Applied Time Series Analysis
FS 2011 – Week 01

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 21, 2011

Presentation of the Lecturer

Education: Dr. Math. ETH

Position: Project Manager R&D @ ZHAW
Lecturer @ ETH Zürich & ZHAW

My connections to time series analysis:

- Research projects with industry: airlines, cargo, marketing
- Academic research on high-frequency financial time series

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

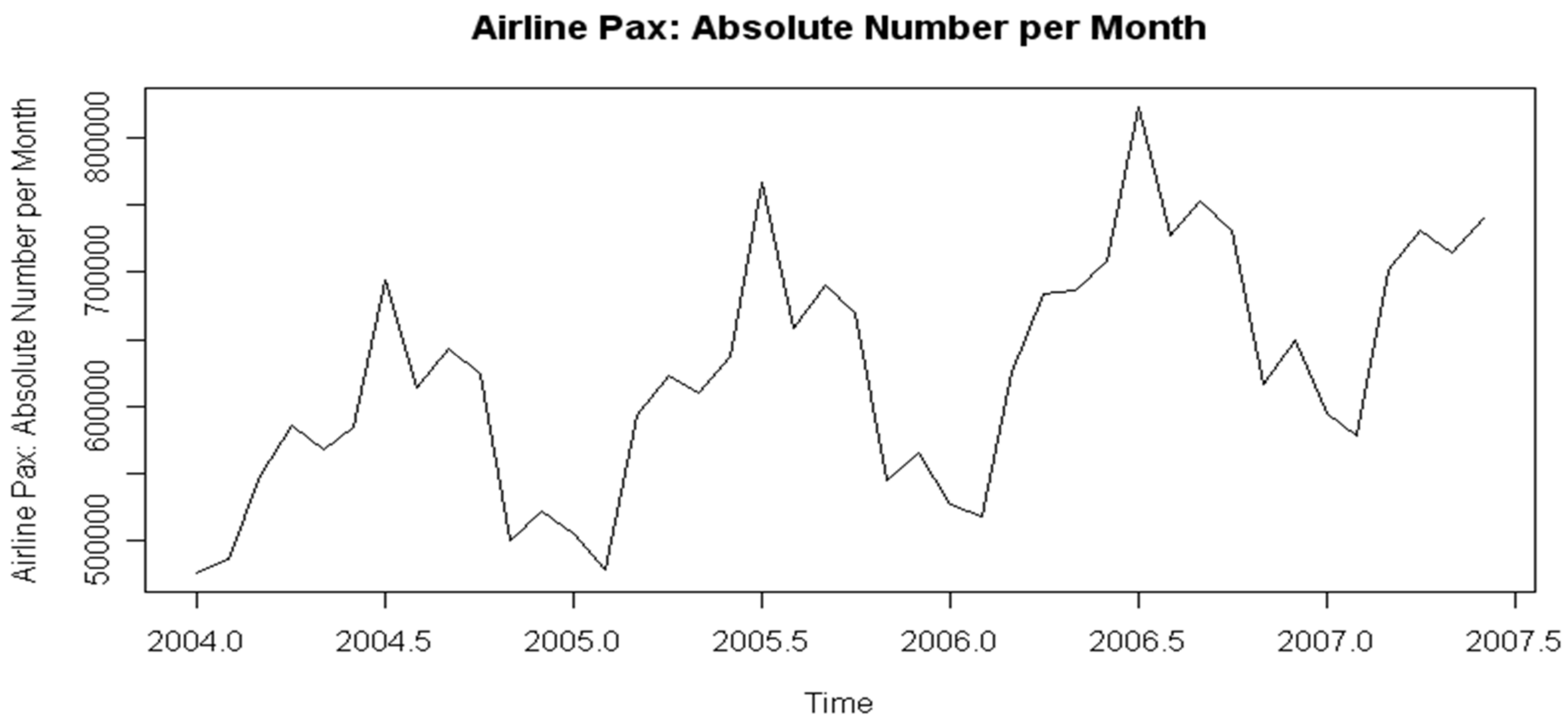
How was this decision taken?

