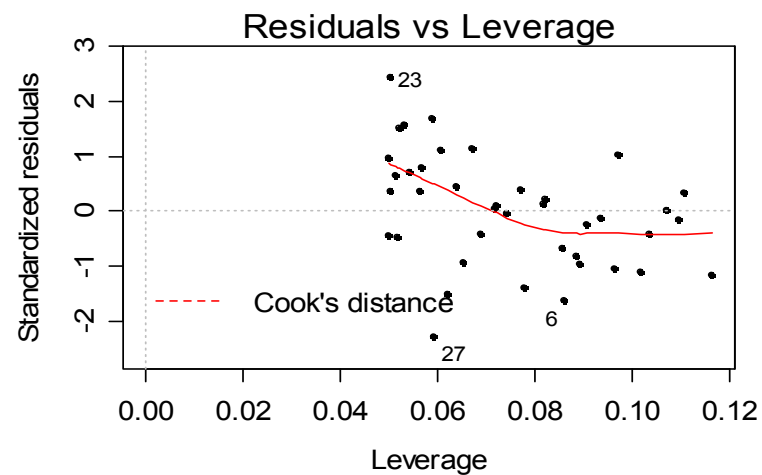
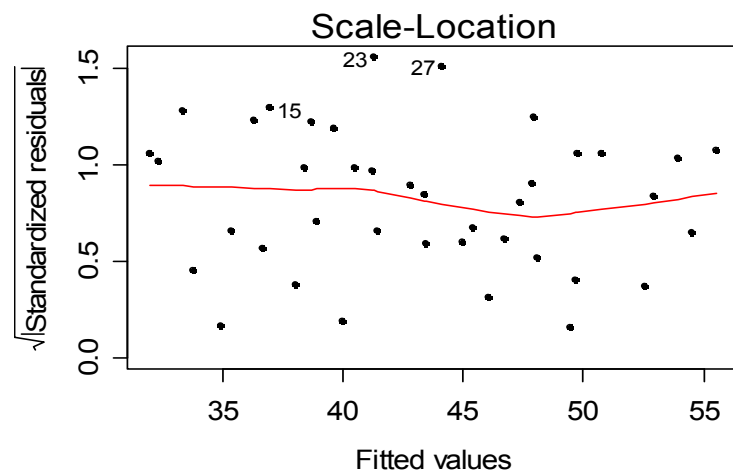
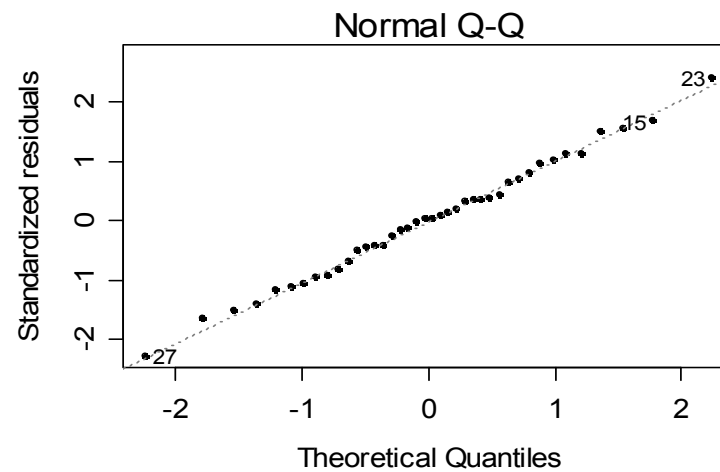
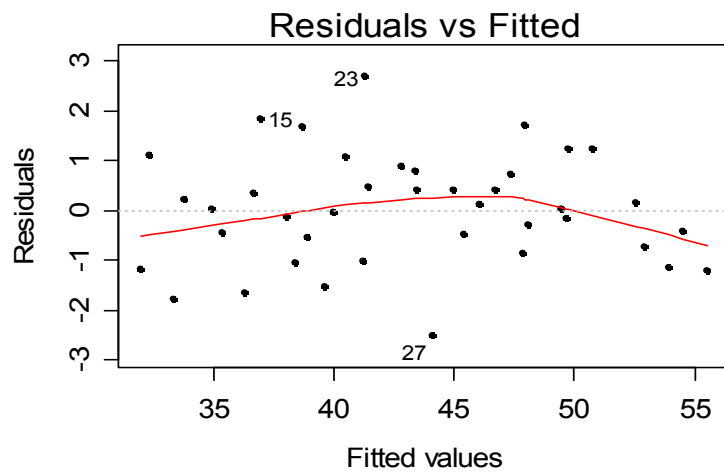
Ski Sales - Winter=1, Summer=0



Applied Time Series Analysis

SS 2015 – Time Series Regression

Residuals from Seasonal Factor Model

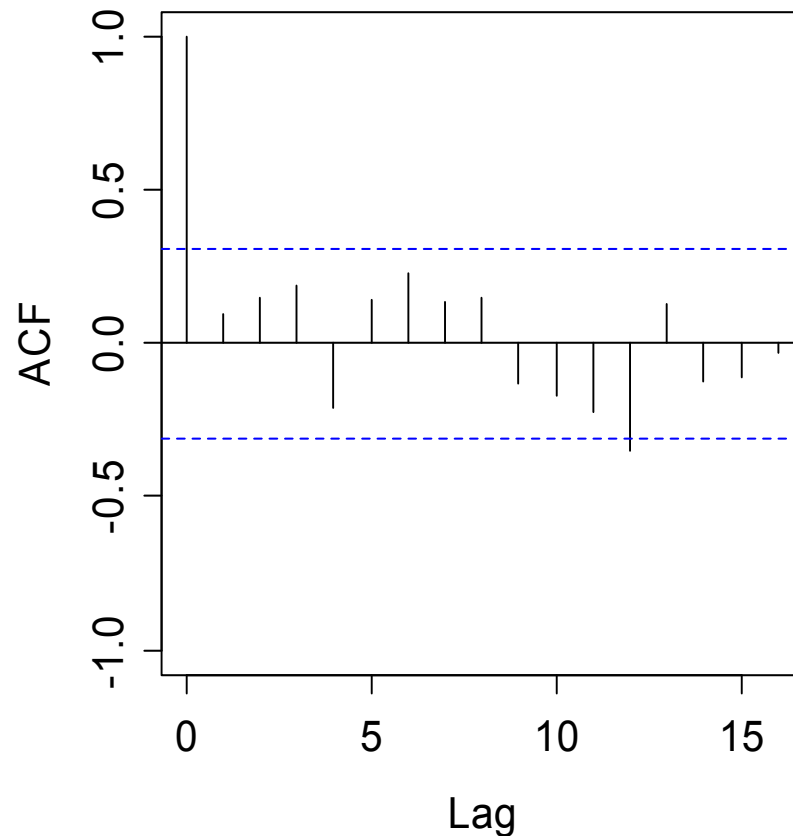


Applied Time Series Analysis

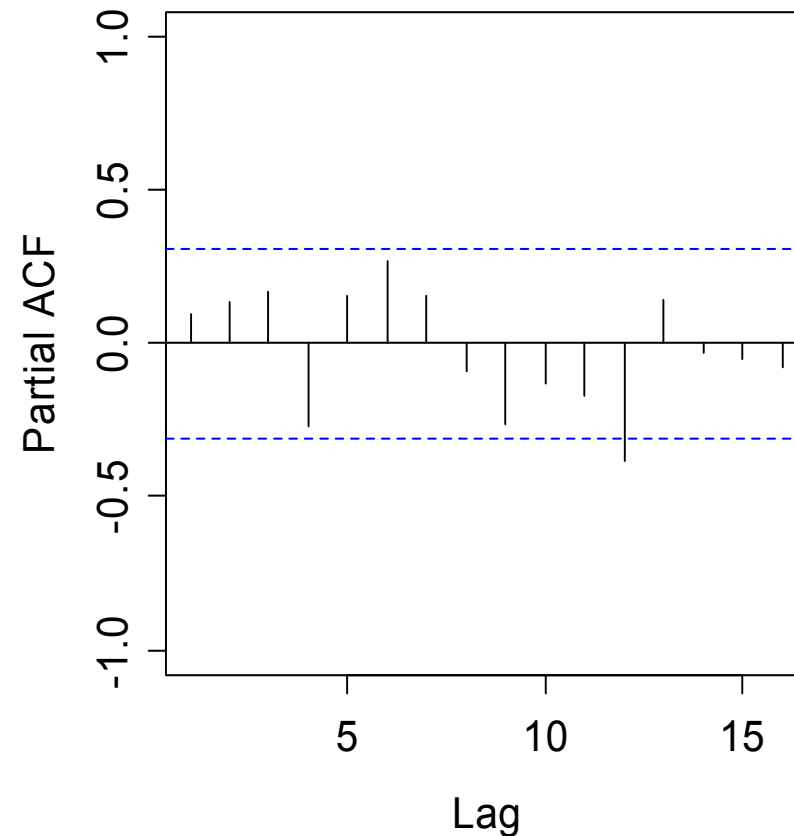
SS 2015 – Time Series Regression

Residuals from Seasonal Factor Model

ACF of Extended Model



PACF of Extended Model



Applied Time Series Analysis

SS 2015 – Time Series Regression

Ski Sales: Summary

- the first model (sales vs. PDI) showed correlated errors
 - the Durbin-Watson test failed to indicate this correlation
 - this apparent correlation is caused by omitting the season
 - adding the season removes all error correlation!
- ***the emergency kit „time series regression“ is, after careful modeling, not even necessary in this example. This is quite often the case!***

Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

Non-Stationary Models: ARIMA and SARIMA

Why?

We have seen that many time series we encounter in practice show trends and/or seasonality. While we could decompose them and model the stationary part, it might also be attractive to directly model a non-stationary series.

How does it work?

There is a mechanism, "the integration" or "the seasonal integration" which takes care of the deterministic features, while the remainder is modeled using an ARMA(p,q).

There are some peculiarities!

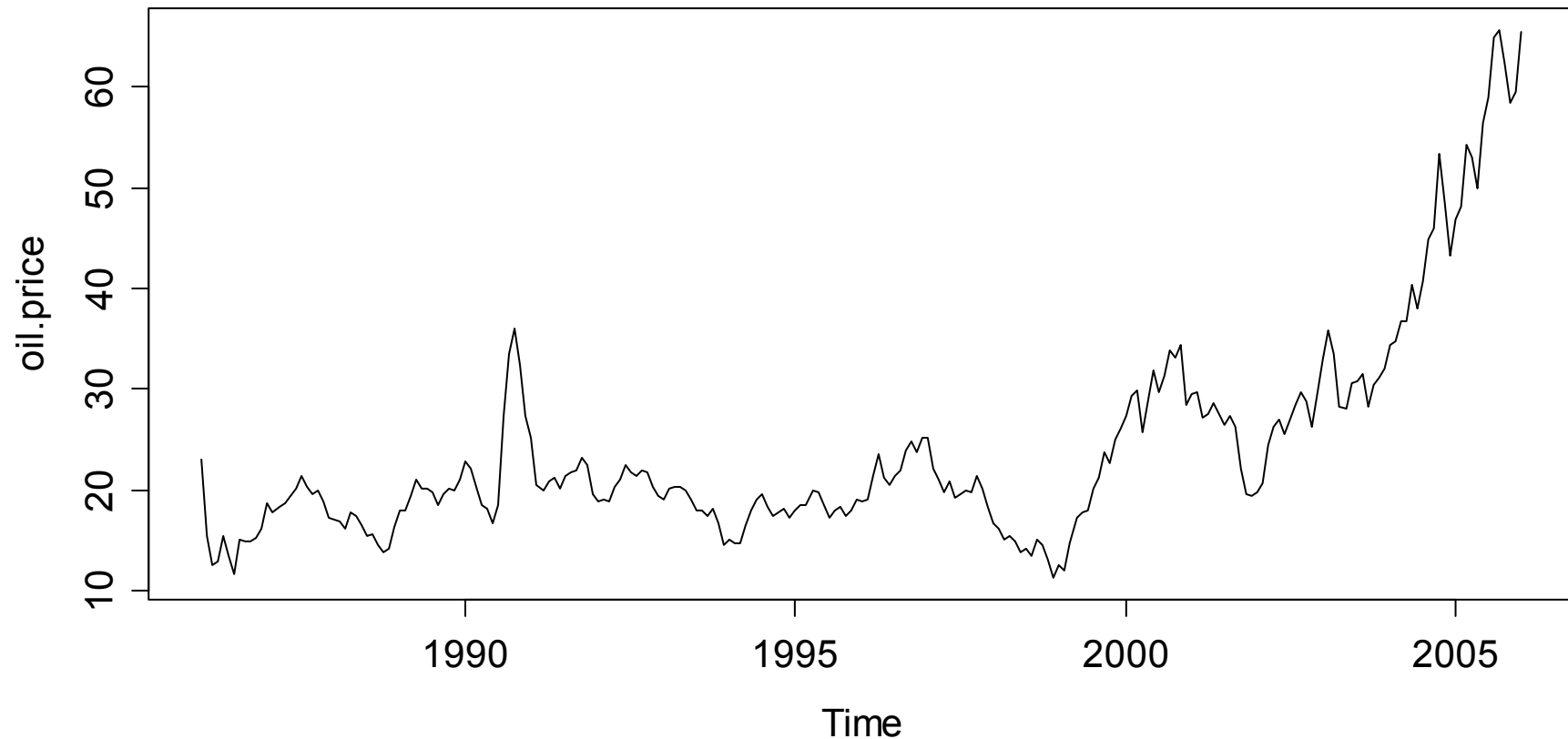
→ **see blackboard!**

Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

Example: Monthly Oil Prices

Monthly Price for a Barrel of Crude Oil

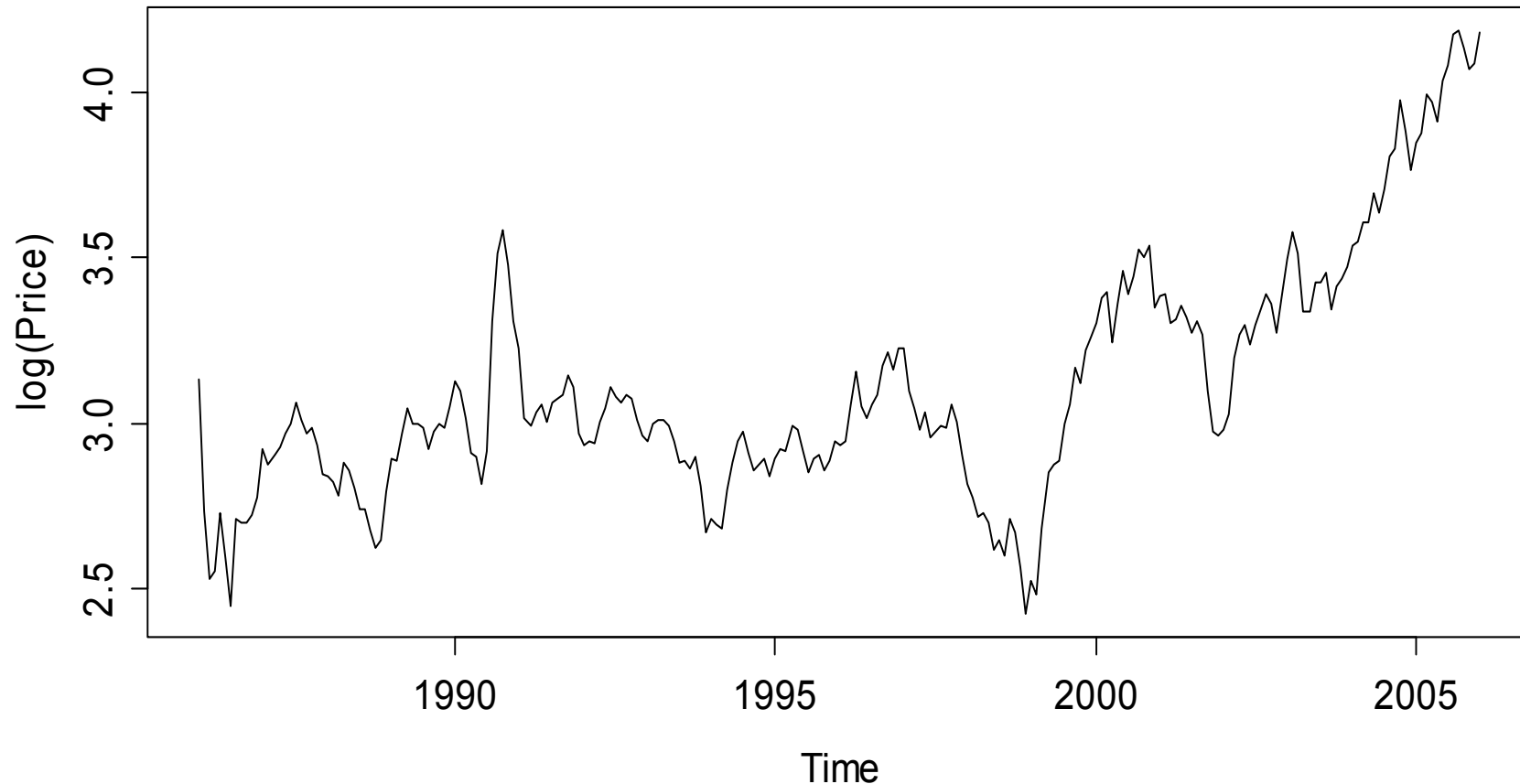


Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

Taking the Logarithm is Key

Logged Monthly Price for a Crude Oil Barrel

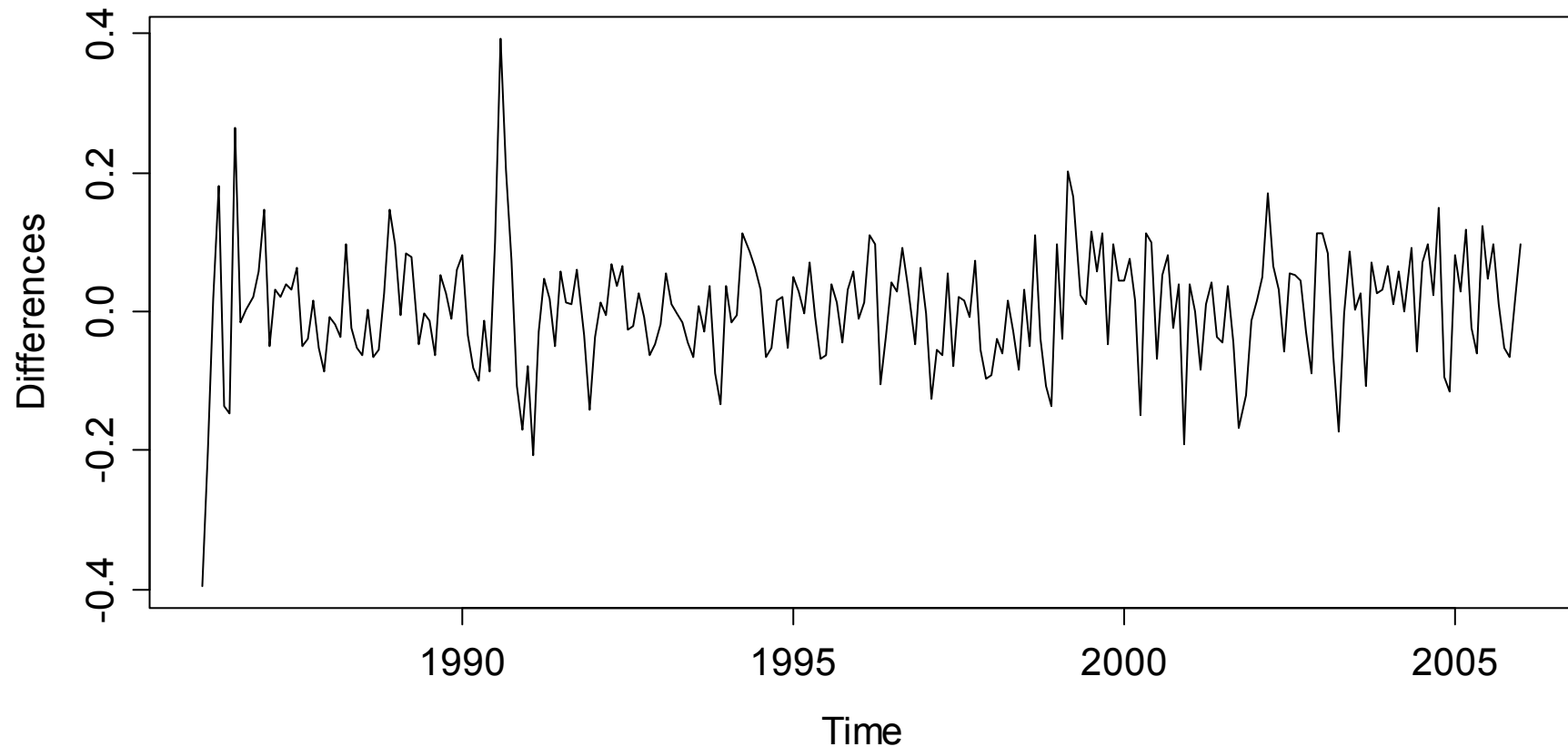


Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

Differencing Yields a Stationary Series

Differences of Logged Monthly Crude Oil Prices



Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

ARIMA(p,d,q)-Models

Idea: Fit an ARMA(p,q) to a time series where the dth order difference with lag 1 was taken before.

Example: If $Y_t = X_t - X_{t-1} = (1-B)X_t \sim ARMA(p, q)$,
then $X_t \sim ARIMA(p, 1, q)$

Notation: With backshift-operator B()

$$\Phi(B)(1-B)^d X_t = \Theta(B)E_t$$

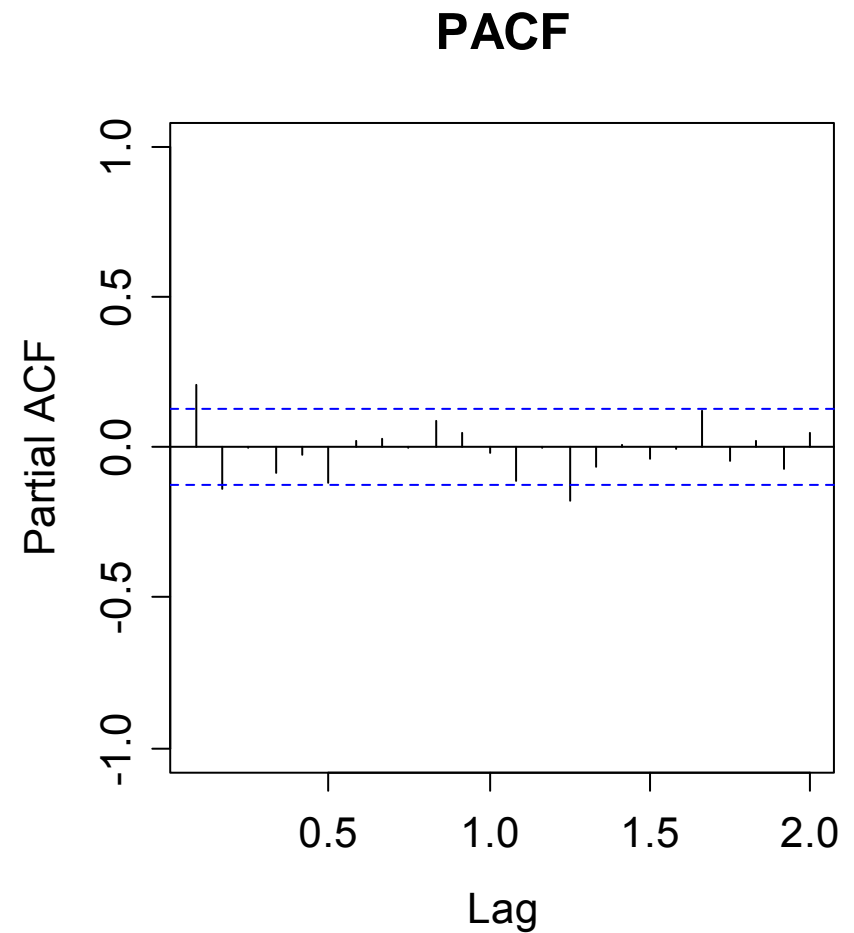
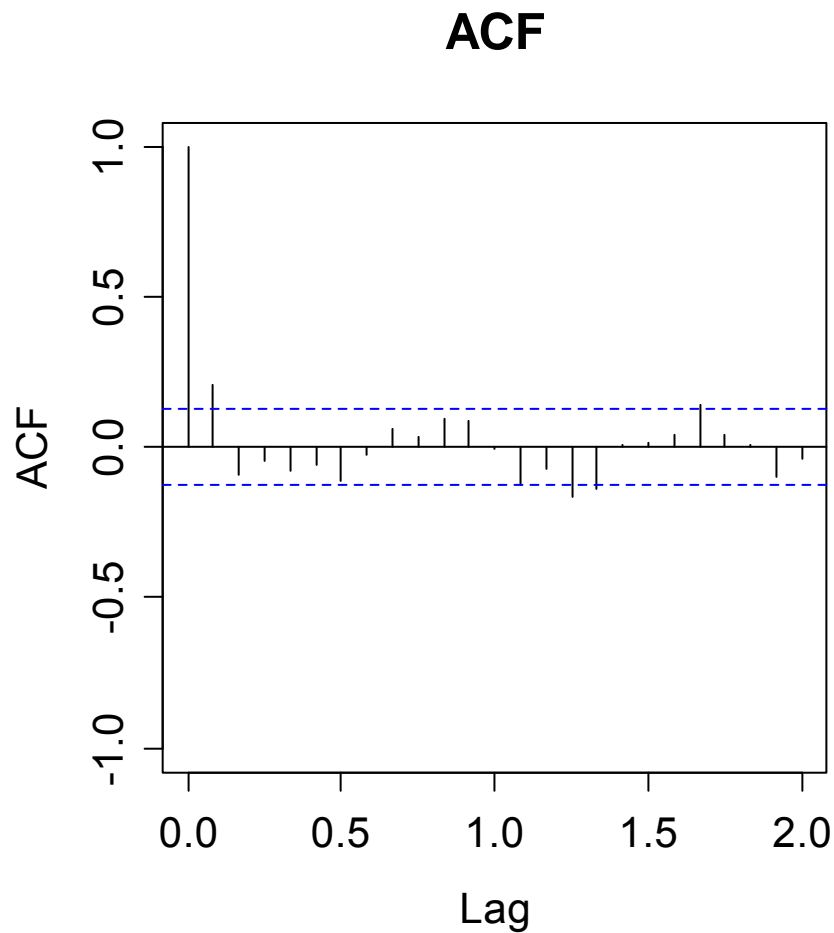
Stationarity: ARIMA-models are usually non-stationary!

Advantage: it's easier to forecast in R!

Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

ACF/PACF of the Differenced Series



Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

Fitting an ARIMA in R

We start by fitting an ARIMA(1,1,2) to the oil series:

```
> arima(lop, order=c(1,1,2))
```

Call:

```
arima(x = lop, order = c(1, 1, 2))
```

Coefficients:

	ar1	ma1	ma2
	0.8429	-0.5730	-0.3104
s.e.	0.1548	0.1594	0.0675

```
sigma^2 = 0.0066: ll = 261.88, aic = -515.75
```

Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

Alternative Fitting

Instead of fitting an ARIMA(1,1,2) to the logged oil series, we can also take the differenced log-oil series and fit an ARMA(1,2) to it.

IMPORTANT:

In this case, we have to do fitting without including an intercept (why?), thus:

```
> arima(diff(log(oil.price)), order=c(1,0,2),  
        include.mean=FALSE)
```

Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

Meaning of the Model / Recipe

We can rewrite the ARIMA(1,1,2) model as an ARMA(2,2),
see blackboard...

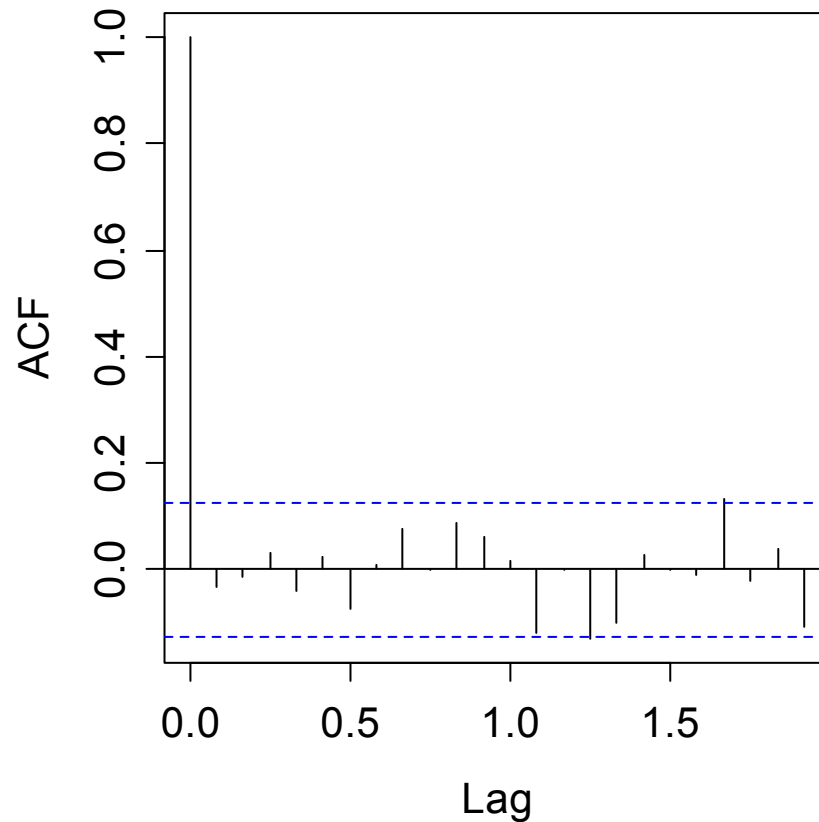
Some guidelines on how to fit ARIMA models to observed time series can also be found **on the blackboard...**

Applied Time Series Analysis

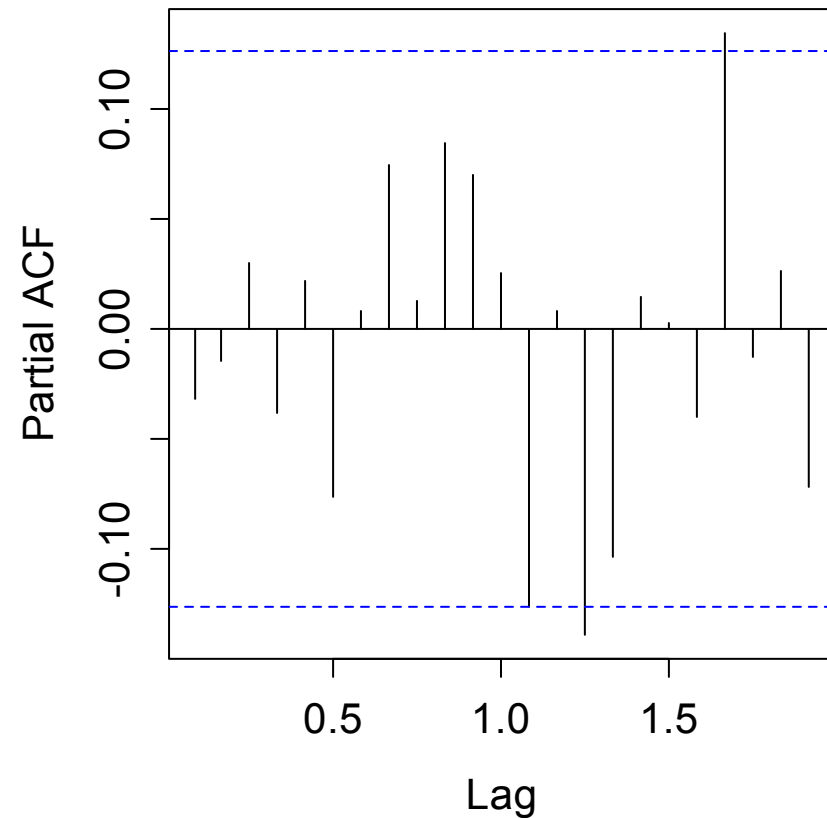
SS 2015 – ARIMA, SARIMA & GARCH

Residual Analysis of the ARIMA(1,1,2)

ACF



PACF



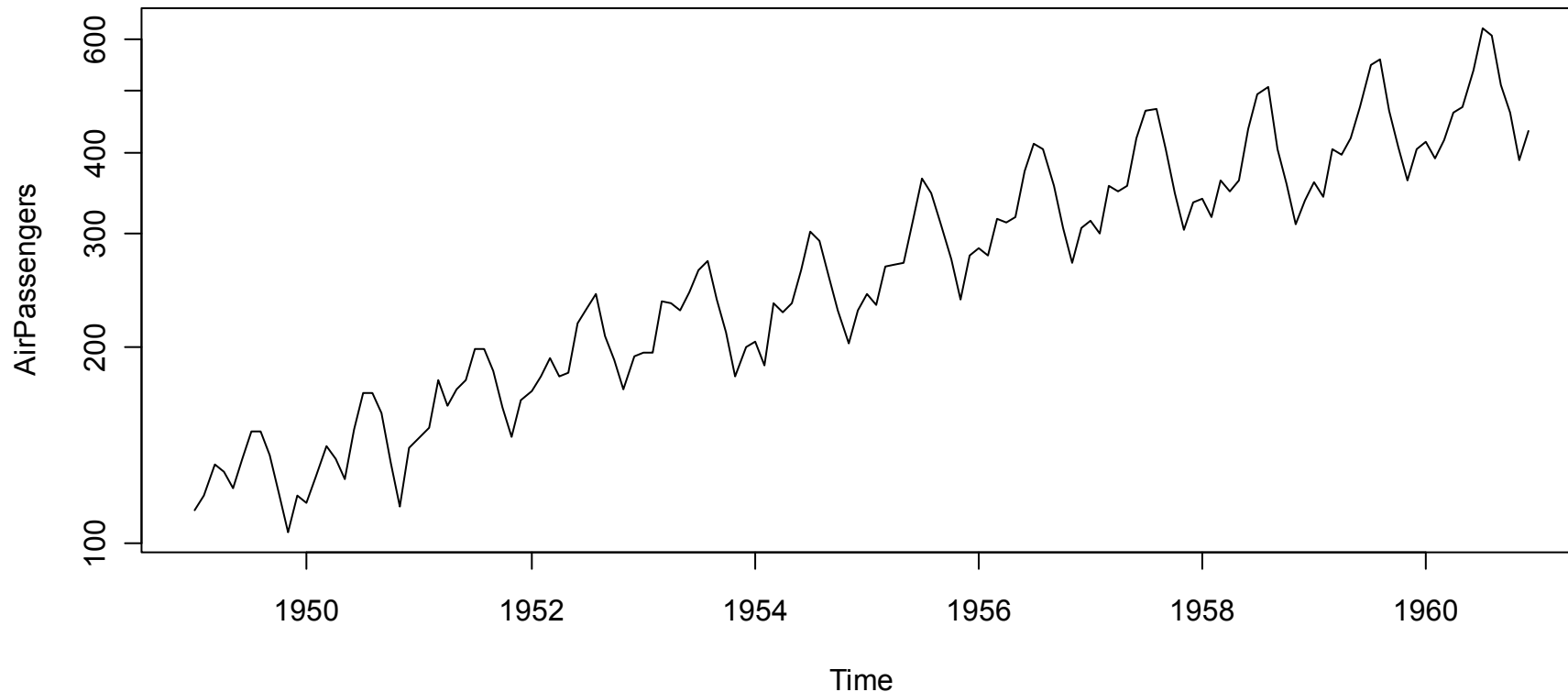
Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

SARIMA(p,d,q)(P,D,Q)^s

= a.k.a. Airline Model. We are looking at the log-trsf. airline data

Log-Transformed Airline Data

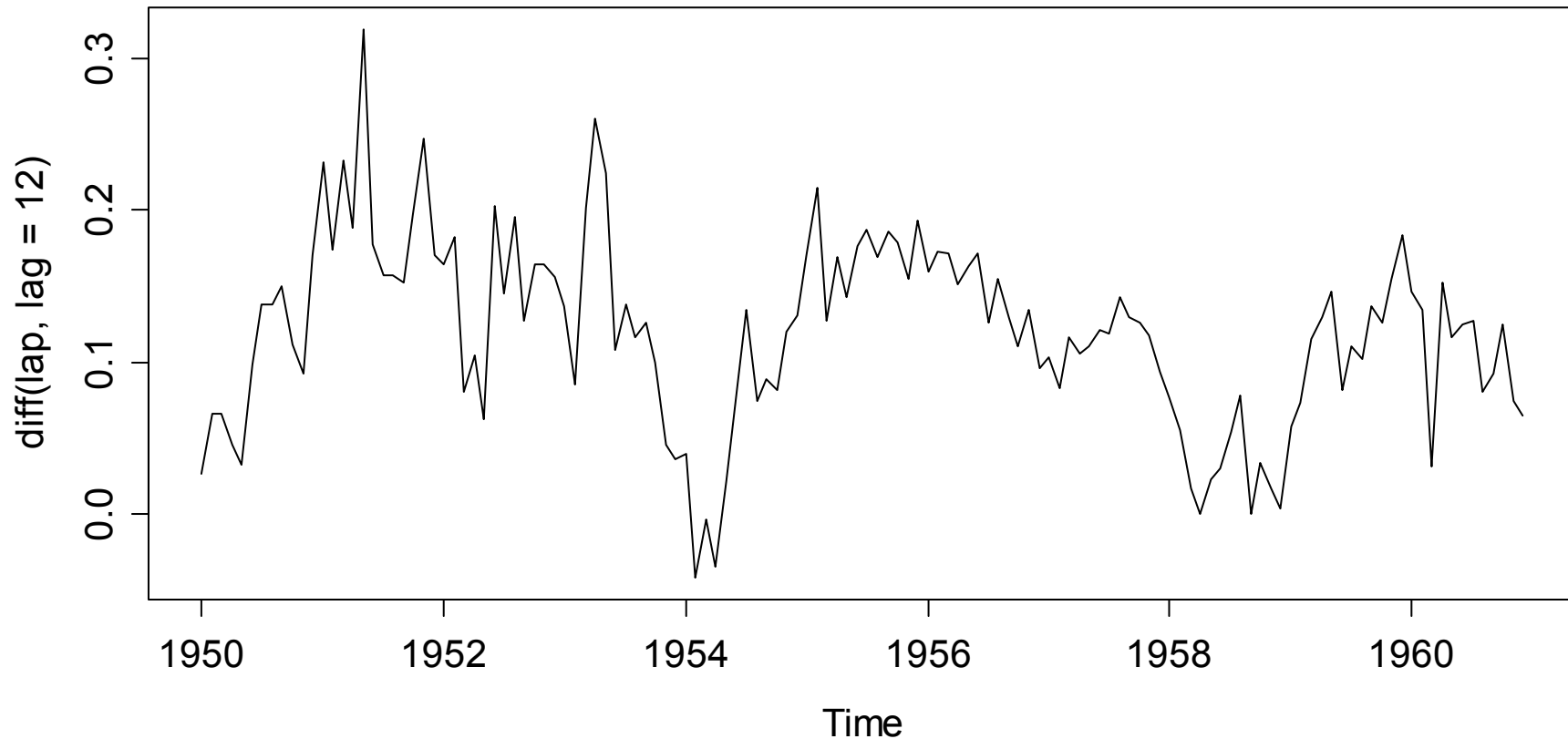


Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

Seasonal Differencing Helps...

Seasonally Differenced Airline Passenger Series

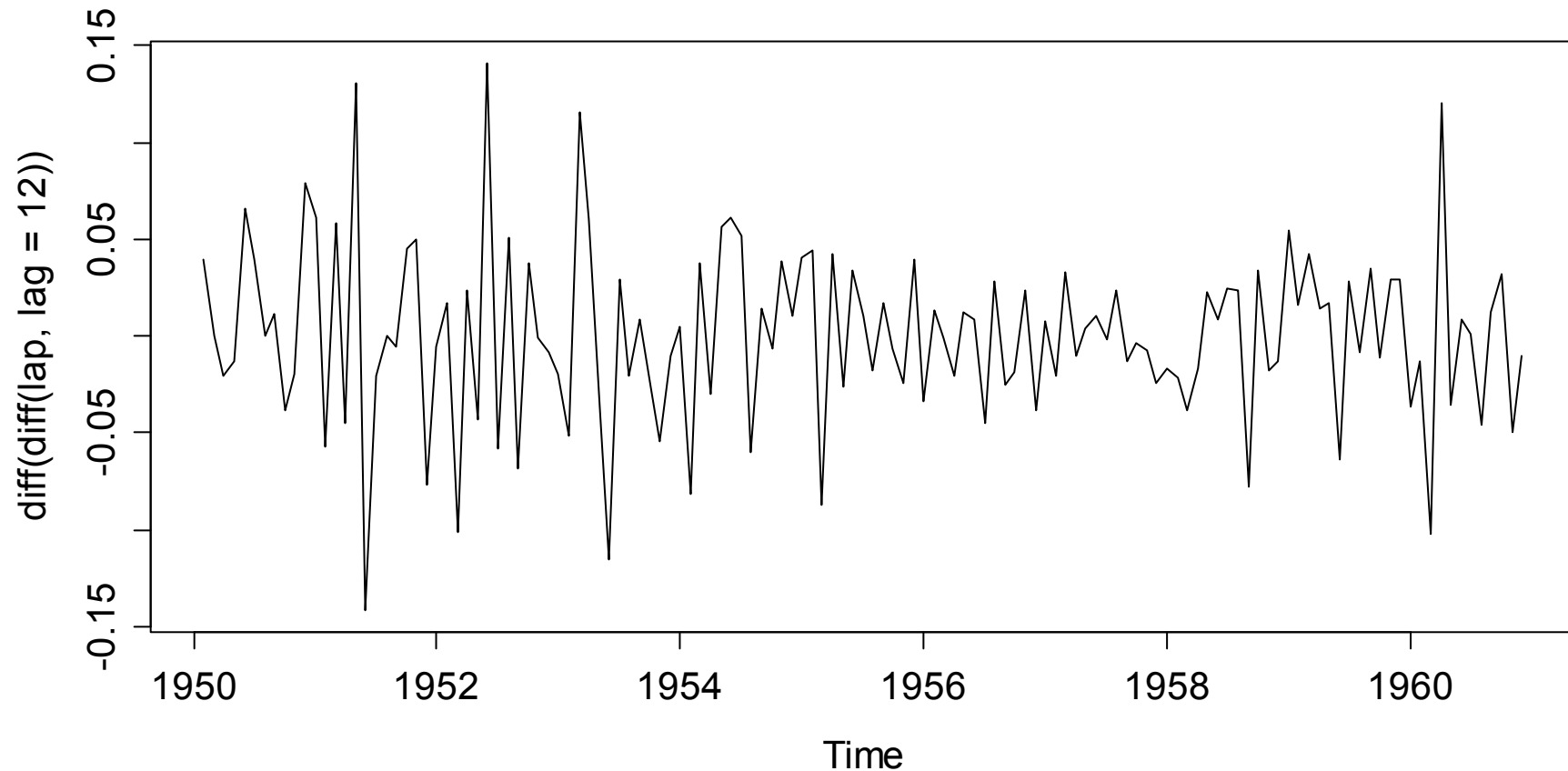


Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

... But More Is Needed!

