# Applied Time Series Analysis
## FS 2012 – Week 09

*Marcel Dettling*

Institute for Data Analysis and Process Design

Zurich University of Applied Sciences

marcel.dettling@zhaw.ch

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

ETH Zürich, April 23, 2012

# *Forecasting with Time Series*

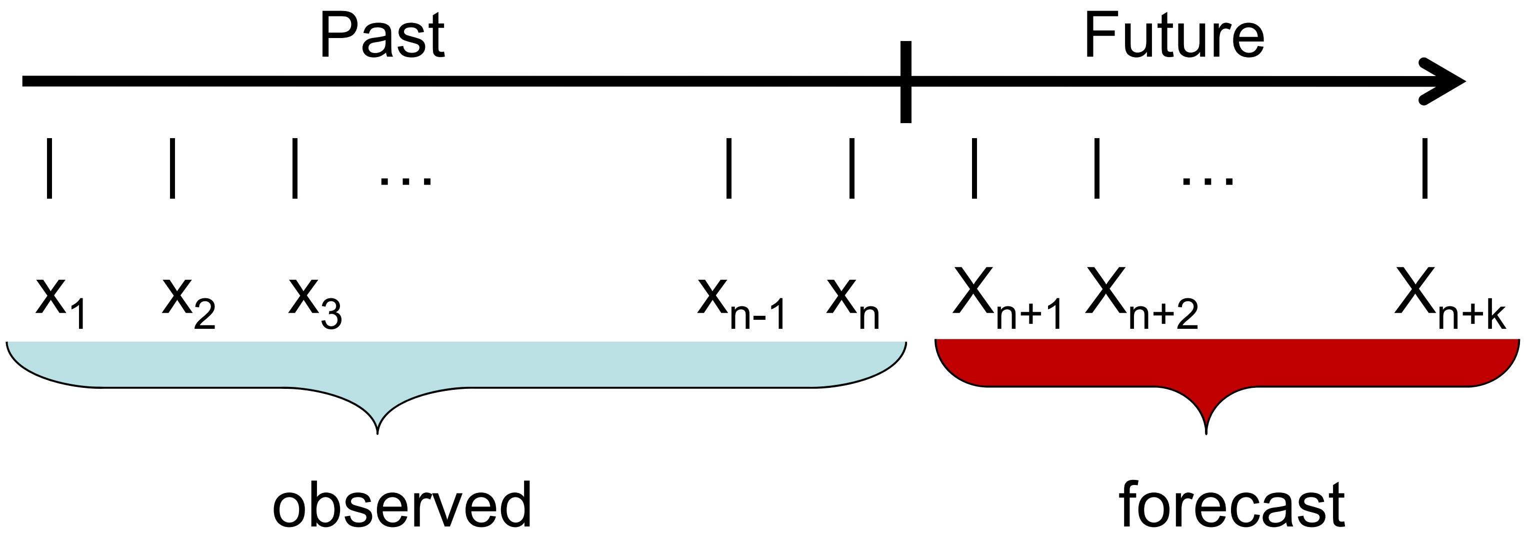
## Goal:

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

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

# Applied Time Series Analysis
## FS 2012 – Week 09

## *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 $\mu$ accurately estimated?

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

# Applied Time Series Analysis
## FS 2012 – Week 09

## *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

# *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
## FS 2012 – Week 09

## *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.

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

## *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:

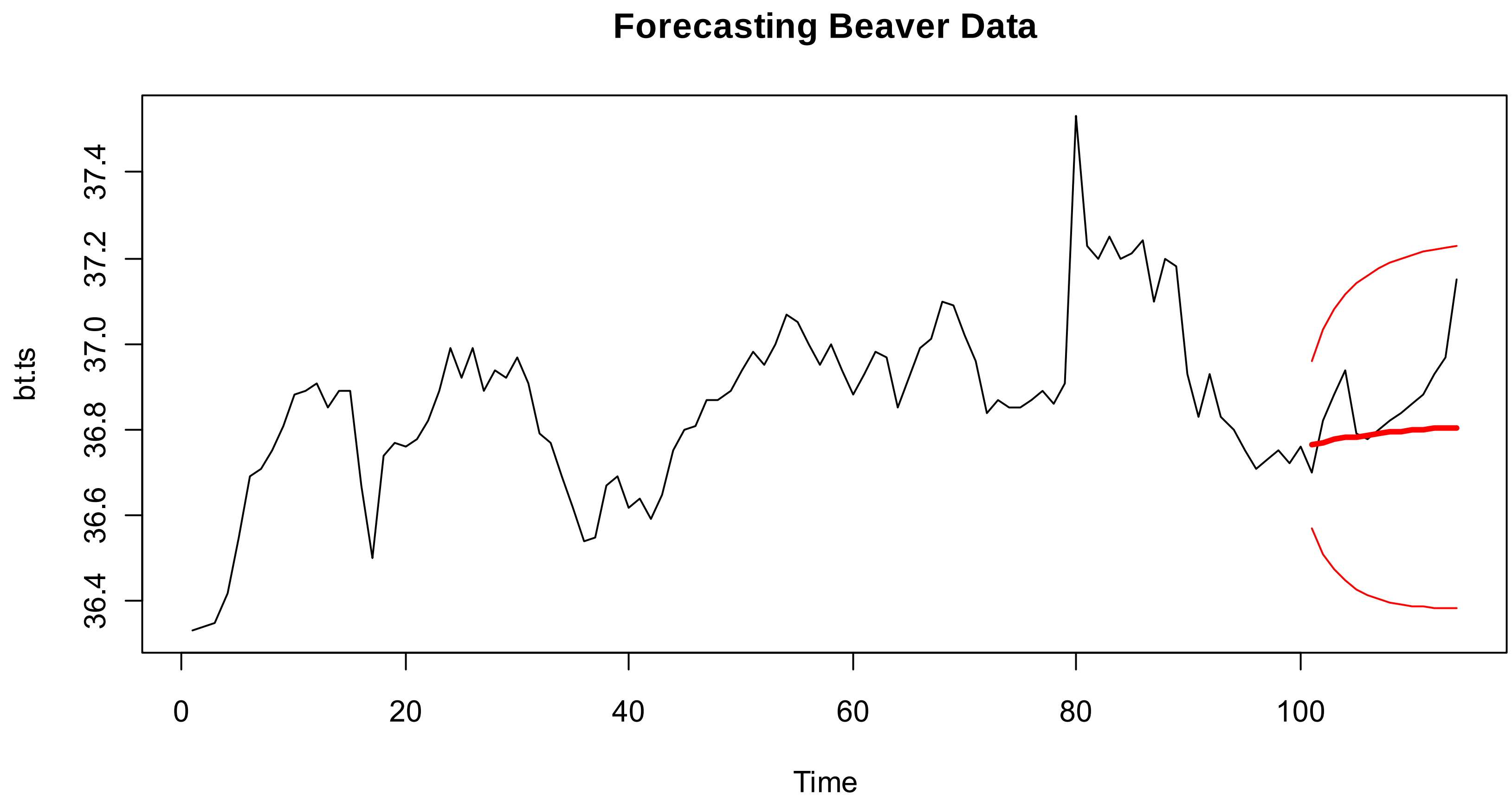
$$\hat{X}_{n+k,n} = \alpha_1^k (x_n - \mu) + \mu$$