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
FS 2011 – Week 01

A Second Example

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



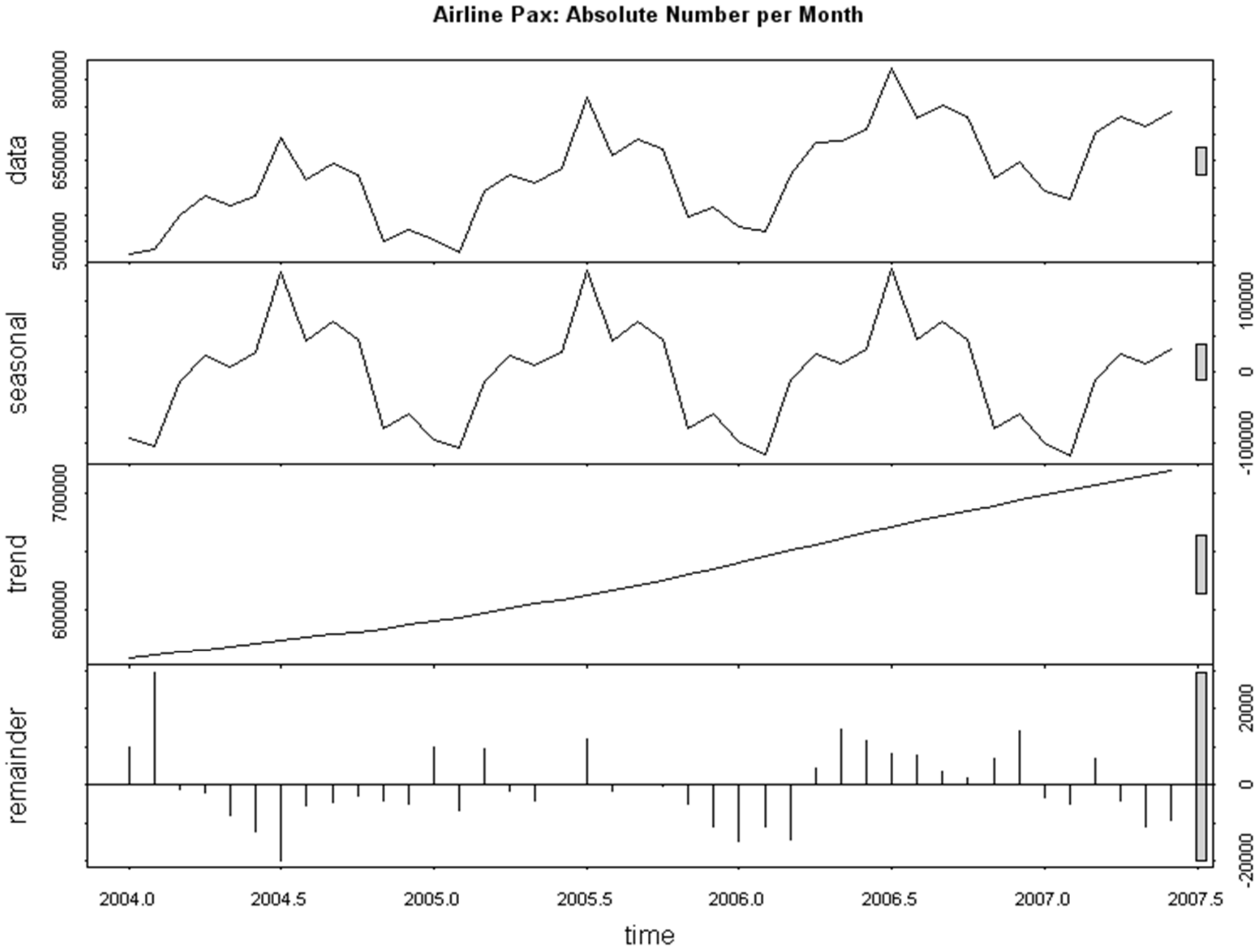
Some Properties of the Series

- 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
FS 2011 – Week 01



Organization of the Course

The goal of the course is:

The students are aware of the basic mathematical concepts behind time series analysis. They are familiar with descriptive techniques, autoregressive and ARMA-modeling and have basic knowledge about more advanced topics such as multivariate time series analysis, state space models and spectral analysis.

They gain extensive 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
FS 2011 – Week 01

Organization of the Course

Applied Time Series Analysis – FS 2011

People:

Lecturer: **Dr. Marcel Dettling** (marcel.dettling@zhaw.ch)

Assistant: **Alain Hauser** (alhauser@ethz.ch)

Course Schedule:

All lectures and exercises will be held at HG E1.2, on Mondays from 10.15-11.55.

Week	Date	L/E	Topics
01	21.02.2011	L/L	Introduction to time series analysis
02	28.02.2011	L/E	Stationarity, decomposition of time series
03	07.03.2011	L/L	Autocorrelation, Correlogram
04	14.03.2011	L/E	Autoregressive Modeling 1
05	21.03.2011	L/L	Autoregressive Modeling 2
06	28.03.2011	L/E	Time series forecasting
07	04.04.2011	L/L	ARMA-Modeling 1
08	11.04.2011	L/E	ARMA-Modeling 2
09	18.04.2011	L/L	Time series regression
10	02.05.2011	L/E	Multivariate time series
11	09.05.2011	L/E	State space models
12	16.05.2011	L/L	Spectral Analysis 1
13	23.05.2011	L/E	Spectral Analysis 2
14	30.05.2011	L/L	Advanced Topics

Exercise Schedule:

The exercises will be held roughly every second week in the lecture room HG E1.2. There is only one group, for which 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	28.02.2011	Time series in R	07.03.2011	14.03.2011
02	14.03.2011	Autocorrelation	21.03.2011	28.03.2011
03	28.03.2011	AR(p)-modeling	04.04.2011	11.04.2011
04	11.04.2011	ARMA(p,q)-modeling	18.04.2011	02.05.2011
05	02.05.2011	Multivariate time series	09.05.2011	16.05.2011
06	09.05.2011	State space modeling	16.05.2011	23.05.2011
07	23.05.2011	Spectral analysis	26.05.2011	30.05.2011

The solved exercises can be handed in in the lectures where an assistant will pick them up. Sending them via e-mail or placing them in the corresponding tray in HG J68 until 11.55am of the due date is another option. Please write down your findings and comments. You can support this with a few plots, but please avoid handing in any R-code or an excessive amount of plots.

→ more details are given on the additional organization sheet

Introduction: What is a Time Series?