Differenced Seasonally Differenced Airline Passenger Series



Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

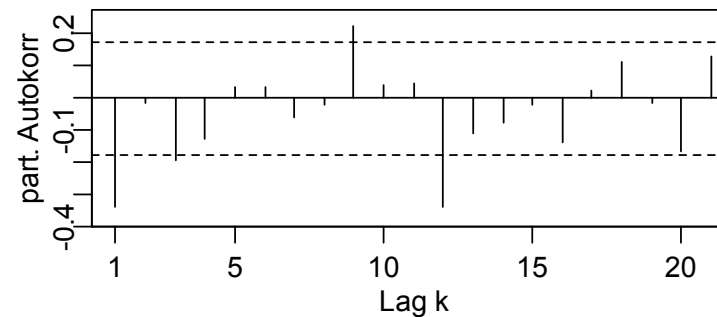
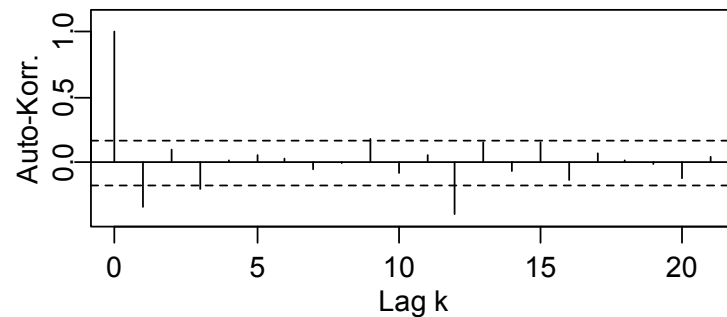
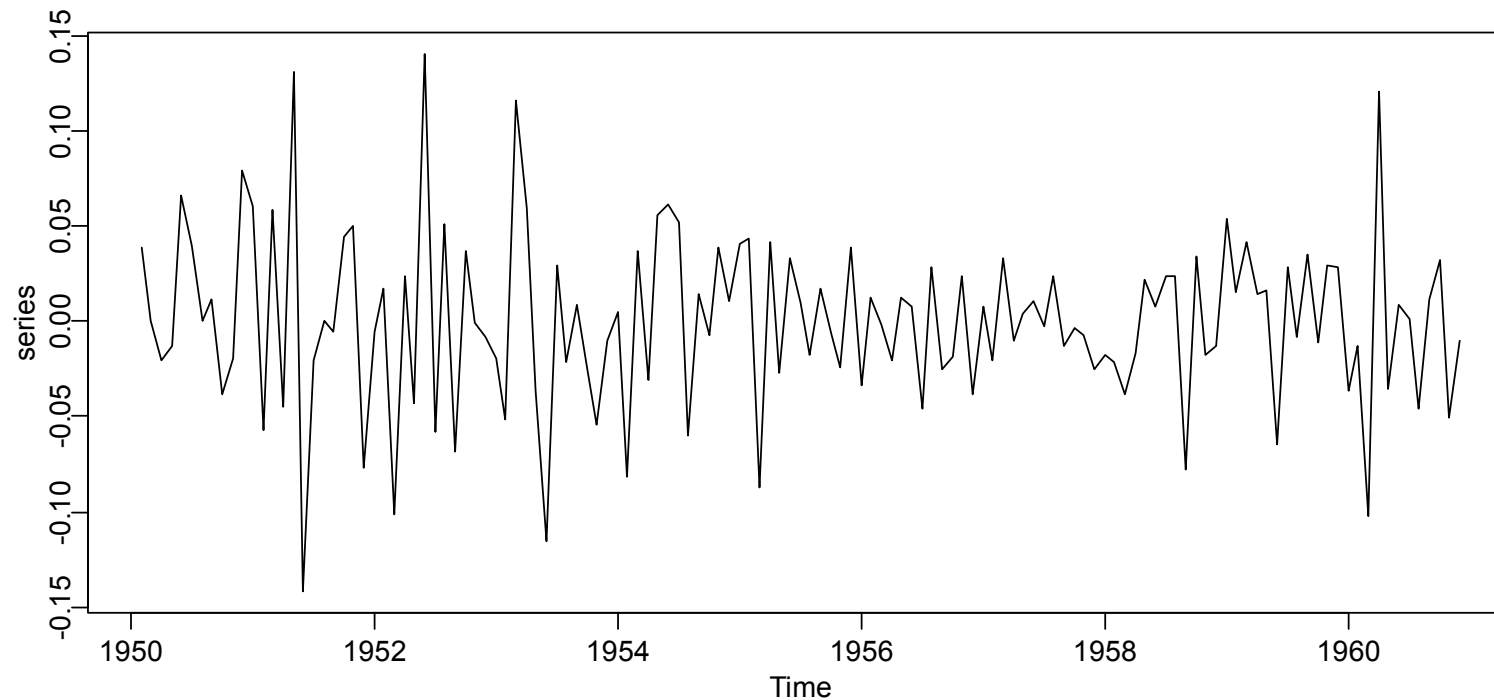
SARIMA(p,d,q)(P,D,Q)^s

We perform some differencing... (→ [see blackboard](#))

Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

ACF/PACF of SARIMA(p,d,q)(P,D,Q)^s



Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

Modeling the Airline Data

Since there are “big gaps” in ACF/PACF:

$$\begin{aligned}Z_t &= (1 + \beta_1 B)(1 + \gamma_1 B^{12})E_t \\ &= E_t + \beta_1 E_{t-1} + \gamma_1 E_{t-12} + \beta_1 \gamma_1 E_{t-13}\end{aligned}$$

This is an MA(13)-model with many coefficients equal to 0, or equivalently, a SARIMA(0,1,1)(0,1,1)¹².

Note: Every SARIMA(p,d,q)(P,D,Q)^s can be written as an ARMA(p+sP,q+sQ), where many coefficients will be equal to 0.

Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

SARIMA(p,d,q)(P,D,Q)^s

The general notation is:

$$Z_t = (1 - B)^d (1 - B^s)^D X_t$$

$$\Phi(B)\Phi_s(B^s)Z_t = \Theta(B)\Theta_s(B^s)E_t$$

Interpretation:

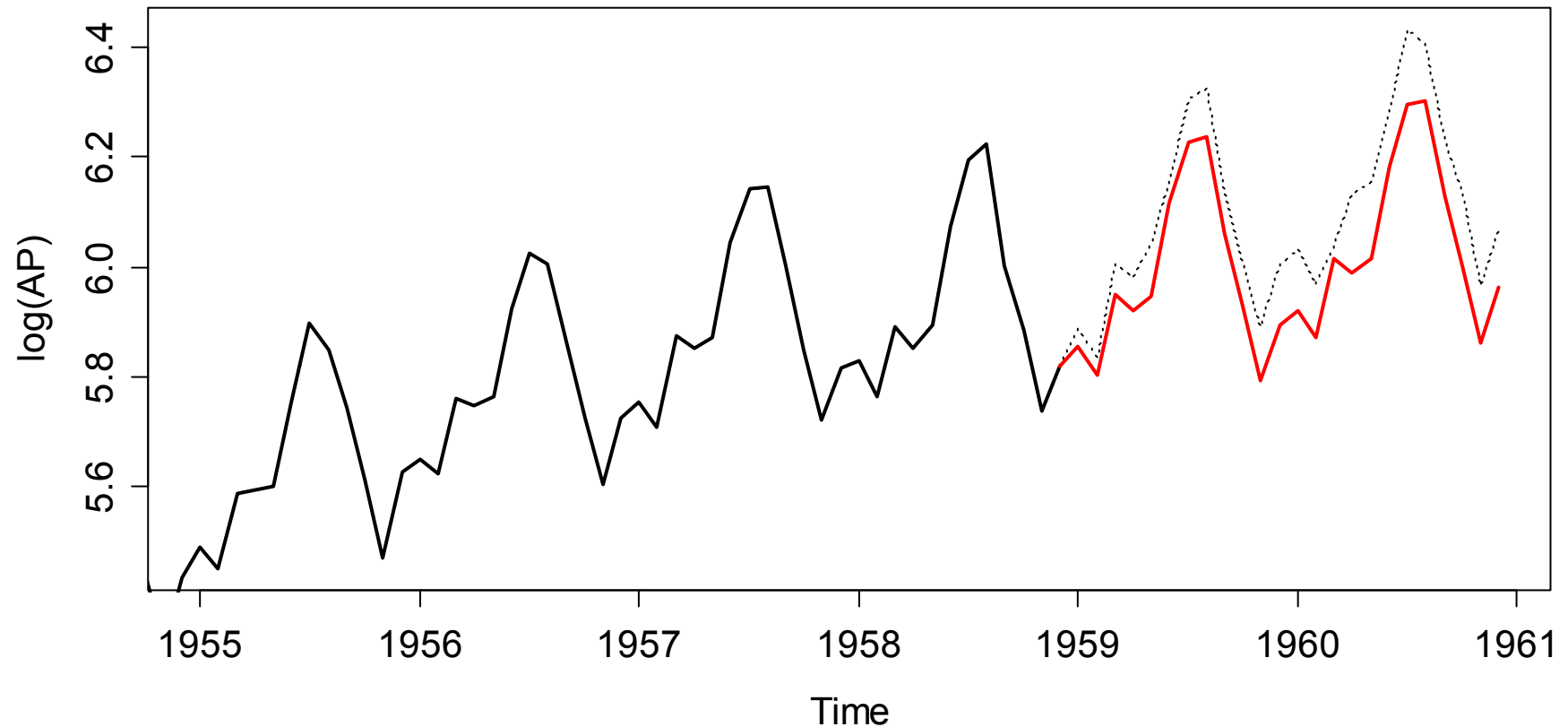
- one typically chooses $d=D=1$
 - s = periodicity in the data (season)
 - P, Q describe the dependency on multiples of the period
- **see blackboard...**

Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

Forecasting Airline Data

Forecast of $\log(\text{AP})$ with SARIMA(0,1,1)(0,1,1)

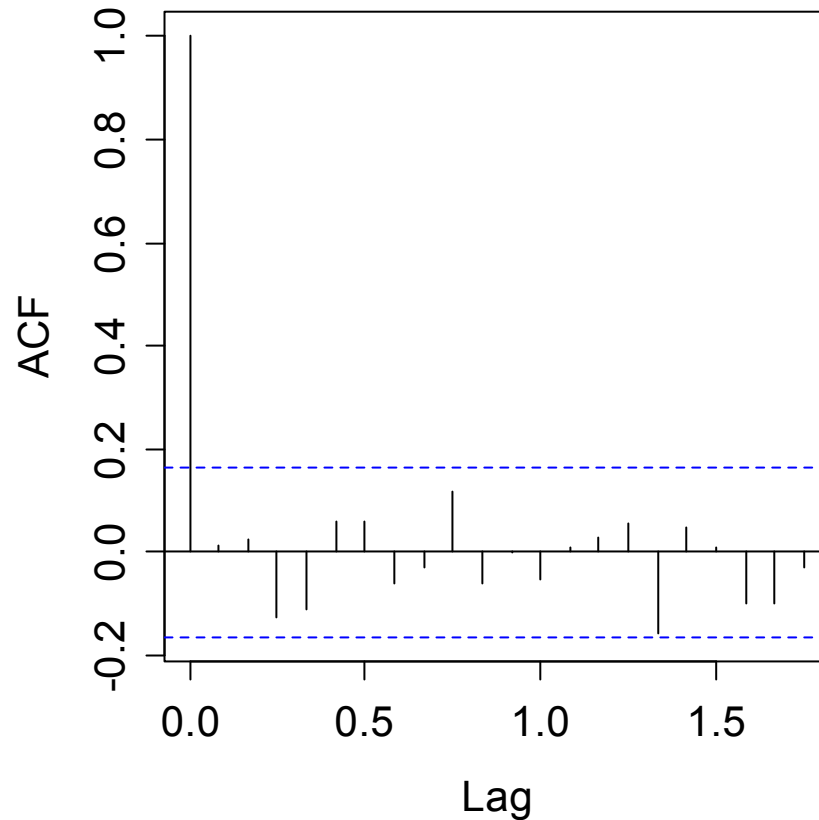


Applied Time Series Analysis

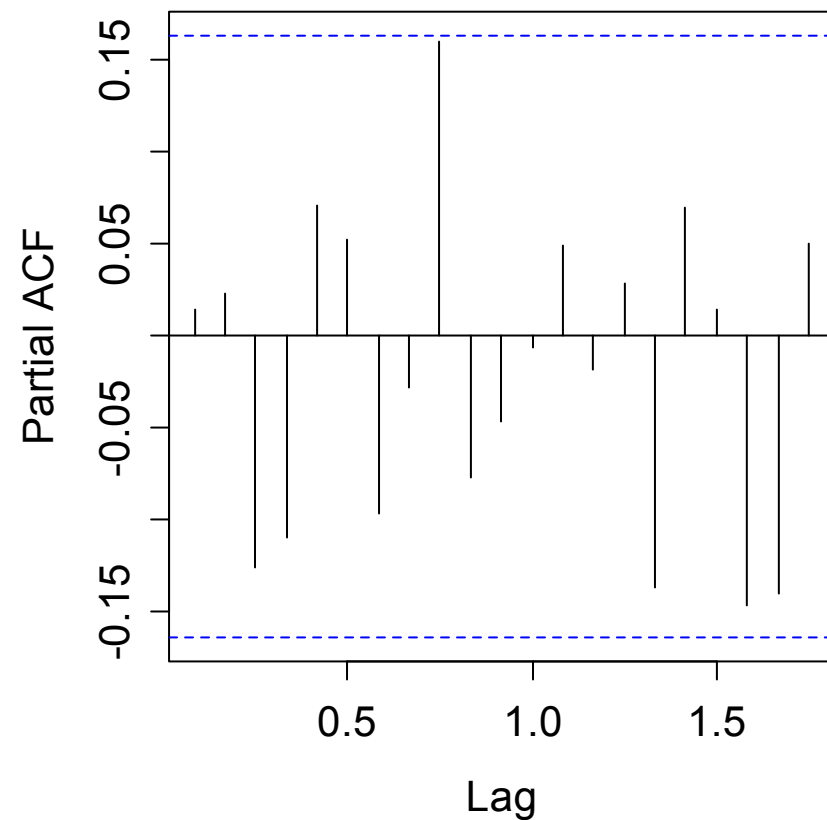
SS 2015 – ARIMA, SARIMA & GARCH

Residual Analysis of SARIMA(0,1,1)(0,1,1)

ACF



PACF



Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

Outlook to Non-Linear Models

What are linear models?

Models which can be written as a linear combination of X_t ,
i.e. all AR-, MA- and ARMA-models

What are non-linear models?

Everything else, e.g. non-linear combinations of X_t ,
terms like X_t^2 in the linear combination, and much more!

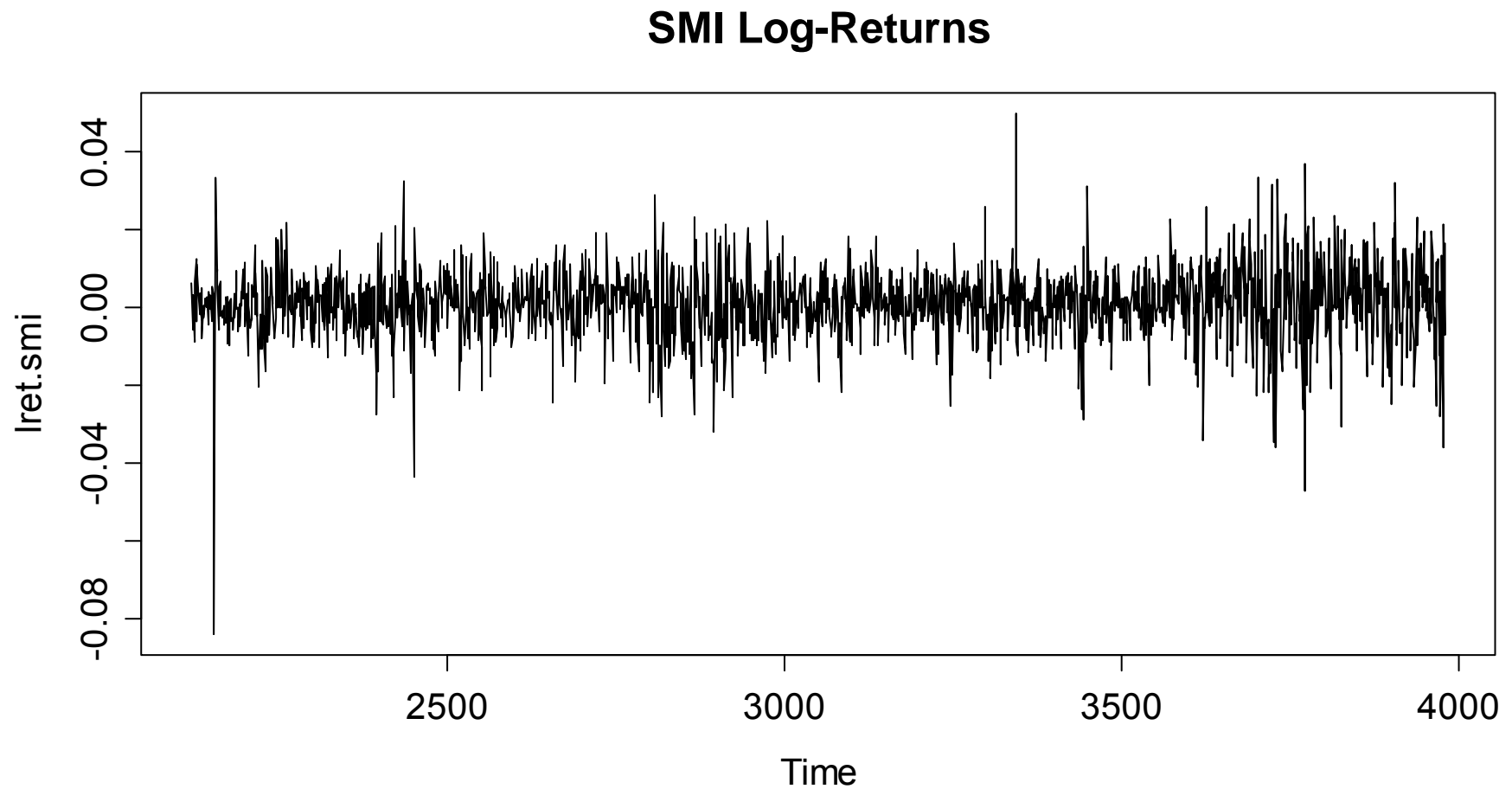
Motivation for non-linear models?

- modeling cyclic behavior with quicker increase than decrease
- non-constant variance, even after transforming the series

Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

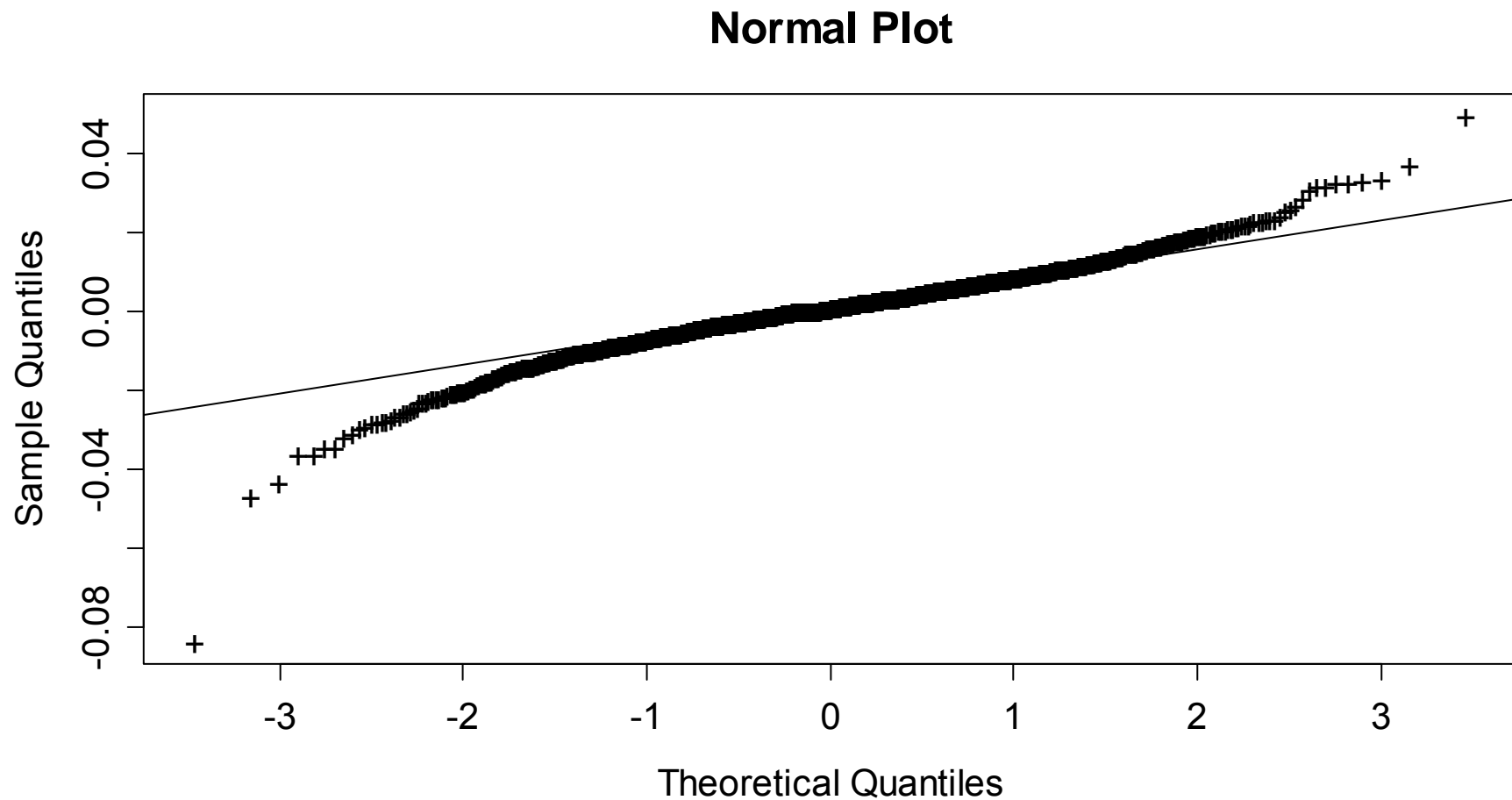
SMI Log-Returns



Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

Normal Plot of SMI Log-Returns

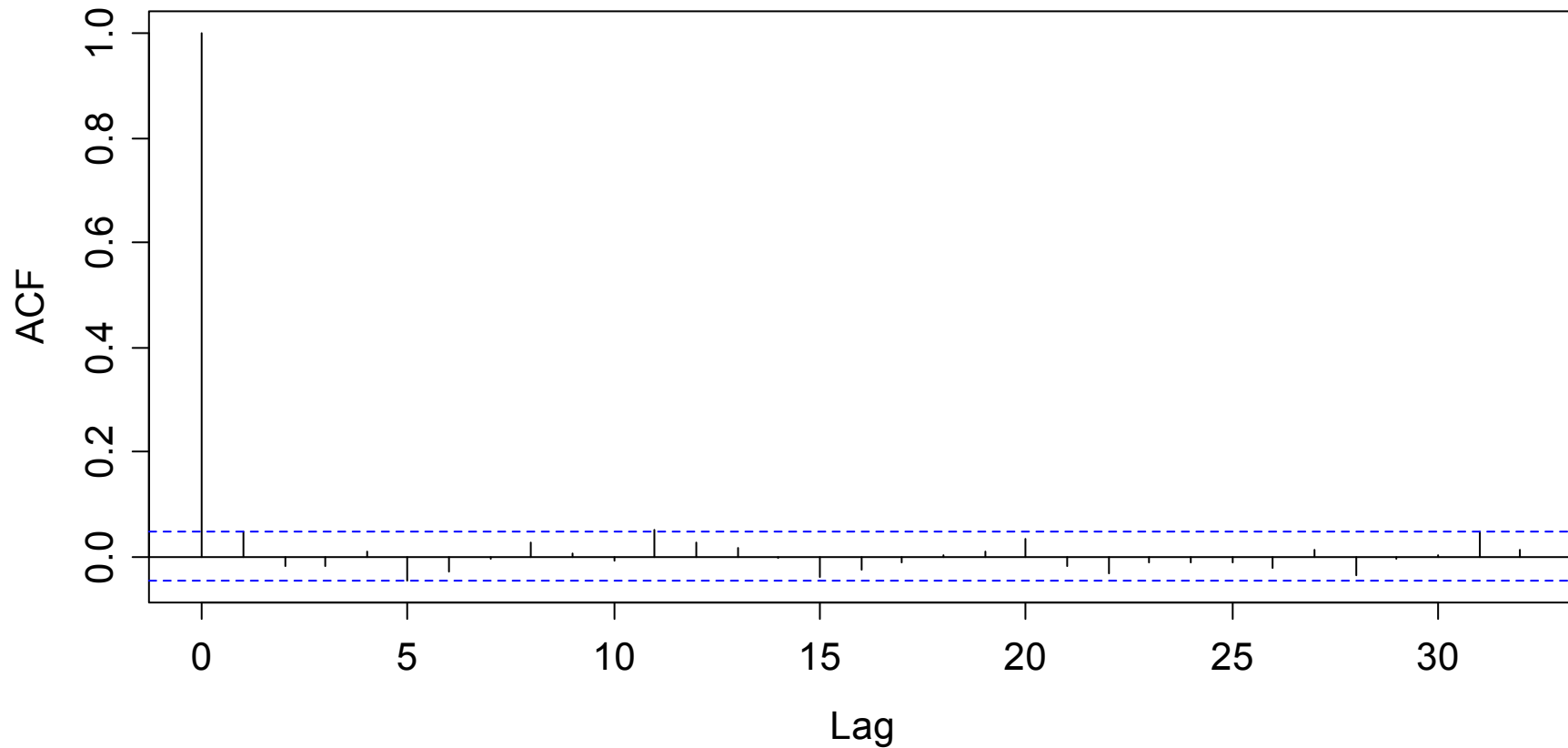


Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

ACF of SMI Log>Returns

ACF of SMI Log>Returns

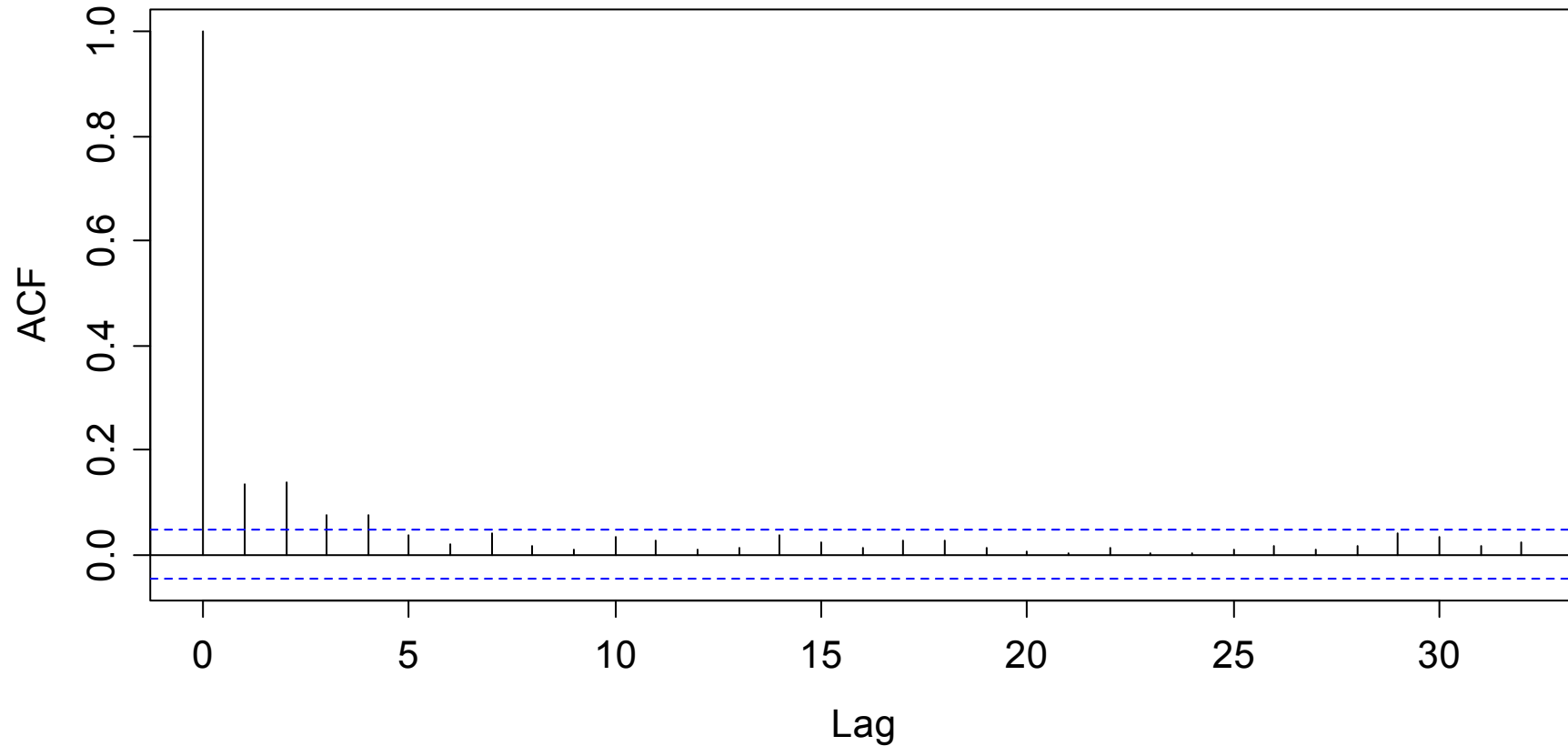


Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

ACF of of Squared SMI Log-Returns

ACF of Squared Log-Returns



Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

The ARCH / GARCH Model

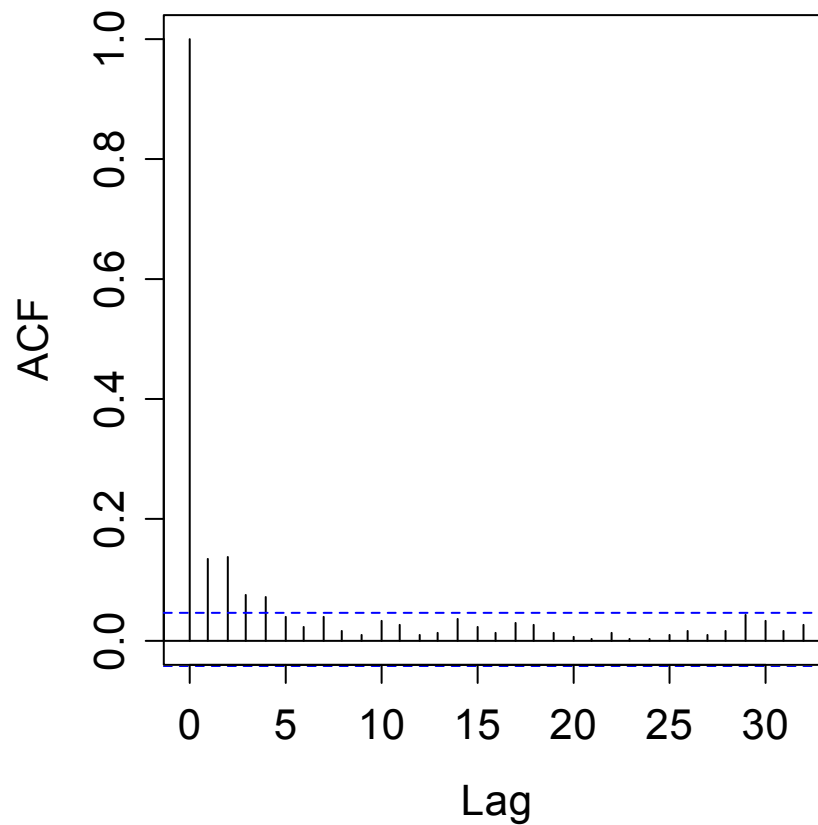
→ See blackboard...

Applied Time Series Analysis

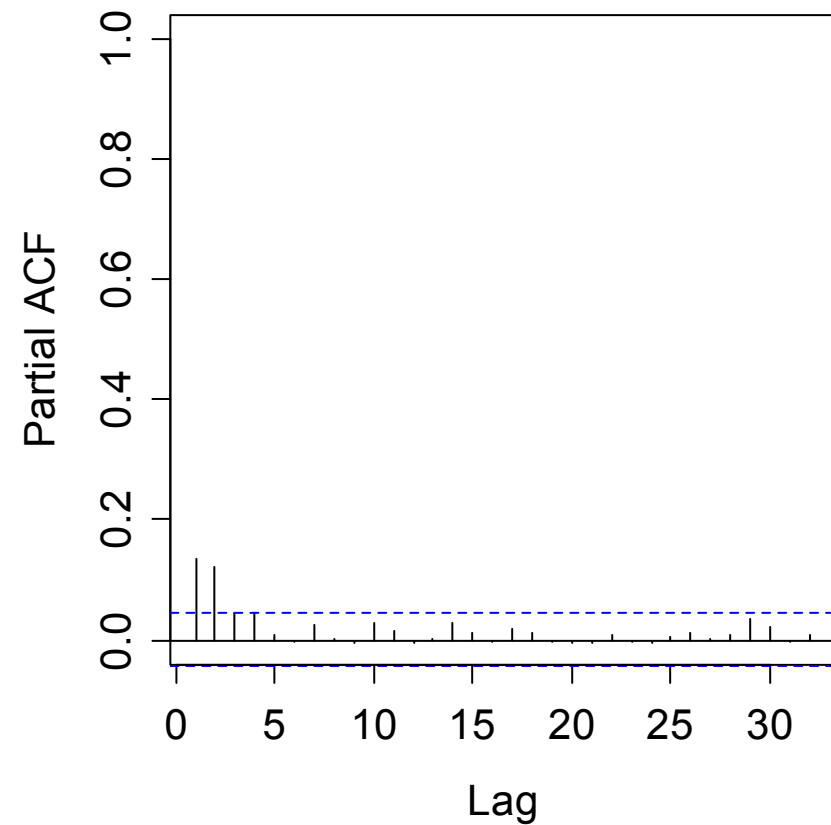
SS 2015 – ARIMA, SARIMA & GARCH

Model Choice

ACF of Squared Log>Returns



PACF of Squared Log>Returns



Applied Time Series Analysis

SS 2015 – ARIMA, SARIMA & GARCH

Fitting an ARCH(2) Model

R allows for convenient fitting...

```
> fit <- garch(lret.smi, order = c(0,2))  
> fit
```

```
Call: garch(x = lret.smi, order = c(0, 2))
```

```
Coefficient(s):
```

a0	a1	a2
6.568e-05	1.309e-01	1.074e-01

Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting with Time Series

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

Note:

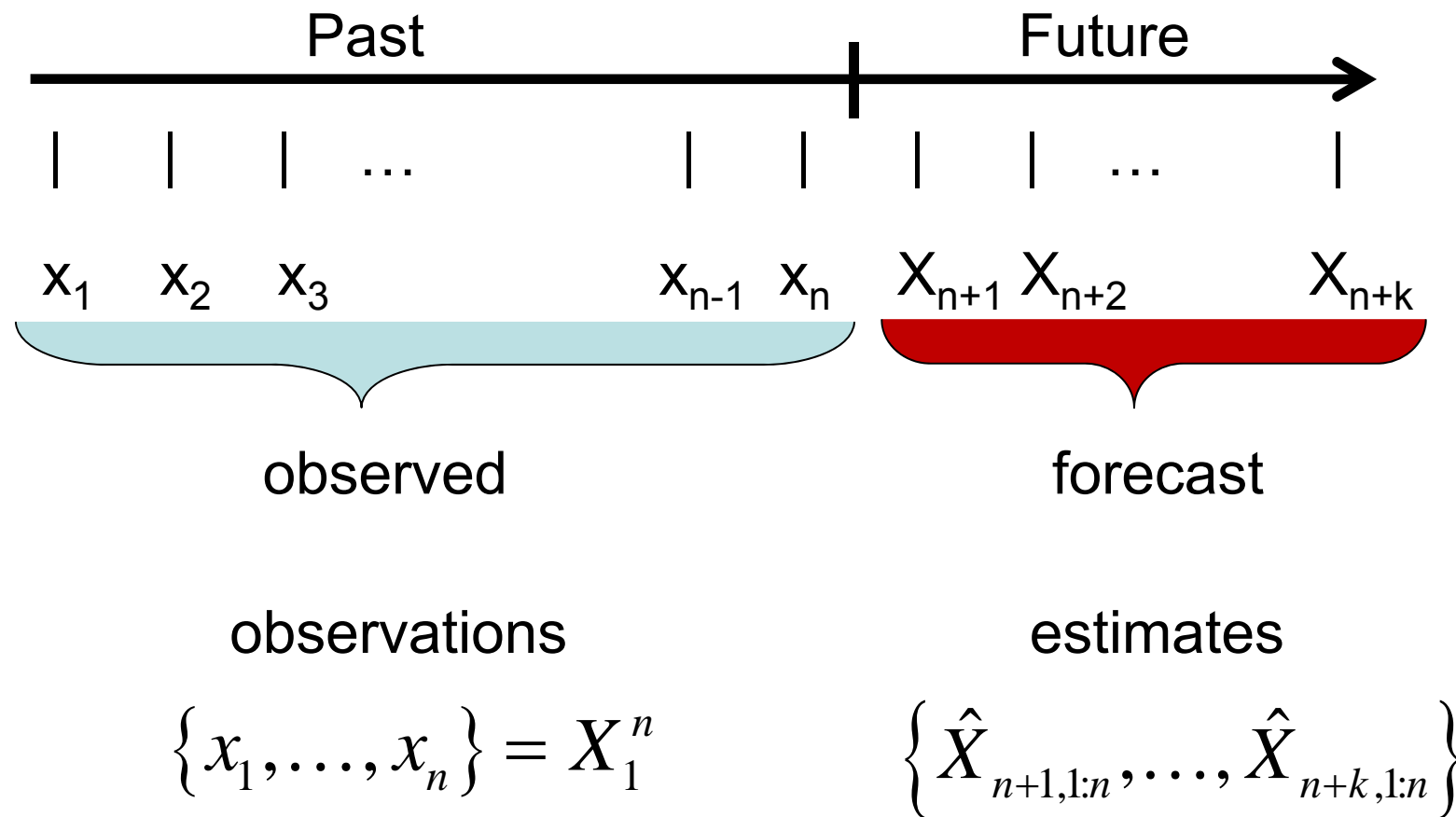
- will be based on a stochastic model
- builds on the dependency structure and past data
- is an extrapolation, thus to take with a grain of salt
- similar to driving a car by using the rear window mirror



Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting, More Technical



Applied Time Series Analysis

SS 2015 – Forecasting

Sources of Uncertainty in Forecasting

There are 4 main sources of uncertainty:

- 1) Does the data generating model from the past also apply in the future? Or are there any breaks?
 - 2) Is the AR(p)-model we fitted to the data $\{x_1, \dots, x_n\}$ correctly chosen? What is the “true” order?
 - 3) Are the parameters $\alpha_1, \dots, \alpha_p, \sigma_E^2$ and μ accurately estimated? How much do they differ from the “truth”?
 - 4) The stochastic variability coming from the innovation E_t
- **we will here restrict to short-term forecasting!**

Applied Time Series Analysis

SS 2015 – Forecasting

How to Forecast?

Probabilistic principle for point forecasts:

$$\hat{X}_{n+k,n} = E \left[X_{n+k} \mid X_1^n \right]$$

→ we forecast the expected value, given our observations

Probabilistic principle for prediction intervals:

$$\text{Var} \left(X_{n+k} \mid X_1^n \right)$$

→ we use the conditional variance

Applied Time Series Analysis

SS 2015 – Forecasting

How to Apply the Principles?

- The principles provide a nice setup, but are only useful and practicable under additional assumptions.
 - For stationary AR(1)-processes with normally distributed innovations, we can apply the principles and derive formulae
- **see blackboard for the derivation!**

Applied Time Series Analysis

SS 2015 – Forecasting

AR(1): 1-Step Forecast

The 1-step forecast for a shifted AR(1) process with mean m is:

$$\hat{X}_{n+1,n} = \alpha_1 (x_n - m) + m$$

with prognosis interval

$$\hat{X}_{n+1,n} \pm 1.96 \cdot \sigma_E$$

Note that when $\hat{\alpha}_1, \hat{\mu}, \hat{\sigma}_E$ are plugged-in, this adds additional uncertainty which is not accounted for in the prognosis interval, i.e.

$$\text{Var}(\hat{X}_{n+1}) > \text{Var}(X_{n+1} | X_1^n)$$

Applied Time Series Analysis

SS 2015 – Forecasting

Simulation Study

We have seen that the usual prognosis interval is too small. But by how much? A simulation study yields some insight:

Generated are 10'000 1-step forecasts on a time series that was generated from an AR(1) process with $\alpha = 0.5$. The series length was variable.

