Applied Time Series Analysis
FS 2011 – Week 06

Marcel Dettling

Institute for Data Analysis and Process Design

Zurich University of Applied Sciences

marcel.dettling@zhaw.ch

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

ETH Zürich, March 28, 2011

Applied Time Series Analysis
FS 2011 – Week 06

Forecasting with Time Series

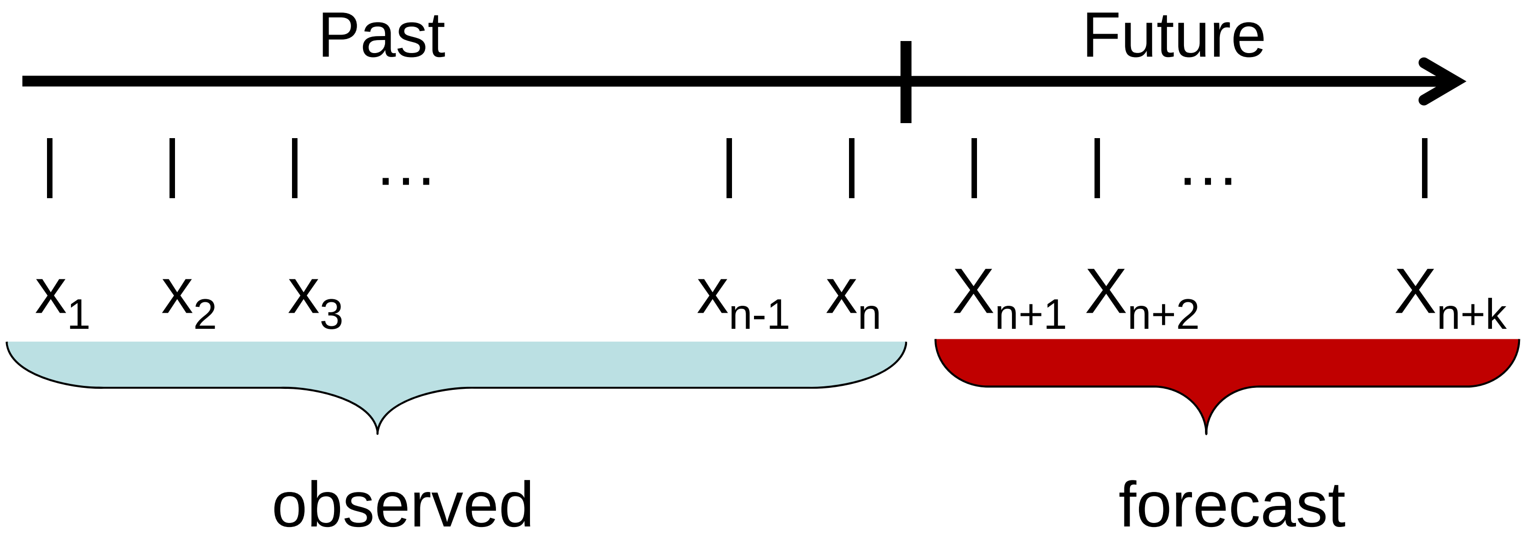
Goal:

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

Important:

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

Forecasting, More Technical



observations

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

estimates

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

Sources of Uncertainty

There are 3 main sources of uncertainty:

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

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

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

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

Zurich University of Applied Sciences

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:

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

→ we use the conditional variance

Applied Time Series Analysis
FS 2011 – Week 06

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!**

AR(1): 1-Step Forecast

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

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

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.