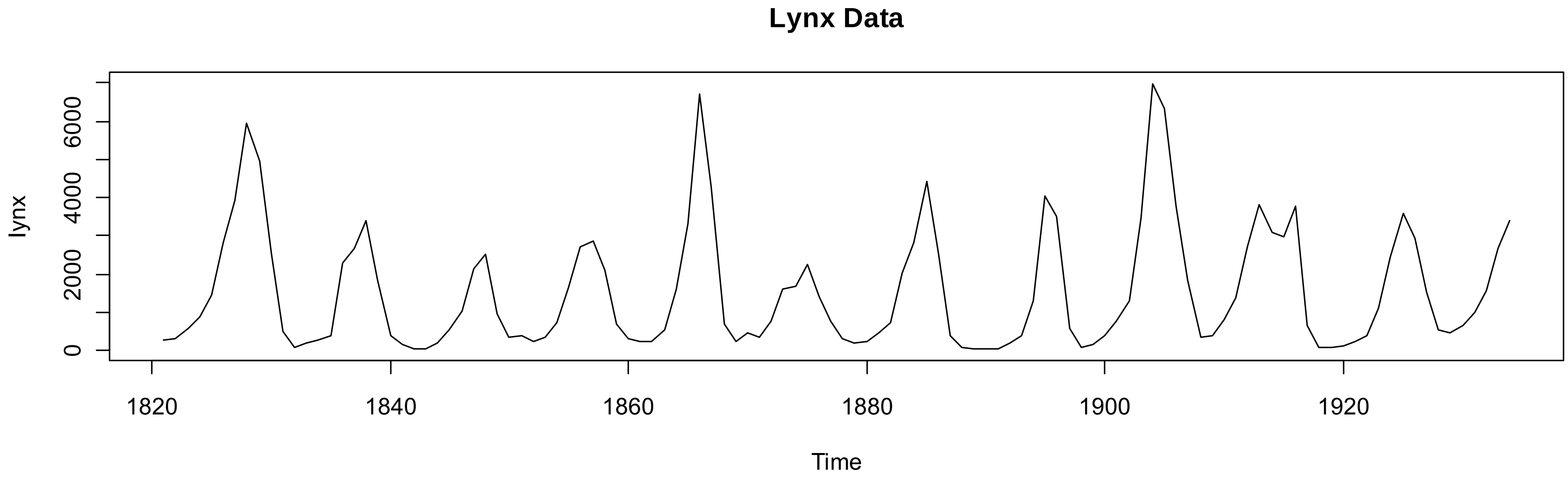
A time series is a set of observations x_t, where each of the observations was made at a specific time t.

- 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.

Example: Lynx Data



Example: Lynx Data

X_t = # of shot lynx in year t at the Mackenzie River district

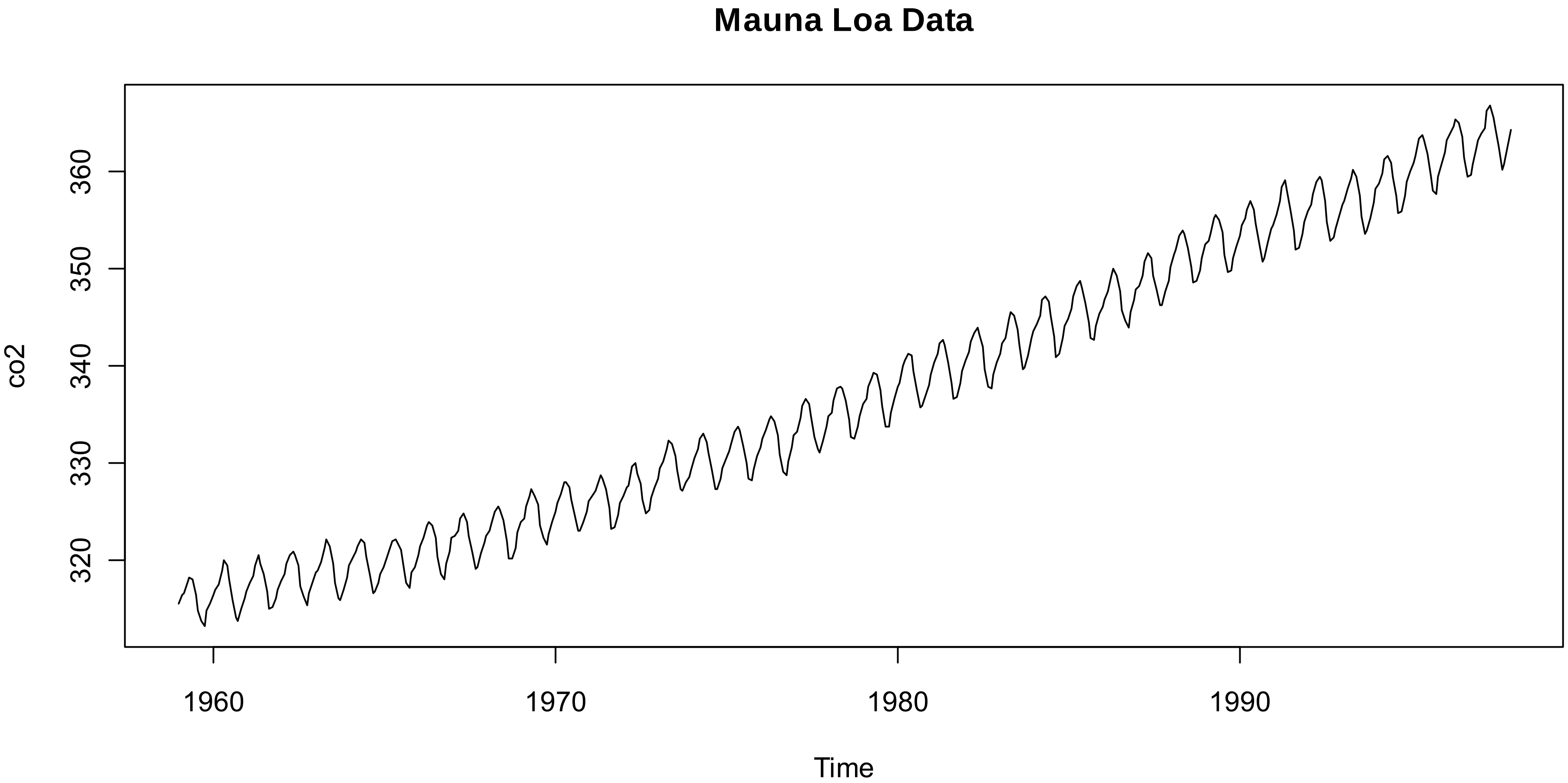
T = {1821, 1822, 1823, ..., 1934}

Observations:

- x_t ε [0, 7000]

- apparent periodicity: “high/low” every ~10 years
  “superhigh” every ~40 years

Example: Mauna Loa Data



Example: Mauna Loa Data

X_t = CO_2 concentrations in ppm on Hawaii in

T = {1959/01, 1959/02, 1959/03, ..., 1997/12}

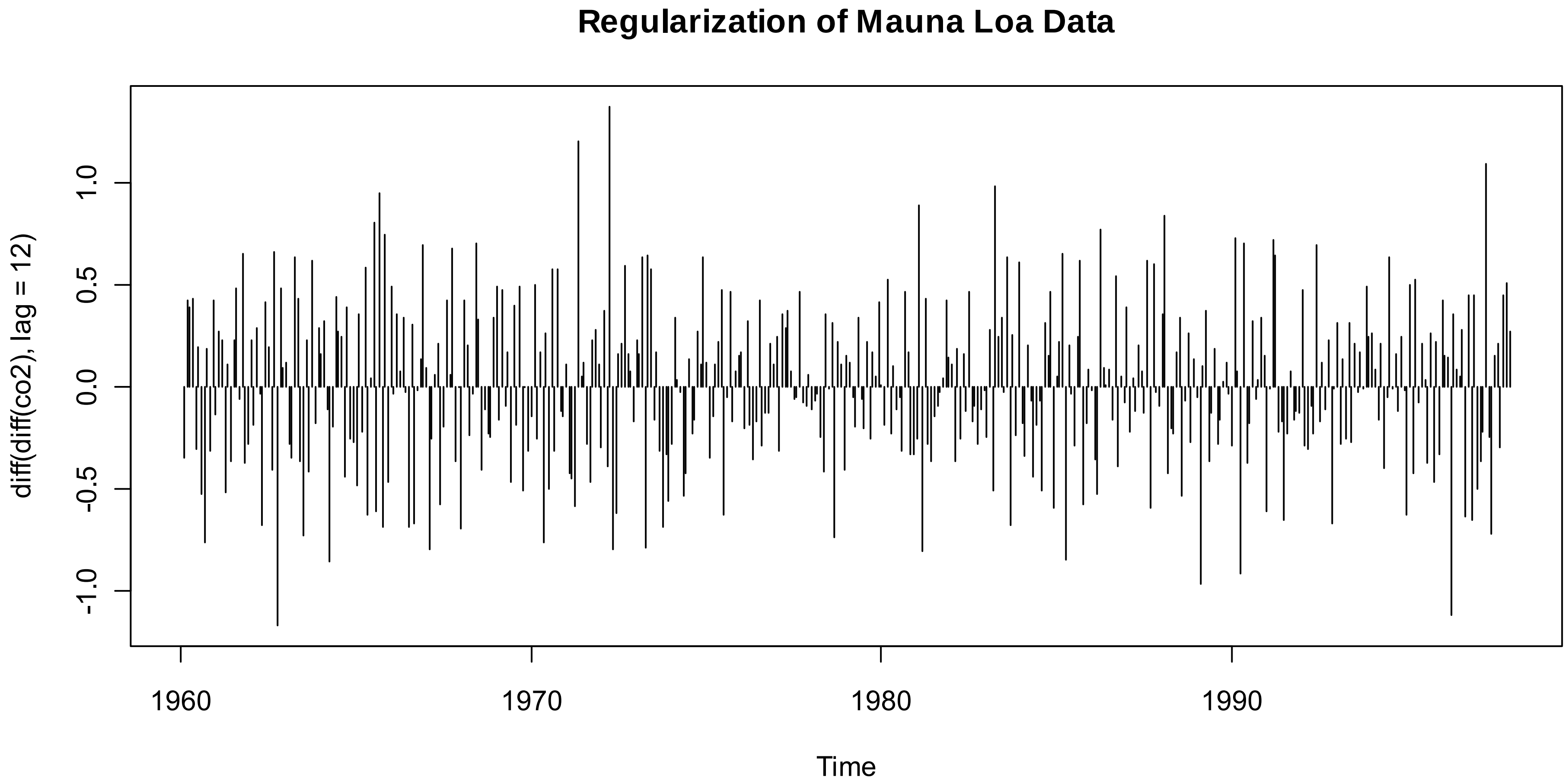
Observations:

- linear, or close to linear increase := “trend”