The 95%-prognosis interval was determined and it was checked whether it included the true value or not. The empirically estimated confidence levels were:

n=20	n=50	n=100	n=200
91.01%	93.18%	94.48%	94.73%

Applied Time Series Analysis

SS 2015 – Forecasting

AR(1): k-Step Forecast

The k-step forecast for an AR(1) process is:

$$\hat{X}_{n+k,n} = \alpha_1^k (x_n - m) + m$$

with prognosis interval based on

$$\text{Var}(X_{n+k,n} | X_1^n) = \left(1 + \sum_{j=1}^{k-1} \alpha_1^{2j} \right) \cdot \sigma_E^2$$

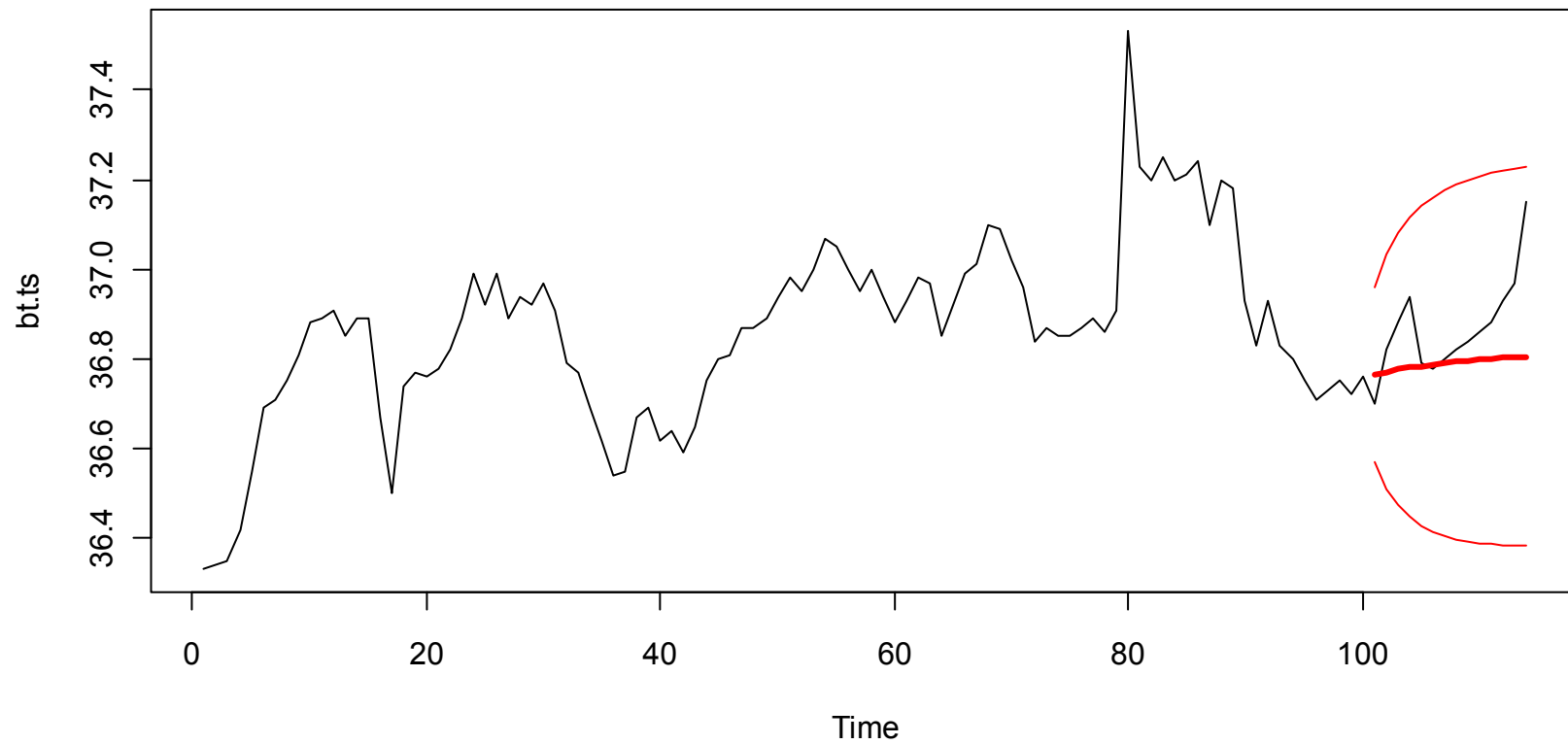
It is important to note that for $k \rightarrow \infty$, the expected value and the variance from above go to μ and σ_X^2 respectively.

Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting the Beaver Data

Forecasting Beaver Data



Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting AR(p)

The principle is the same, forecast and prognosis interval are:

$$E[X_{n+k} | X_1^n] \text{ and } Var(X_{n+k} | X_1^n)$$

The computations are more complicated, but do not yield any further insight. We are thus doing without.

$$\text{1-step-forecast: } \hat{X}_{n+1,1:n} = \alpha_1(x_n - m) + \dots + \alpha_p(x_{n+1-p} - m) + m$$

$$\text{k-step-forecast: } \hat{X}_{n+k,1:n} = \alpha_1(\hat{X}_{n+k-1,1:n} - m) + \dots + \alpha_p(\hat{X}_{n+k-p,1:n} - m) + m$$

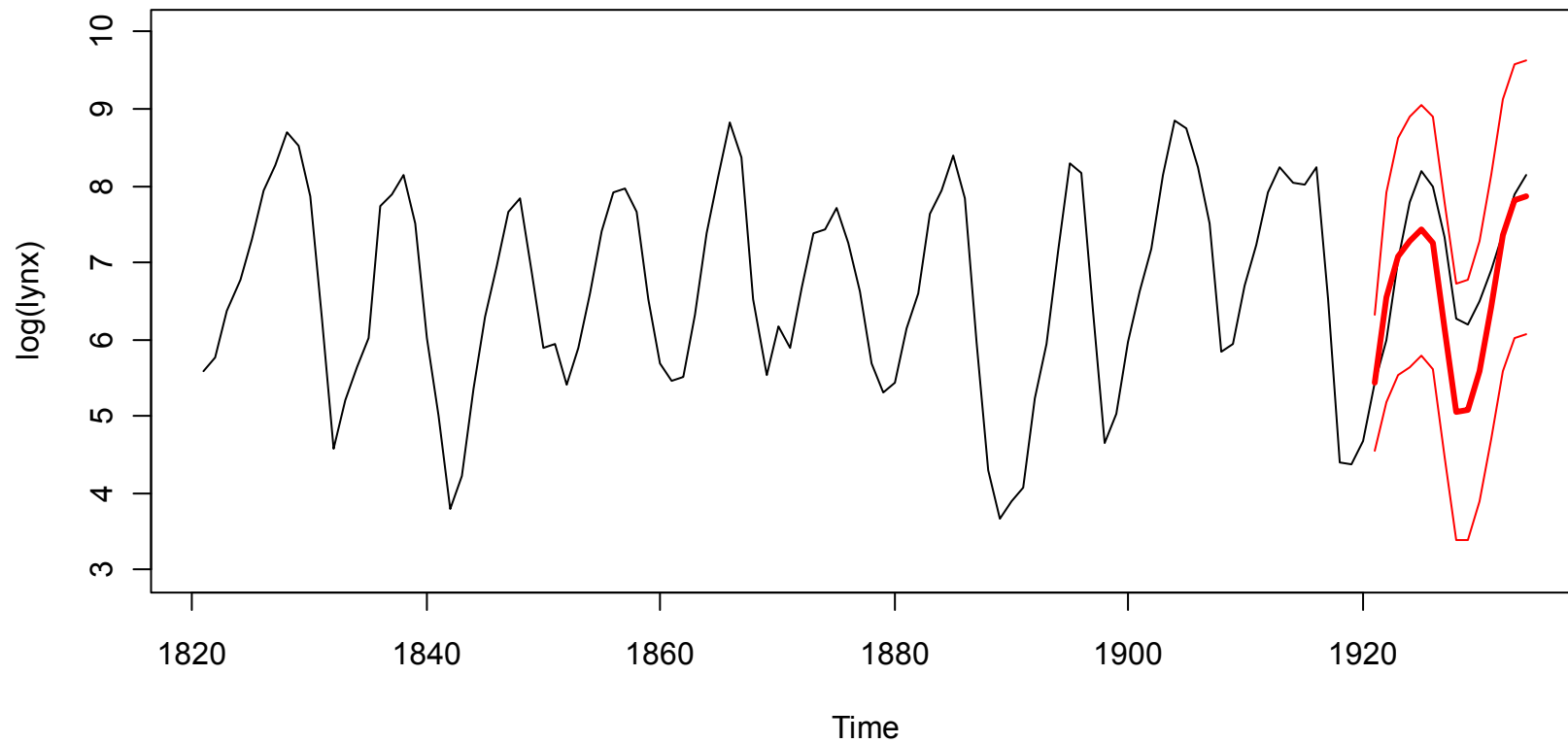
If an observed value is available, we plug it in. Else, the forecast is determined in a recursive manner.

Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting the Lynx Data

Forecasting log(Lynx) Data



Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting: Remarks

- AR(p) processes have a Markov property. Given the model parameters, we only need the last p observations to compute the forecast.
- The prognosis intervals are not simultaneous prognosis intervals, and they are generally too small. However, simulation studies show that this is not excessively so.
- Retaining the final part of the series, and predicting it with several competing models may give hints which one yields the best forecasts. This can be an alternative approach for choosing the model order p .

Applied Time Series Analysis

SS 2015 – Forecasting

Sources of Uncertainty in Forecasting

There are 4 main sources of uncertainty:

- 1) Does the data generating model from the past also apply in the future? Or are there any breaks?
 - 2) Is the ARMA(p,q)-model we fitted to the data $\{x_1, \dots, x_n\}$ correctly chosen? What is the “true” order?
 - 3) Are the parameters $\alpha, \beta, \sigma_E^2$ and μ accurately estimated? How much do they differ from the “truth”?
 - 4) The stochastic variability coming from the innovation E_t
- **we will here restrict to short-term forecasting!**

Applied Time Series Analysis

SS 2015 – Forecasting

How to Forecast?

Probabilistic principle for point forecasts:

$$\hat{X}_{n+k,n} = E \left[X_{n+k} \mid X_1^n \right]$$

→ we forecast the expected value, given our observations

Probabilistic principle for prediction intervals:

$$\text{Var} \left(X_{n+k} \mid X_1^n \right)$$

→ we use the conditional variance

Applied Time Series Analysis

SS 2015 – Forecasting

How to Apply the Principles?

- The principles provide a nice setup, but are only useful and practicable under additional assumptions.
 - Whereas for $AR(p)$, knowing the last p observations is sufficient for coming up with a forecast, $ARMA(p,q)$ models require knowledge about the infinite past.
 - In practice, one is using recursive formulae
- **see blackboard for the derivation in the $MA(1)$ case!**

Applied Time Series Analysis

SS 2015 – Forecasting

MA(1) Forecasting: Summary

- We have seen that for an MA(1)-process, the k-step forecast for $k > 1$ is equal to m .
- In case of $k=1$, we obtain for the MA(1)-forecast:

$$\hat{X}_{n+1,n} = \mu + \beta_1 \cdot E[E_n | X_1^n]$$

The conditional expectation is (too) difficult to compute

- As a trick, we not only condition on observations $1, \dots, n$, but on the infinite past:

$$e_n := E[E_n | X_{-\infty}^n]$$

Applied Time Series Analysis

SS 2015 – Forecasting

MA(1) Forecasting: Summary

- We then write the MA(1) as an AR(∞) and solve the model equation for E_n :

$$E_n = \sum_{j=0}^{\infty} (-\beta_1)^j \cdot (X_{n-j} - m)$$

- In practice, we plug-in the time series observations x_{n-j} where available. For the „early“ times, where we don't have observations, we plug-in \hat{m} .
- This is of course only an approximation to the true MA(1)-forecast, but it works well in practice, because of:

$$|\beta_1| < 1$$

Applied Time Series Analysis

SS 2015 – Forecasting

ARMA(p,q) Forecasting

As with MA(1)/MA(q) forecasting, we face problems with

$$E[E_{n+1-j} | X_{-\infty}^n]$$

which is difficult to compute. We use the same tricks as for MA(1) and obtain

$$\begin{aligned} \hat{X}_{n+k,n} = & \mu + \sum_{i=1}^p \alpha_i (E[X_{n+k-i} | X_{-\infty}^n] - \mu) \\ & + E[E_{n+k} | X_{-\infty}^n] - \sum_{j=1}^q \beta_j E[E_{n+k-j} | X_{-\infty}^n] \end{aligned}$$

where ...

Applied Time Series Analysis

SS 2015 – Forecasting

ARMA(p,q) Forecasting

...where

$$E[X_t | X_{-\infty}^n] = \begin{cases} x_t & \text{if } t \leq n \\ \hat{X}_{t,n} & \text{if } t > n \end{cases}$$

and

$$E[E_t | X_{-\infty}^n] = \begin{cases} e_t & \text{if } t \leq n \\ 0 & \text{if } t > n \end{cases}$$

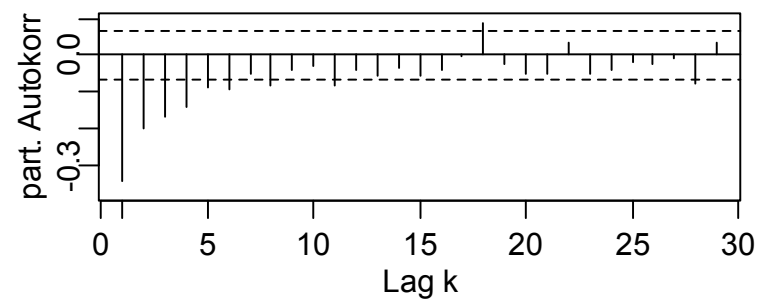
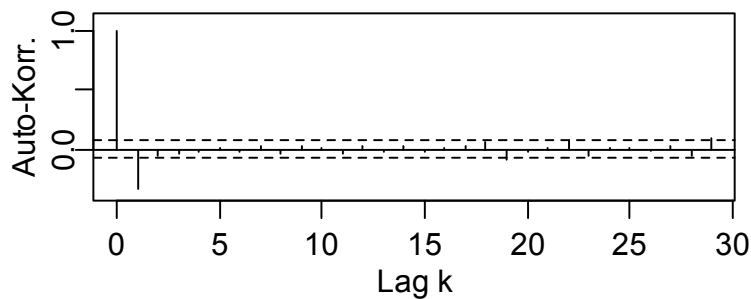
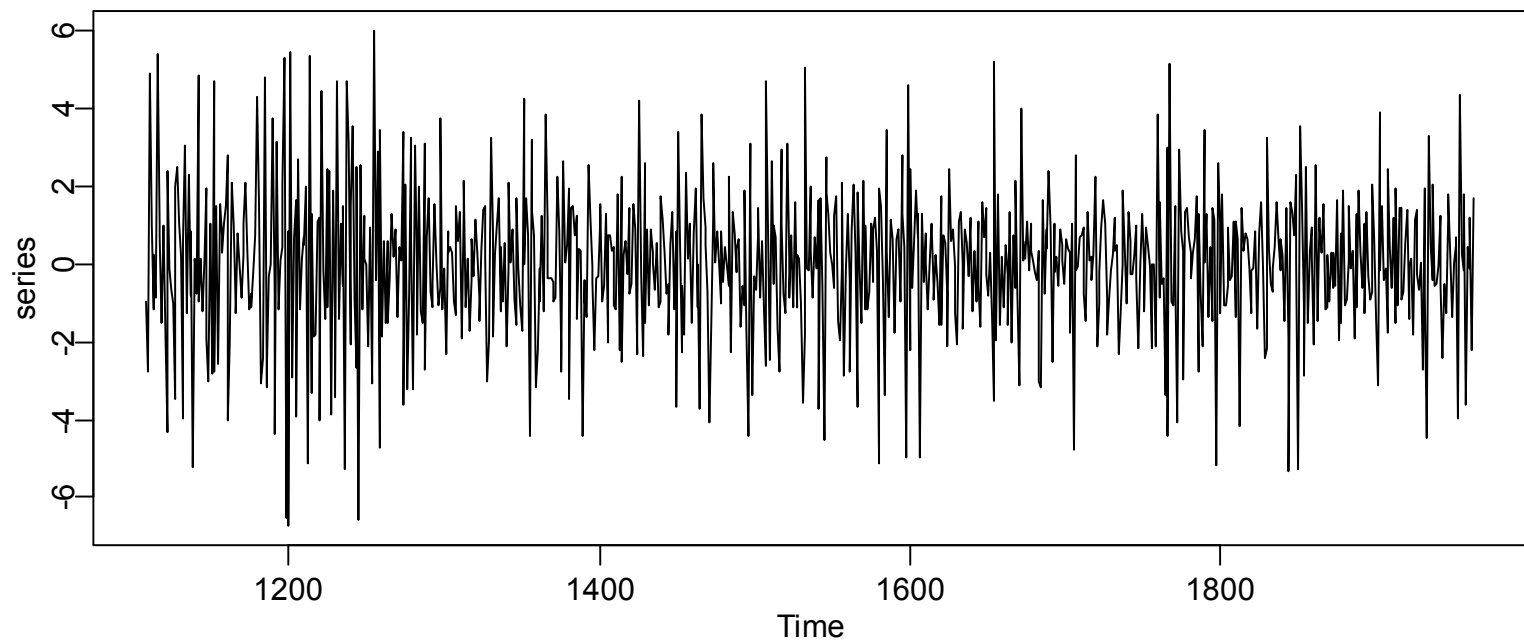
with

$$e_t = x_t - \mu - \sum_{i=1}^p \alpha_i (x_{t-i} - \mu) + \sum_{j=1}^q \beta_j e_{t-j}$$

Applied Time Series Analysis

SS 2015 – Forecasting

ARMA(p,q) Forecasting: Douglas Fir

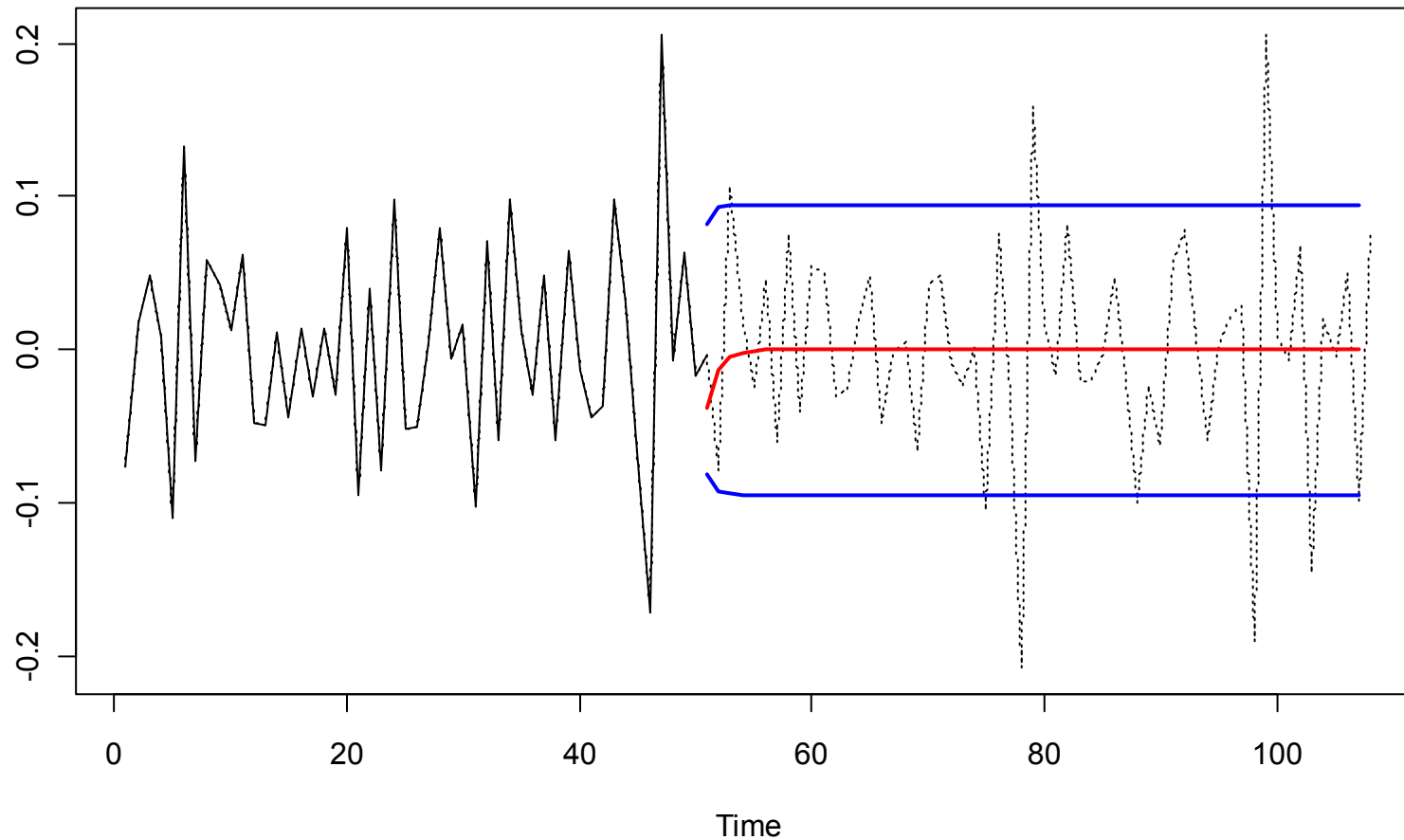


Applied Time Series Analysis

SS 2015 – Forecasting

ARMA(p,q) Forecasting: Example

Forecasting the Differenced Douglas Fir Series



Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting with SARIMA

Time series with a trend and/or seasonal effect can either be predicted after decomposing or with exponential smoothing. It is also very easy and quick to predict from a SARIMA model.

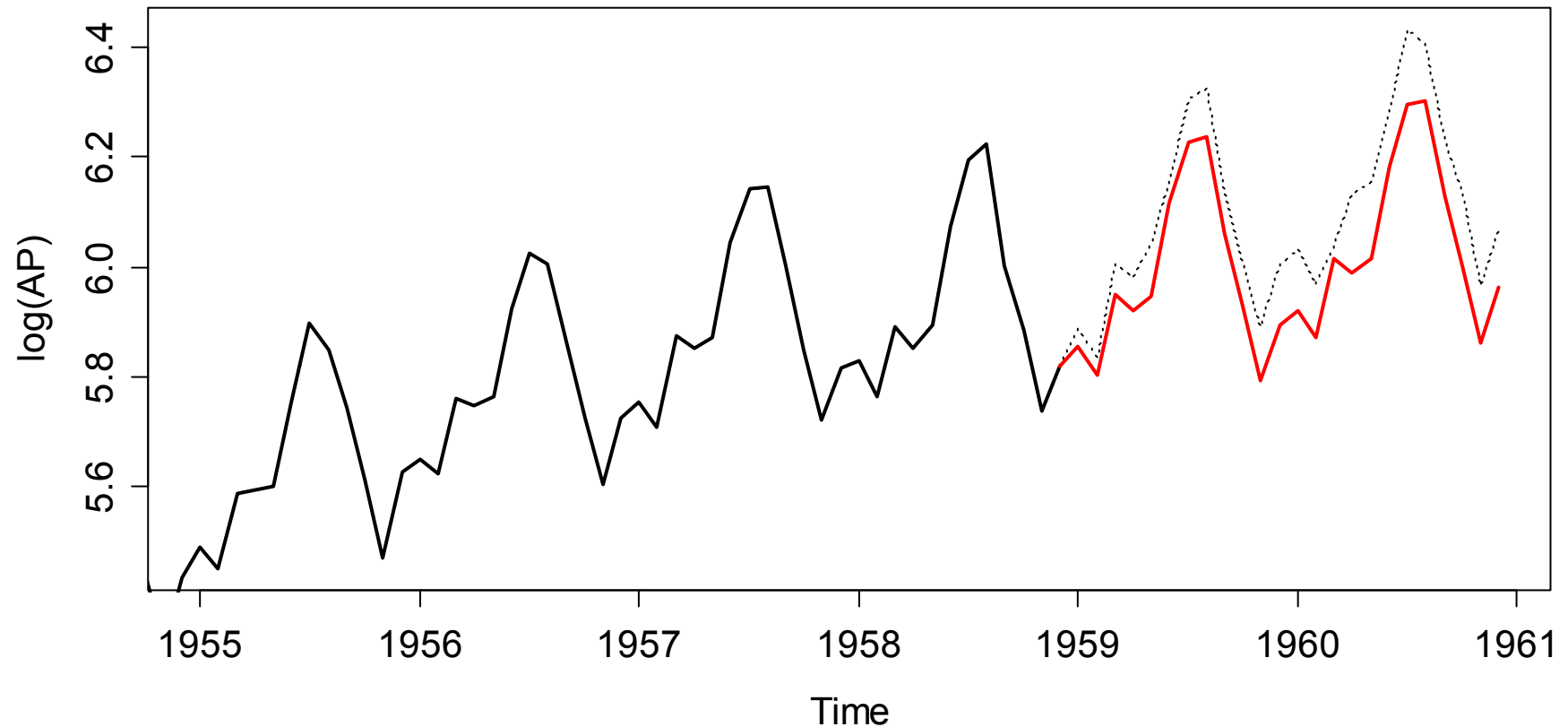
- The SARIMA model is fitted in R as usual. Then, we can simply employ the `predict()` command and obtain the forecast plus a prediction interval.
- Technically, the forecast comes from the non-stationary ARMA(p,q)-formulation of the SARIMA model.
- The disadvantage of working with SARIMA forecasts is that it has somewhat the touch of a **black box approach**.

Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting with SARIMA: Example

Forecast of $\log(\text{AP})$ with SARIMA(0,1,1)(0,1,1)



Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series

The principle for forecasting time series that are decomposed into trend, seasonal effect and remainder is:

1) Stationary Remainder

Is usually modelled with an $ARMA(p,q)$, so we can generate a time series forecast with the methodology from before.

2) Seasonal Effect

Is assumed as remaining “as is”, or “as it was last” (in the case of evolving seasonal effect) and extrapolated.

3) Trend

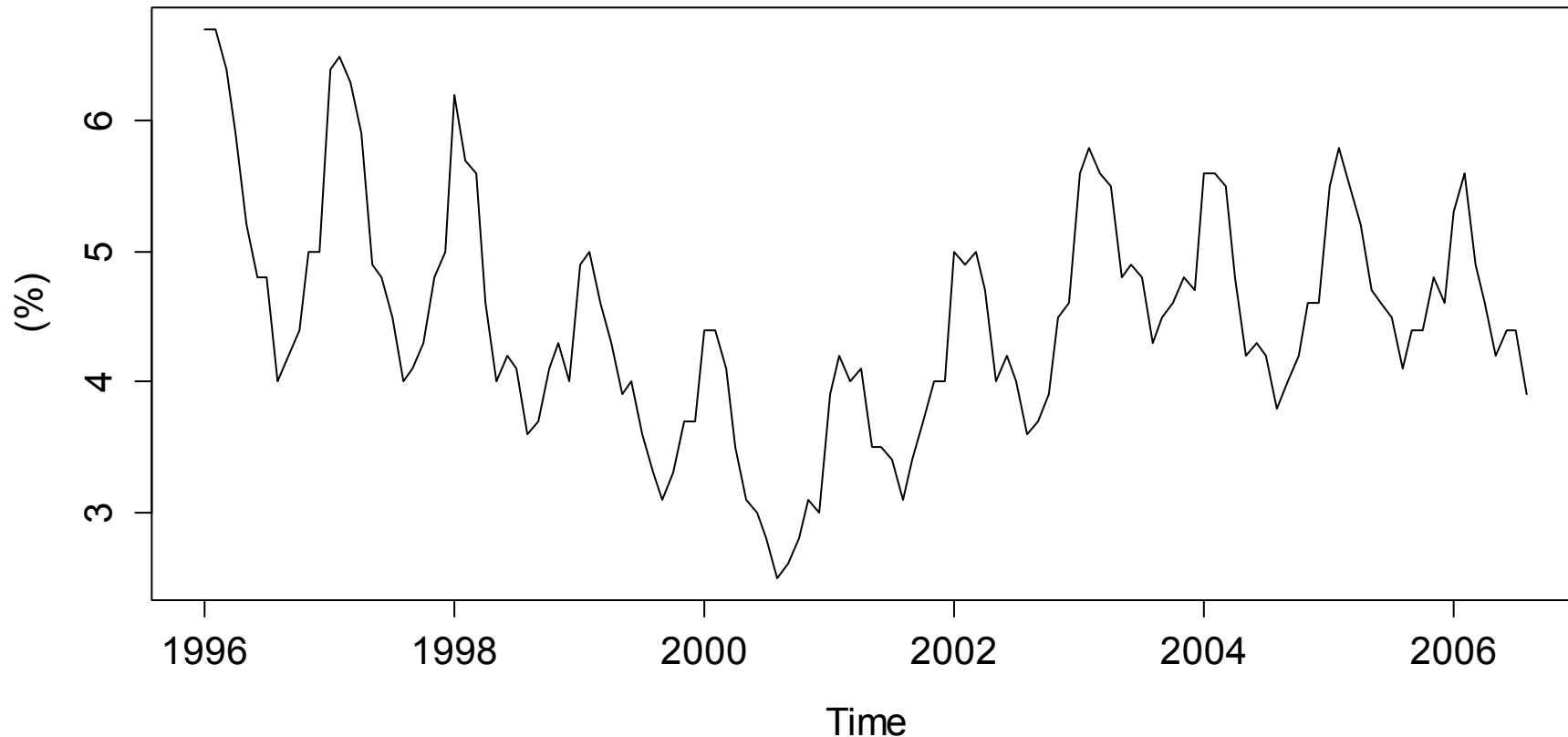
Is either extrapolated linearly, or sometimes even manually.

Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series: Example

Unemployment in Maine

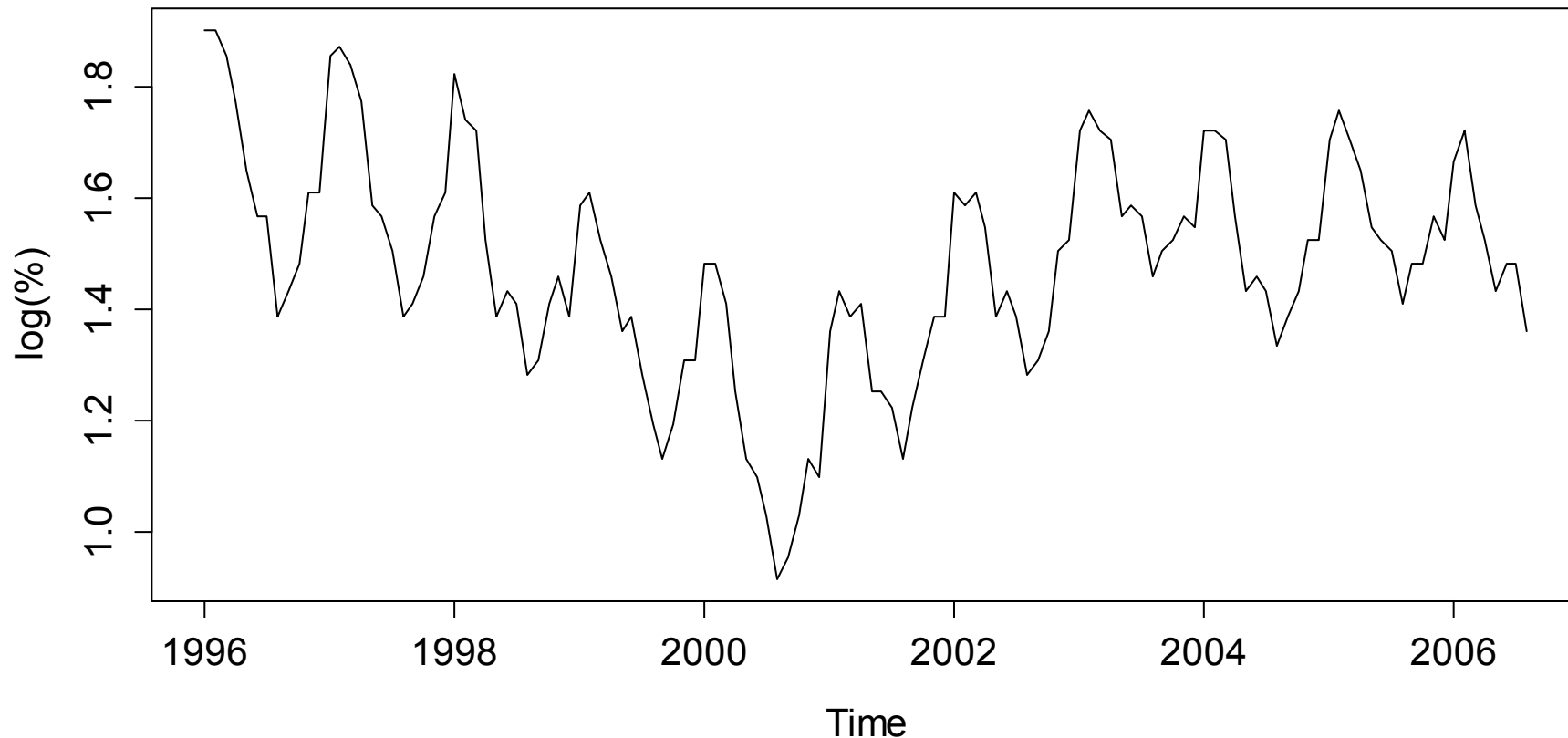


Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series: Example

Logged Unemployment in Maine

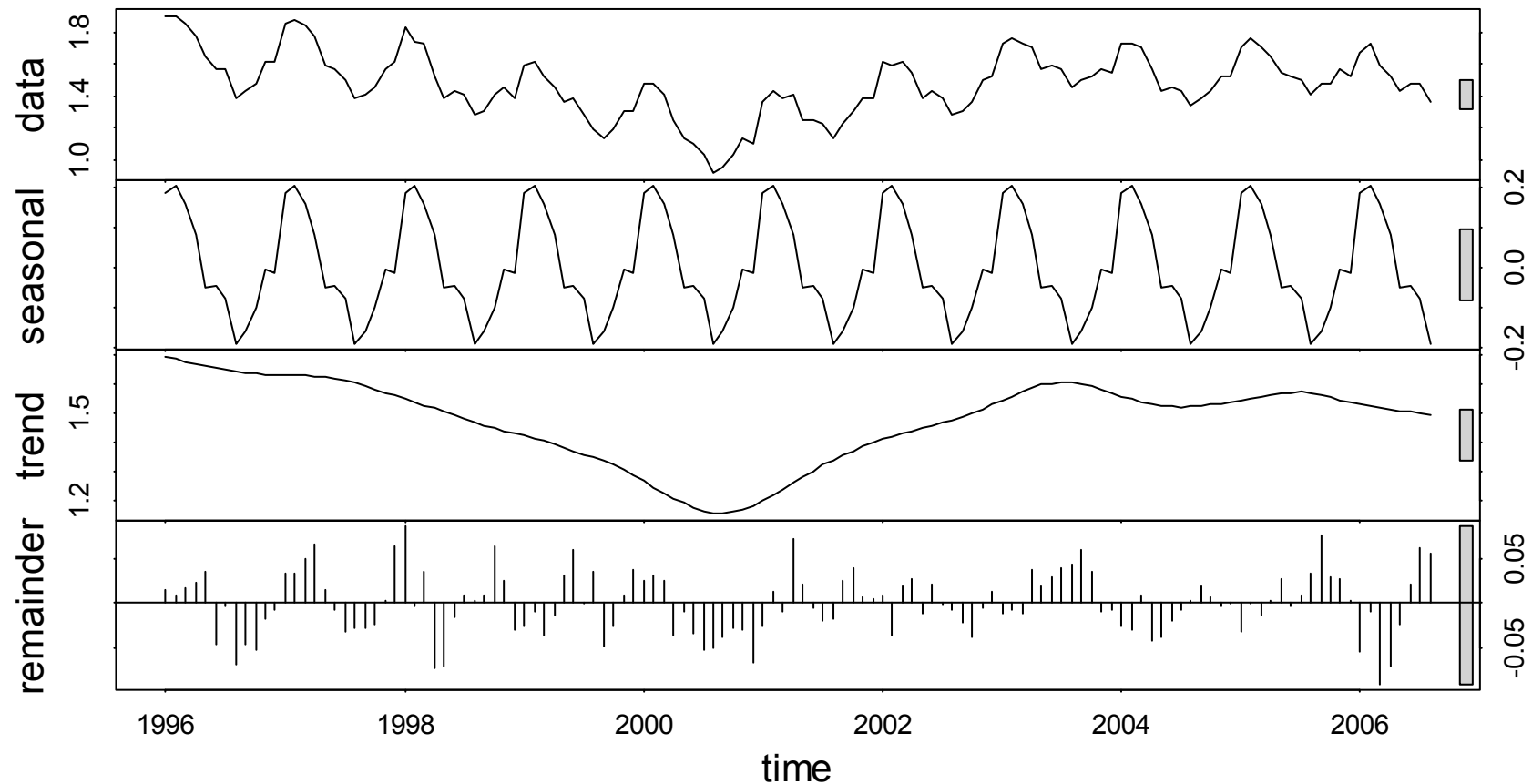


Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series: Example

STL-Decomposition of Logged Maine Unemployment Series

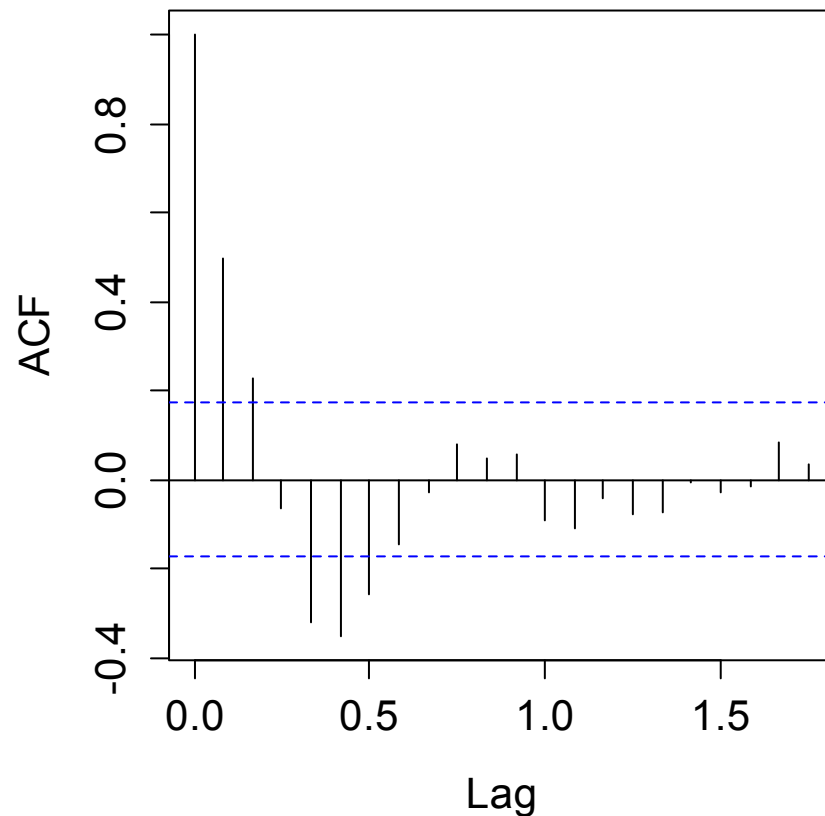


Applied Time Series Analysis

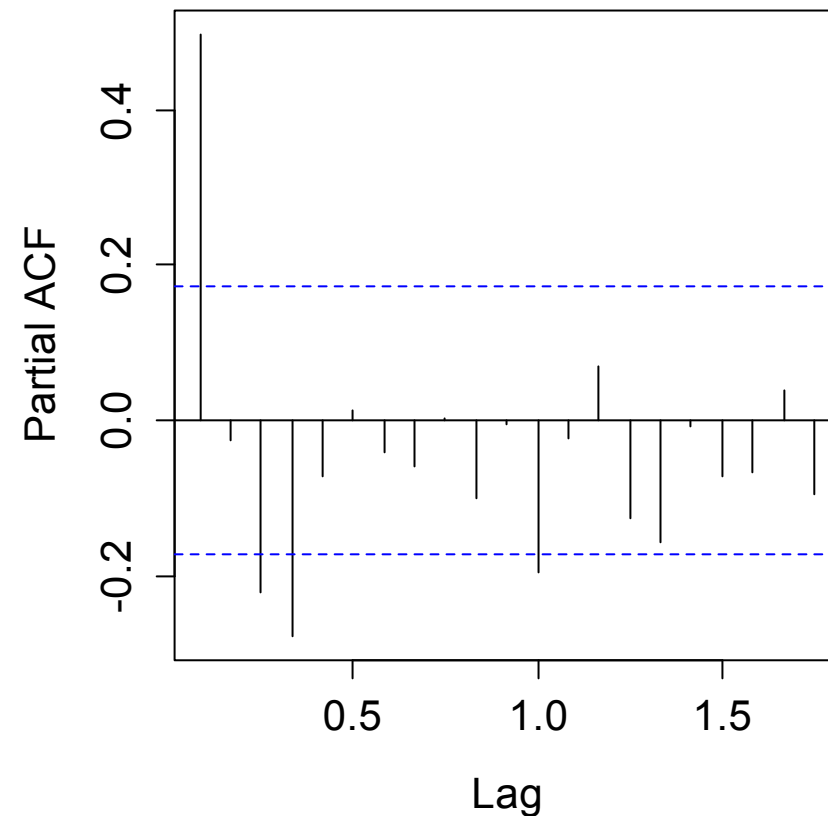
SS 2015 – Forecasting

Forecasting Decomposed Series: Example

ACF of Remainder Series



PACF of Remainder Series

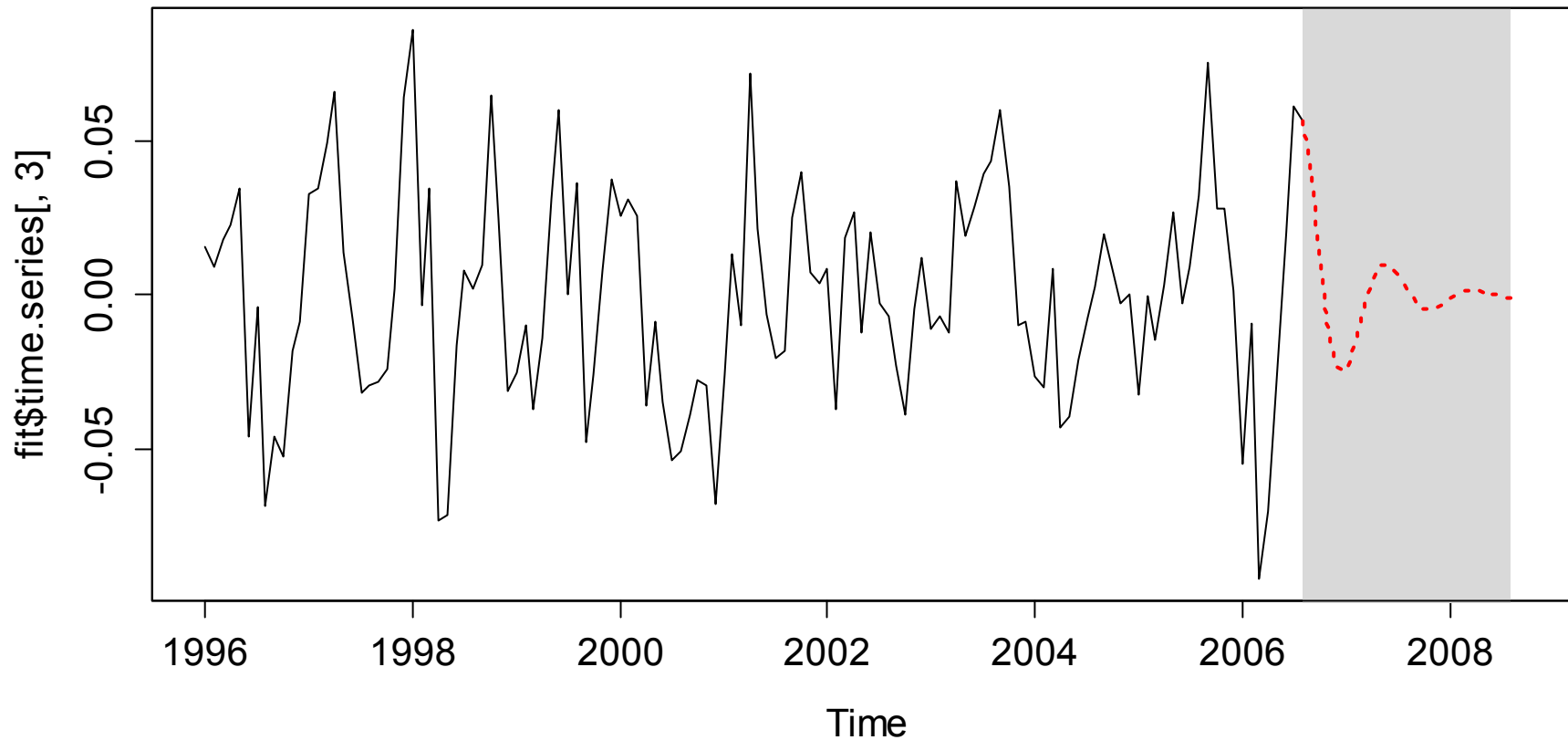


Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series: Example

AR(4) Forecast for Remainder Series

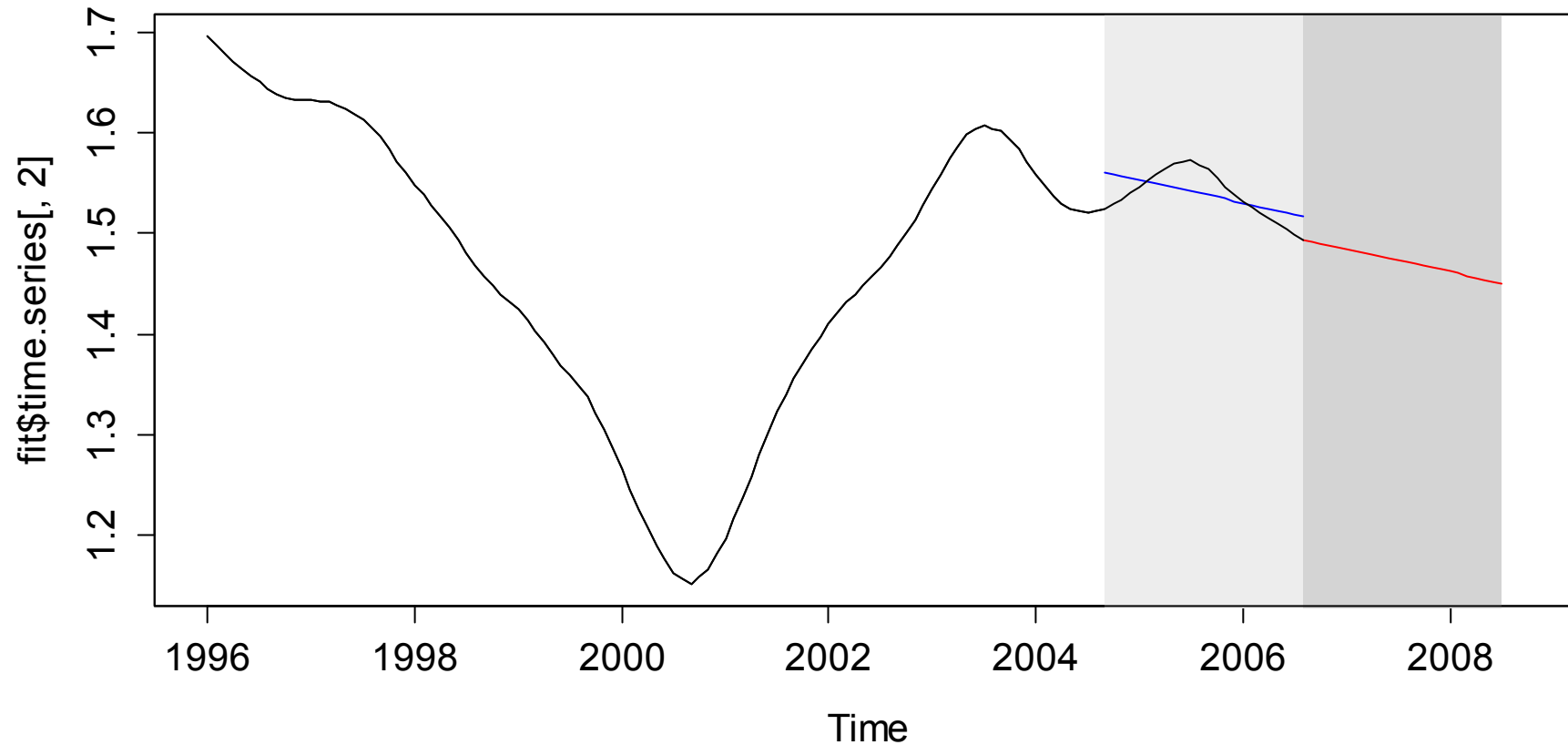


Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series: Example

Trend Forecast by Linear Extrapolation

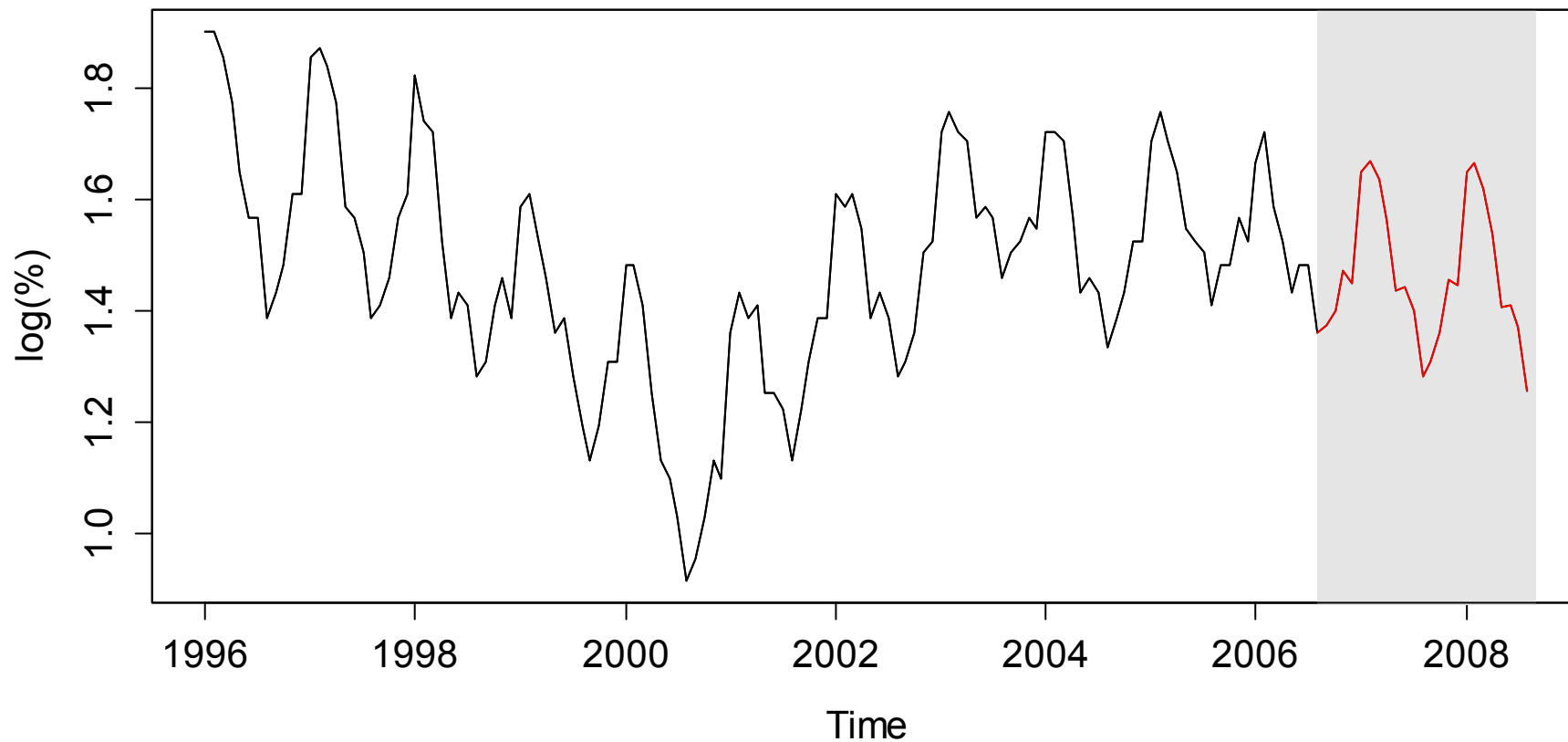


Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series: Example

Forecast of Logged Unemployment in Maine



Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series

The principle for forecasting time series that are decomposed into trend, seasonal effect and remainder is:

1) Stationary Remainder

Is usually modelled with an $ARMA(p,q)$, so we can generate a time series forecast with the methodology from before.

2) Seasonal Effect

Is assumed as remaining “as is”, or “as it was last” (in the case of evolving seasonal effect) and extrapolated.

3) Trend

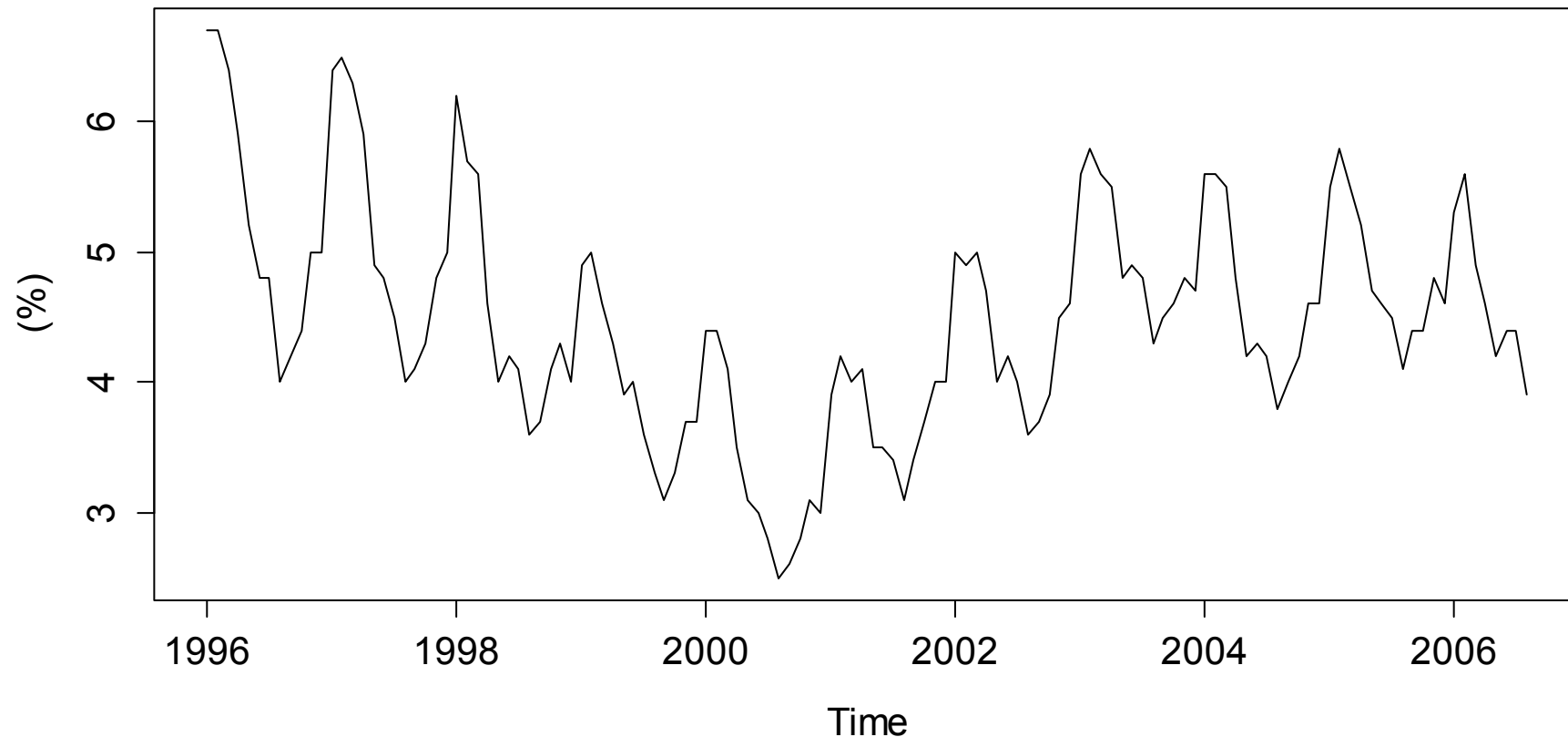
Is either extrapolated linearly, or sometimes even manually.

Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series: Example

Unemployment in Maine

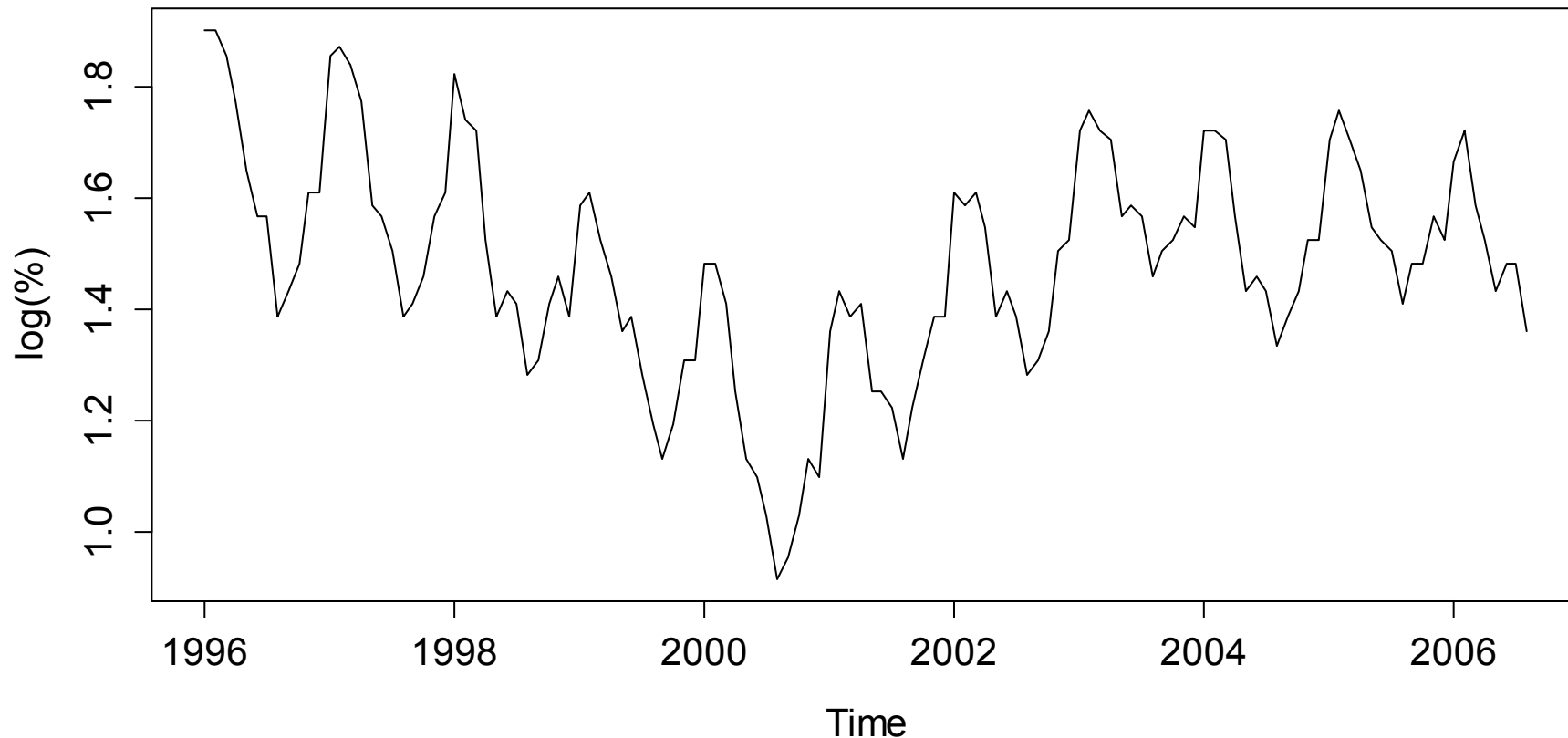


Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series: Example

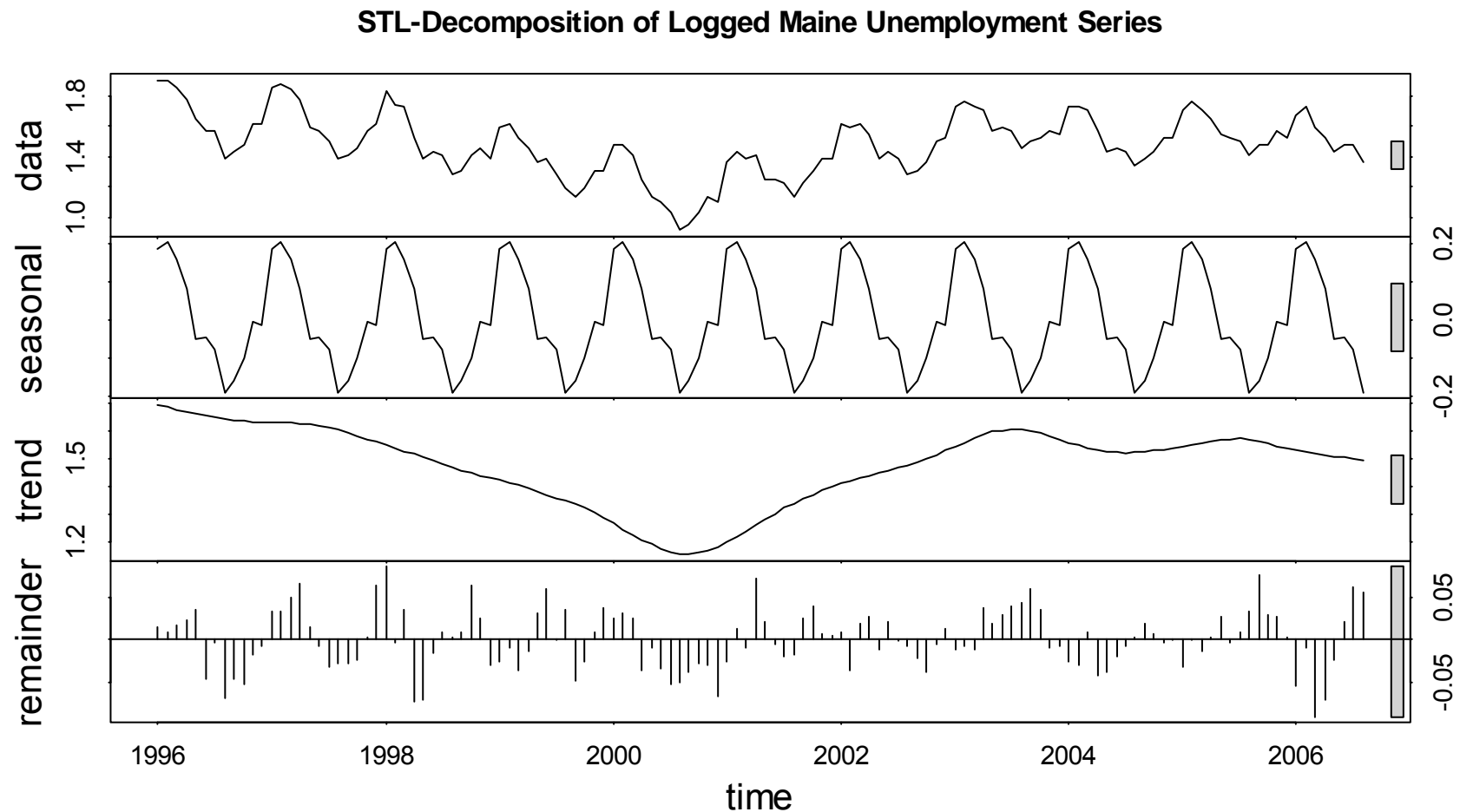
Logged Unemployment in Maine



Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series: Example

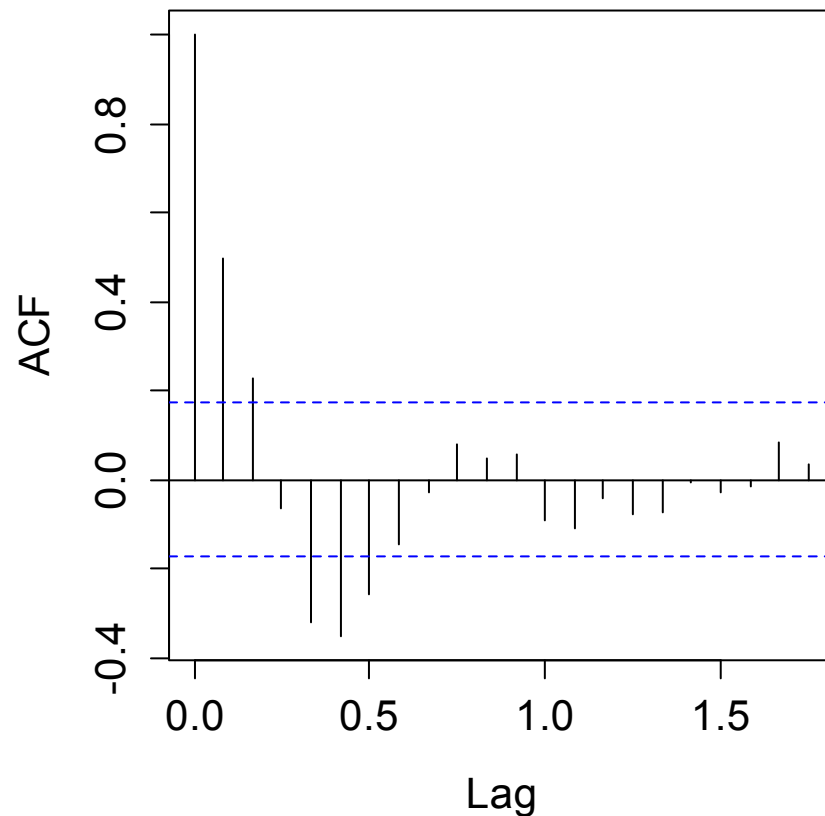


Applied Time Series Analysis

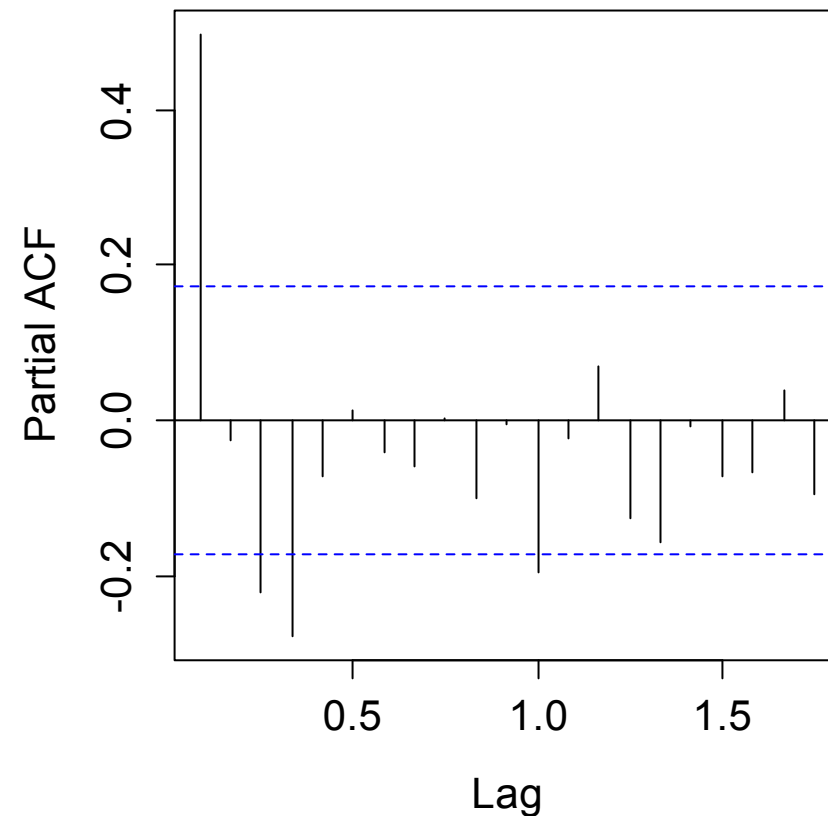
SS 2015 – Forecasting

Forecasting Decomposed Series: Example

ACF of Remainder Series



PACF of Remainder Series

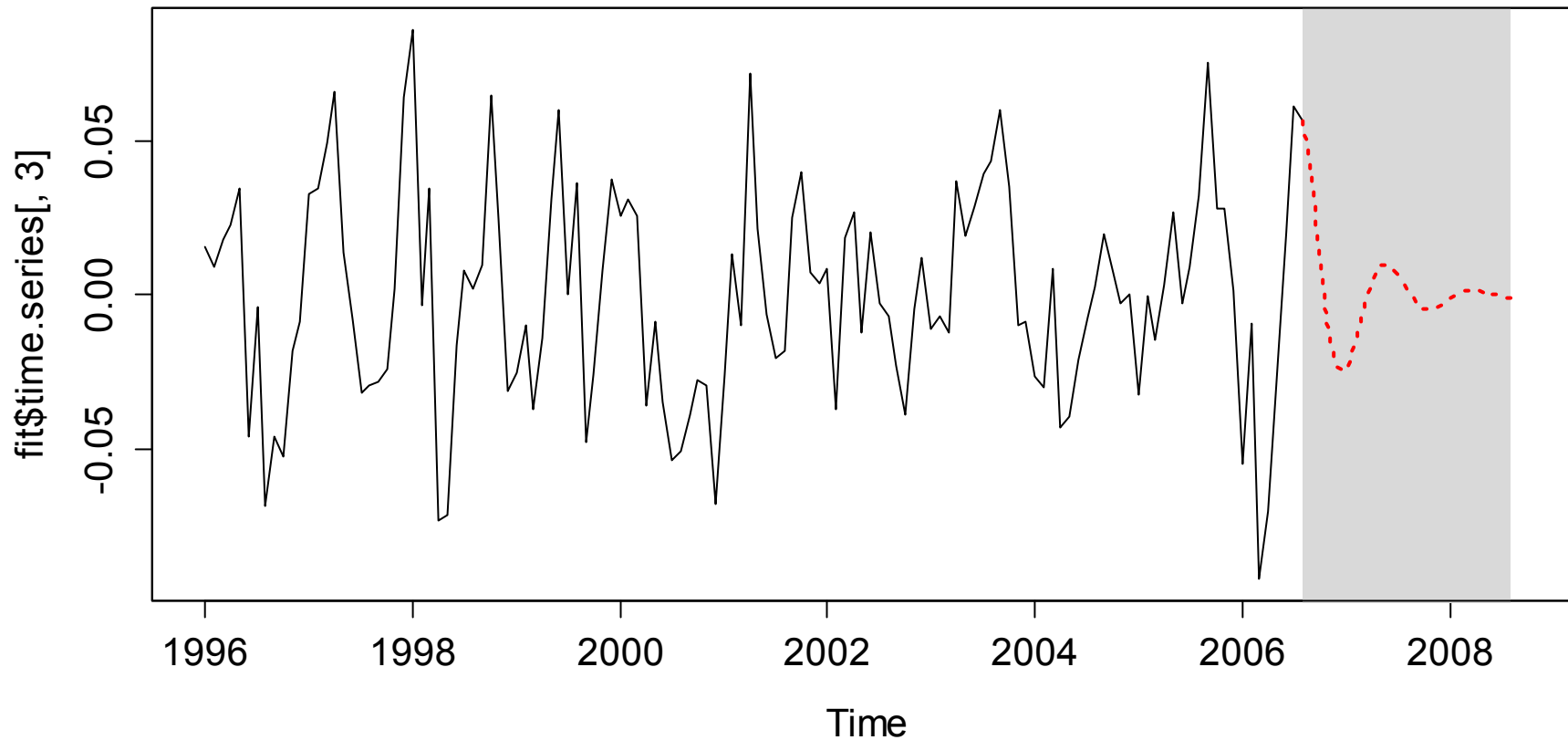


Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series: Example

AR(4) Forecast for Remainder Series

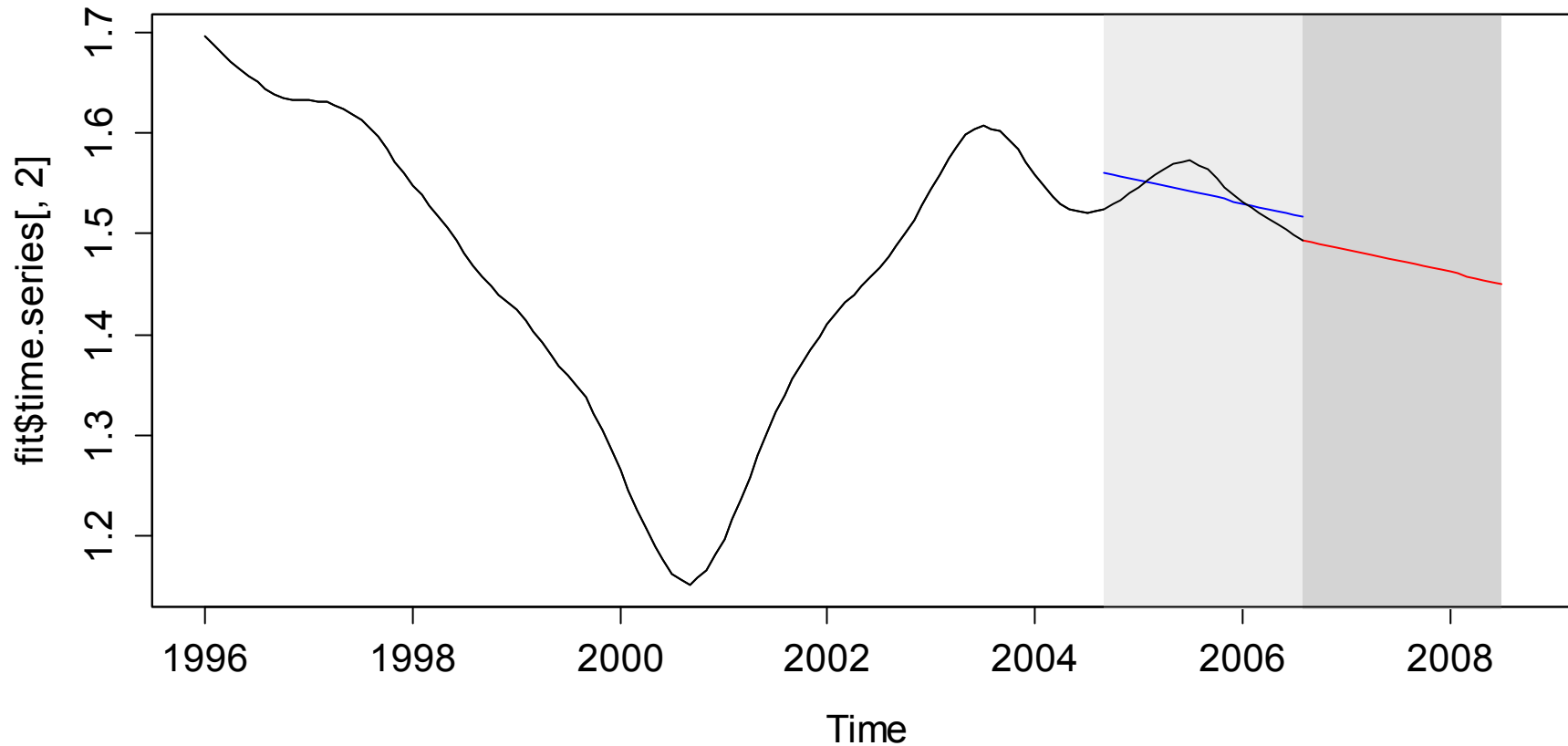


Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series: Example

Trend Forecast by Linear Extrapolation

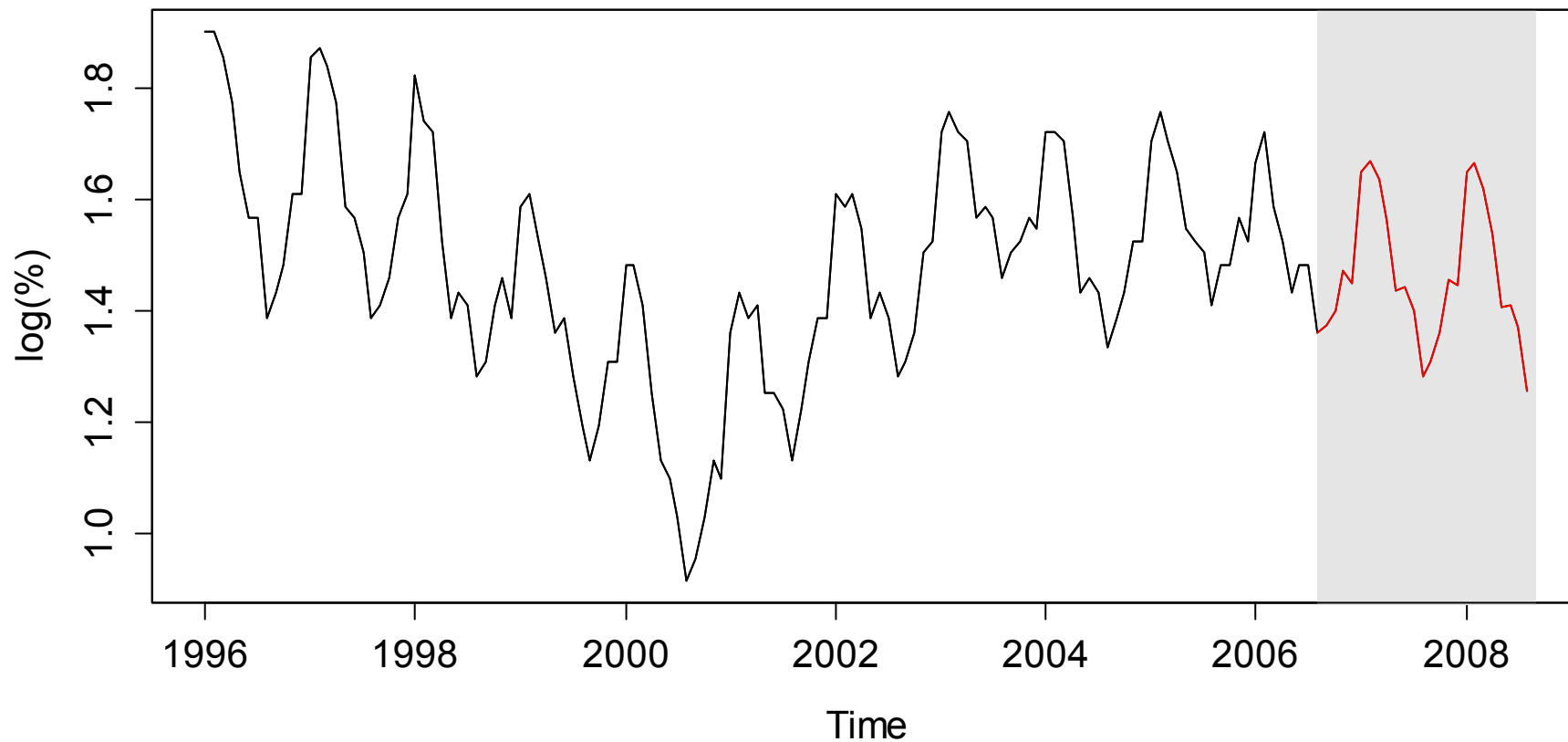


Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting Decomposed Series: Example

Forecast of Logged Unemployment in Maine



Applied Time Series Analysis

SS 2015 – Forecasting

Simple Exponential Smoothing

This is a quick approach for estimating the current level of a time series, as well as for forecasting future values. It works for any stationary time series without a trend and season.

The **simple, intuitive idea behind** is:

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

The weights are often chosen to be exponentially decaying, two examples with different parameters are on the next slide. However, there is also a deeper mathematical notion of ExpSmo.

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

Applied Time Series Analysis

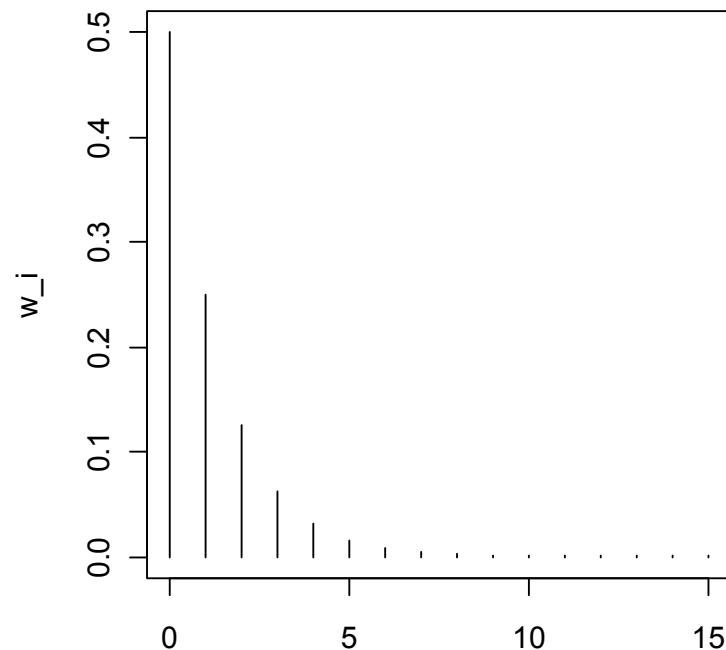
SS 2015 – Forecasting

Choice of Weights

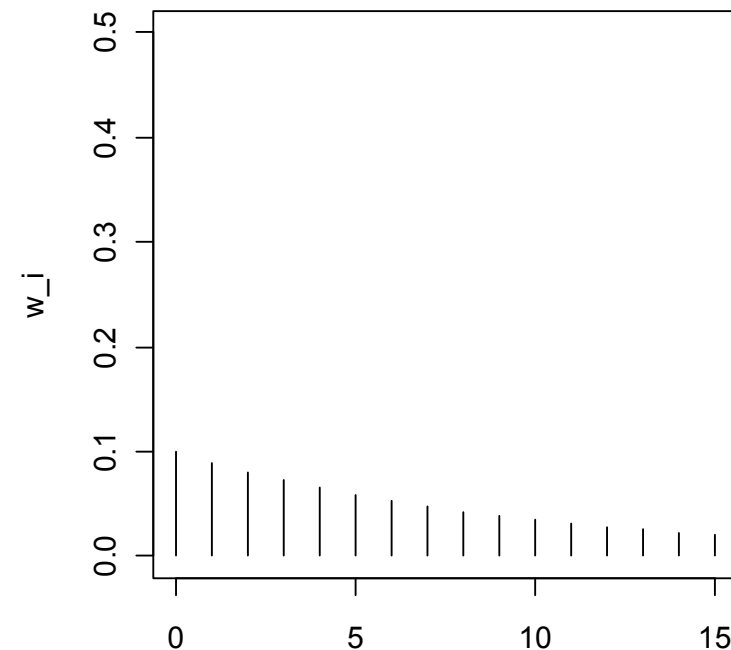
An usual choice are exponentially decaying weights:

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

a=0.5



a=0.1



Applied Time Series Analysis

SS 2015 – Forecasting

Simple Exponential Smoothing: Summary

What is it?