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 $\mu$ and $\sigma_X^2$ respectively.

# *Forecasting the Beaver Data*



# Applied Time Series Analysis
## FS 2012 – Week 09

# *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.

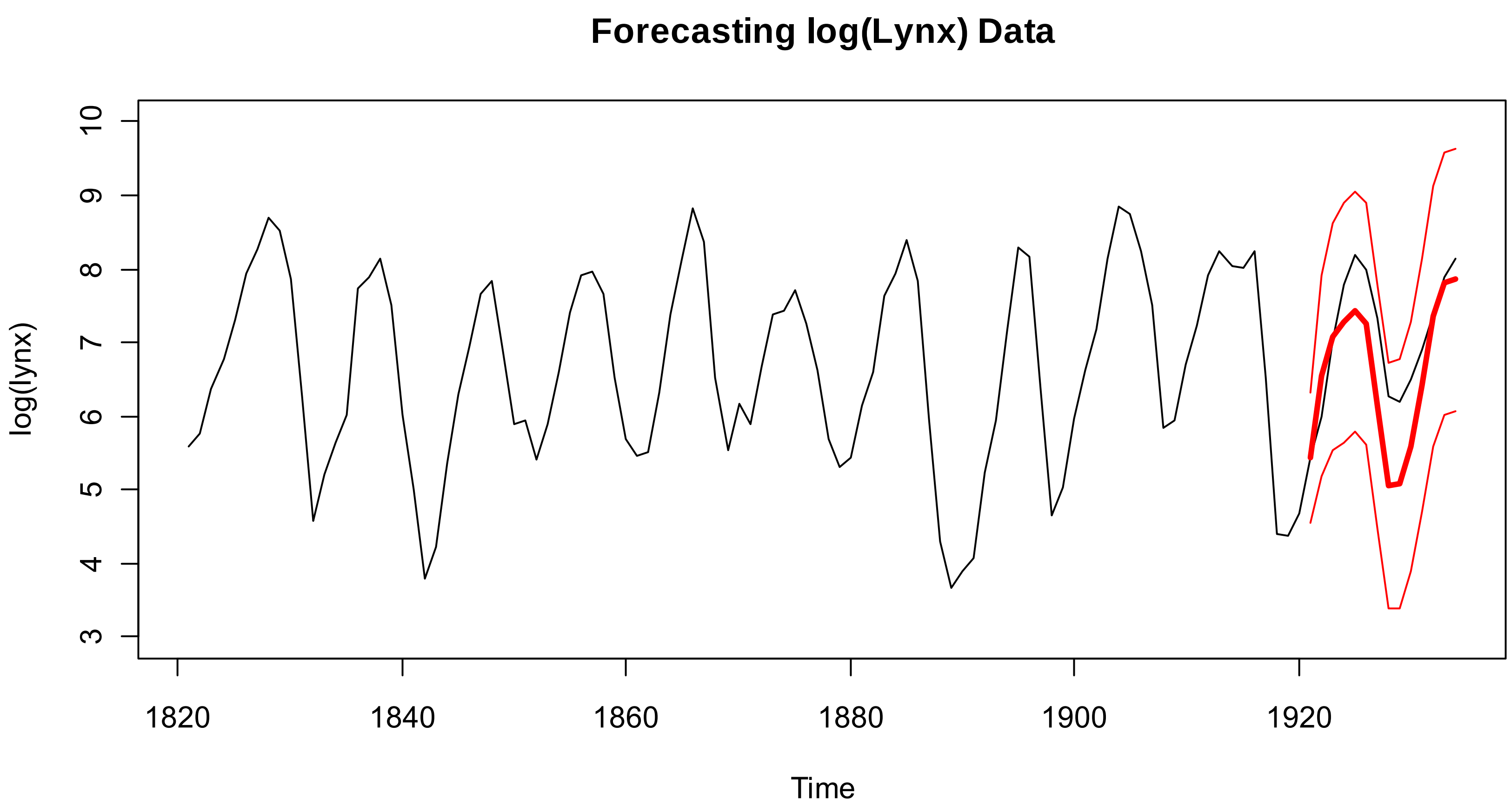
1-step-forecast: $\hat{X}_{n+1,n} = \alpha_1(x_n - \mu) + ... + \alpha_p(x_{n+1-p} - \mu) + \mu$

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

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

# Applied Time Series Analysis
## FS 2012 – Week 09

# *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 $p$.

# *Forecasting with ARMA(p,q)*

**There are 3 main sources of uncertainty:**

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

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

3) Are the parameters $\alpha$ , $\beta$ , $\sigma_E^2$ and $\mu$ accurately estimated?