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

Applied Time Series Analysis
FS 2011 – Week 06

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%

AR(1): k-Step Forecast

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

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

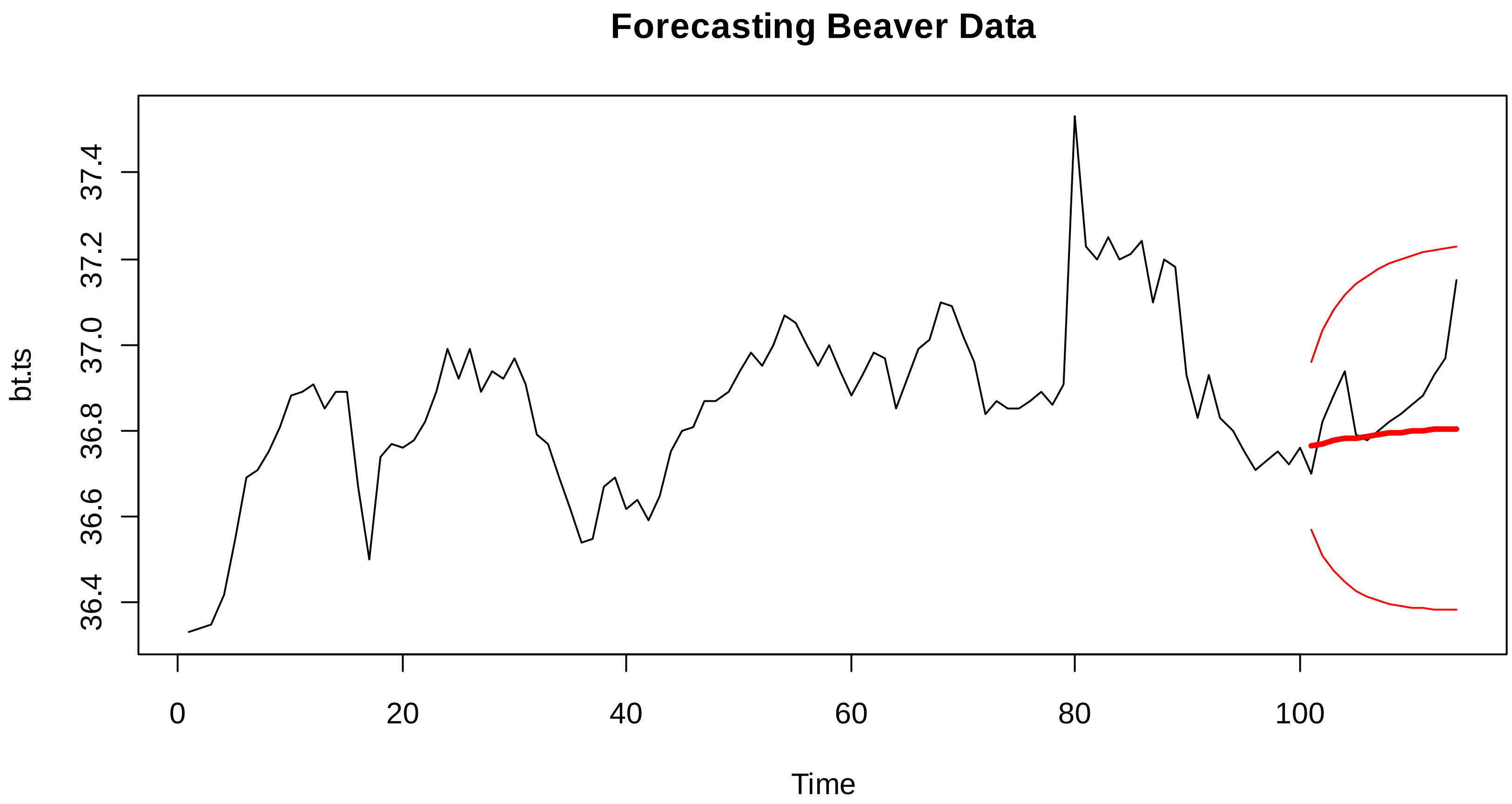
with prognosis interval based on

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

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

Applied Time Series Analysis
FS 2011 – Week 06

Forecasting the Beaver Data



Forecasting AR(p)