- A method for estimating and forecasting the conditional mean

Basic notion: $X_t = \mu_t + E_t$

- μ_t is the conditional expectation, which we try to estimate from the data. The estimate a_t is called level of the series.
- E_t is a completely random innovation term.

Estimation of the level: two notions exist...

- Weighted updating: $a_t = \alpha x_t + (1 - \alpha) a_{t-1}$
- Exponential smoothing: $a_t = \sum_{i=0}^{\infty} \alpha (1 - \alpha)^i x_{t-i}$

Applied Time Series Analysis

SS 2015 – Forecasting

Forecasting with Exponential Smoothing

The forecast, for any horizon $k > 0$ is:

$$\hat{X}_{n+k,1:n} = a_n$$

Hence, the forecast is given by the current level, and it is constant for all horizons k . However, it does depend on the choice of the smoothing parameter α . In R, a data-adaptive solution is available by minimizing SS1PE:

$$\text{1-step-prediction-error: } e_t = x_t - \hat{X}_{t;1:(t-1)} = x_t - a_{t-1}$$

$$\hat{\alpha} = \arg \min_{\alpha} \sum_{i=2}^n e_i^2$$

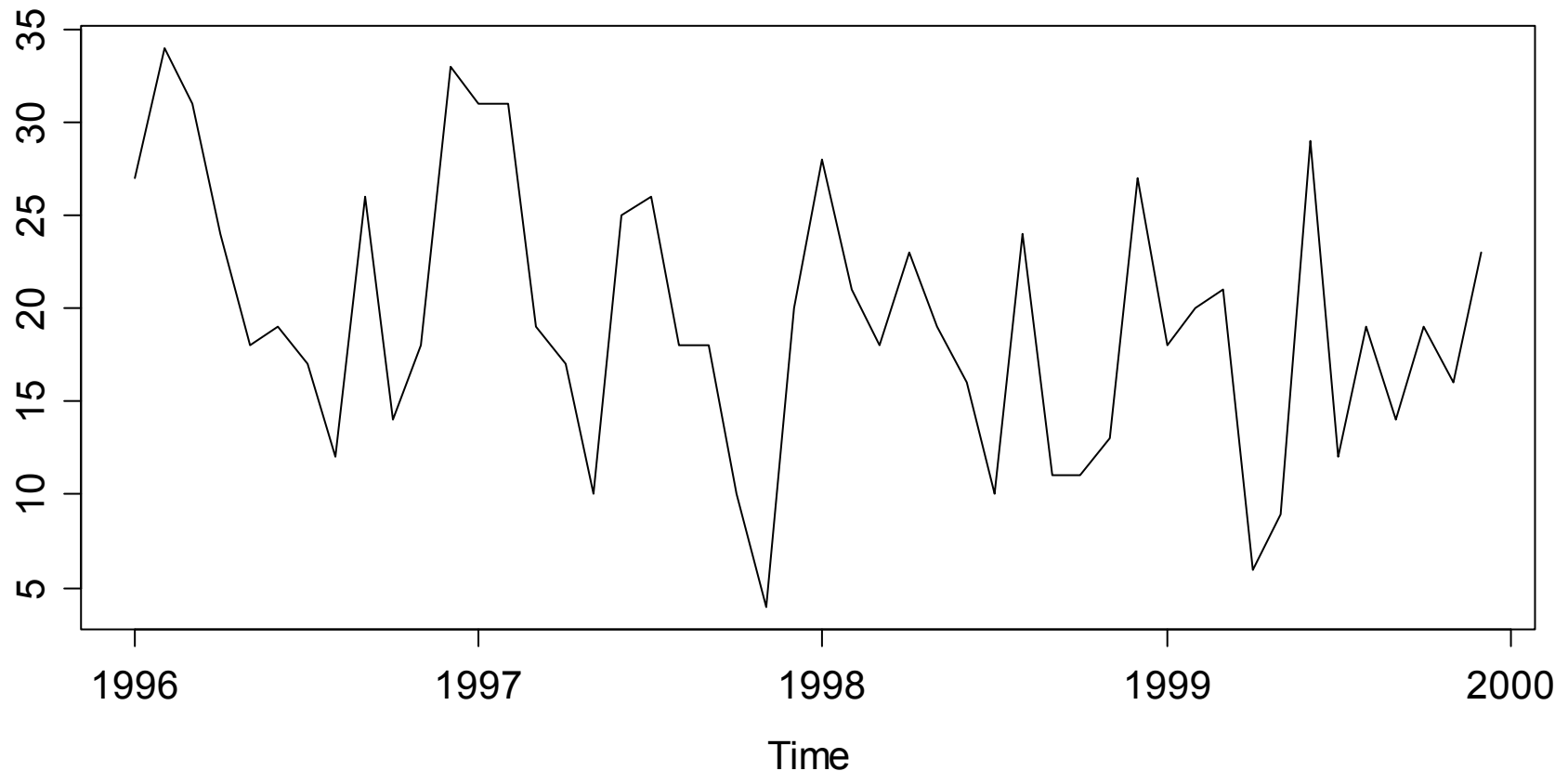
The solution needs to be found with numerical optimization.

Applied Time Series Analysis

SS 2015 – Forecasting

Exponential Smoothing: Example

Complaints to a Motorizing Organization



Applied Time Series Analysis

SS 2015 – Forecasting

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]
```

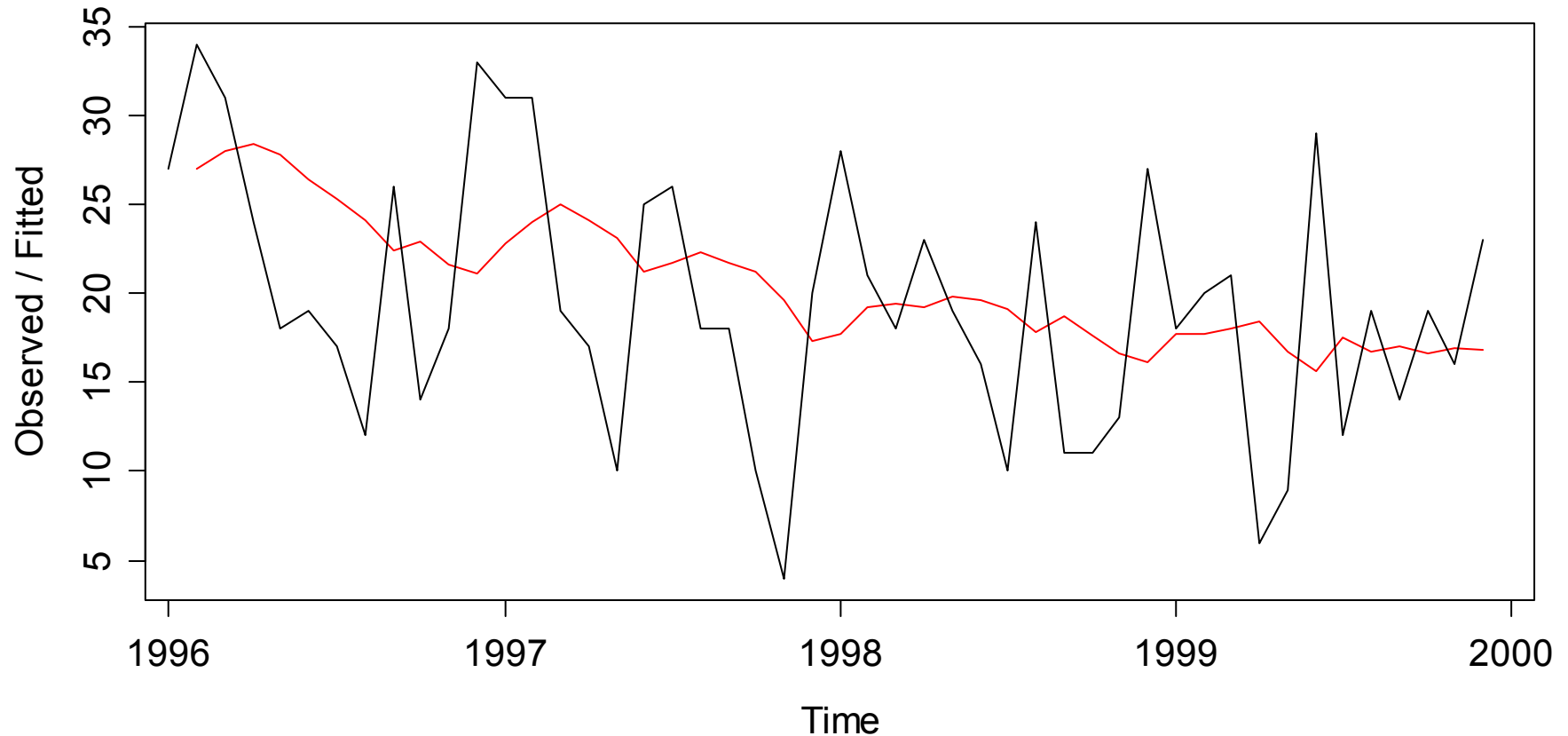
```
a 17.70343
```

Applied Time Series Analysis

SS 2015 – Forecasting

Exponential Smoothing: Example

Holt-Winters filtering



Applied Time Series Analysis

SS 2015 – Forecasting

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:

$$a_t = \alpha(x_t - s_{t-p}) + (1 - \alpha)(a_{t-1} + b_{t-1})$$

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$$

$$s_t = \gamma(x_t - a_t) + (1 - \gamma)s_{t-p}$$

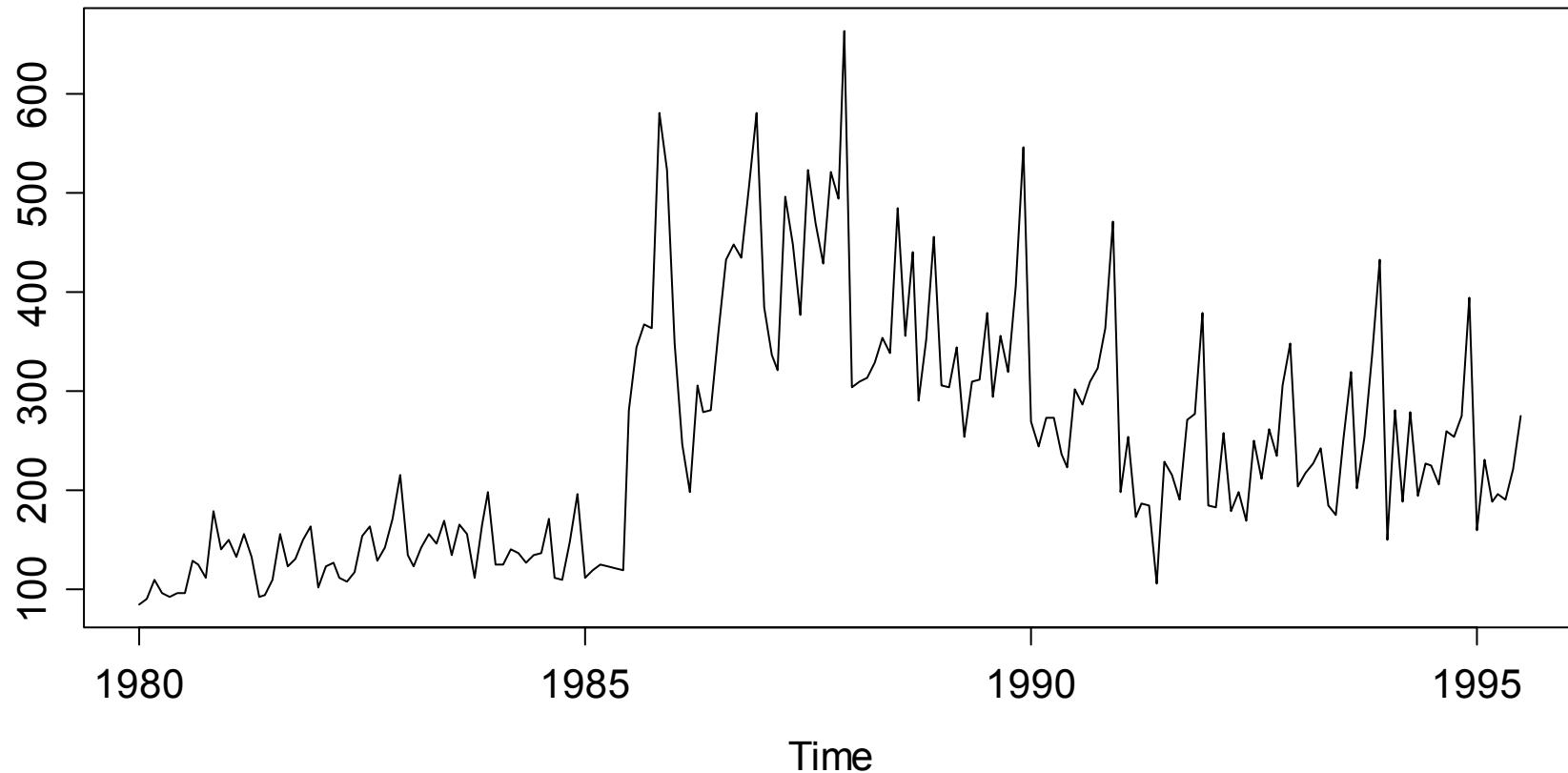
→ **See the blackboard for the derivation...**

Applied Time Series Analysis

SS 2015 – Forecasting

Holt-Winters: Example

Sales of Australian White Wine

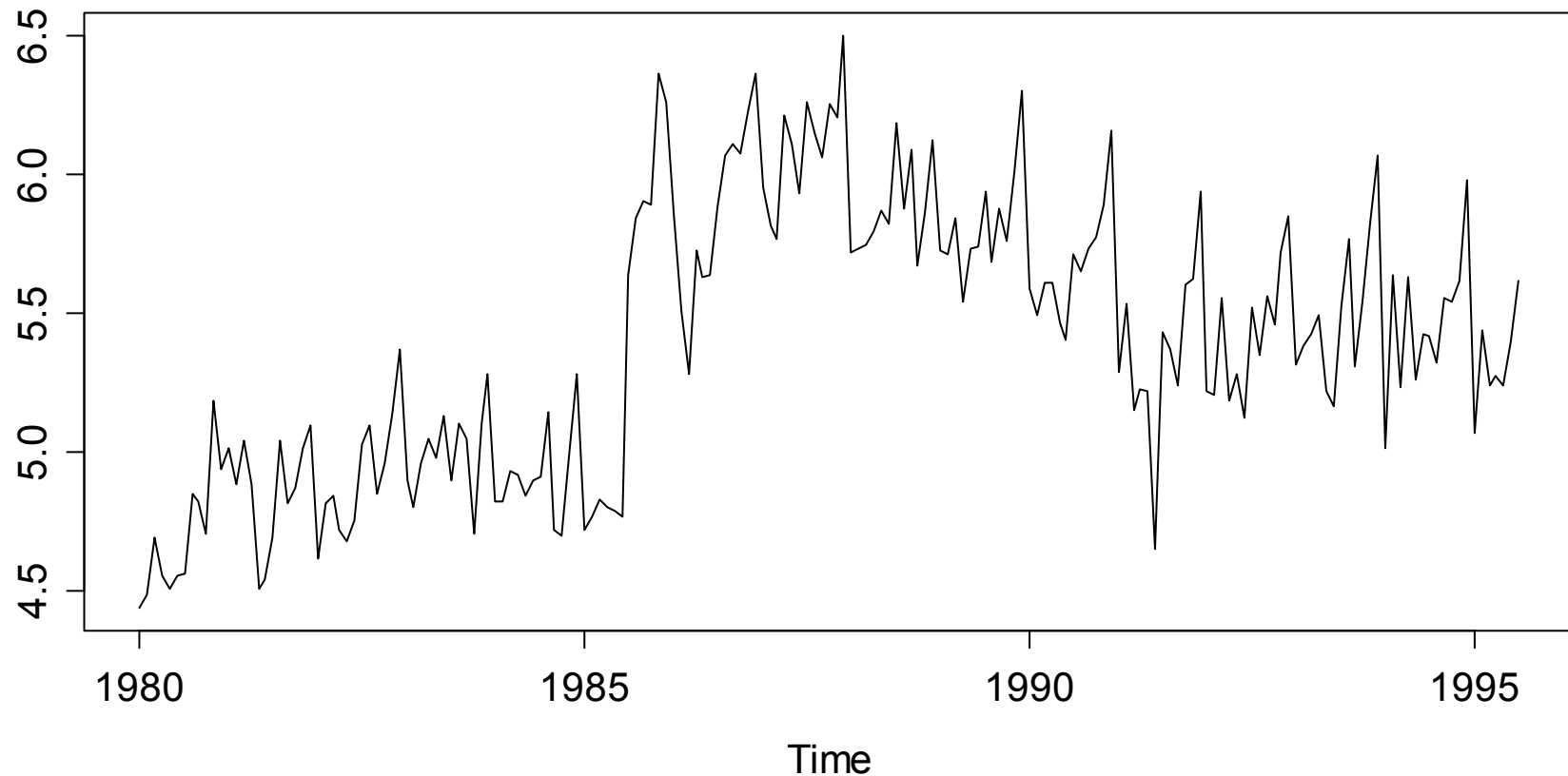


Applied Time Series Analysis

SS 2015 – Forecasting

Holt-Winters: Example

Logged Sales of Australian White Wine



Applied Time Series Analysis

SS 2015 – Forecasting

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:

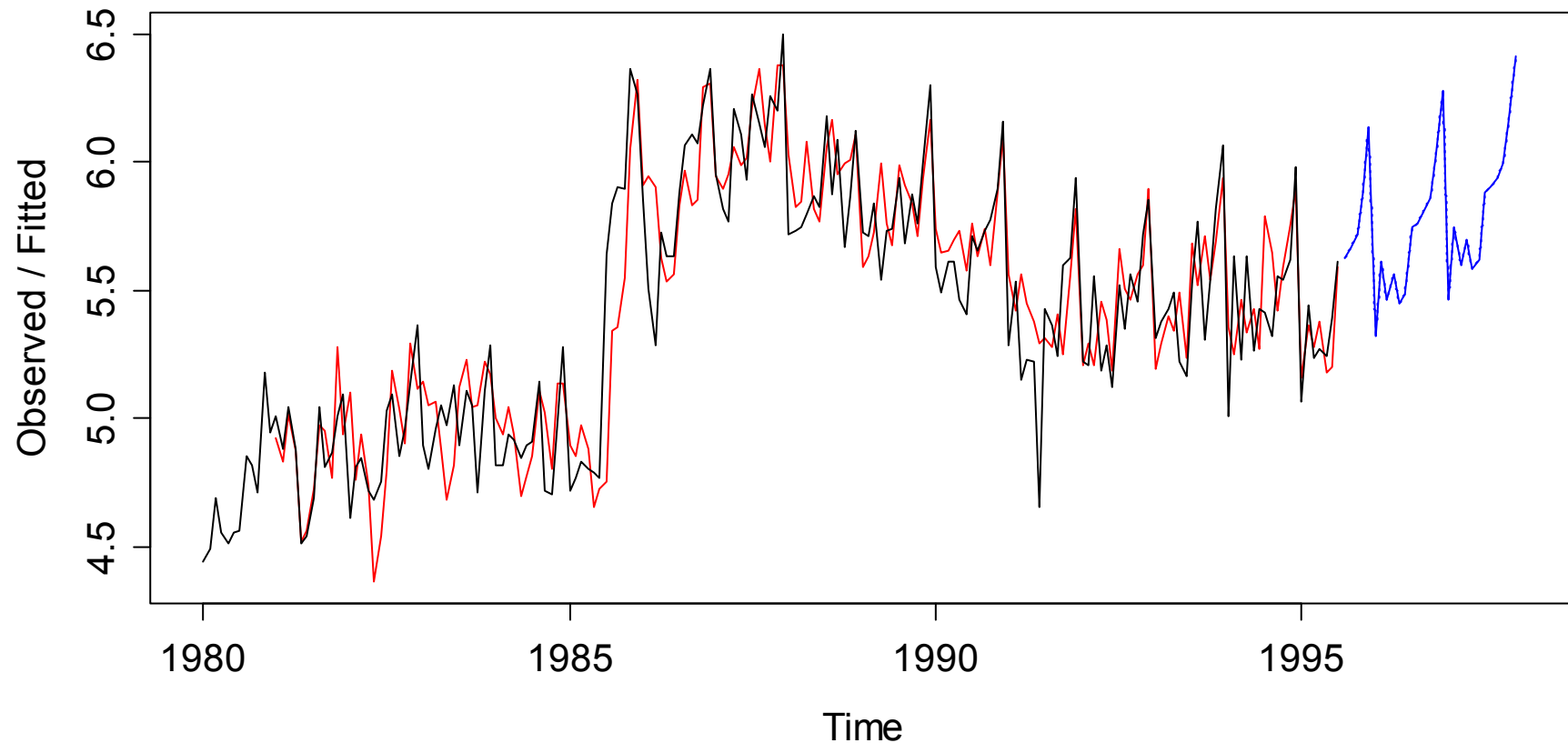
```
a      5.62591329; b      0.01148402
s1     -0.01230437; s2     0.01344762; s3     0.06000025
s4     0.20894897; s5     0.45515787; s6     -0.37315236
s7     -0.09709593; s8     -0.25718994; s9     -0.17107682
s10    -0.29304652; s11    -0.26986816; s12    -0.01984965
```


Applied Time Series Analysis

SS 2015 – Forecasting

Holt-Winters: Fitted Values & Predictions

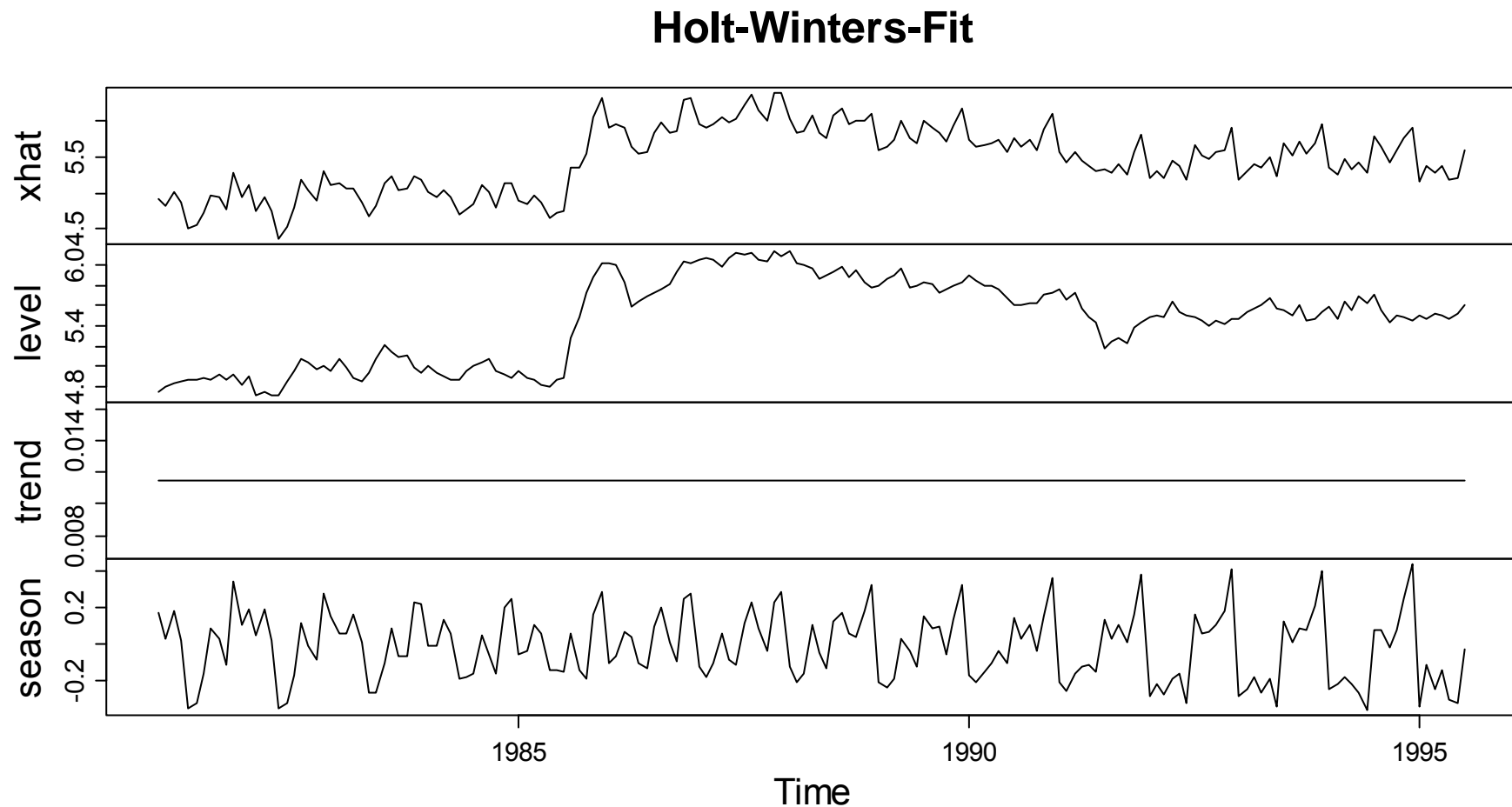
Holt-Winters filtering



Applied Time Series Analysis

SS 2015 – Forecasting

Holt-Winters: In-Sample Analysis

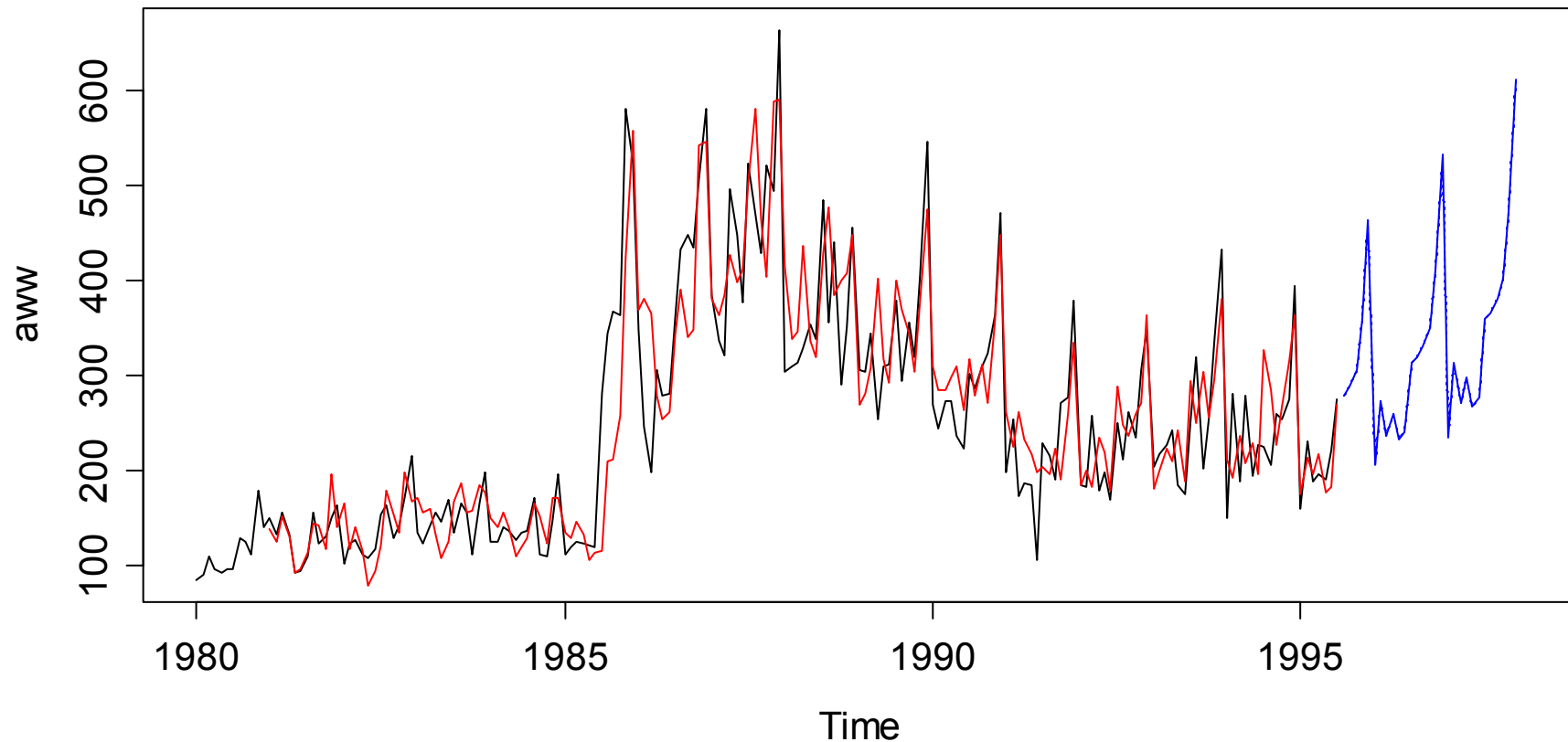


Applied Time Series Analysis

SS 2015 – Forecasting

Holt-Winters: Predictions on Original Scale

Holt-Winters-Forecast for the Original Series



Applied Time Series Analysis

SS 2015 – Forecasting

Exercise

Data:

- *use the Australian white wine sales data...*
- *... or any other dataset you like*

Goal:

- *Find a good model describing these data*
- *Evaluate which model yields the best predictions*
- *Generate a 29-month forecast from this model*

Method:

- *Remove the last 29 observations and mimic oos-forecasting*

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Multivariate Time Series Analysis

Idea: Infer the relation between two time series

$$X_1 = (X_{1,t}) \text{ and } X_2 = (X_{2,t}).$$

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.

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Example: Permafrost Boreholes



A collaboration between the Swiss Institute for Snow and Avalanche Research with the Zurich University of Applied Sciences:

Evelyn Zenklusen Mutter & Marcel Dettling

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Example: Permafrost Boreholes

- Given is a bivariate time series with 2×92 observations
- 2 measurements were made everyday in summer 2006
- Series 1: air temperature at Plathorn 3345m
- Series 2: soil temperature at Hörnli hut 3295m

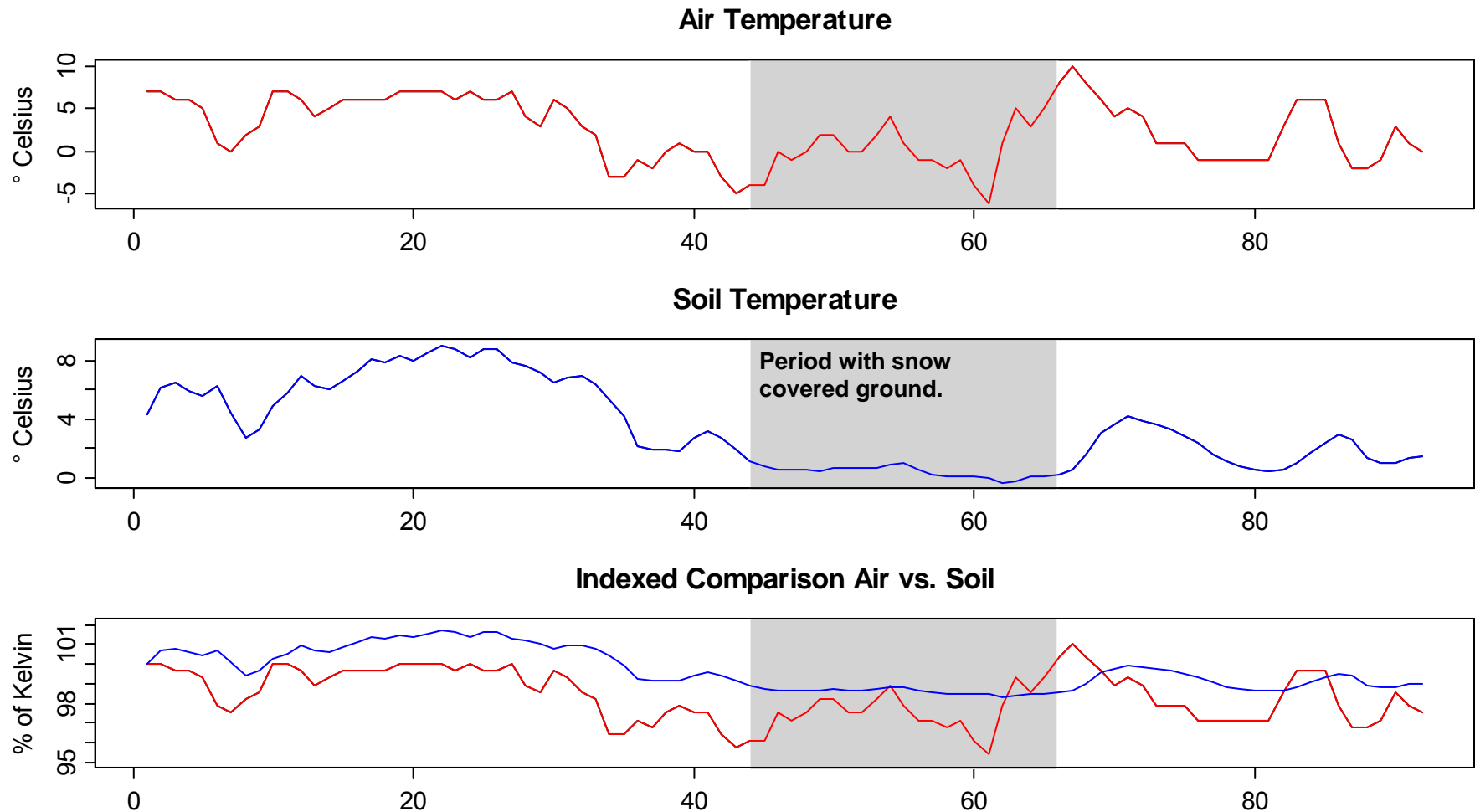
Goal of the analysis:

- 1) Answer whether changes in the air temperature are correlated with changes in the soil temperature.
- 2) If a correlation is present, what is the delay?

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Air & Soil Temperature Comparison

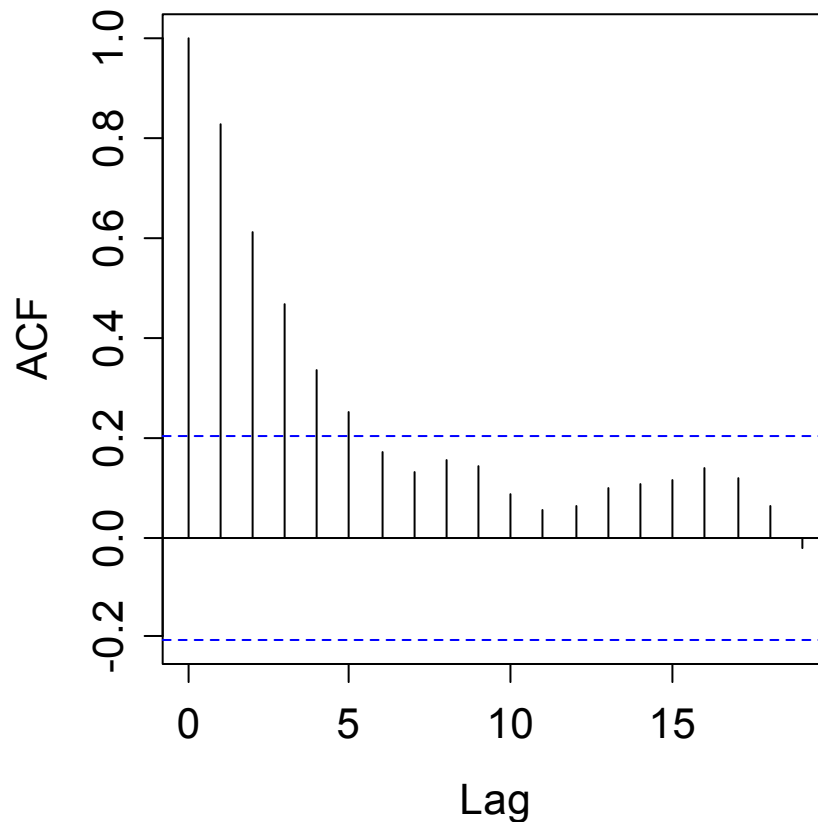


Applied Time Series Analysis

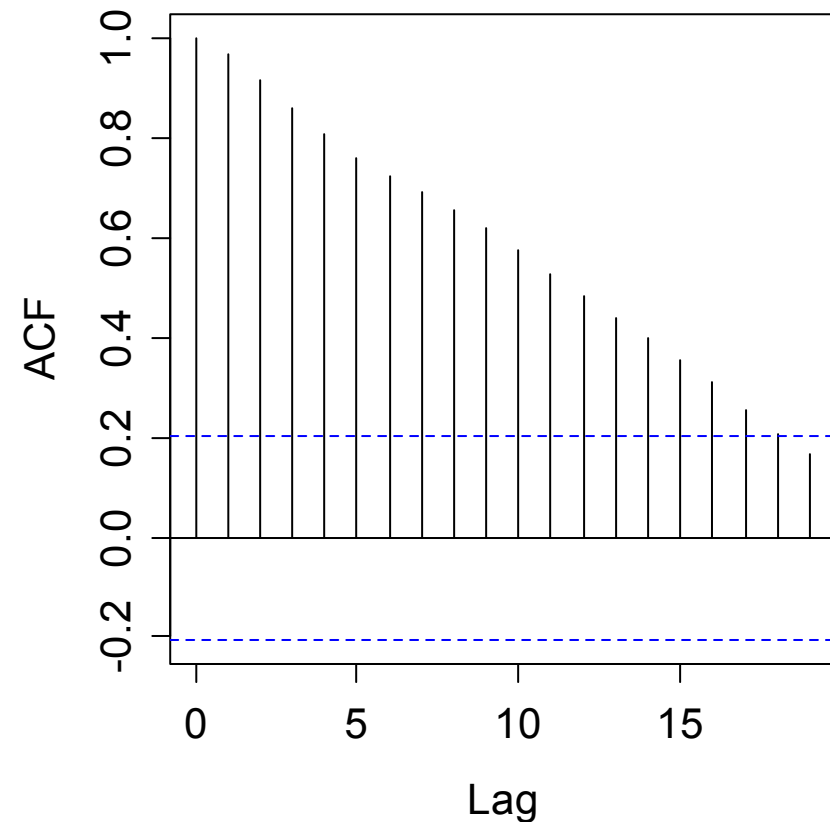
SS 2015 – Multivariate Time Series Analysis

Are the Series Stationary?

ACF of Air Temperature



ACF of Soil Temperature



Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

How to Proceed?

- 1) The series seem to have „long memory“
 - 2) Pure AR/MA/ARMA do not fit the data well
- Differencing may help with this

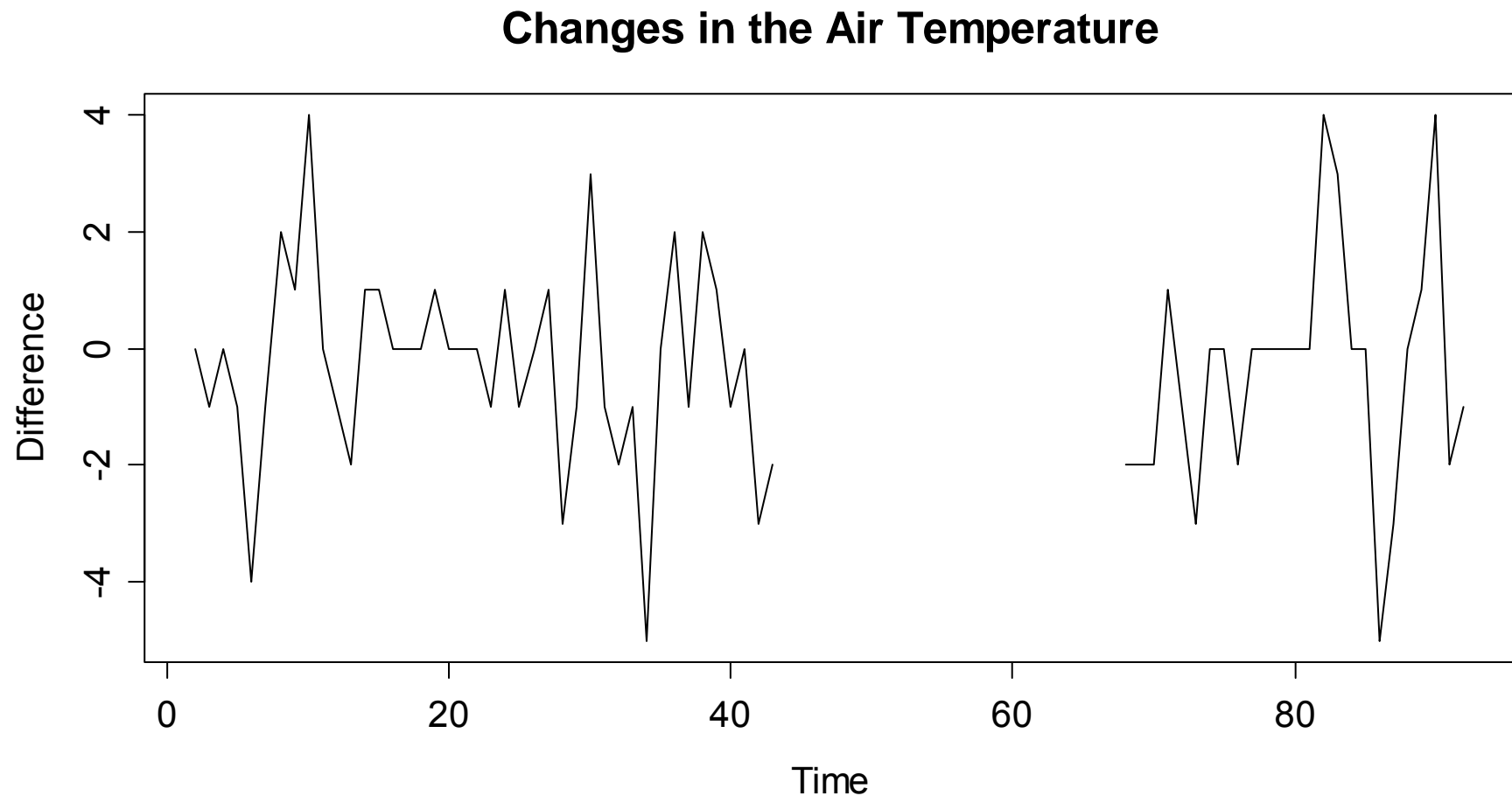
Another advantage of taking differences:

- we infer, whether there is a relation between the changes in the air temperatures, and the changes in the soil temperatures.

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Changes in the Air Temperature

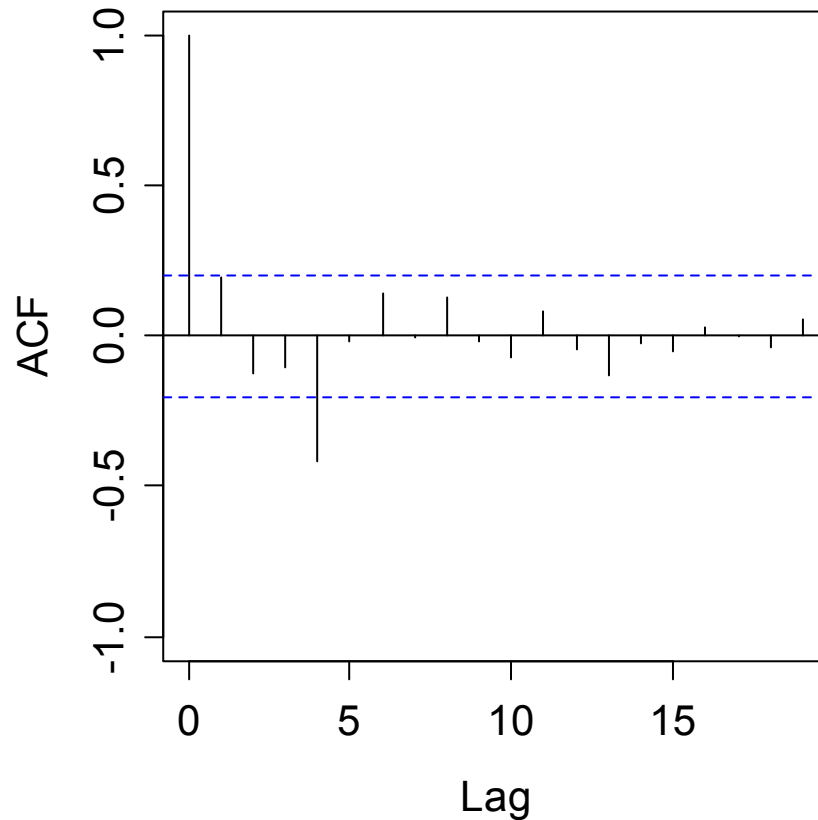


Applied Time Series Analysis

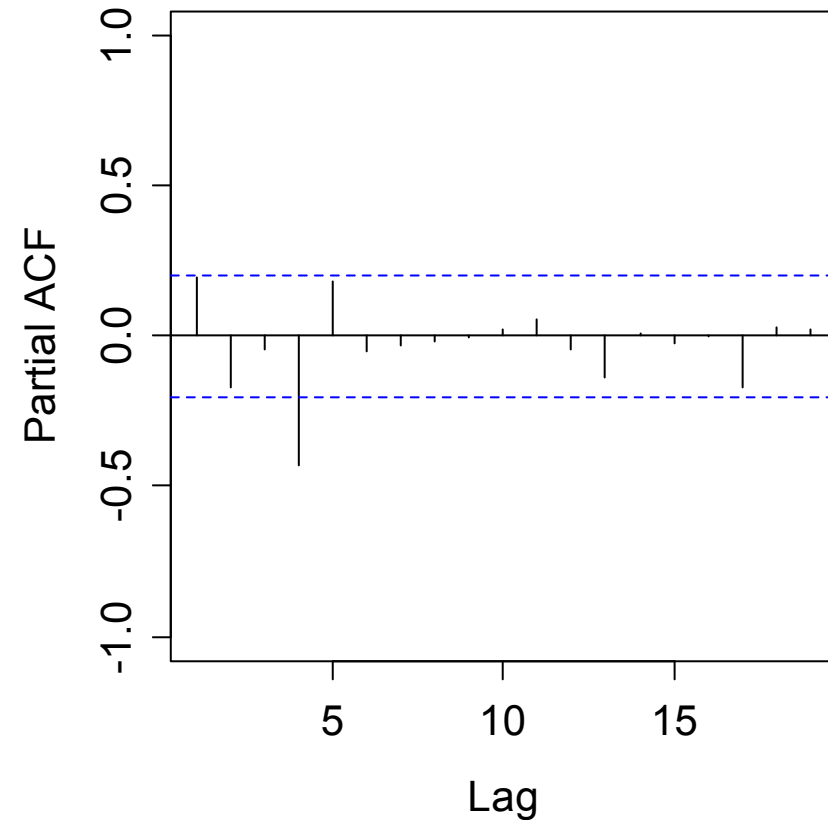
SS 2015 – Multivariate Time Series Analysis

ACF/PACF for Air Temperature Changes

ACF



PACF

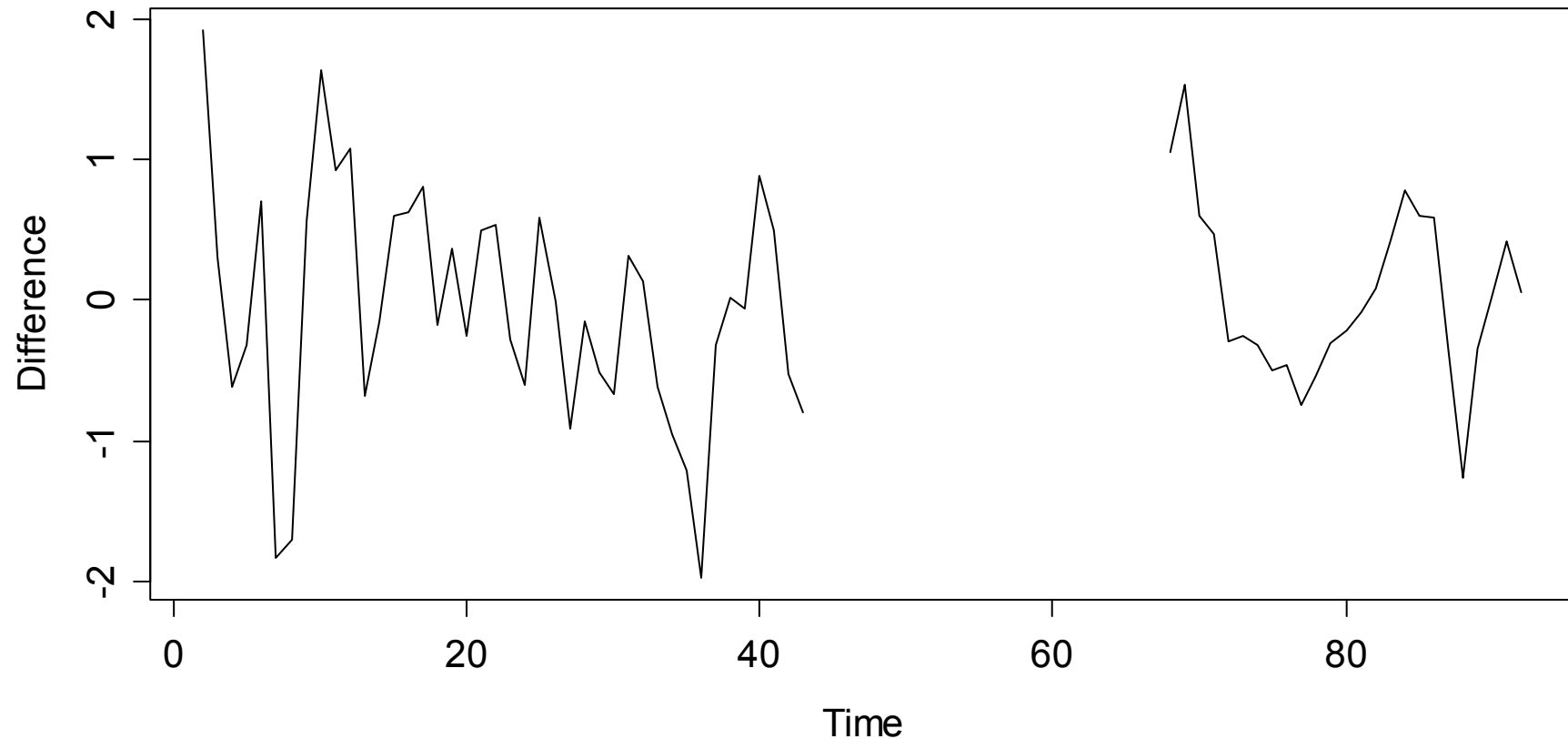


Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Changes in the Soil Temperature

Changes in the Soil Temperature

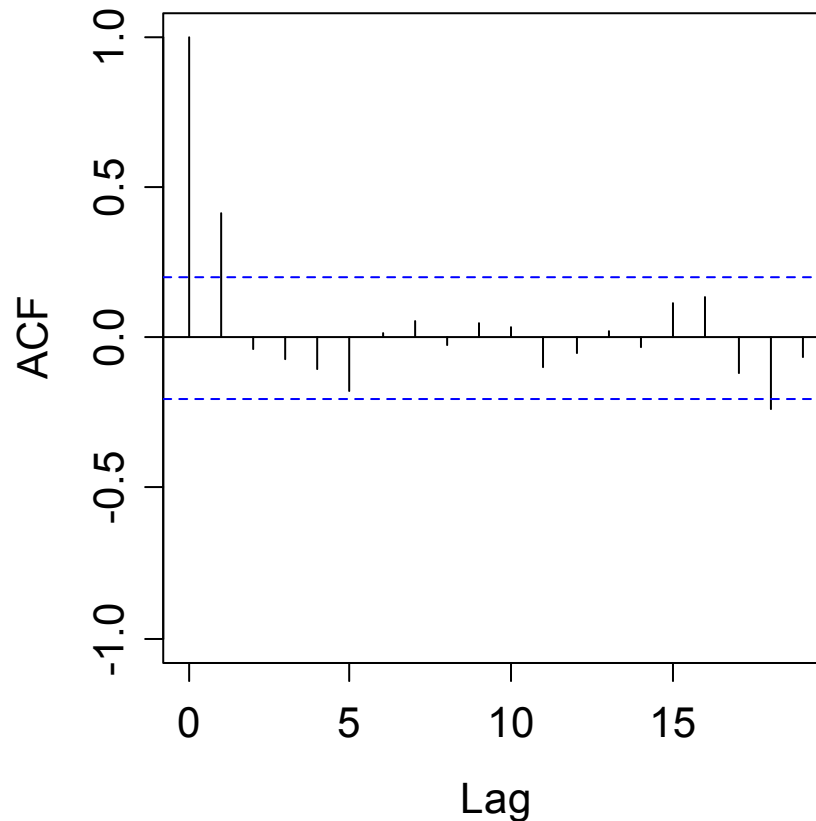


Applied Time Series Analysis

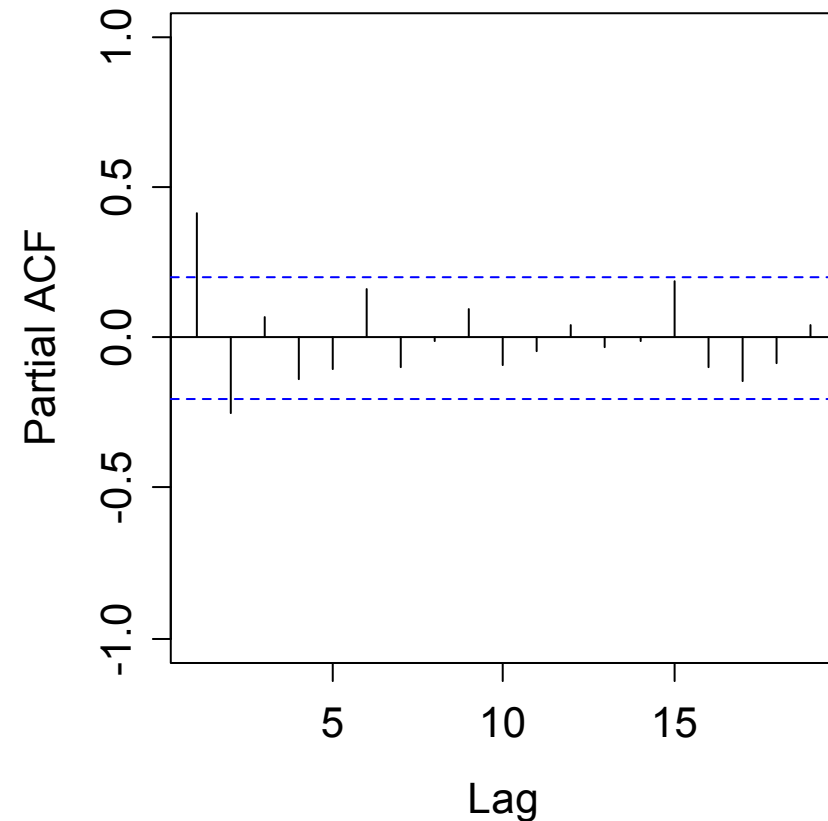
SS 2015 – Multivariate Time Series Analysis

ACF/PACF for Soil Temperature Changes