→ **we will here restrict to short-term 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:

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

→ we use the conditional variance

# *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!**

# *MA(1) Forecasting: Summary*

- We have seen that for an MA(1)-process, the k-step forecast for k>1 is equal to $\mu$.

- In case of k=1, we obtain for the MA(1)-forecast:

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

The conditional expectation is (too) difficult to compute

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

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

# *MA(1) Forecasting: Summary*

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

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

- 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 $\hat{\mu}$.

- 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
## FS 2012 – Week 09

# ARMA(p,q) Forecasting

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

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

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

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

where …

## *ARMA(p,q) Forecasting*

…where

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

and

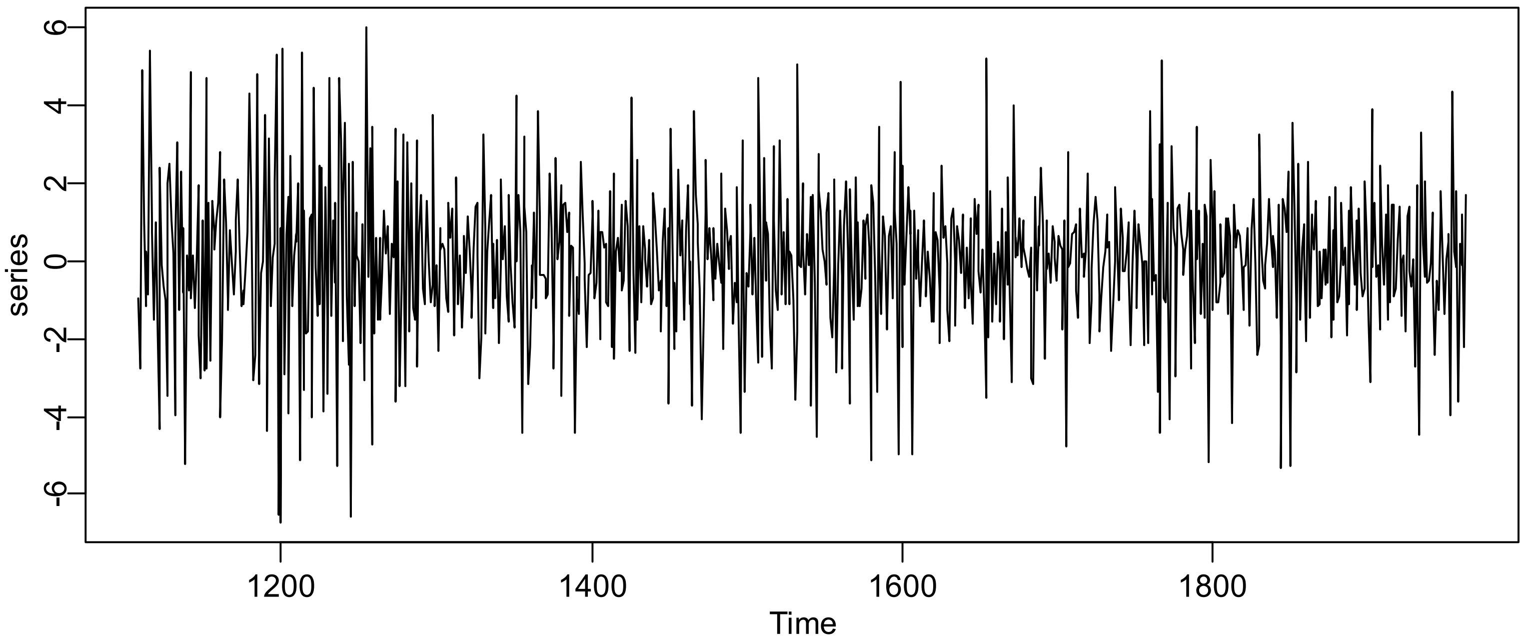
$$E[E_t \mid X_{-\infty}^{n}] = \begin{cases} e_t & \text{if } t \le n \\ 0 & \text{if } t > n \end{cases}$$

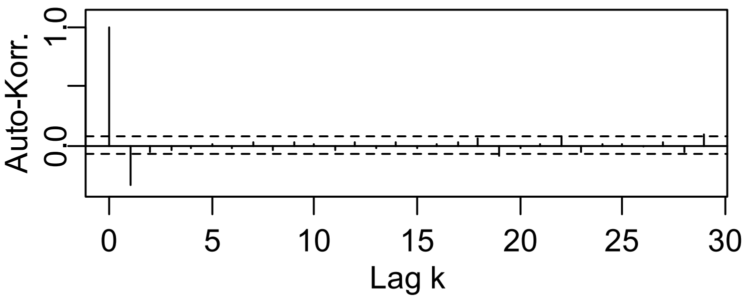
with

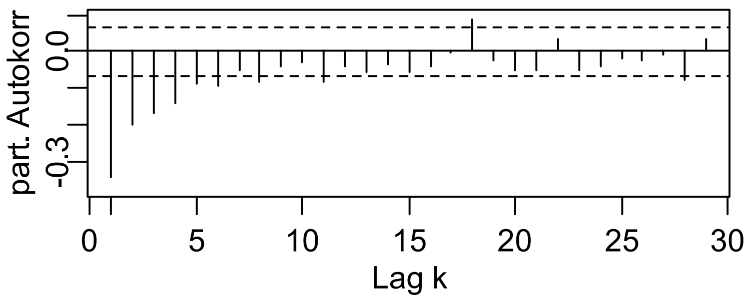
$$e_t = x_t - \mu - \sum_{i=1}^{p} \alpha_i (x_{t-i} - \mu) + \sum_{j=1}^{q} \beta_j e_{t-j}$$