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

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

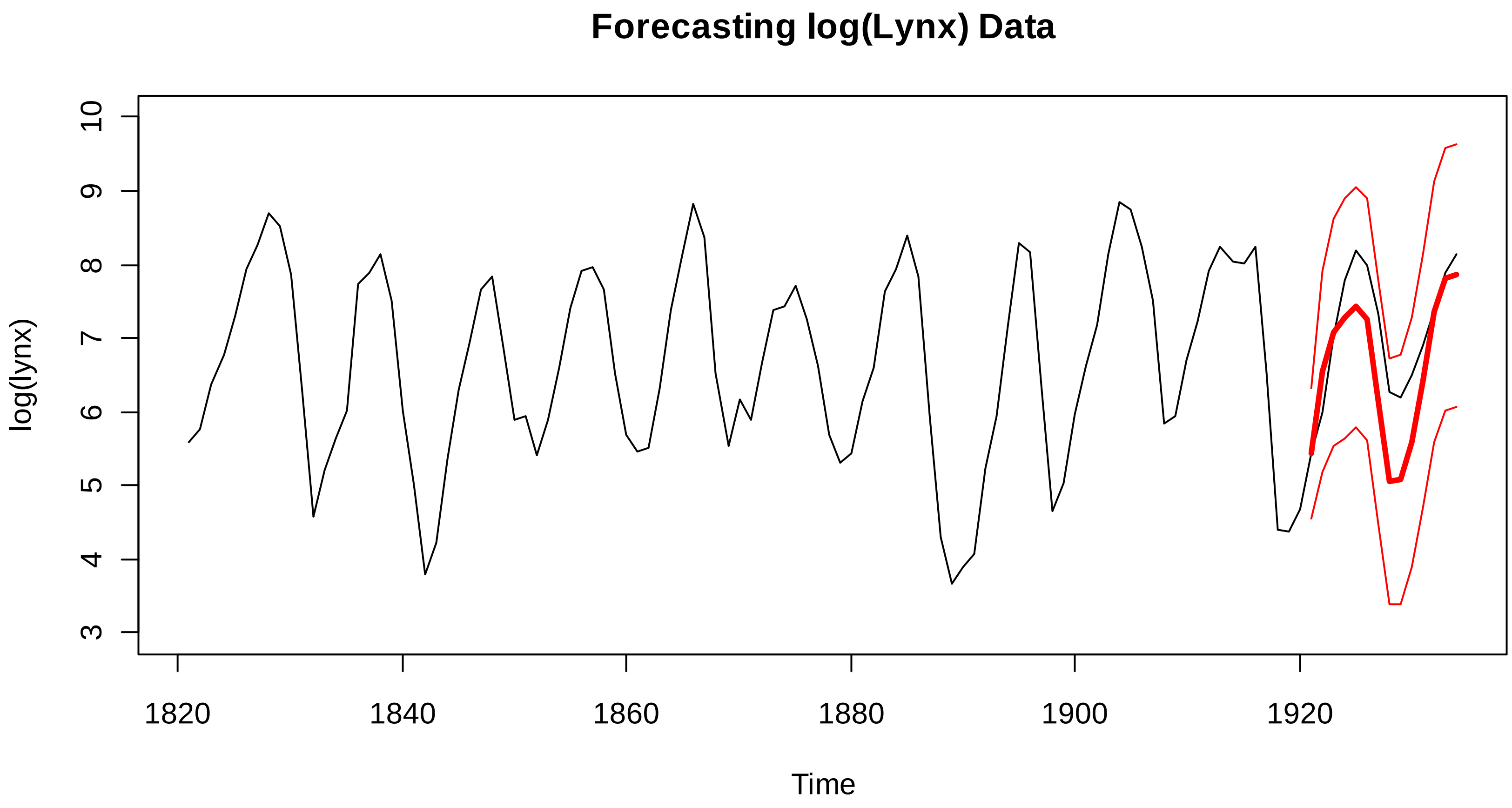
The computations are more complicated, but do not yield any further insight. We are thus doing without.

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

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

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

Forecasting the Lynx Data



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
FS 2011 – Week 06

Exponential Smoothing

Simple exponential smoothing:

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

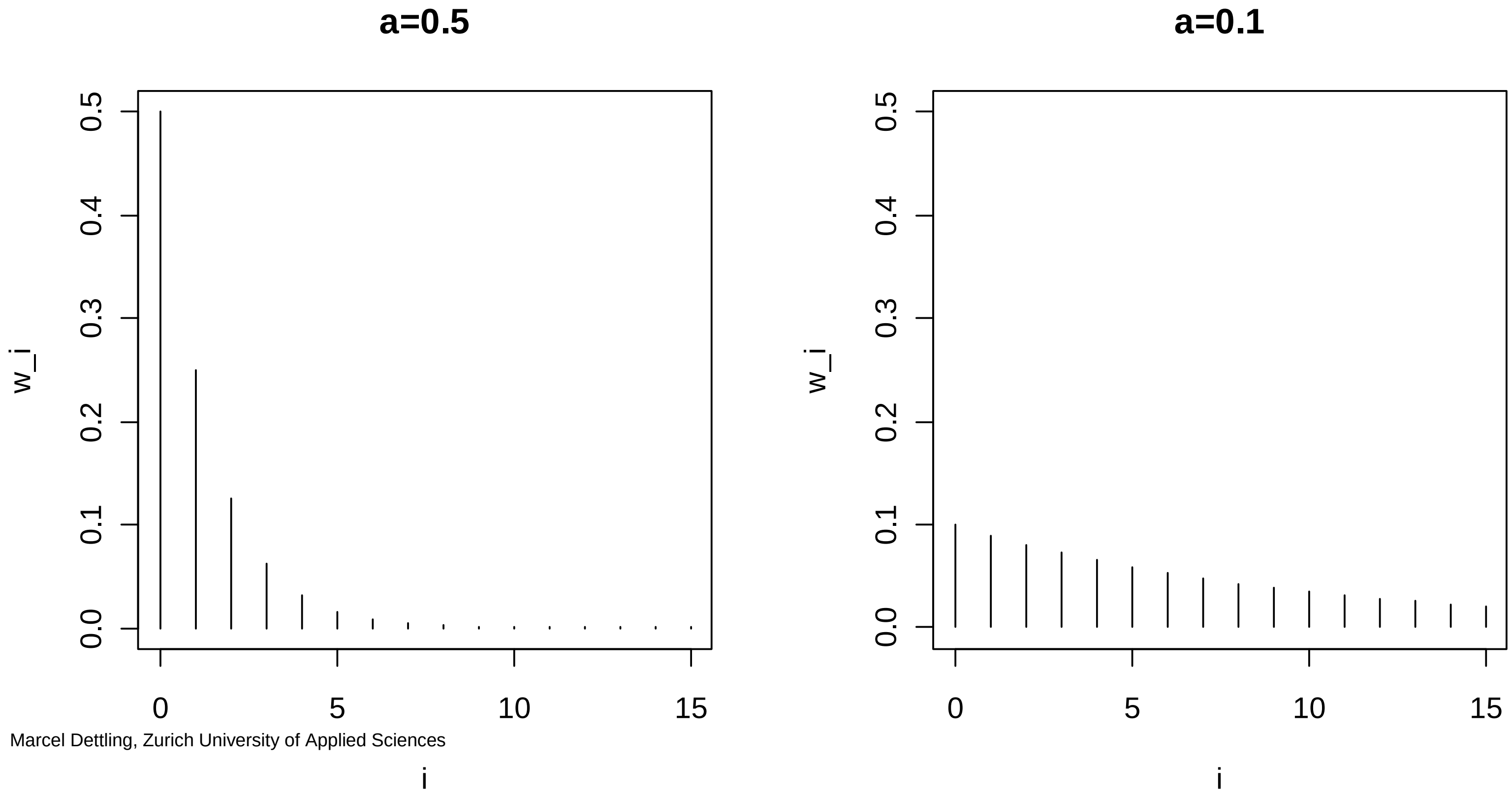
$$\hat{X}_{n+1,n} = \sum_{i=0}^{n-1} w_i x_{n-i} \quad \text{where } w_0 \geq w_1 \geq w_2 \geq \ldots \geq 0 \text{ and } \sum_{i=0}^{n-1} w_i = 1$$

Note that this is a weighted mean over all available, past observations. This is fundamentally different from the AR(p) forecasting scheme!

Choice of Weights

An usual choice are exponentially decaying weights:

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



Forecasting with Exponential Smoothing

The 1-step forecast is:

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

Remarks:

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

Applied Time Series Analysis
FS 2011 – Week 06

Exponential Smoothing: Remarks

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

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

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

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

Review of Previous Topics

- What follows now is a repetition with a short review of the most important previous topics.
- These slides will not be discussed during the lectures, because the content was covered before.
- You should by now be well familiar with what is written here, if not, please study it on your own.

Applied Time Series Analysis
FS 2011 – Week 06

(Weak)Stationarity

- The mathematical concepts that enables time series analysis
- Probabilistic character of the series remains constant over time
- Moments: $\mu_t = E[X_t]; \ \sigma_t^2 = Var(X_t); \ \gamma(s,t) = Cov(X_s, X_t)$
- There is no proof for stationarity, only evidence against it

→ Stationarity is a hypothesis that permanently needs to be questioned during an analysis. If there is evidence against it, it needs to be rejected, i.e. corrections need to be made.

Applied Time Series Analysis
FS 2011 – Week 06

How to Check for Stationarity?

- Time series plot
- Decomposition with stl()
- Analysis of ACF/PACF
- Residual analysis after fitting a (stationary) model
- Simulation from a (stationary) model fit
- Forecasting accuracy

Applied Time Series Analysis
FS 2011 – Week 06

How to Achieve Stationarity?

- (Semi-)parametric modeling
- Non-parametric smooting
- Differencing
- STL-decomposition

AR(p)-Modeling

- Observation X_t is a linear combination of the *p* past ones
- Probabilistic character of the series remains constant over time
- Model choice is a difficult topic…

→ The decision for a specific model order p is a hypothesis that permanently needs to be questioned during an analysis. If there is evidence against it, it needs to be rejected, i.e. corrections need to be made.

Choice of the Order p

- Generally by using ACF/PACF
- Choose the simplest model, i.e. lowest *p*, such that:
  - residuals look satisfactory (ts-plot, ACF/PACF)
  - check AIC/BIC, should support the decision made
  - good forecasting accuracy out-of-sample

Forecasting with Time Series

Goal:

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

Important:

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