ACF



PACF



Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Cross Covariance

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

$$\left. \begin{aligned} \gamma_{11}(k) &= \text{Cov}(X_{1,t+k}, X_{1,t}) \\ \gamma_{22}(k) &= \text{Cov}(X_{2,t+k}, X_{2,t}) \end{aligned} \right\} \begin{array}{l} \text{usual „within series“} \\ \text{covariance} \end{array}$$

$$\left. \begin{aligned} \gamma_{12}(k) &= \text{Cov}(X_{1,t+k}, X_{2,t}) \\ \gamma_{21}(k) &= \text{Cov}(X_{2,t+k}, X_{1,t}) \end{aligned} \right\} \begin{array}{l} \text{cross covariance,} \\ \text{independent from } t \end{array}$$

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

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

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

SS 2015 – Multivariate Time Series Analysis

Estimation

Cross covariances and correlations are estimated as follows:

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

and

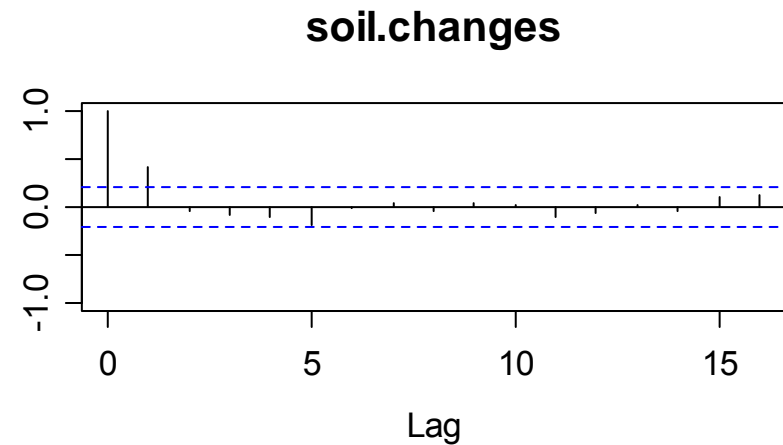
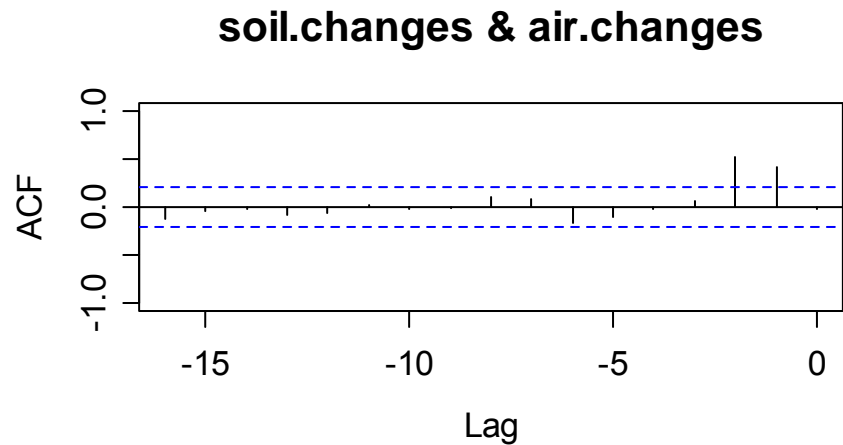
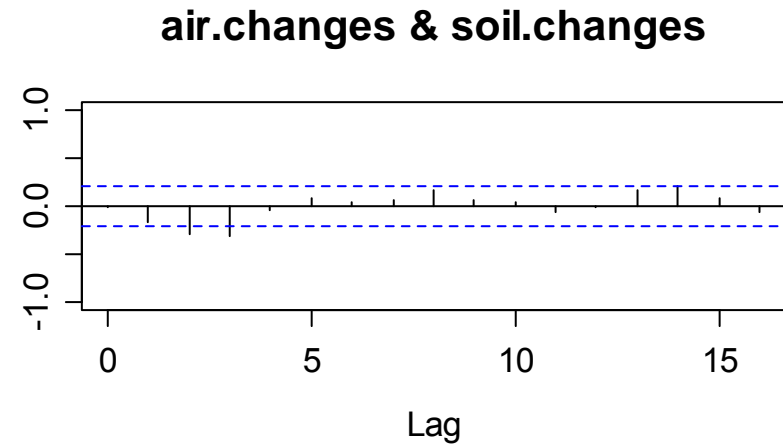
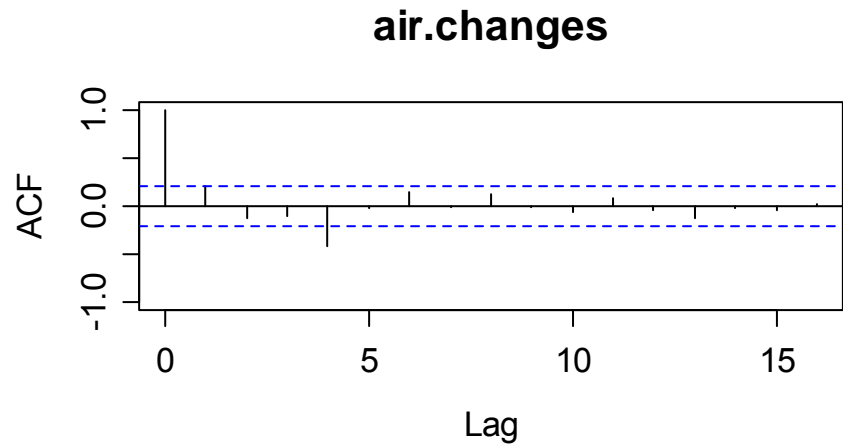
$$\hat{\rho}_{12}(k) = \frac{\hat{\gamma}_{12}(k)}{\sqrt{\hat{\gamma}_{11}(0)\hat{\gamma}_{22}(0)}}, \quad \text{respectively.}$$

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

SS 2015 – Multivariate Time Series Analysis

Sample Cross Correlation



Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

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

SS 2015 – Multivariate Time Series Analysis

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 $|k| \geq m$:

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

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.

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Special Case 2

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

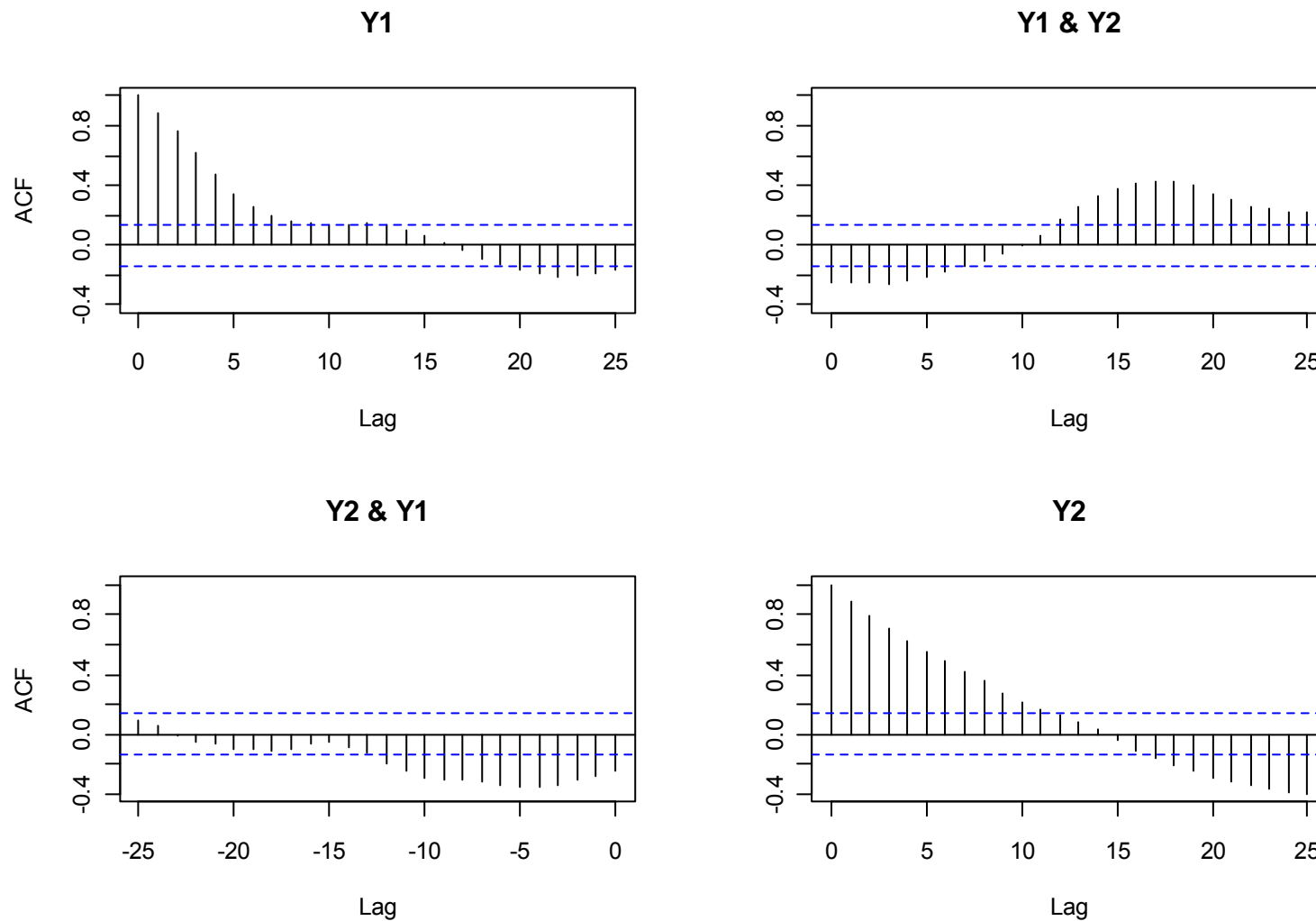
$$\text{Var}(\hat{\rho}_{12}(k)) \approx \frac{1}{n} \sum_{j=-\infty}^{\infty} \rho_{11}(j) \rho_{22}(j)$$

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

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Special Case 2: Simulation Example



Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

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:

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

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.

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

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.

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Cross Correlation of Prewhitened Series

The cross correlation between U_t and V_t can be derived from the one between X_1 and X_2 :

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

Thus we have:

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

Now: generate U_t, V_t ; estimate cross correlations; and, by using the confidence bands, check whether they are significant

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Simulation Example

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

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

In practice, we don't know the AR-coefficients, but plug-in the respective estimates:

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

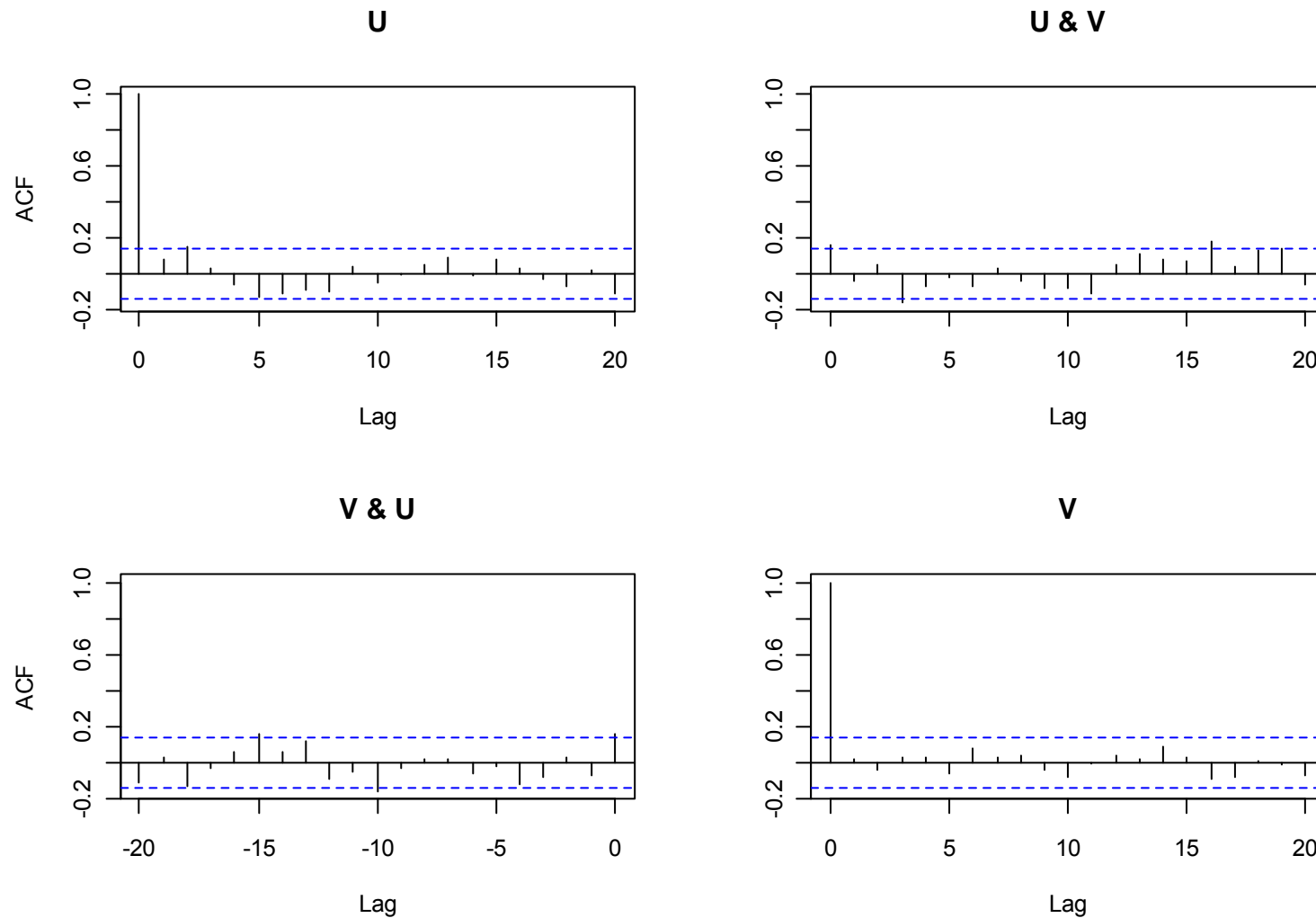
$$V_t = X_{2,t} - \hat{\alpha}_{2,1} X_{2,t-1} \quad \text{with } \hat{\alpha}_{2,1} = 0.822$$

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 .

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Cross Correlation in the Simulation Example



Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

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.

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Prewhitening the Borehole Data

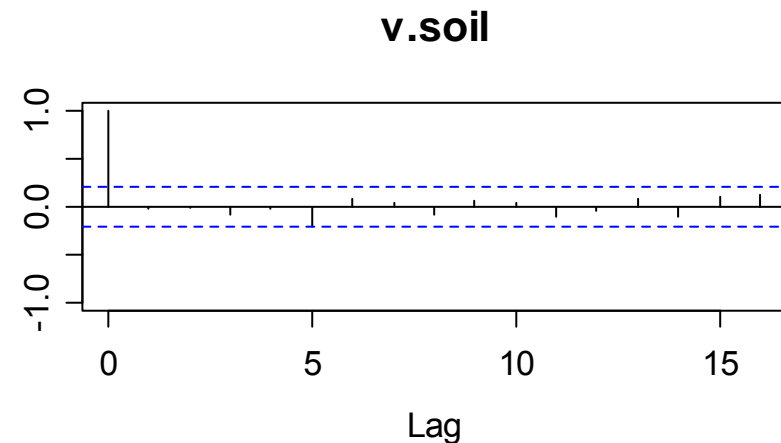
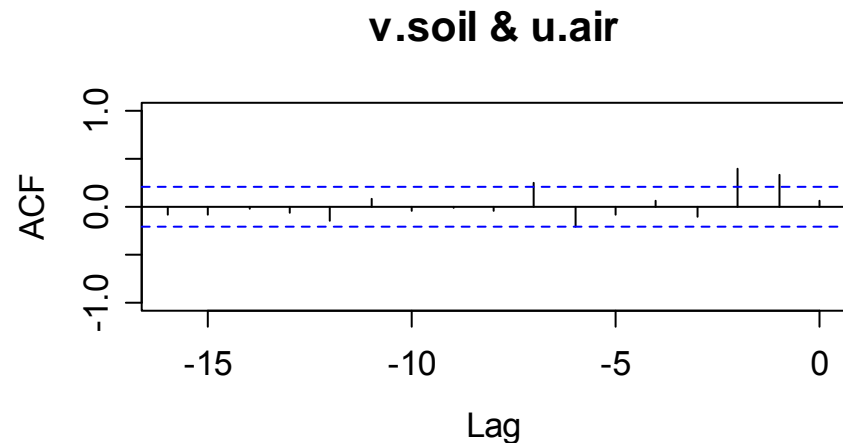
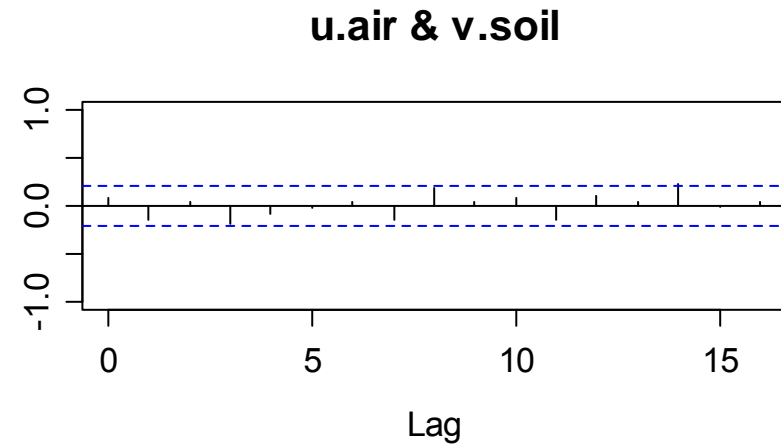
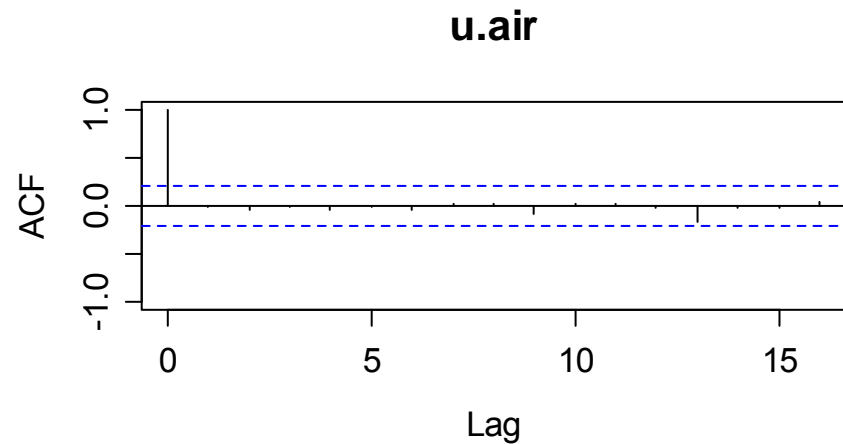
What to do:

- ARMA(p,q)-models are fitted to the differenced series
- Best choice: AR(5) for the air temperature differences
MA(1) for the soil temperature differences
- The residual time series are U_t and W_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.

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Prewhitening the Borehole Data



Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

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:

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

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Transfer Function Models

The model is:

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

- $E[E_t] = 0$
- 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

SS 2015 – Multivariate Time Series Analysis

Cross Covariance

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

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

- 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{v}_j .
- **This is not a statistically efficient estimation method.**

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

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:

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

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

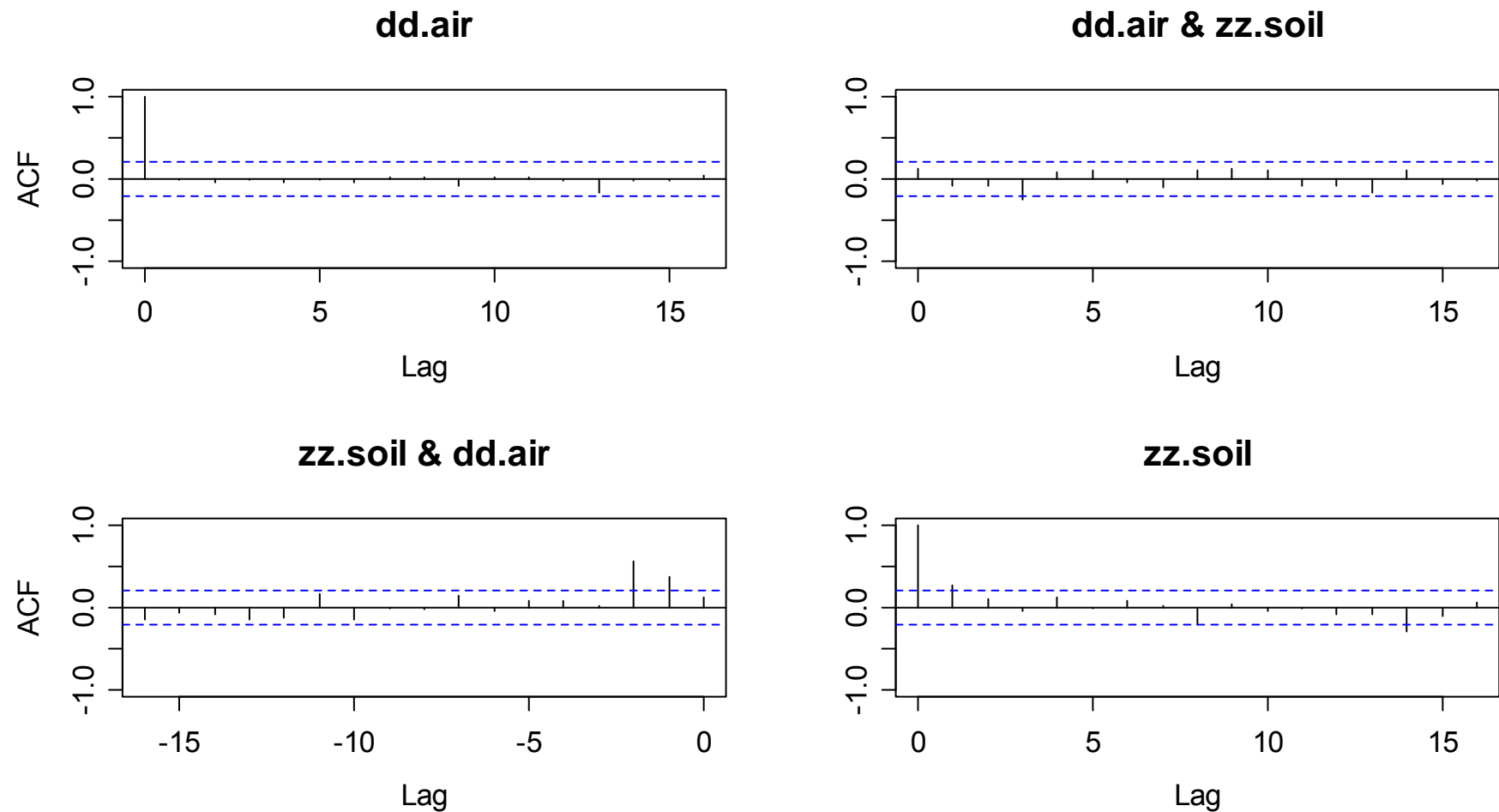
- transform all series in a clever way
- the transfer function model has identical coefficients
- the new, transformed input series is uncorrelated

→ see blackboard for the transformation

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Borehole Transformed



Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Borehole: 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 1-2 days.
- The formula for the transfer function model coefficients is:

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

Applied Time Series Analysis

SS 2015 – Multivariate Time Series Analysis

Borehole: R-Code and Results