# Applied Time Series Analysis
## FS 2012 – Week 09

## *ARMA(p,q) Forecasting: Douglas Fir*



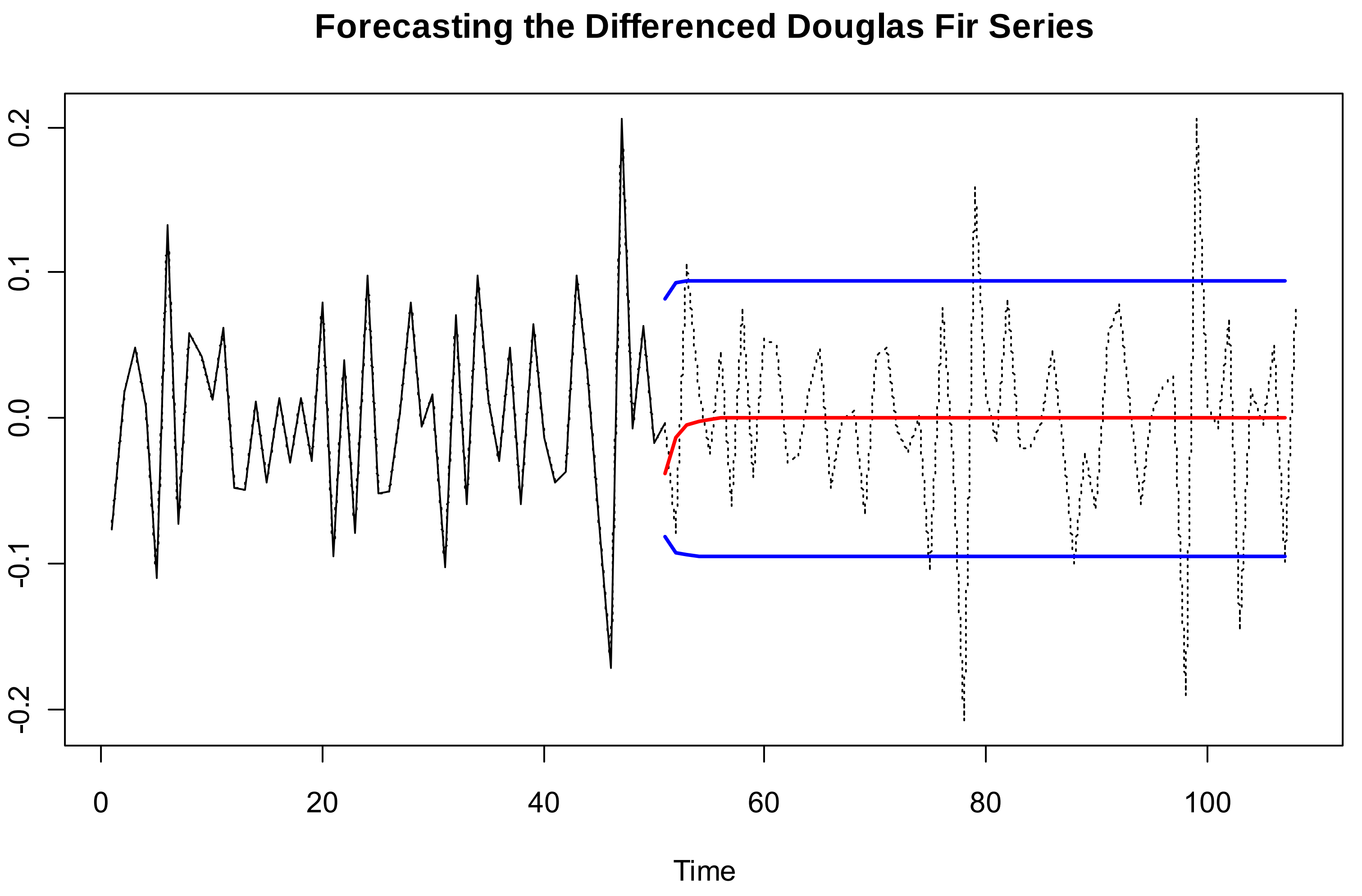



Zurich University
of Applied Sciences

# Applied Time Series Analysis
## FS 2012 – Week 09

# *ARMA(p,q) Forecasting: Example*



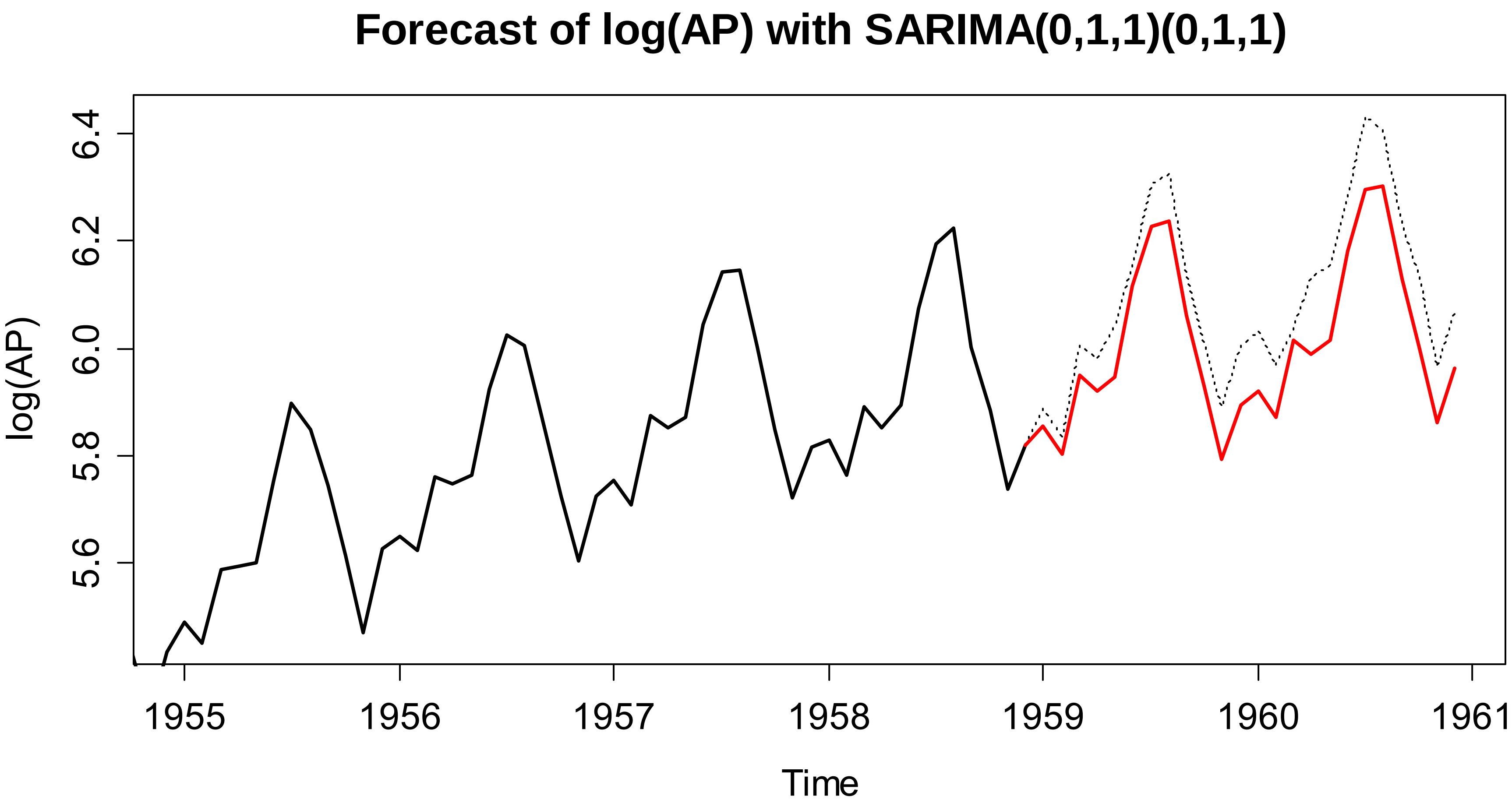
# Applied Time Series Analysis
## FS 2012 – Week 09

# *Forecasting with SARIMA*

Some general remarks about forecasting with ARIMA/SARIMA:

...