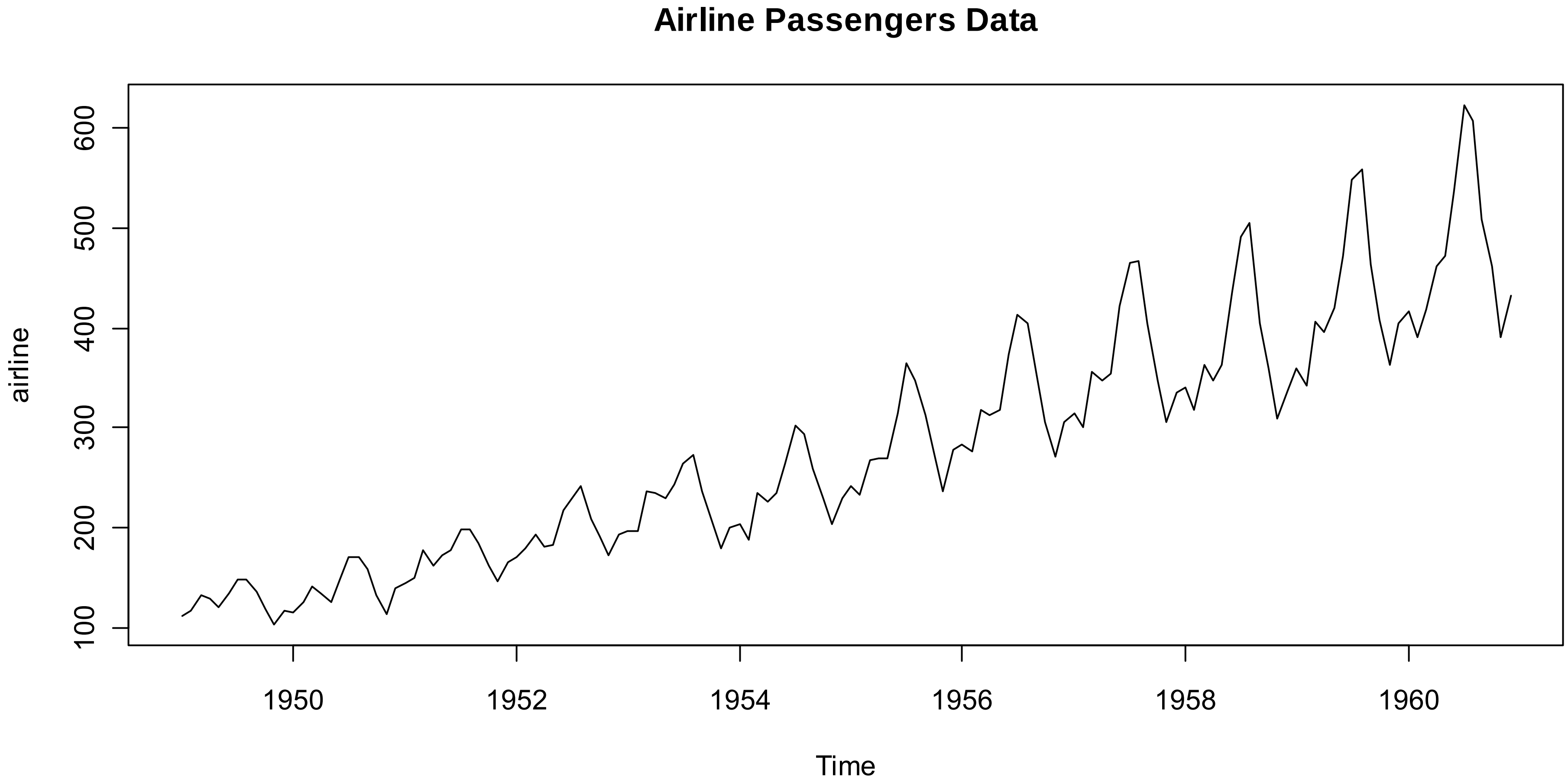
- yearly recurring seasonal pattern

- constant within-year variation

Example: Mauna Loa Data



Example: Airline Passengers



Example: Airline Passengers

X_t = # of international passengers of an unknown airline

T = {1949/01, 1949/02, 1949/03, ..., 1960/12}

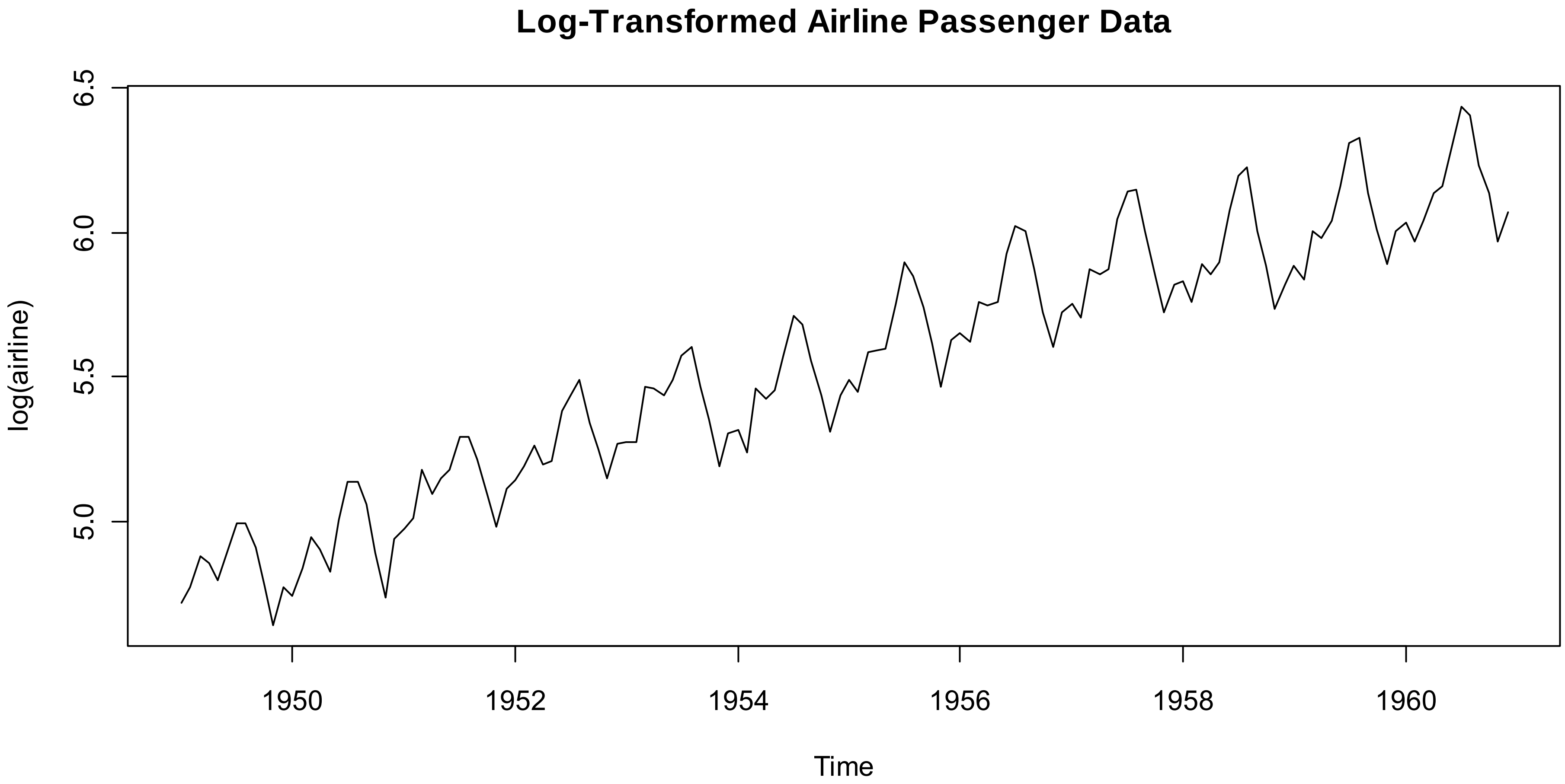
Observations:

- Increasing within-year variation (log transform!)

- yearly recurring seasonal pattern

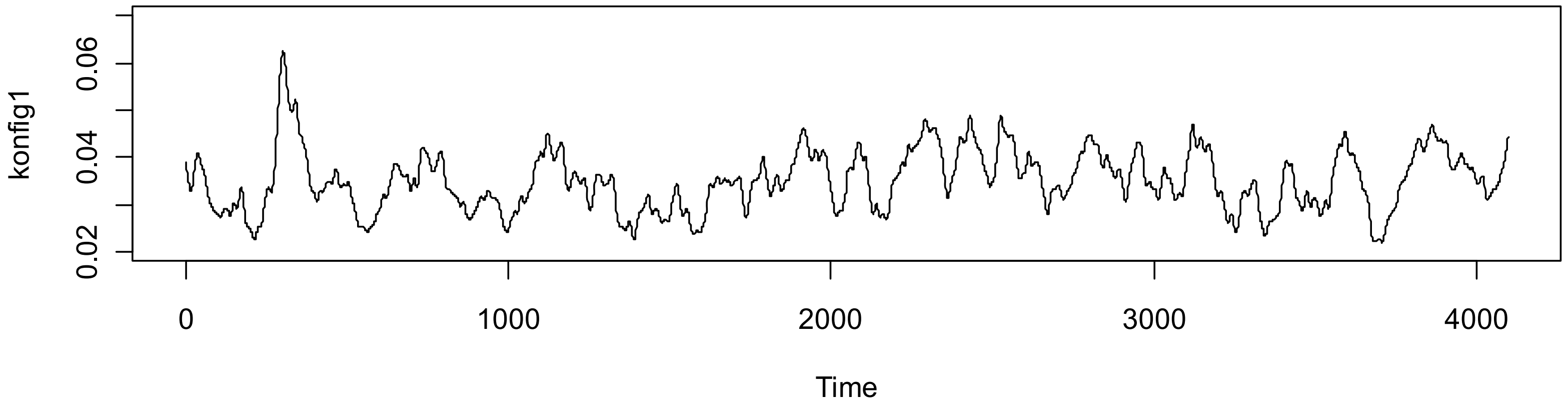
- a trend is visible, it seems close to linear