```
> dd.air <- resid(fit.air)
> coefs <- coef(fit.air)[1:5]
> zz.soil <- filter(diff(soil.na), c(1, -coefs, sides=1))
> as.int <- ts.intersect(dd.air, zz.soil)
> acf.val <- acf(as.int, na.action=na.pass)
```

Transfer Function Model Coefficients:

```
> multip <- sd(zz.soil, na.rm=..)/sd(dd.air, na.rm=..)
> multip*acf.val$acf[,2,1]
[1] 0.054305137 0.165729551 0.250648114 0.008416697
[5] 0.036091971 0.042582917 -0.014780751 0.065008411
[9] -0.002900099 -0.001487220 -0.062670672 0.073479065
[13] -0.049352348 -0.060899602 -0.032943583 -0.025975790
[17] -0.057824007
```

Applied Time Series Analysis

SS 2015 – Spectral Analysis

Spectral Analysis

Idea: Time series are interpreted as a combination of cyclic components, and thus, a linear combination of harmonic oscillations.

Why: As a descriptive means, showing the character and the dependency structure within the series.

What: It is in spirit, but also mathematically, closely related to the correlogram

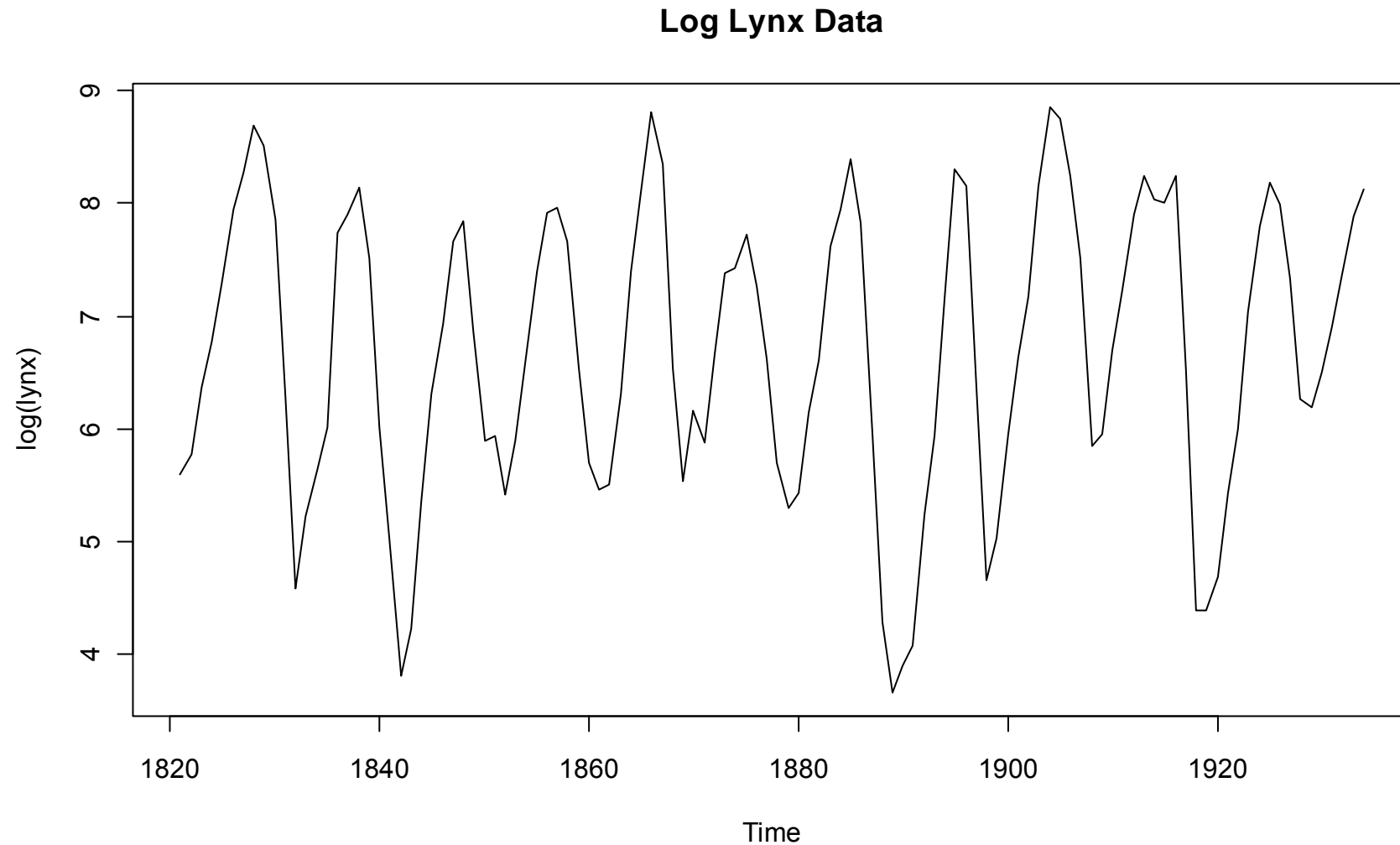
Where:

- engineering
- economics
- biology/medicine

Applied Time Series Analysis

SS 2015 – Spectral Analysis

Lynx Data

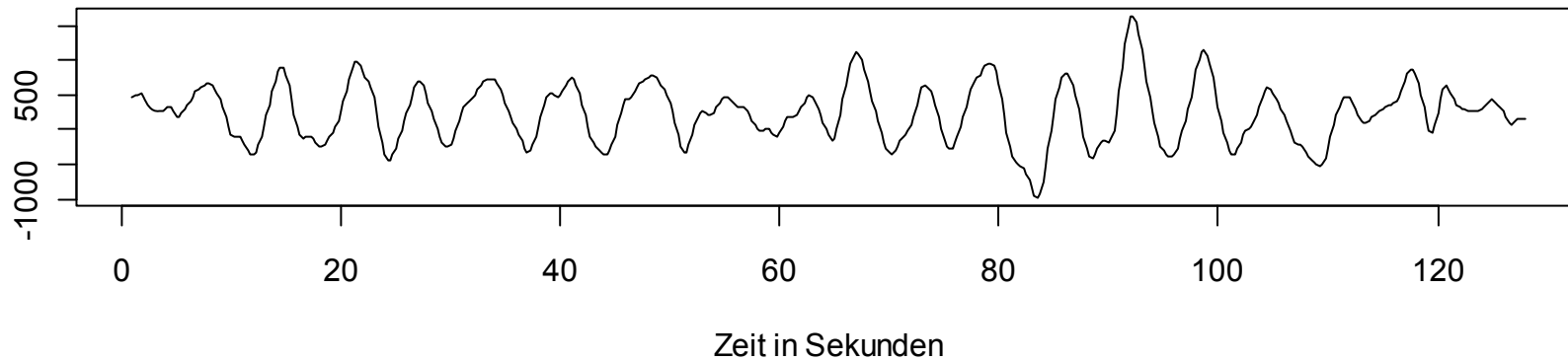


Applied Time Series Analysis

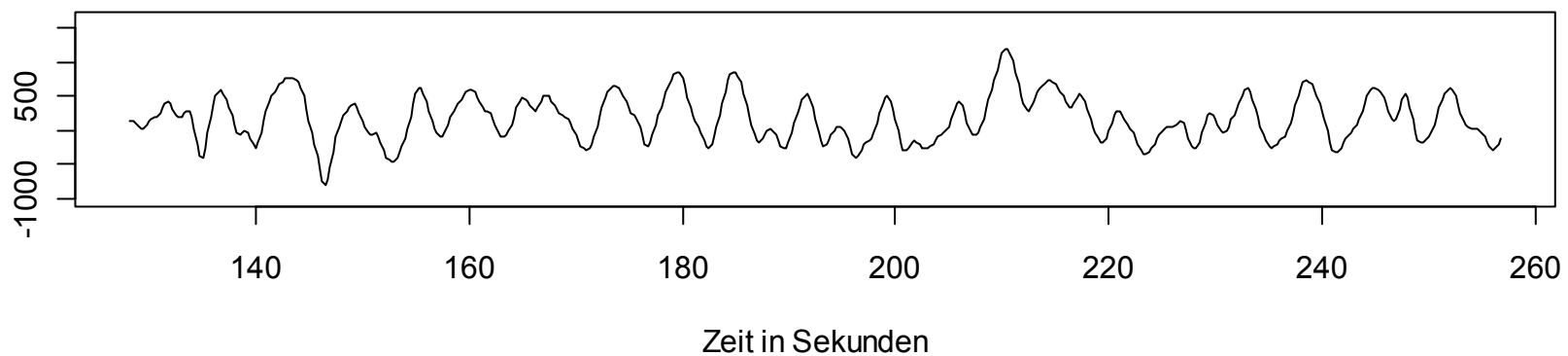
SS 2015 – Spectral Analysis

Ocean Wave Data

Ocean Wave Height Data, Part 1



Ocean Wave Height Data, Part 2

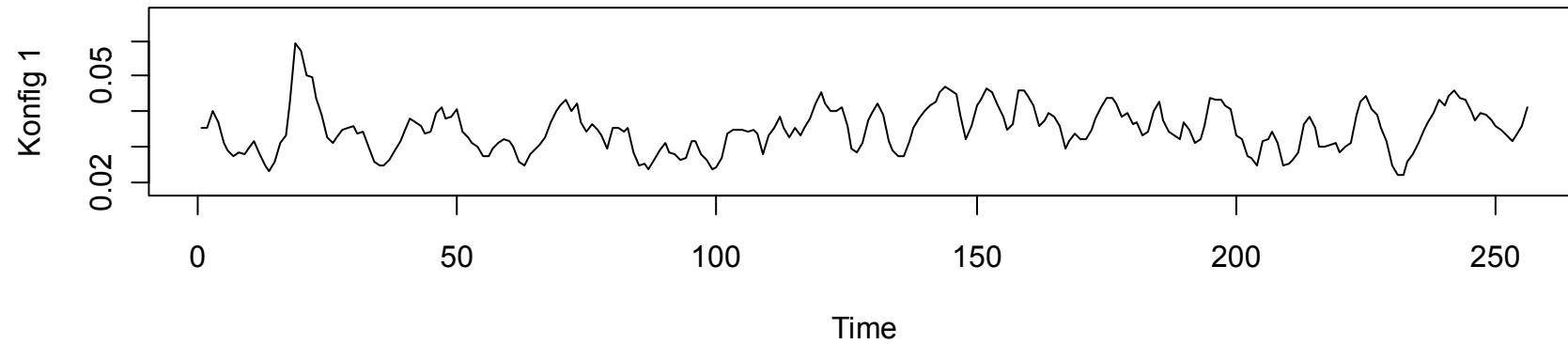


Applied Time Series Analysis

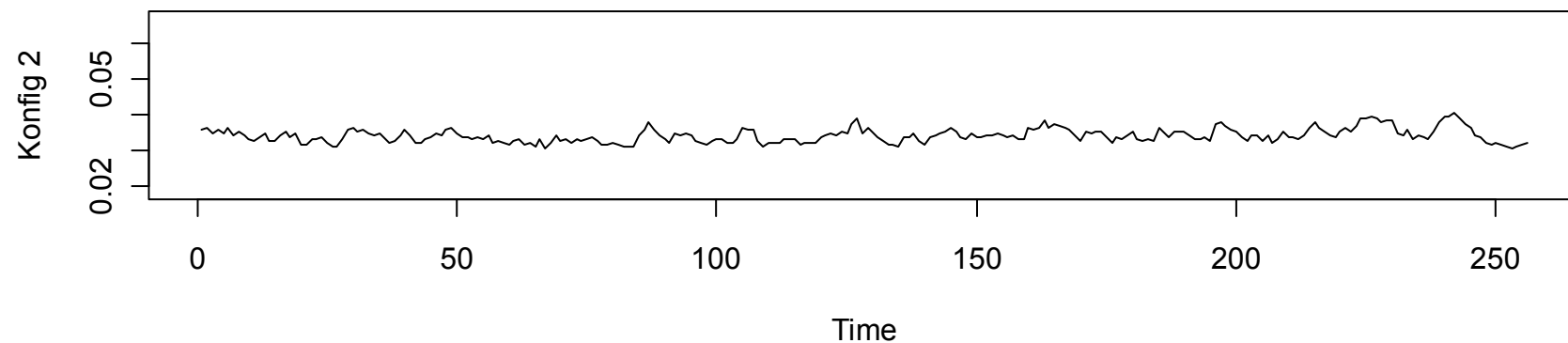
SS 2015 – Spectral Analysis

2-Component-Mixture Data

2-Component-Mixture: Series 1



2-Component-Mixture: Series 2



Applied Time Series Analysis

SS 2015 – Spectral Analysis

Harmonic Oscillations

The most simple periodic functions are sine and cosine, which we will use as the basis of our analysis.

A harmonic oscillation has the following form:

$$y(t) = \alpha \cos(2\pi\nu t) + \beta \sin(2\pi\nu t)$$

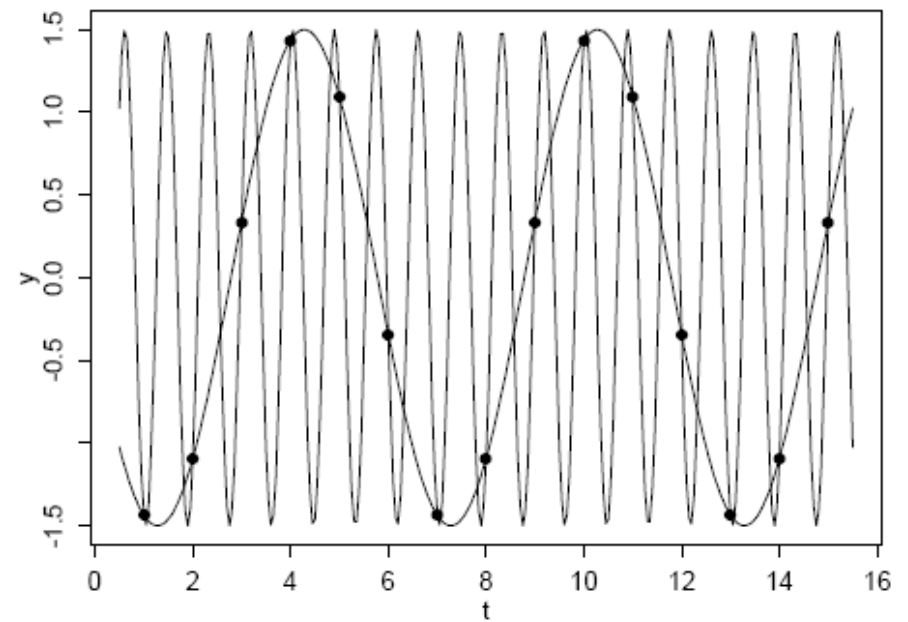
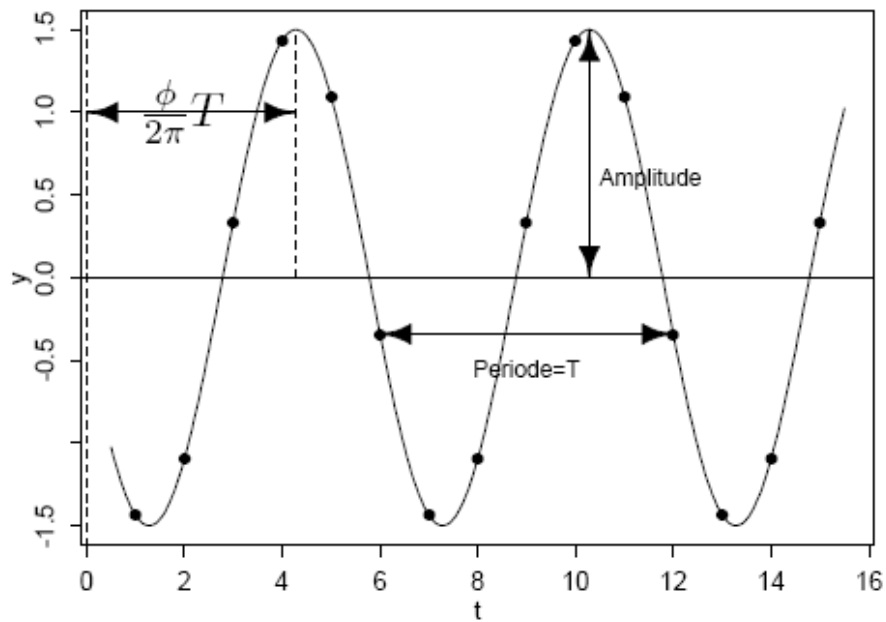
For the derivation, [see the blackboard...](#)

- In discrete time, we have aliasing, i.e. some frequencies cannot be distinguished (\rightarrow see next slide).
- The periodic analysis is limited to frequencies between 0 and 0.5, i.e. things we observe at least twice.

Applied Time Series Analysis

SS 2015 – Spectral Analysis

Aliasing



Applied Time Series Analysis

SS 2015 – Spectral Analysis

Regression Model & Periodogram

We try to write a time series with a regression equation containing sine and cosine terms at the fourier frequencies.

→ **see the blackboard**

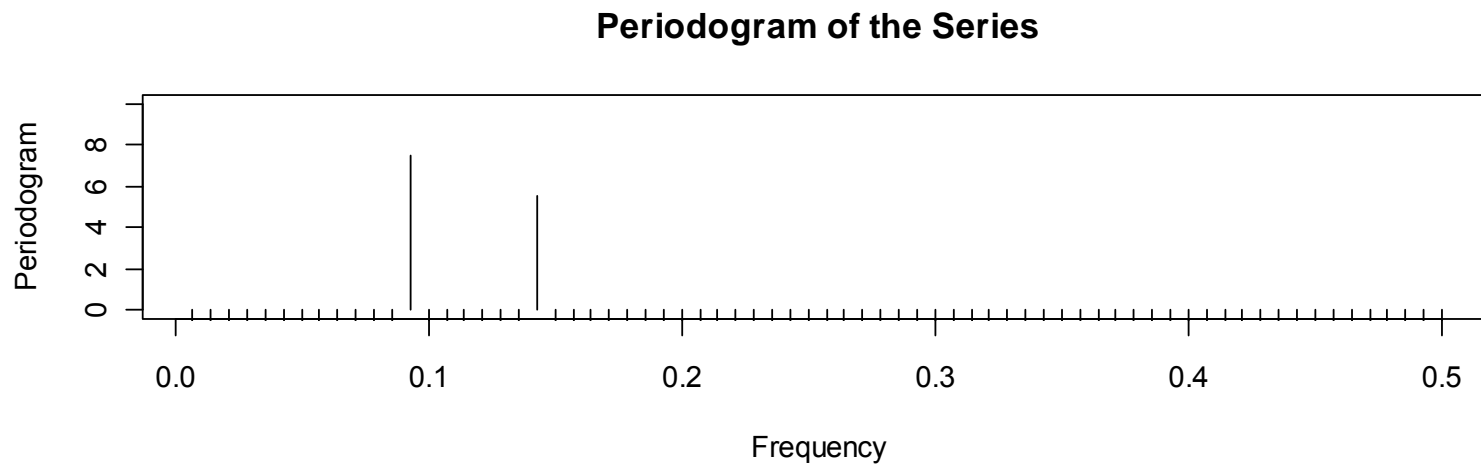
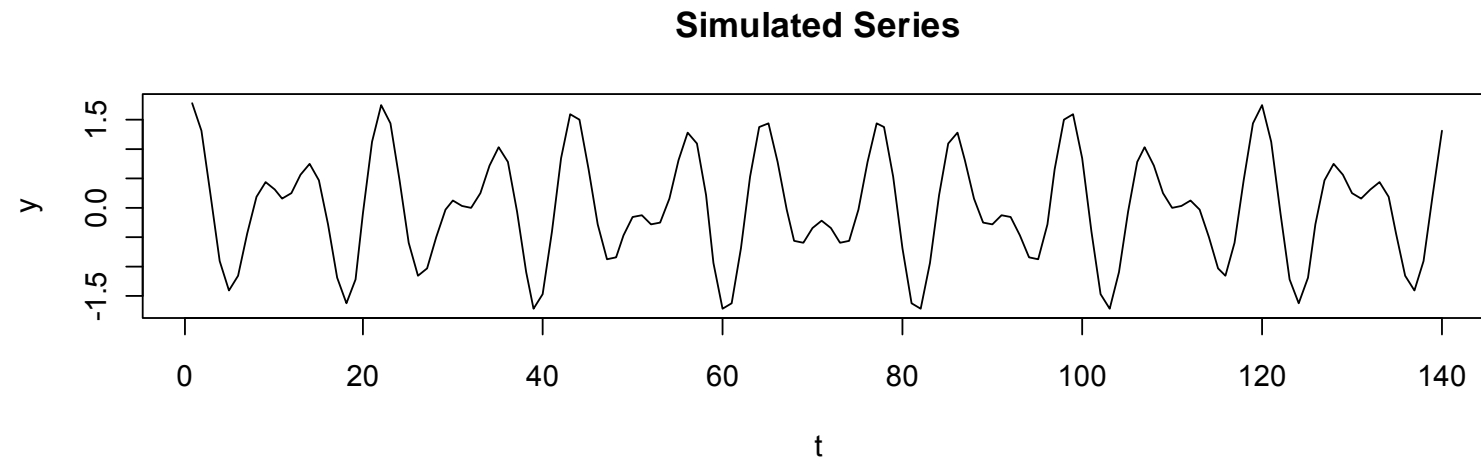
The most important frequencies within the series, which when omitted, lead to pronounced increase in goodness-of-fit.

- This idea is used as a proxy for the periodogram,
→ **see the blackboard...**
- However, if the „true“ frequency is not a fourier frequency, we have leakage (→ see next 2 slides).

Applied Time Series Analysis

SS 2015 – Spectral Analysis

Periodogram of a Simulated Series

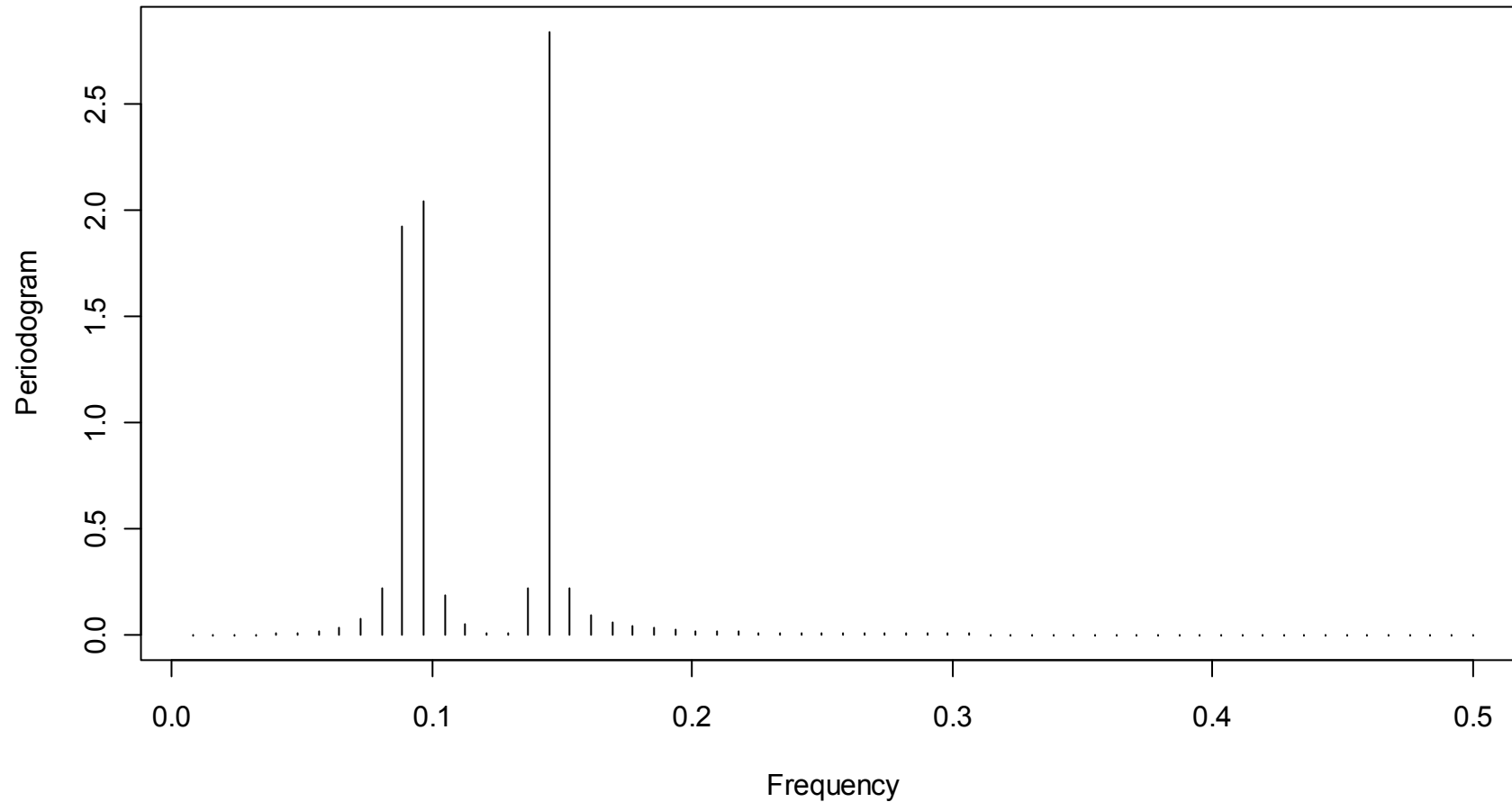


Applied Time Series Analysis

SS 2015 – Spectral Analysis

Periodogram of the Shortened Series

Periodogram of the Shortened Series



Applied Time Series Analysis

SS 2015 – Spectral Analysis

Properties of the Periodogram

Periodogram and correlogram are mathematically equivalent, the former is the **fourier transform** of the latter.

→ **see the blackboard** for the derivation

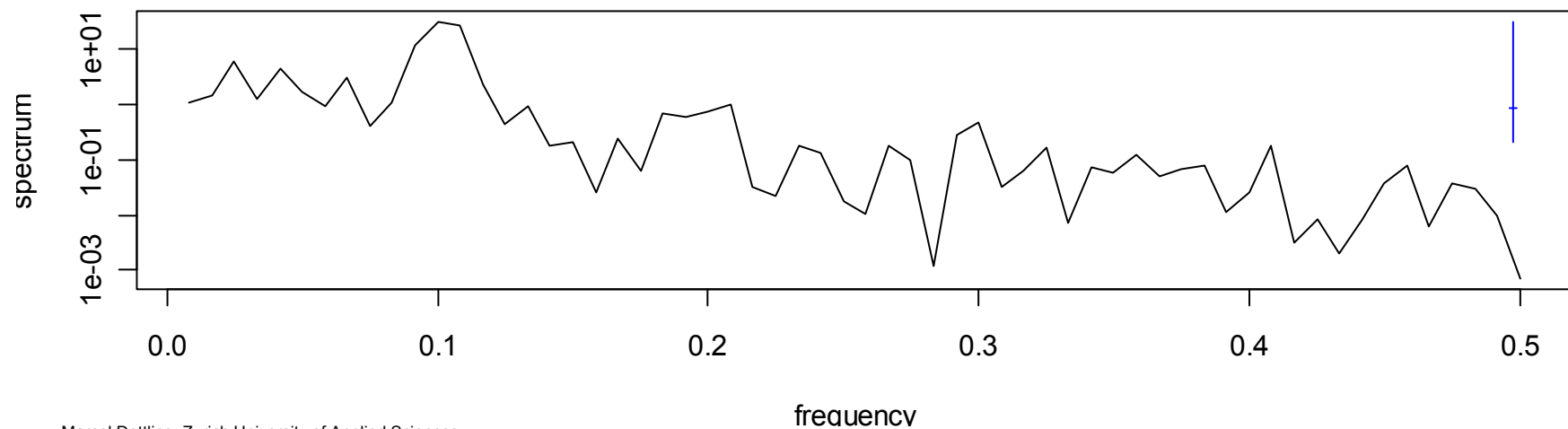
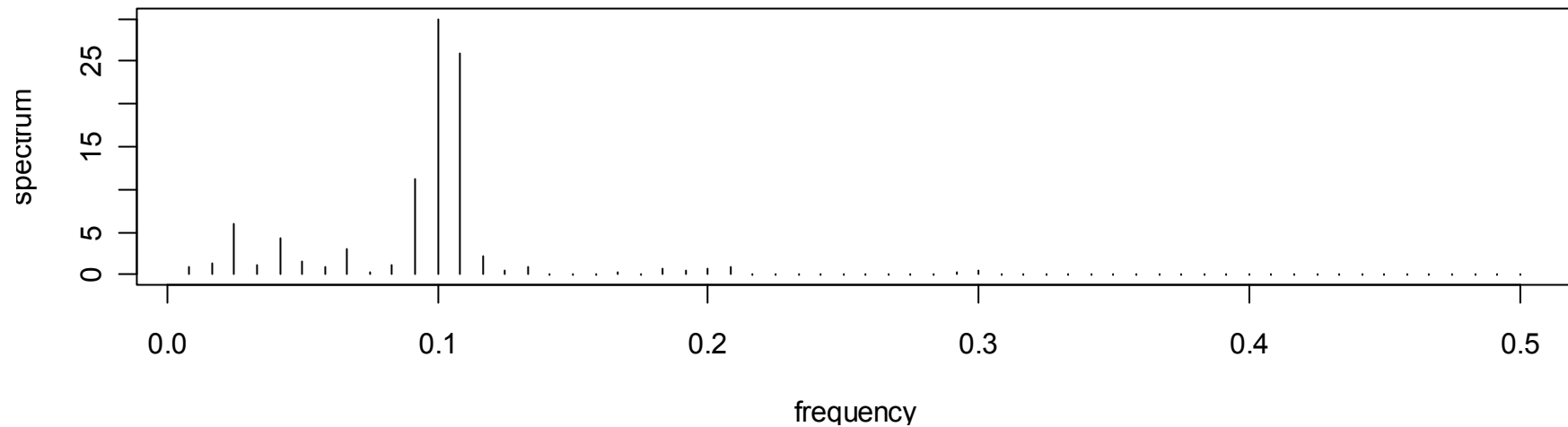
Note: this is a reason why we divided by $1/n$ in the ACV.

- $I(\nu_k)$ or $\log(I(\nu_k))$ are plotted against k/n
- Estimates seem rather instable and noisy
- On the log-scale, most frequencies are present
- It seems as if smoothing is required for interpretation.

Applied Time Series Analysis

SS 2015 – Spectral Analysis

Periodogram of the Log Lynx Data

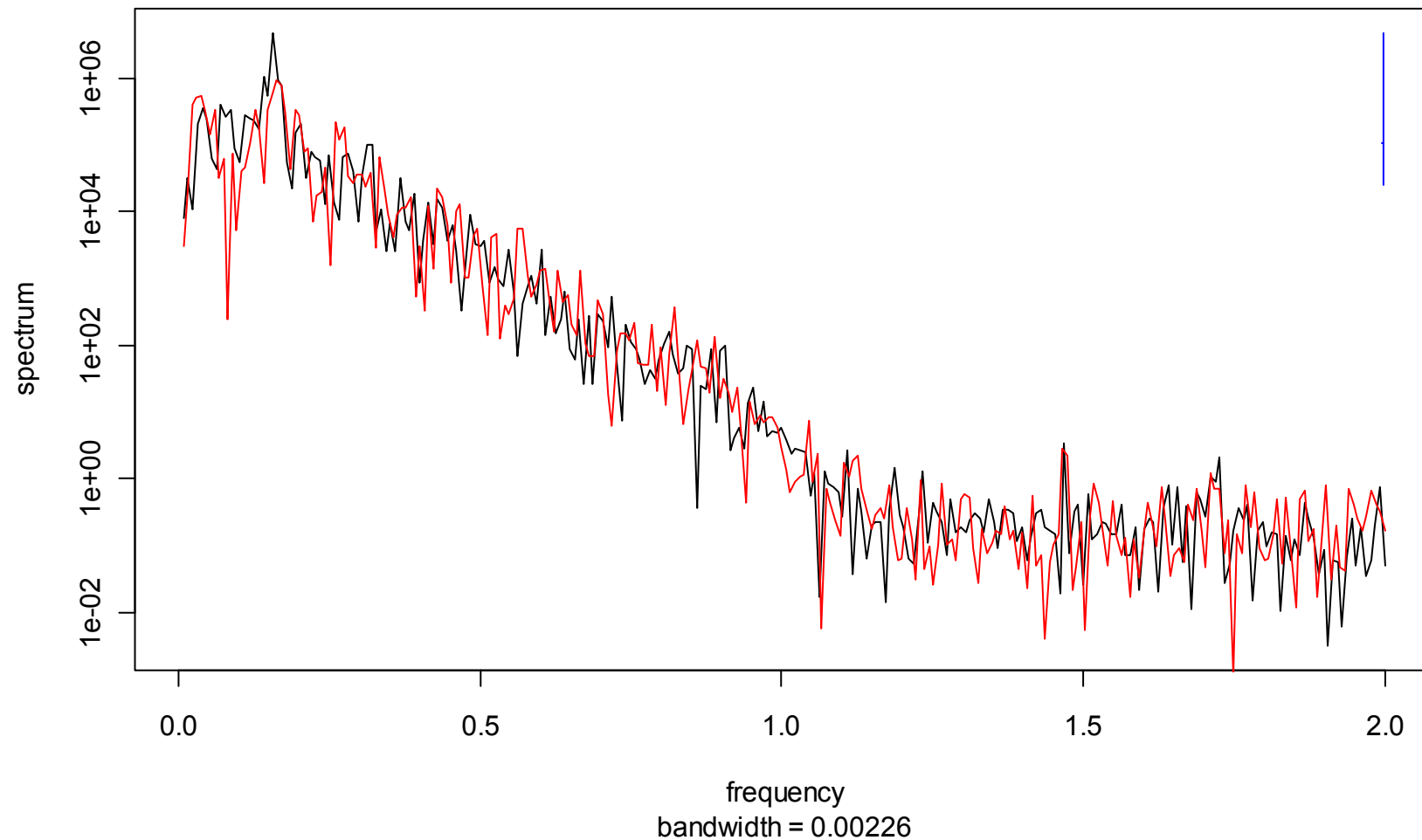


Applied Time Series Analysis

SS 2015 – Spectral Analysis

Periodogram of the Ocean Wave Data

Periodogram of the Ocean Wave Data

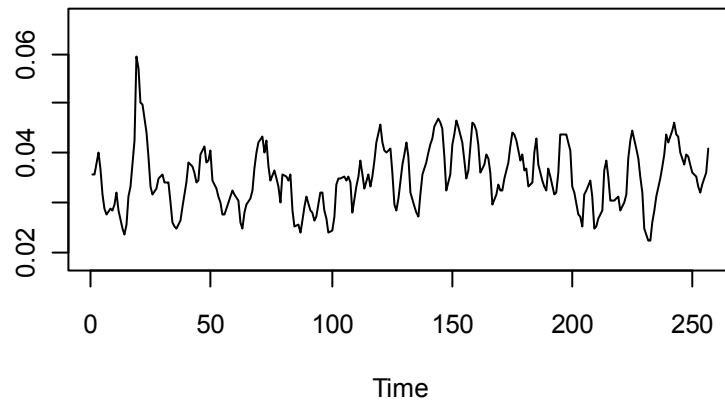


Applied Time Series Analysis

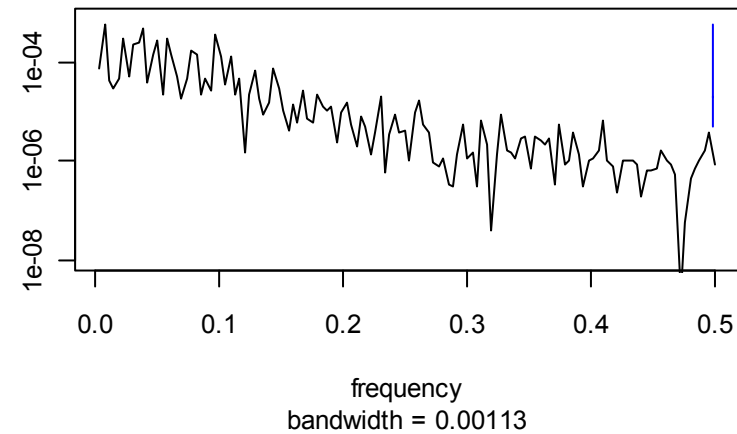
SS 2015 – Spectral Analysis

Periodogram of the 2-Component-Mixture

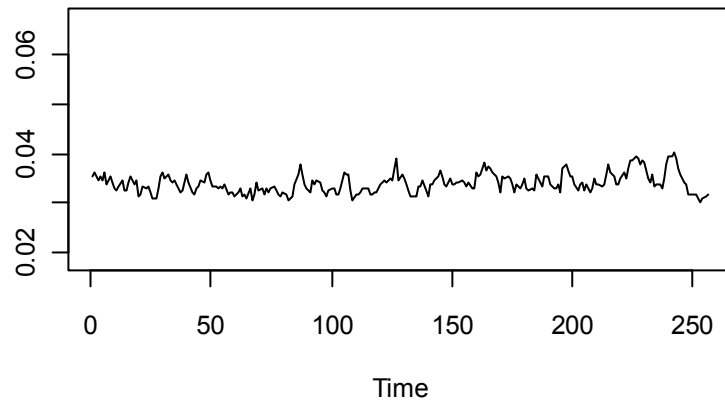
2-Component-Mixture: Config 1



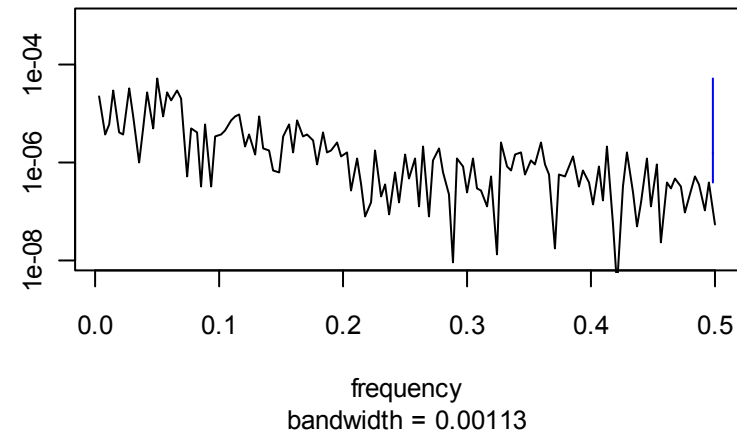
Periodogram of Config 1



2-Component-Mixture: Config 2



Periodogram of Config 2



Applied Time Series Analysis

SS 2015 – Spectral Analysis

The Spectrum

Observed time series	↔	Stochastic process
Empirical ACF	↔	Theoretical ACF
Periodogram	↔	Spectrum

There is a link between ACF and periodogram/spectrum

$$f(\nu) = \sum_{k=-\infty}^{+\infty} \gamma(k) \cos(2\pi\nu k)$$

and

$$\gamma(k) = \int_{-0.5}^{+0.5} f(\nu) \cos(2\pi\nu k) d\nu$$

respectively. The spectrum is thus the Fourier transformation of the ACF.

Applied Time Series Analysis

SS 2015 – Spectral Analysis

What's the Spectrum Good For?

Theorem: **Cramer Representation**

Every stationary process can be written as the limit of a linear combination consisting of harmonic oscillations with random, uncorrelated amplitudes.

- The spectrum characterizes the variance of all these random amplitudes.
- Or vice versa: $\int_{\nu_1}^{\nu_2} f(\nu) d\nu$ is the variance between the frequencies that make the integration limits.
- The spectrum takes only positive values. Thus, not every ACF sequence defines a stationary series.

Applied Time Series Analysis

SS 2015 – Spectral Analysis

A Few Particular Spectra

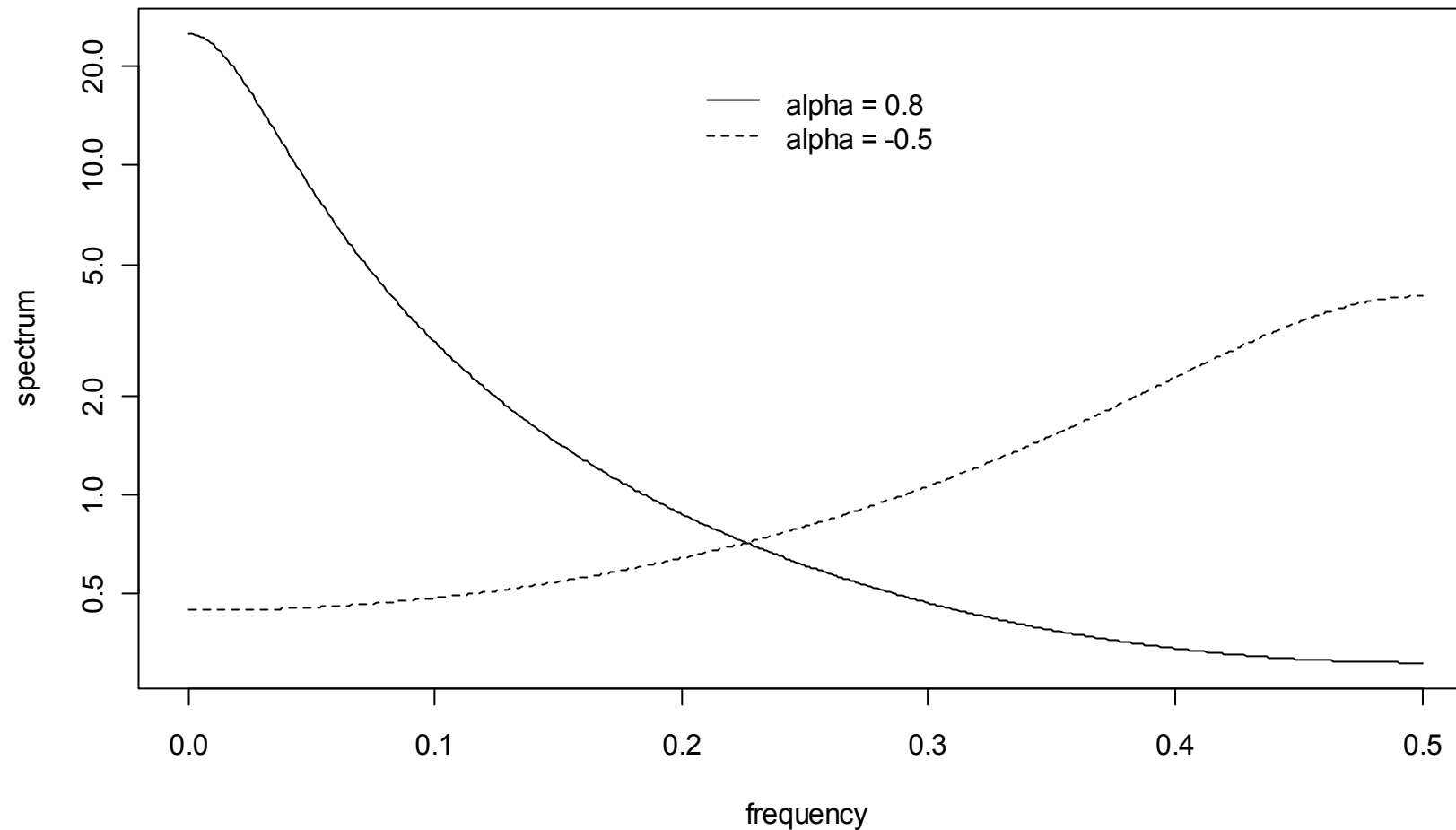
- **White noise**
→ the spectrum is constant over all frequencies.
- **AR(1)**, [see next slide](#)
→ already quite a complicated function α_1
- **ARMA (p,q)**
→ the characteristic polynoms determine the spectrum
$$f(\nu) = \sigma_E^2 \frac{|\Theta(\exp(-i2\pi\nu))|}{|\Phi(\exp(-i2\pi\nu))|}$$
- Note: to generate m maxima in the spectrum, we require an AR-model, where the order is at least $2m$.

Applied Time Series Analysis

SS 2015 – Spectral Analysis

Spectrum of AR(1)-Processes

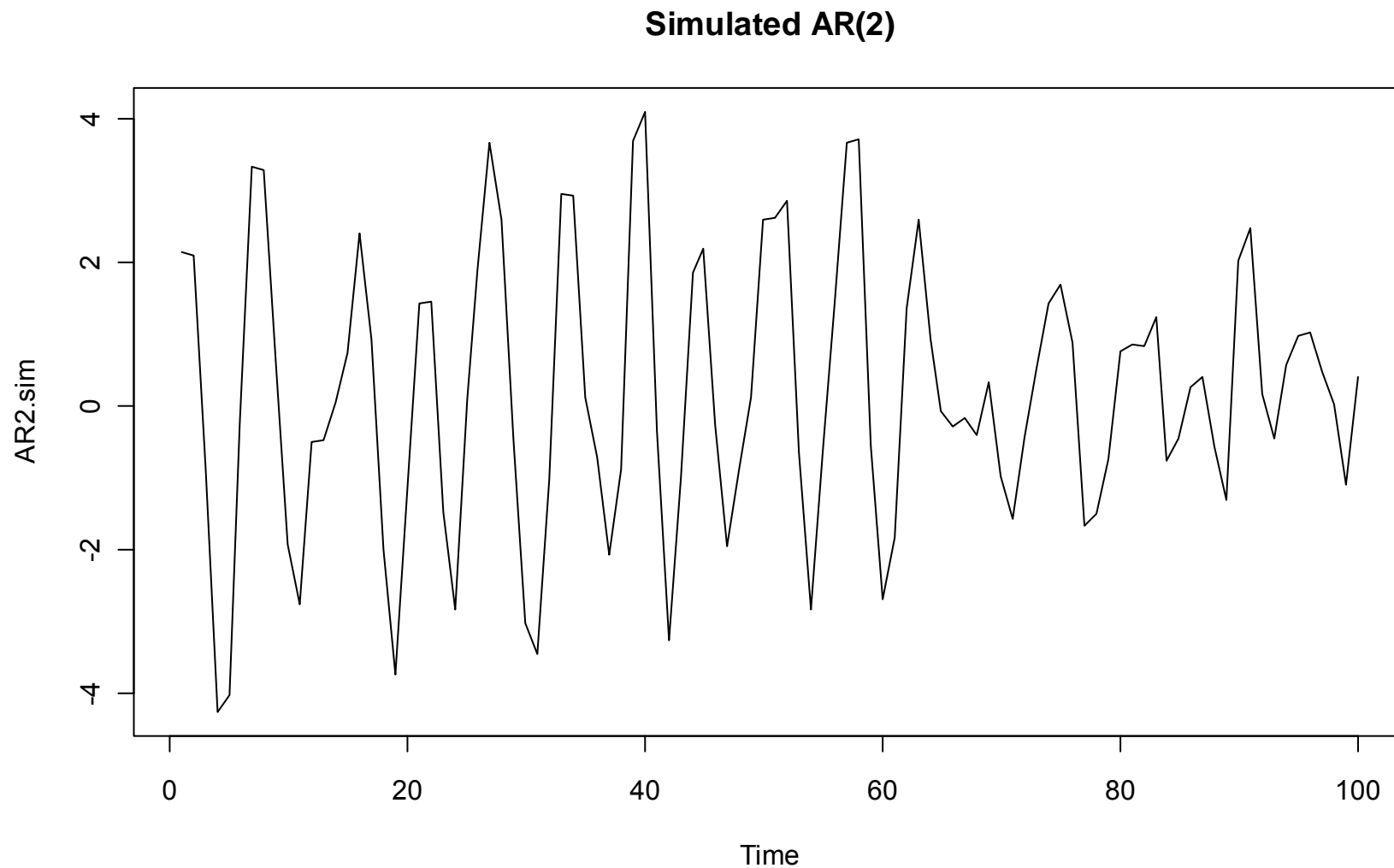
Spectrum of Simulated AR(1)-Processes



Applied Time Series Analysis

SS 2015 – Spectral Analysis

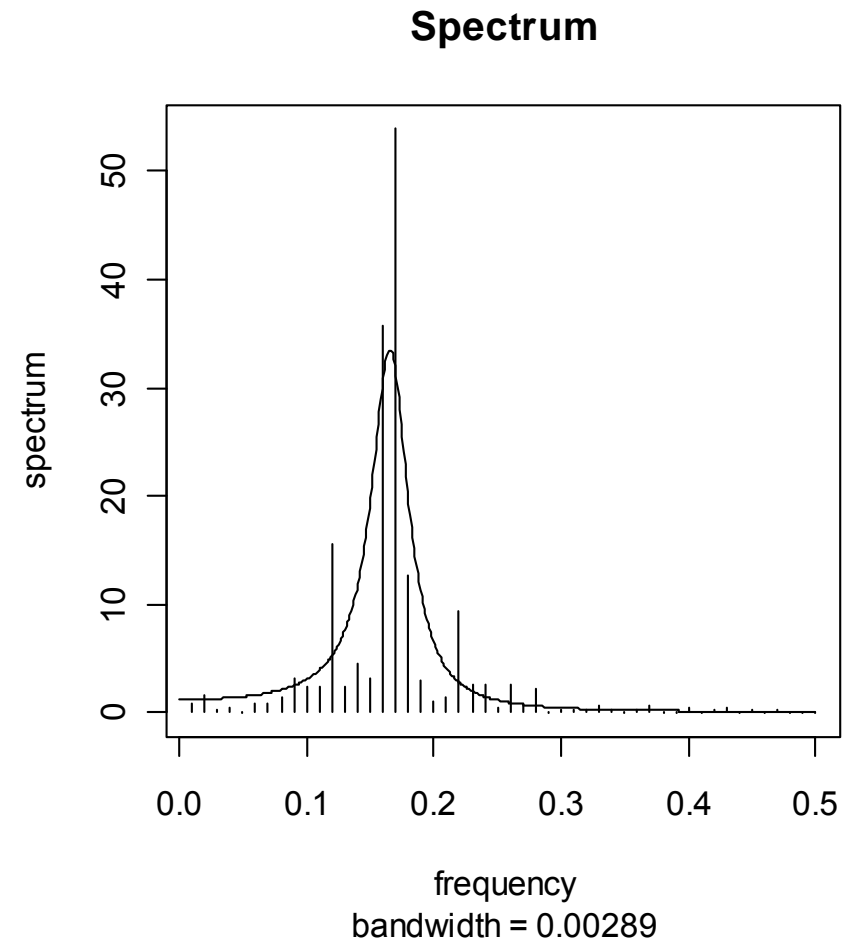
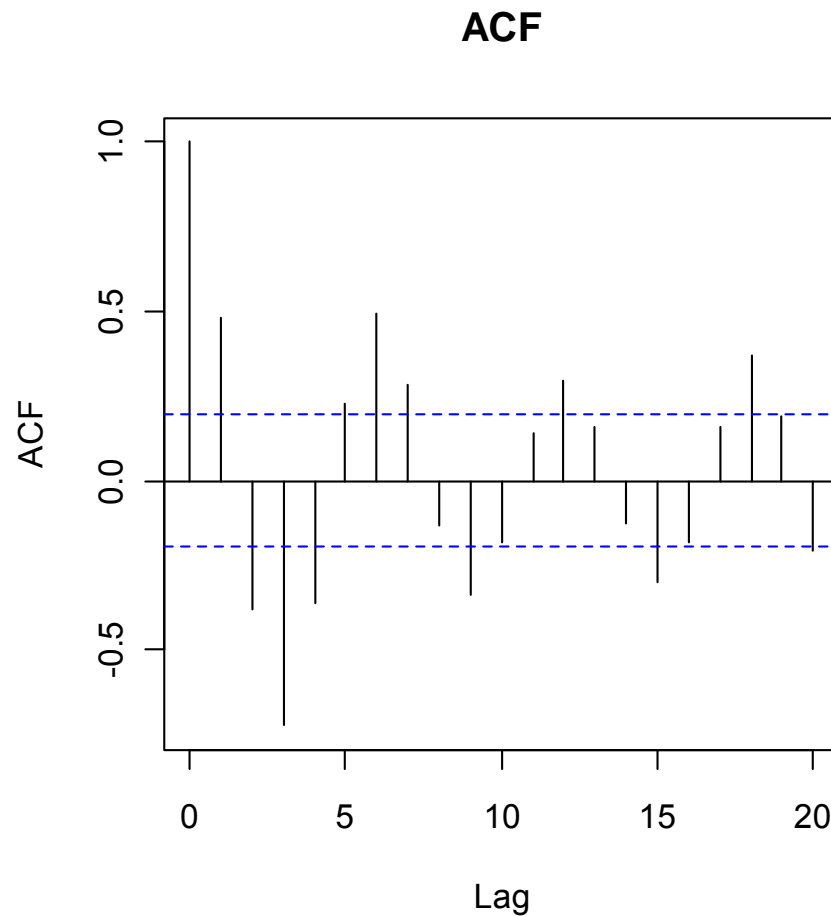
Simulated AR(2)-Process



Applied Time Series Analysis

SS 2015 – Spectral Analysis

ACF/Spectrum of Simulated AR(2)-Process



Applied Time Series Analysis

SS 2015 – Spectral Analysis

Spectral Analysis

- **Spectral analysis** is a descriptive technique, where the time series is interpreted as a linear combination of harmonic oscillations.
- The **periodogram** shows empirically, which frequencies are „important“, i.e. lead to a substantial increase in RSS when omitted from the linear combination.
- The **spectrum** is the theoretical counterpart to the periodogram. It can also be seen as the Fourier transformation of the theoretical autocovariances.
- The periodogram is a poor **estimator** for the spectrum: it's not smooth and inconsistent.

Applied Time Series Analysis

SS 2015 – Spectral Analysis

Improving the Raw Periodogram

1) Smoothing with a running mean

- + simple approach
- choice of the bandwidth

2) Smoothing with a weighted running mean

- + choice of the bandwidth is less critical
- difficulties shift to the choice of weights

3) Weighted plug-in estimation

- + weighted Fourier trsf. of estimated autocovariances
- choice of weights

4) Piecewise periodogram estimation with averaging

- + can serve as a check for stationarity, too

Applied Time Series Analysis

SS 2015 – Spectral Analysis

Improving the Raw Periodogram

5) Spectrum of an estimated model

- + fundamentally different from 1)-4)
- only works for „small“ orders p

6) Tapering

- + further modification of periodogram estimation
- + reduces the bias in the periodogram
- + should always be applied

7) Prewhitening and Rescoloring

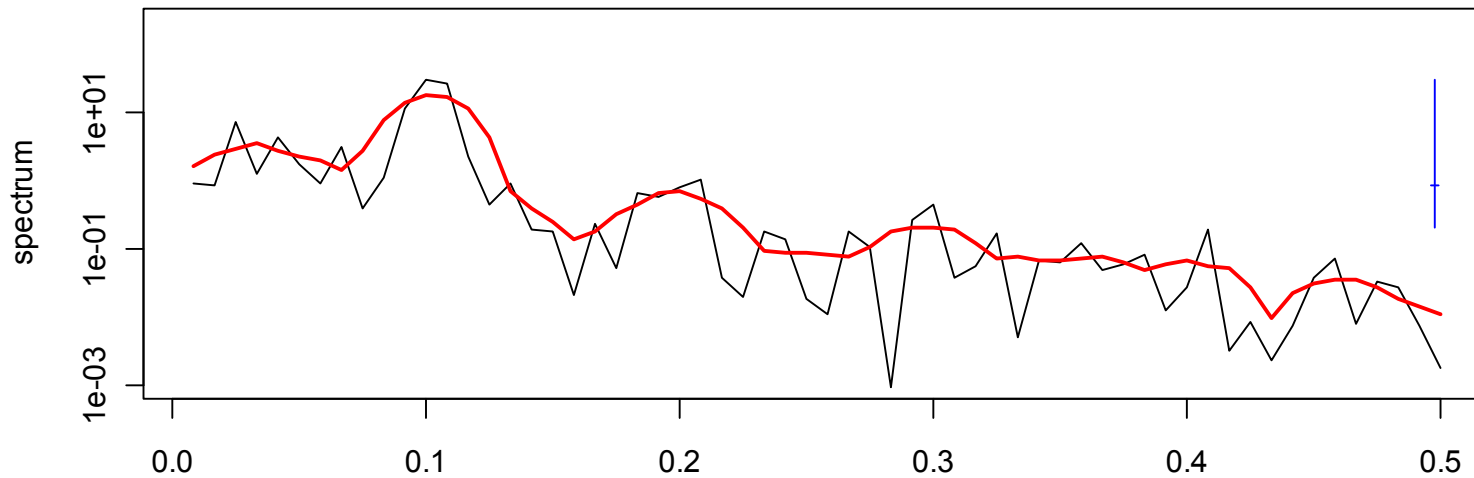
- + model fit and periodogram estimation on residuals
- + the effect of the model will be added again

Applied Time Series Analysis

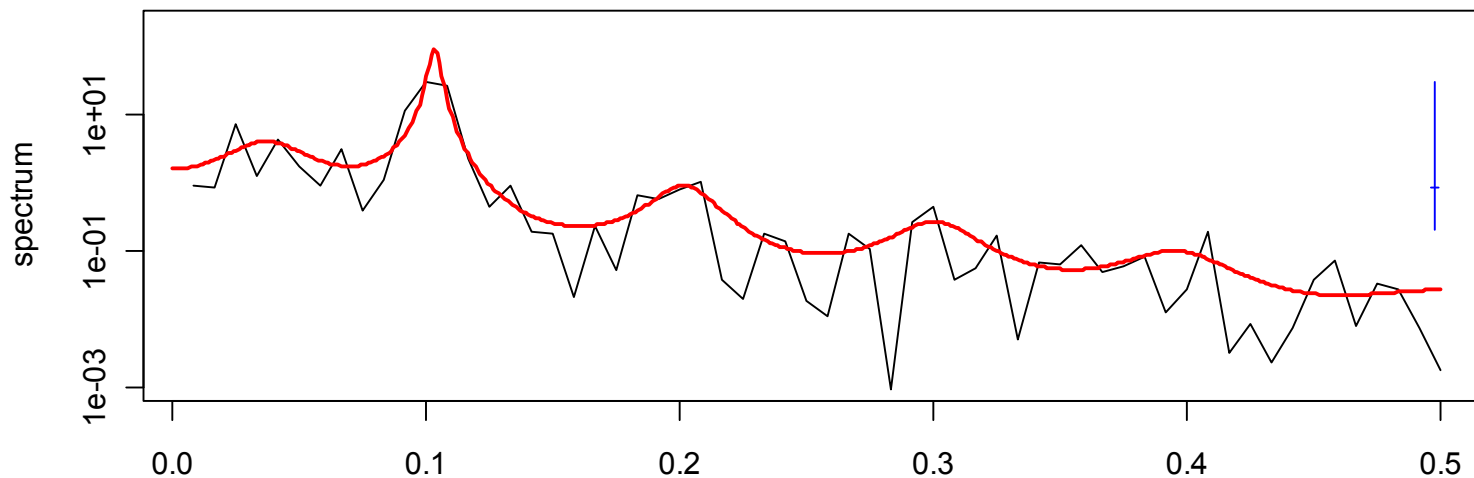
SS 2015 – Spectral Analysis

Modified Periodogram of log(Lynx) Data

Raw and Smoothed Periodogram



Raw and Model Based Periodogram

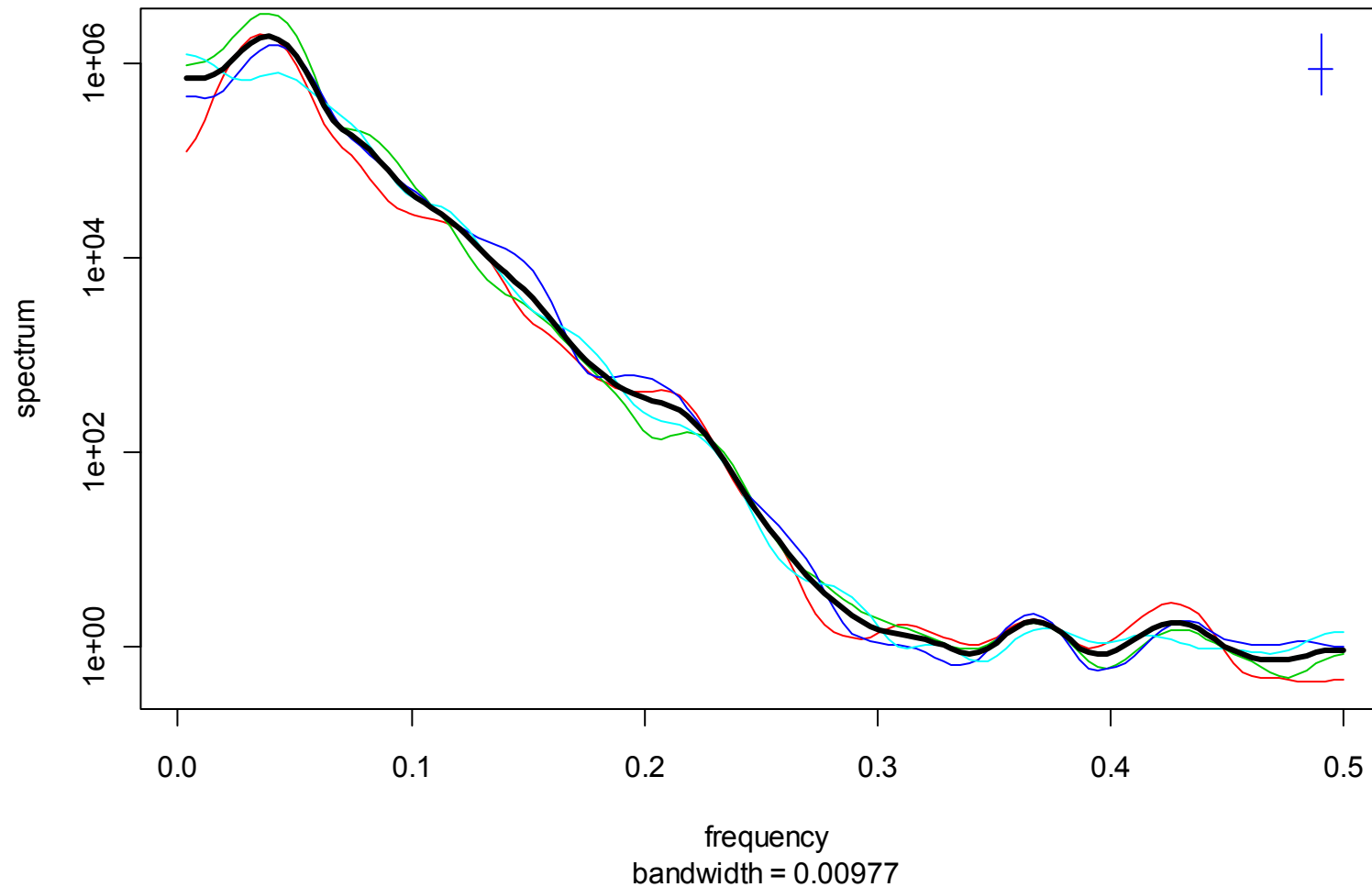


Applied Time Series Analysis

SS 2015 – Spectral Analysis

Modified Periodogram of $\log(\text{Lynx})$ Data

Piecewise periodogram of ocean wave data



Applied Time Series Analysis

SS 2015 – State Space Models

State Space Models

Basic idea: There is a stochastic process/time series X_t which we cannot directly observe, but only under the addition of some measurement noise.

Thus: We observe the time series $Y_t = X_t + V_t$,
with iid measurement errors $V_t \sim N(0, \sigma_V^2)$

Example: $X_t = \#$ of fish in a lake
 $Y_t = \#$ estimated number of fish from a sample

Other:

- Dynamic linear modeling
- Regression with time-varying coefficients

Applied Time Series Analysis

SS 2015 – State Space Models

State Space Formulation

State space models are always built on two different equations, one of which aims for the process, and the other for the measurement noise:

State Equation: $X_t = G_t X_{t-1} + W_t$, where $W_t \sim N(0, w_t)$

Observation Equation: $Y_t = F_t X_t + V_t$, where $V_t \sim N(0, v_t)$

All matrices in this model, i.e. G_t, F_t, w_t, v_t can be time-varying. However, often they are time-constant, if anything, then F_t is adapting over time.

Note: such models are usually estimated with the Kalman filter.

Applied Time Series Analysis

SS 2015 – State Space Models

AR(1) with Measurement Noise

We assume that the true underlying process is an AR(1), i.e.

$$X_t = \alpha_1 X_{t-1} + W_t,$$

where

$W_t \sim N(0, \sigma_W^2)$ are i.i.d. innovations, „process noise“.

In practice, we only observe y_t , as realizations of the process

$$Y_t = X_t + V_t, \text{ with } V_t \sim N(0, \sigma_V^2), \text{ i.i.d.}$$

and additionally, the V_t are independent of X_s, W_s for all s, t , thus they are independent „observation white noise“.

Applied Time Series Analysis

SS 2015 – State Space Models

More Terminology

We call

$$X_t = \alpha_1 X_{t-1} + W_t$$

the „state equation“, and

$$Y_t = X_t + V_t$$

the „observation equation“.

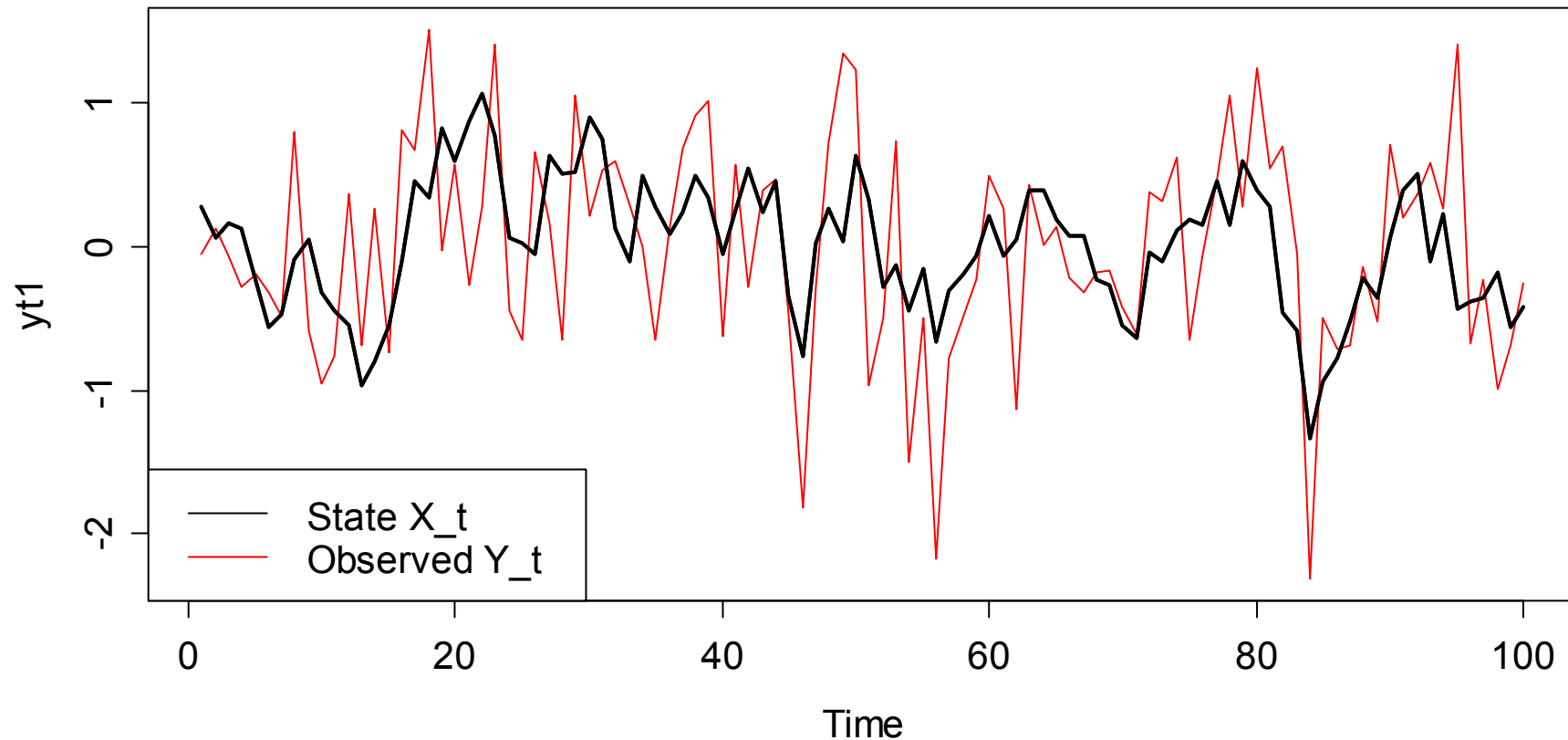
On top of that, we remember once again that the „process noise“ W_t is an innovation that affects all future values X_{t+k} and thus also Y_{t+k} , whereas V_t only influences the current observation Y_t , but no future ones.

Applied Time Series Analysis

SS 2015 – State Space Models

AR(1)-Example with $\alpha=0.7$

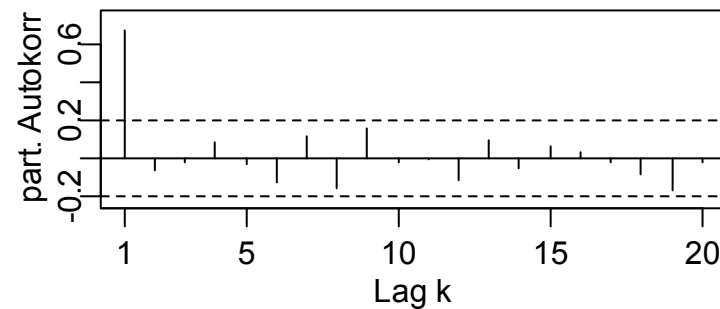
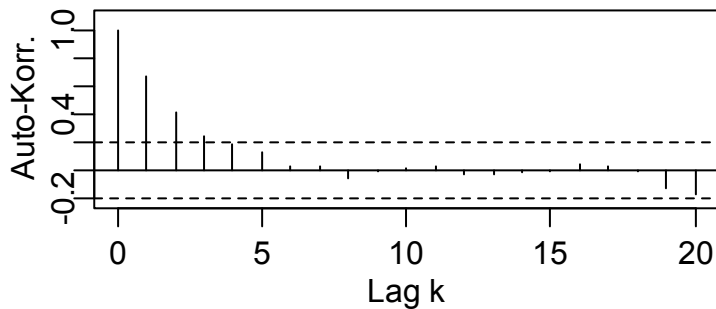
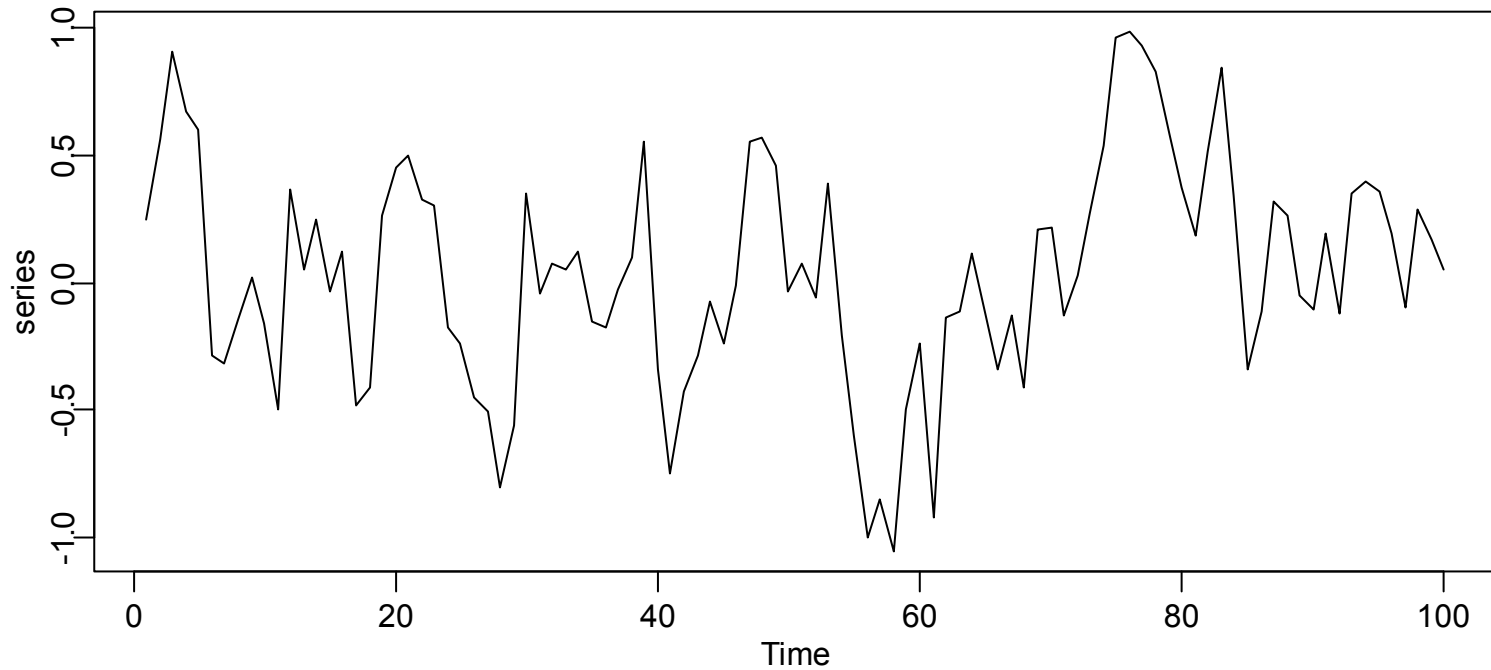
AR(1) Simulation Example



Applied Time Series Analysis

SS 2015 – State Space Models

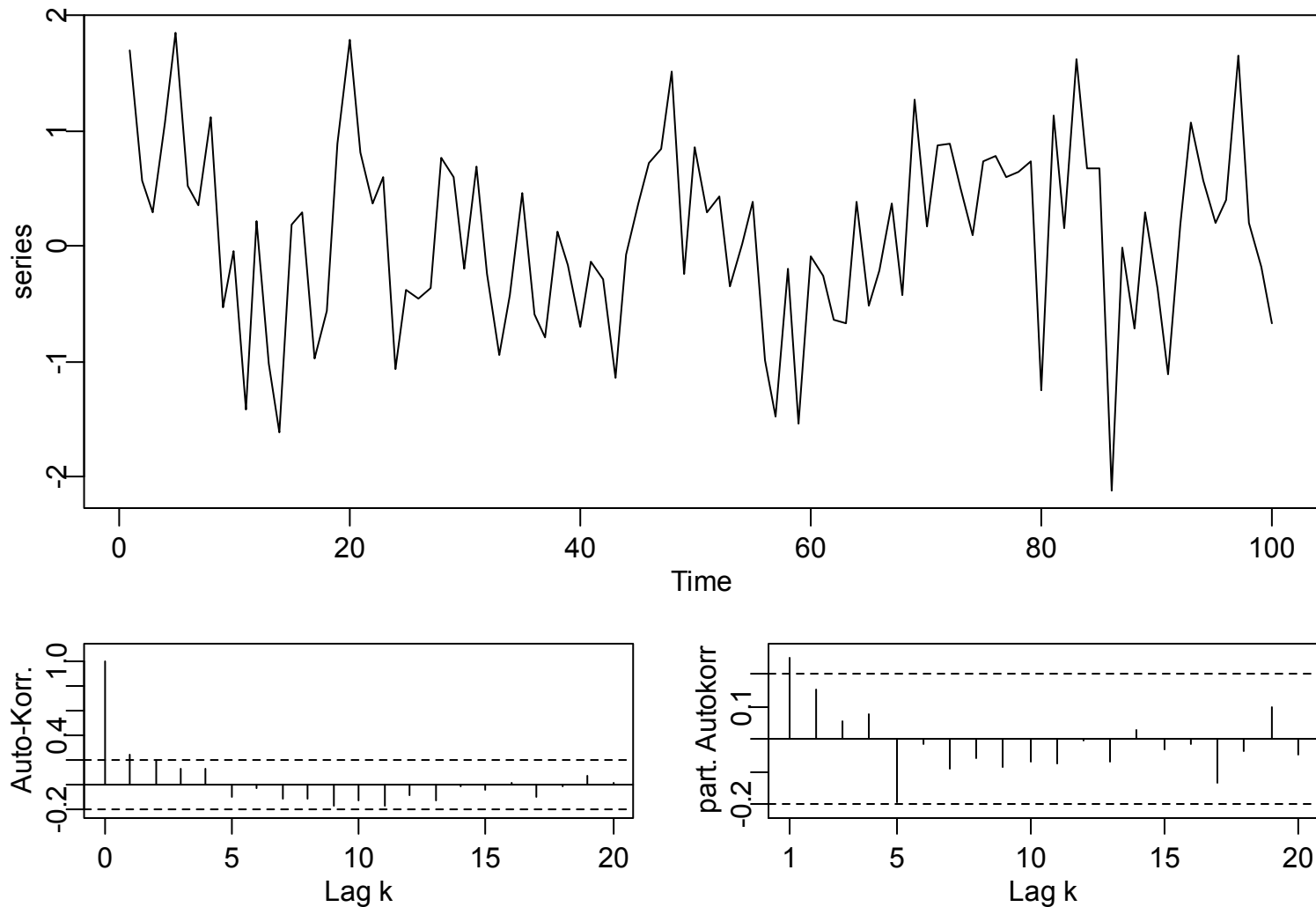
ACF/PACF of X_t



Applied Time Series Analysis

SS 2015 – State Space Models

ACF/PACF of Y_t



Applied Time Series Analysis

SS 2015 – State Space Models

What is the goal?

The goal of State Space Modeling/Kalman Filtering is:

To uncover the „de-noised“ process X_t from the observed process Y_t .

- The algorithm of Kalman Filtering works with non-stationary time series, too.
- The algorithm is based on a maximum-likelihood-principle where one assume normal distortions.
- There are extensions to multi-dimensional state space models. **See blackboard** for an example how the state space formulation of an AR(2) is set up .

Applied Time Series Analysis

SS 2015 – State Space Models

State Space and Kalman Filtering in R