# Forecasting with SARIMA: Example



# *Forecasting Decomposed Series*

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

...

# Example: Swissport

- Budgeting for 2008 is done in August 2007.

  - Forecasts are generated on a month by month basis

  - Data are available from January 2005 to May 2007

- Modeling and prediction for the effort based on

  - Number of passengers

  - Number of aircraft handled

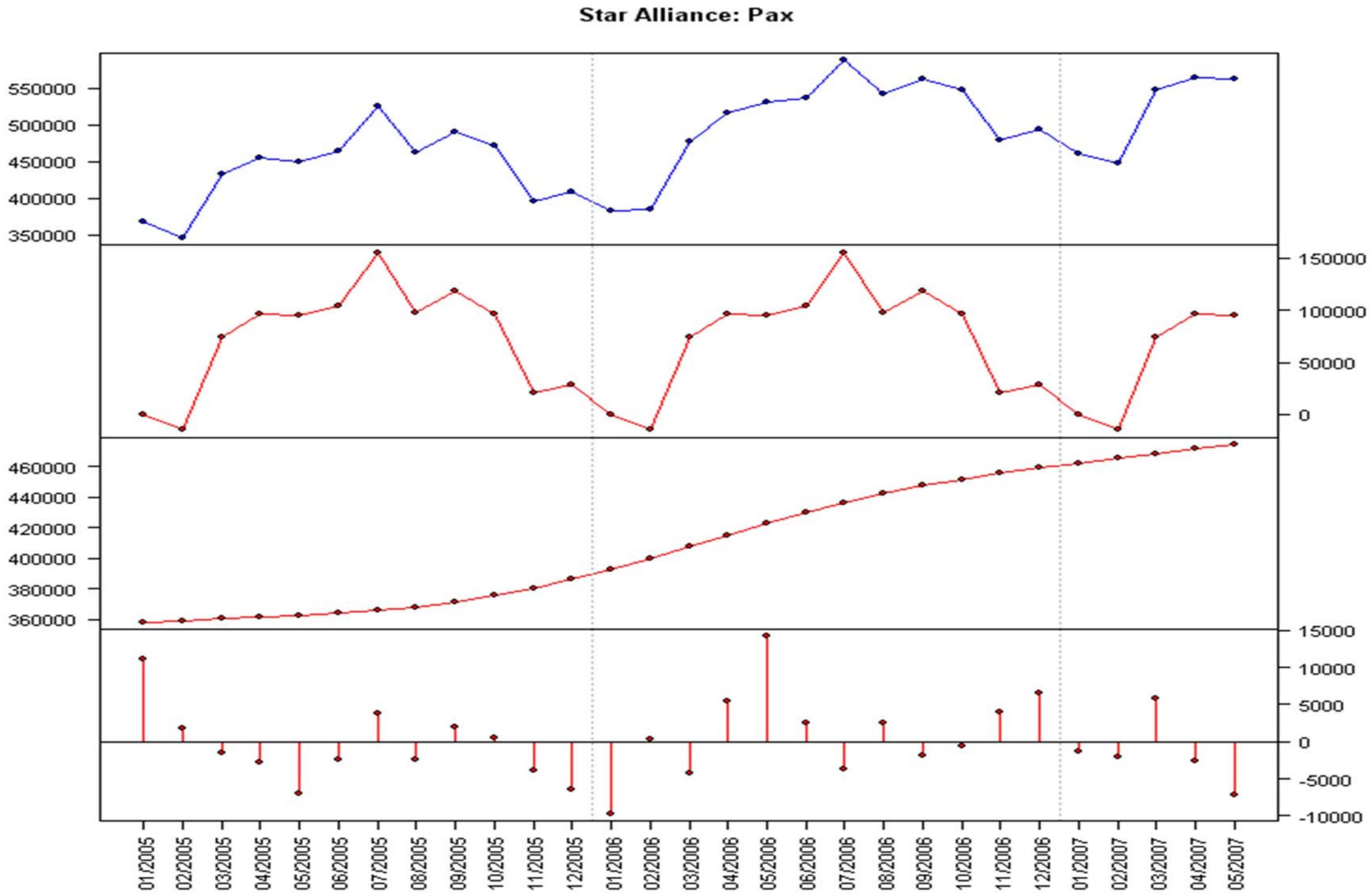
- Modeling and prediction of the revenue