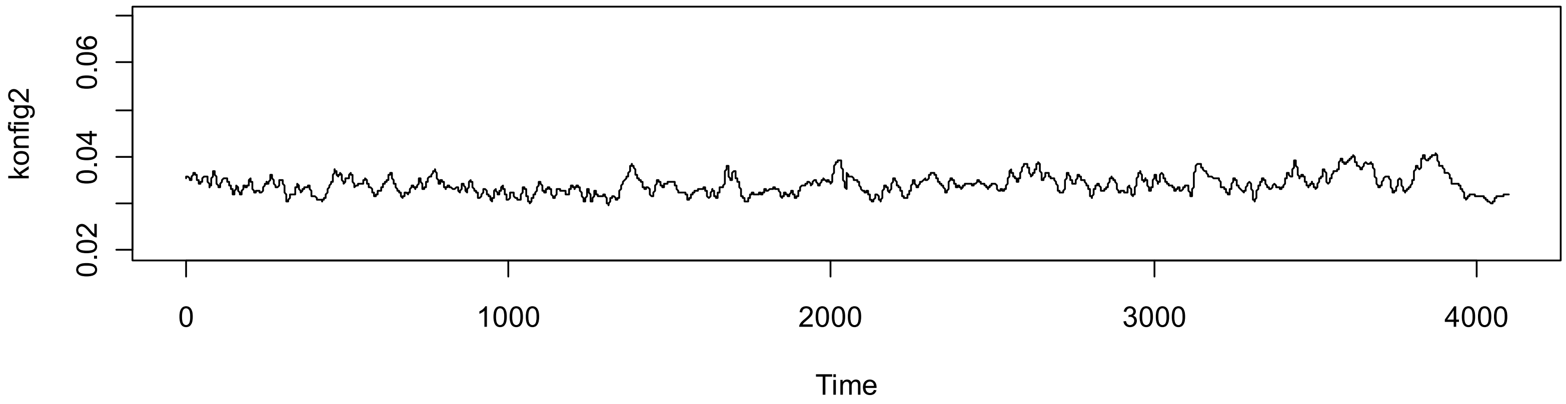
Example: Airline Passengers



Applied Time Series Analysis
FS 2011 – Week 01

Example: 2-Component Mixture





Example: 2-Component Mixture

X_t = proportion of component 1 in mixture configuration 1

Y_t = proportion of component 1 in mixture configuration 2

T = {1, 2, 3, ..., 4096}

Observations:

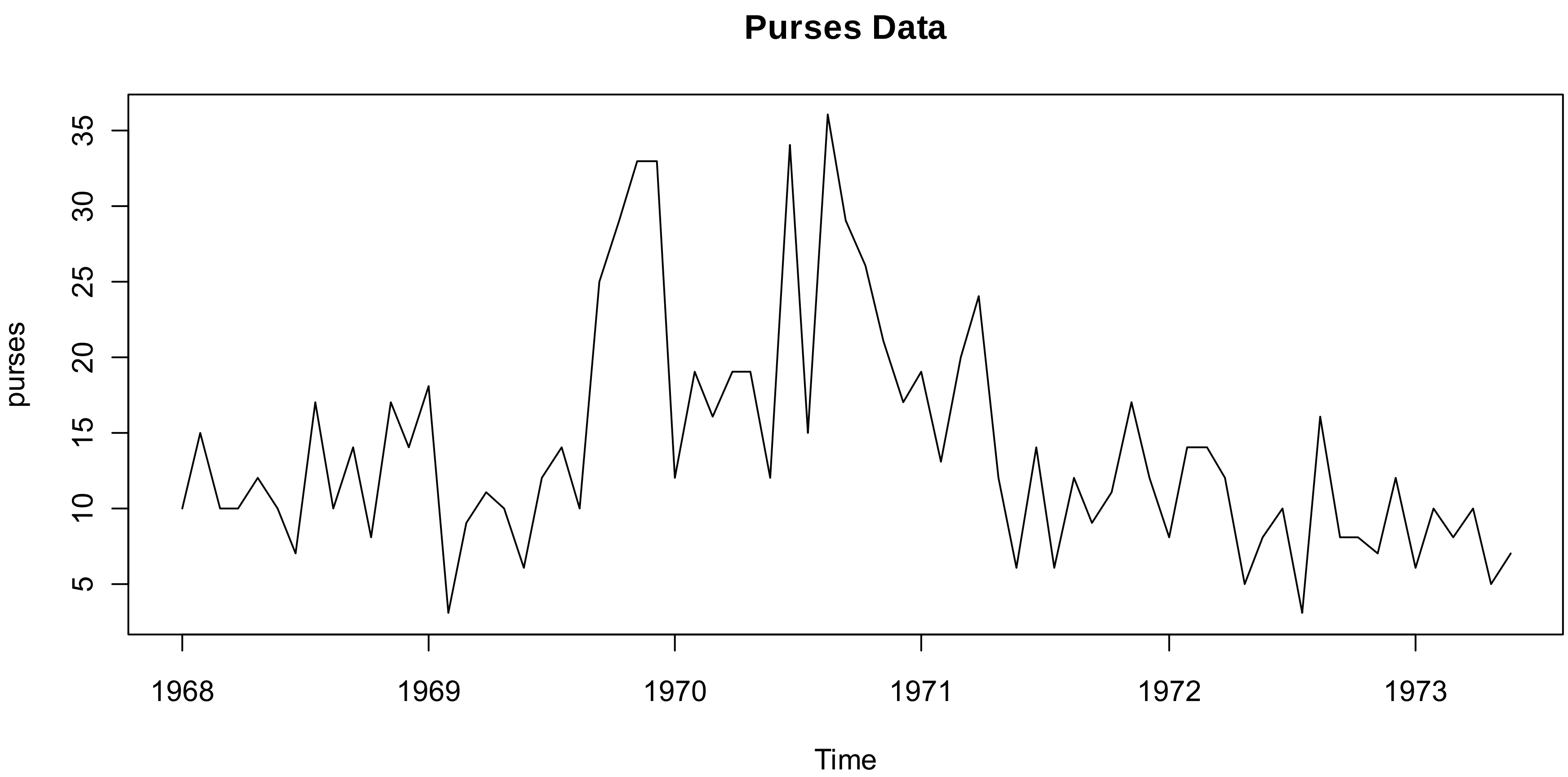
- neither trend nor seasonal pattern are apparent