```
## Load the package for Kalman filtering  
library(sspir)
```

```
## State Space Formulation
```

```
ssf <- SS(y = as.matrix(obs),  
         Fmat = function(tt,x,phi) { return(matrix(1)) },  
         Gmat = function(tt,x,phi) { return(matrix(0.7)) },  
         Vmat = function(tt,x,phi) { return(matrix(0.5)) },  
         Wmat = function(tt,x,phi) { return(matrix(0.1)) },  
         m0 = matrix(0), C0 = matrix(0.1))
```

```
## Kalman Filtering
```

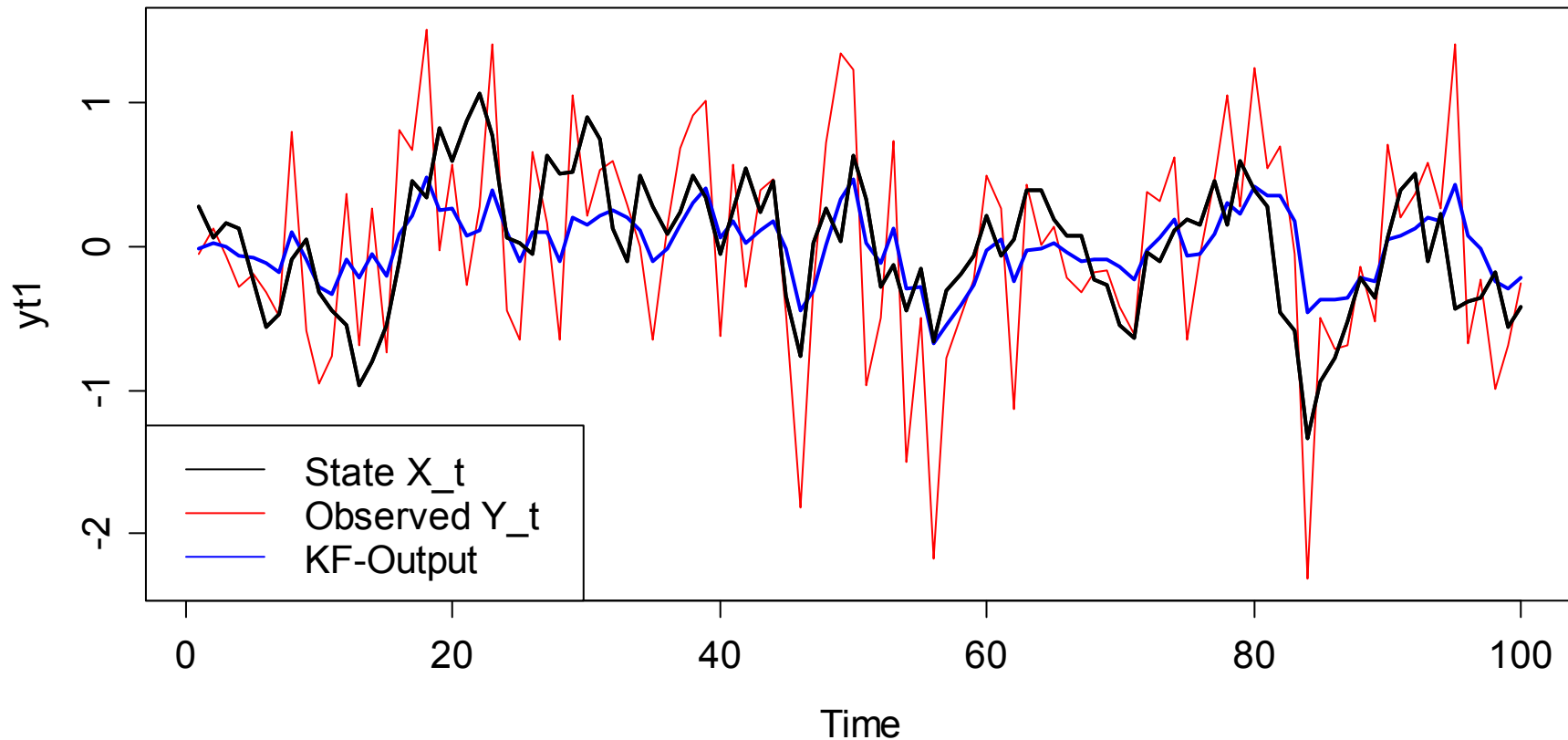
```
fit <- kfilter(ssf)
```

Applied Time Series Analysis

SS 2015 – State Space Models

Kalman Filter Solution

AR(1) Simulation Example with Kalman Filter Output



Applied Time Series Analysis

SS 2015 – State Space Models

State Space Formulation of an AR(2)

→ see blackboard...

Applied Time Series Analysis

SS 2015 – State Space Models

Dynamic Linear Models

In particular: *regression models with time-varying coefficients*

Example: the sales of a housing company depend on the general level of sales in that area at time t , and on the pricing policy at time t .

$$S_t = L_t + \beta_t P_t + V_t$$

This is a regression model with price as the predictor, and the general sales level as the intercept. They are time-varying:

$$L_t = L_{t-1} + \Delta L_t \quad \beta_t = \beta_{t-1} + \Delta \beta_t$$

Here, $V_t, \Delta L_t, \Delta \beta_t$ are random elements, noise & perturbations

Applied Time Series Analysis

SS 2015 – State Space Models

Simulation Example

→ see blackboard...

Applied Time Series Analysis

SS 2015 – State Space Models

Kalman Filtering for Regression

```
#### State Space Formulation
ssf <- SS(y=y.mat, x=x.mat,
        Fmat=function(tt,x,phi) return(matrix(c(x[tt,1],x[tt,2]),2,1)),
        Gmat=function(tt,x,phi) return(diag(2)),
        Wmat=function(tt,x,phi) return(0.1*diag(2)),
        Vmat=function(tt,x,phi) return(matrix(1)),
        m0=matrix(c(5,3),1,2),C0=10*diag(2))

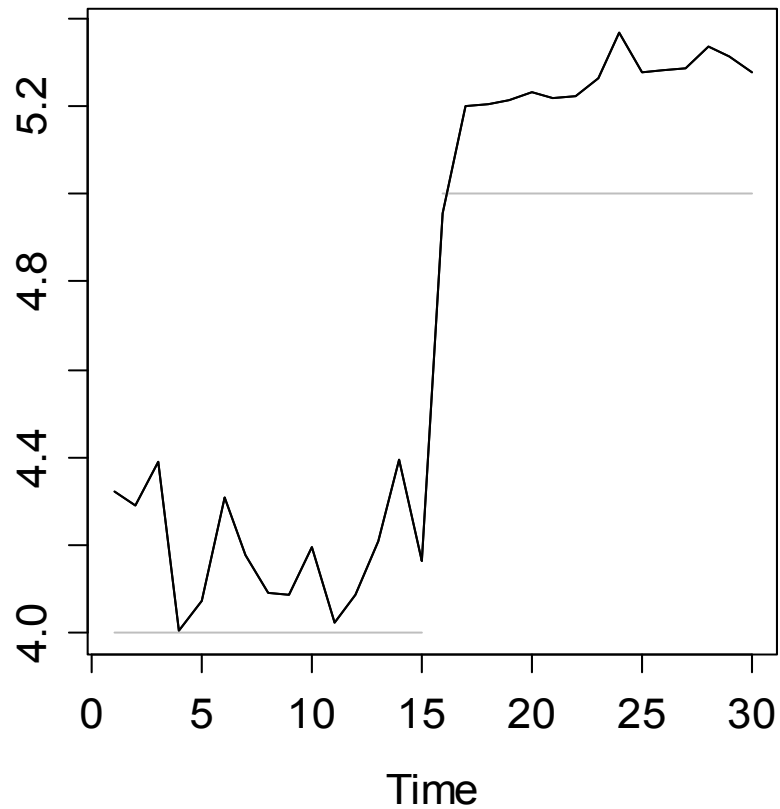
## Kalman-Filtering
fit <- kfilter(ssf)
plot(fit$m[,1], type="l", xlab="Time", ylab="Intercept")
plot(fit$m[,2], type="l", xlab="Time", ylab="Slope")
```

Applied Time Series Analysis

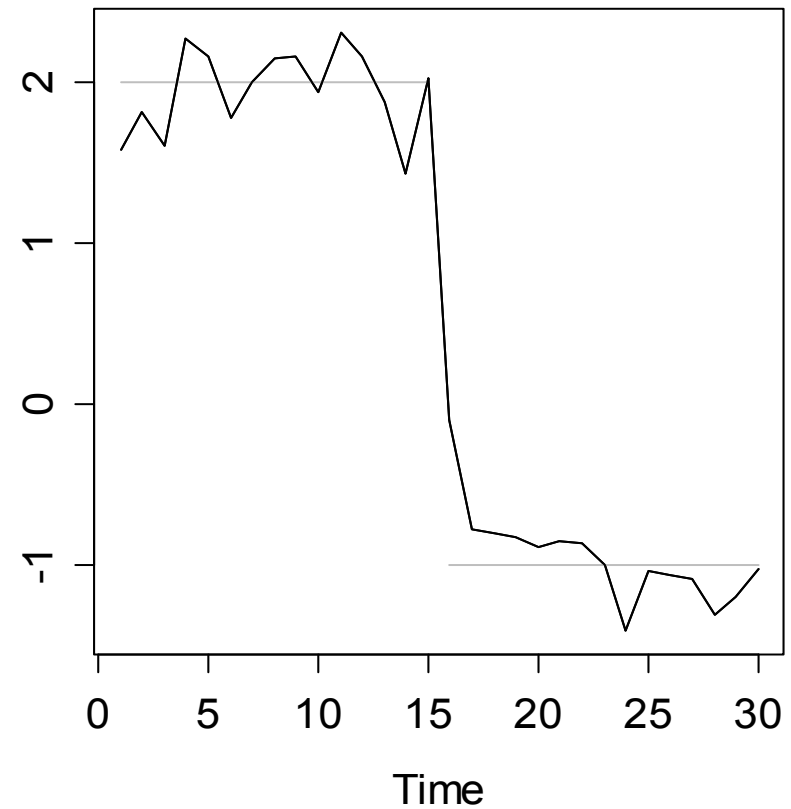
SS 2015 – State Space Models

Kalman Filter Solution

Kalman Filtered Intercept



Kalman Filtered Slope



Applied Time Series Analysis

SS 2015 – State Space Models

Summary of Kalman Filtering

Summary:

- 1) The Kalman Filter is a recursive algorithm
- 2) It relies on an update idea, i.e. we update the forecast $\hat{X}_{t+1,t}$ with the difference $(y_{t+1} - \hat{Y}_{t+1,t})$.
- 3) The weight of the update is determined by the relation between the process variance σ_W^2 and the measurement noise σ_V^2 .
- 4) This relies on the knowledge of G , F , σ_W^2 , σ_V^2 . In R we have procedures where everything is estimated simultaneously.

Applied Time Series Analysis

SS 2015 – State Space Models

Additional Remarks

- 1) For the recursive approach of Kalman filtering, initial values are necessary. Their choice is not crucial, their influence cancels out rapidly.
- 2) The procedures yield forecast and filter intervals:
$$\hat{X}_{t+1,t} \pm 1.96 \cdot \sqrt{R_{t+1,t}} \quad \text{and} \quad \hat{X}_{t+1,t+1} \pm 1.96 \cdot \sqrt{R_{t+1,t+1}}$$
- 3) State space models are a very rich class. Every ARIMA(p,d,q) can be written in state space form, and the Kalman filter can be used for estimating the coefficients.