  - Volume depends on the number of aircraft handled

  - Prices are fixed with airlines, but can change

# Applied Time Series Analysis
## FS 2012 – Week 09

# *Swissport: STL-Decomposition*



# *Forecasting Strategy for Swissport*

- Decomposition of the data into:

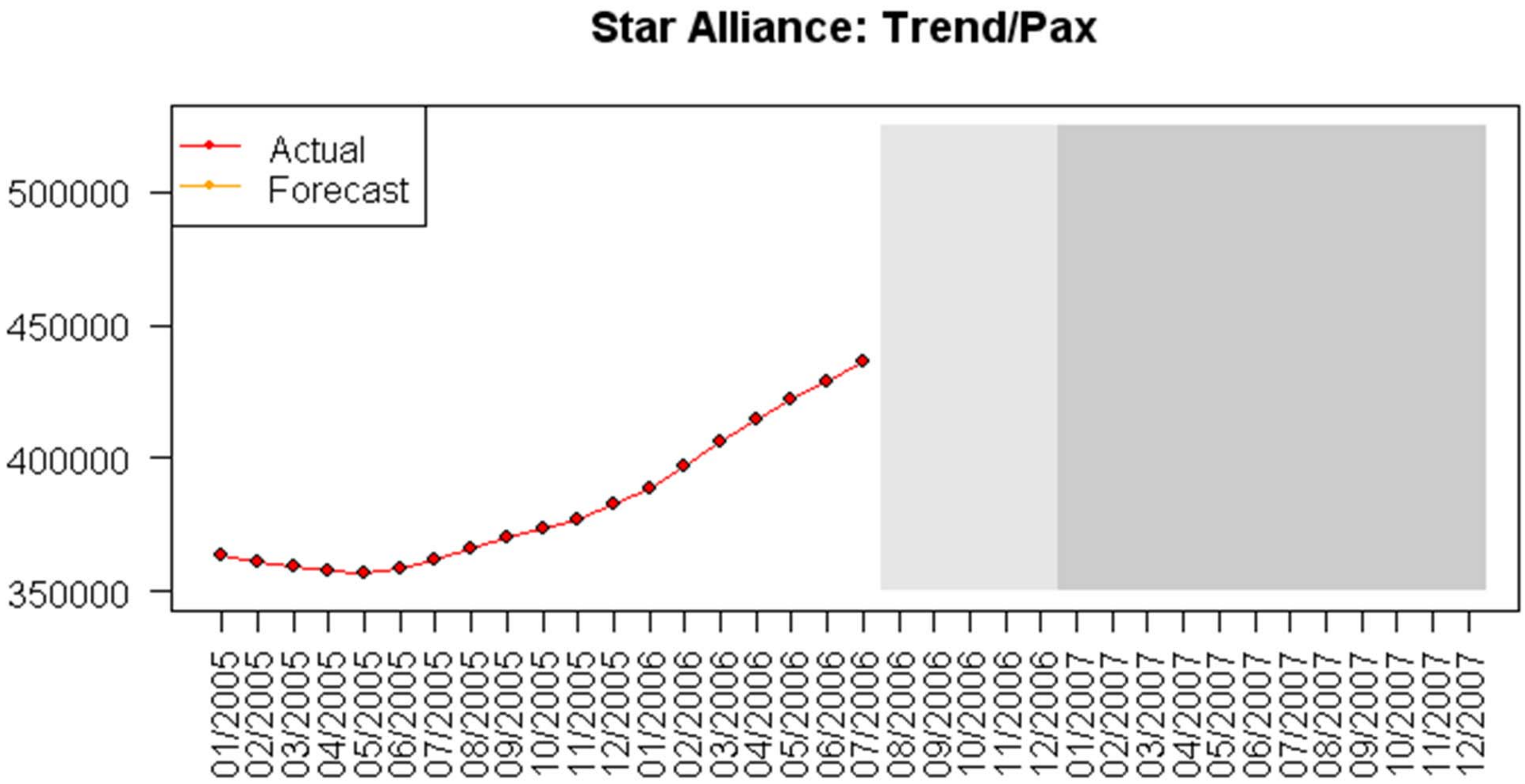
  - trend → smooth
  - seasonal pattern → stable
  - remainder → stationary & small

- Forecasts are generated with the following method:

  - keep the seasonal pattern constant
  - suggest linear extrapolation of the trend
  - trend suggestion can be altered by management
  - fit a time series model for the remainder and predict
  - aggregate the forecasts of all 3 components