- “outlying” observations in time series 1

- different variabilty in time series 1 vs. time series 2

Applied Time Series Analysis
FS 2011 – Week 01

Example: Purses Data



Example: Purses Data

X_t = # of purses reported as stolen at Hyde Park, Chicago

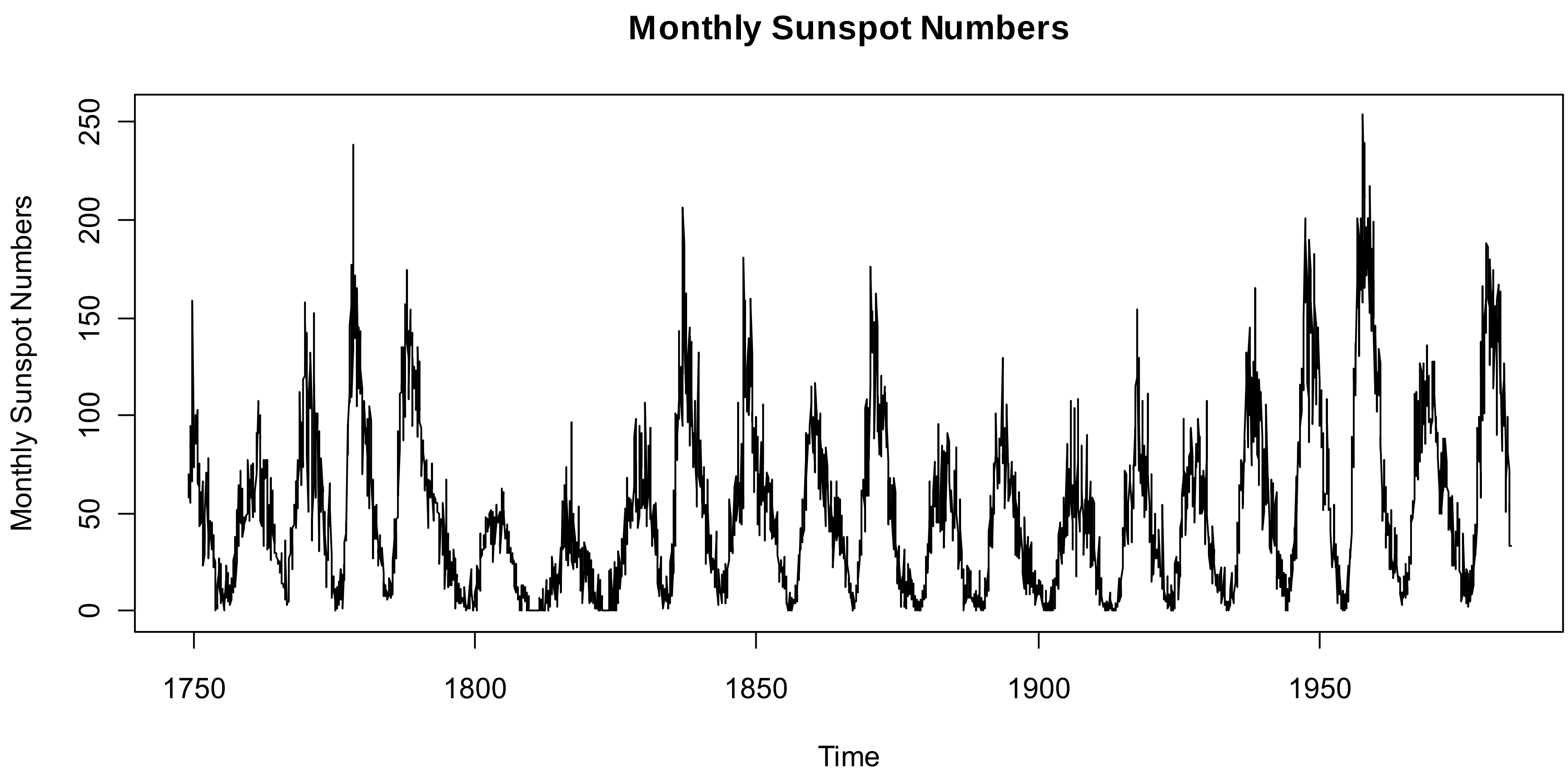
T = {1968.1, 1968.2, ..., 1968.13, ..., 1973.6}

Observations:

- reported figures are aggregated thefts over 28 days
- neither a trend nor a seasonal pattern is apparent
- no further peculiarities

Applied Time Series Analysis
FS 2011 – Week 01

Example: Sunspot Data



Example: Sunspot Data

X_t = # of sunspots observed per month

T = {1700/01, 1700/02, 1700/03, ..., 1988/12}

Observations:

- periodicity ~10 years, but period seems fluctuating
- the increase is quicker than the decline
- aspect ratio is important: banking to 45°

Goals in Time Series Analysis

1) Descriptive Analysis

Visualization of the properties of the series

- time series plot
- decomposition into trend/seasonal pattern/random error
- correlogram for understanding the dependency structure
- spectral analysis for visualizing the periodicity

Goals in Time Series Analysis

2) Modeling

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

- can happen in an explorative way

- use of previous knowledge (if available) is adequate, too

- parameter estimation is a crucial task

- inference: how good does the model fit the data?

Goals in Time Series Analysis

3) Forecasting

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

- will be based on a stochastic model

- builds on the dependency structure and past data

- is an extrapolation, which are always to take with a grain of salt

- similar to driving a car by looking in the rear window mirror

Applied Time Series Analysis
FS 2011 – Week 01

Goals in Time Series Analysis

4) Times Series Regression

Modeling a response time series using one or more input series

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

where ε_t is independent from u_t and v_t, but is not i.i.d.

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

=> **Fitting this model under i.i.d. error assumption leads to unbiased estimates, but:**

- standard errors are grossly wrong
- confindence intervals and tests are misleading

Goals in Time Series Analysis

5) Process Control

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

- fitting a stochastic model

- prediction of future data

- can serve as a monitoring and regulation tool

Applied Time Series Analysis
FS 2011 – Week 01

Probabilistic Notation and Stationarity

Are presented and discussed on the blackboard…