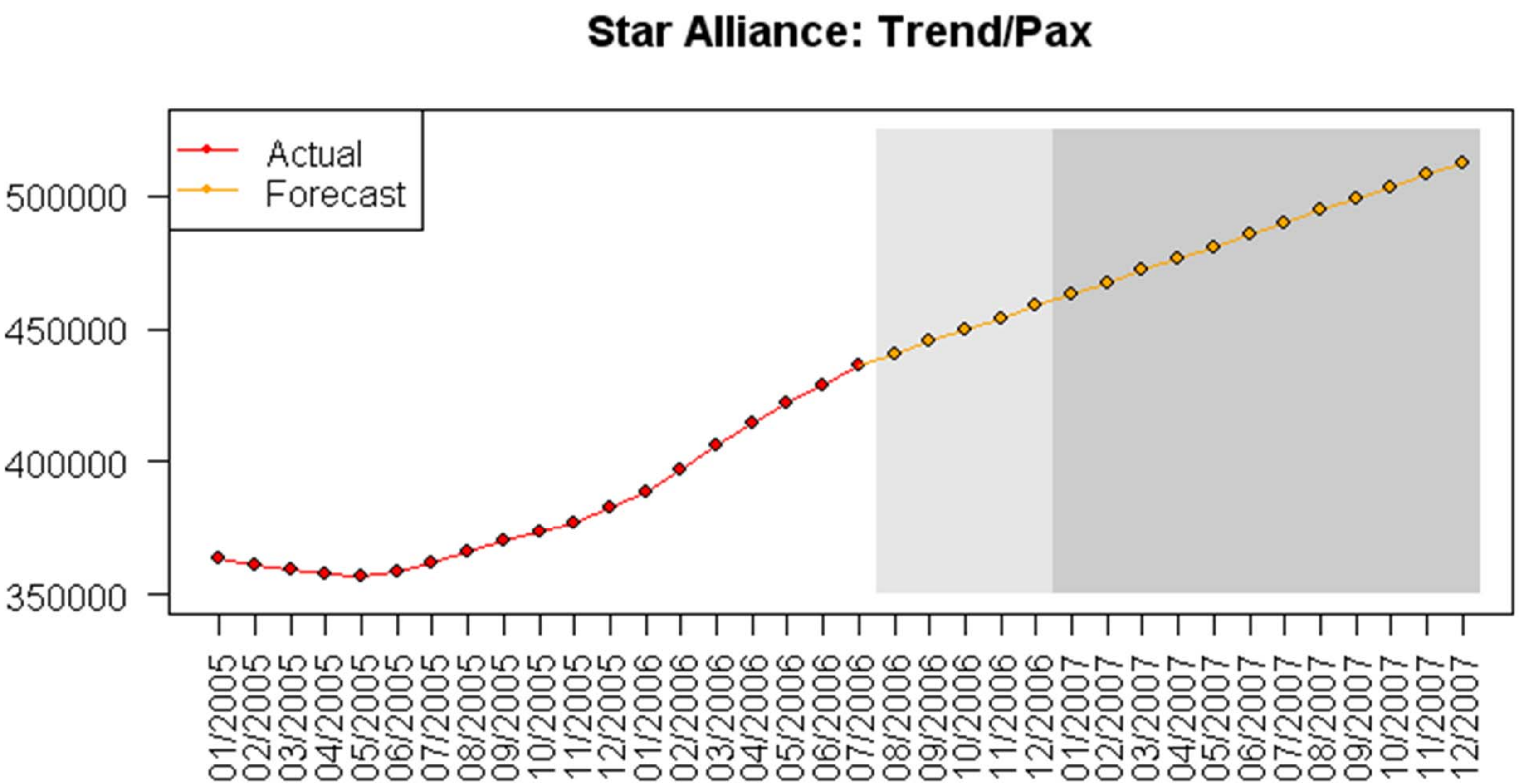
# Applied Time Series Analysis
## FS 2012 – Week 09

# *Swissport: Trend Forecasting*



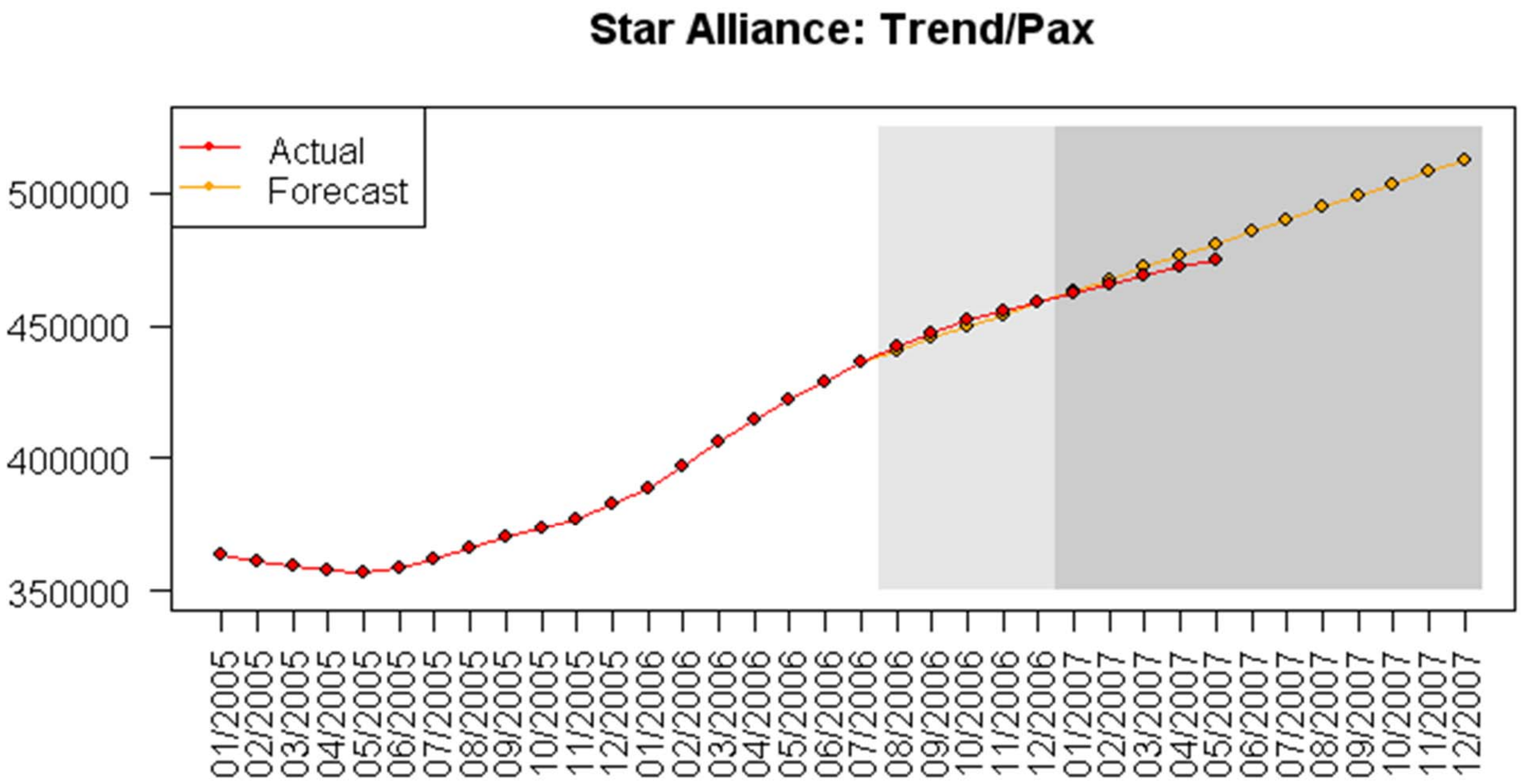
# Applied Time Series Analysis
## FS 2012 – Week 09

# *Swissport: Trend Forecasting*



# Applied Time Series Analysis
## FS 2012 – Week 09

# *Swissport: Trend Forecasting*



# *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

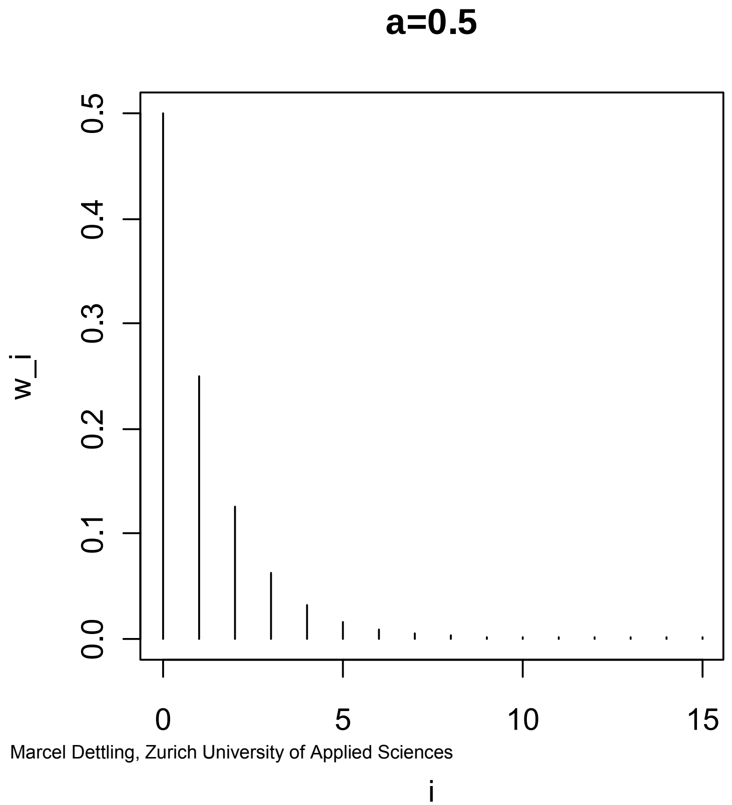
$$\hat{X}_{n+1,n} = \sum_{i=0}^{n-1} w_i x_{n-i} \quad \text{where } w_0 \geq w_1 \geq w_2 \geq \ldots \geq 0 \text{ and } \sum_{i=0}^{n-1} w_i = 1$$

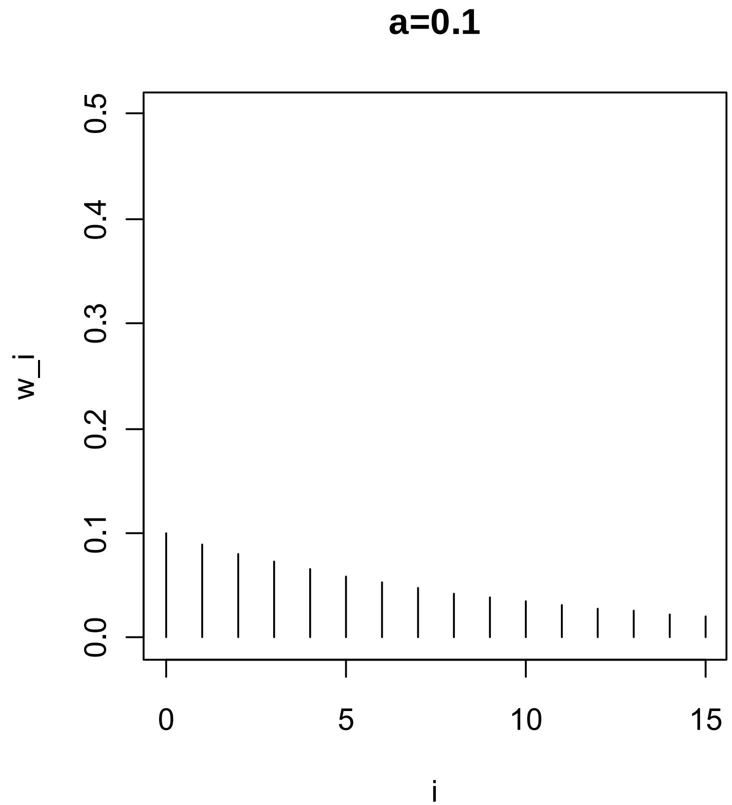
**Note that this is a weighted mean over all available, past observations. This is fundamentally different from the AR(p) forecasting scheme!**

# 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:

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

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.

# *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.