# Applied Time Series Analysis
## FS 2012 – Week 05

*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 18, 2012

# *AR(p)-Model*

We here introduce the AR(p)-model

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

where again

$E_t$ is i.i.d with $E[E_t] = 0$ and $Var(E_t) = \sigma_E^2$

Under these conditions, $E_t$ is a white noise process, and we additionally require **causality**, i.e. $E_t$ being an **innovation**:

$E_t$ is independent of $X_s, s < t$

# ***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
## FS 2012 – Week 05

***AR-Modelling***

1

**Identification of the Order p**

- ACF/PACF
- AIC/BIC

2

**Parameter Estimation**

- Regression
- Yule-Walker
- MLE
- Burg

3

**Model Diagnostics**

- Residual Analysis
- Simulation

# *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.

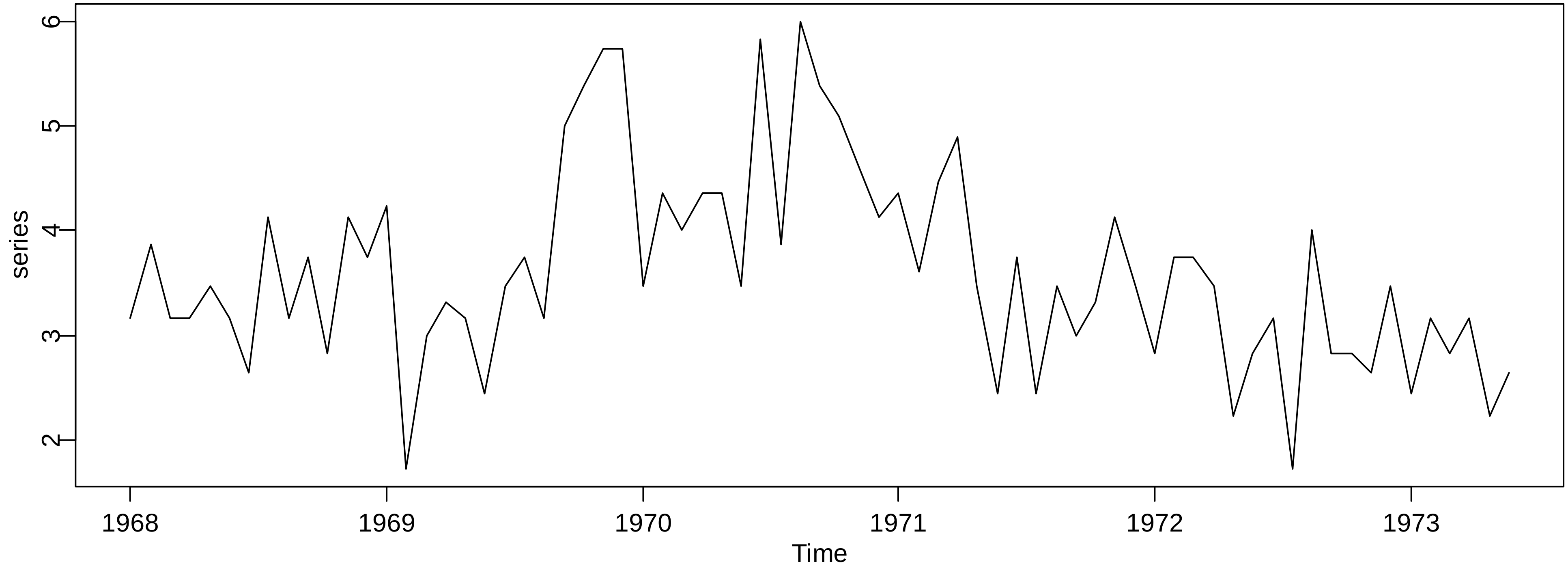
- For all AR(p)-models, the **PACF** is equal to zero for all lags k>p.

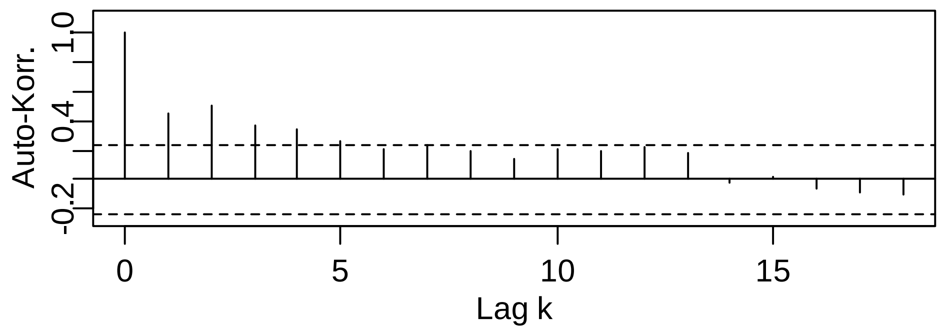
If what we observe is fundamentally different from the above, it is unlikely that the series was generated from an AR(p)-process. We thus need other models, maybe more sophisticated ones.

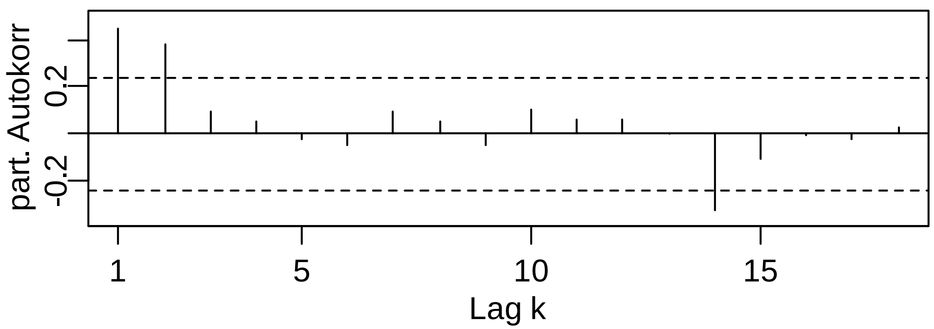
**Remember that the sample ACF has a few peculiarities and is tricky to interpret!!!**

# Applied Time Series Analysis
## FS 2012 – Week 05

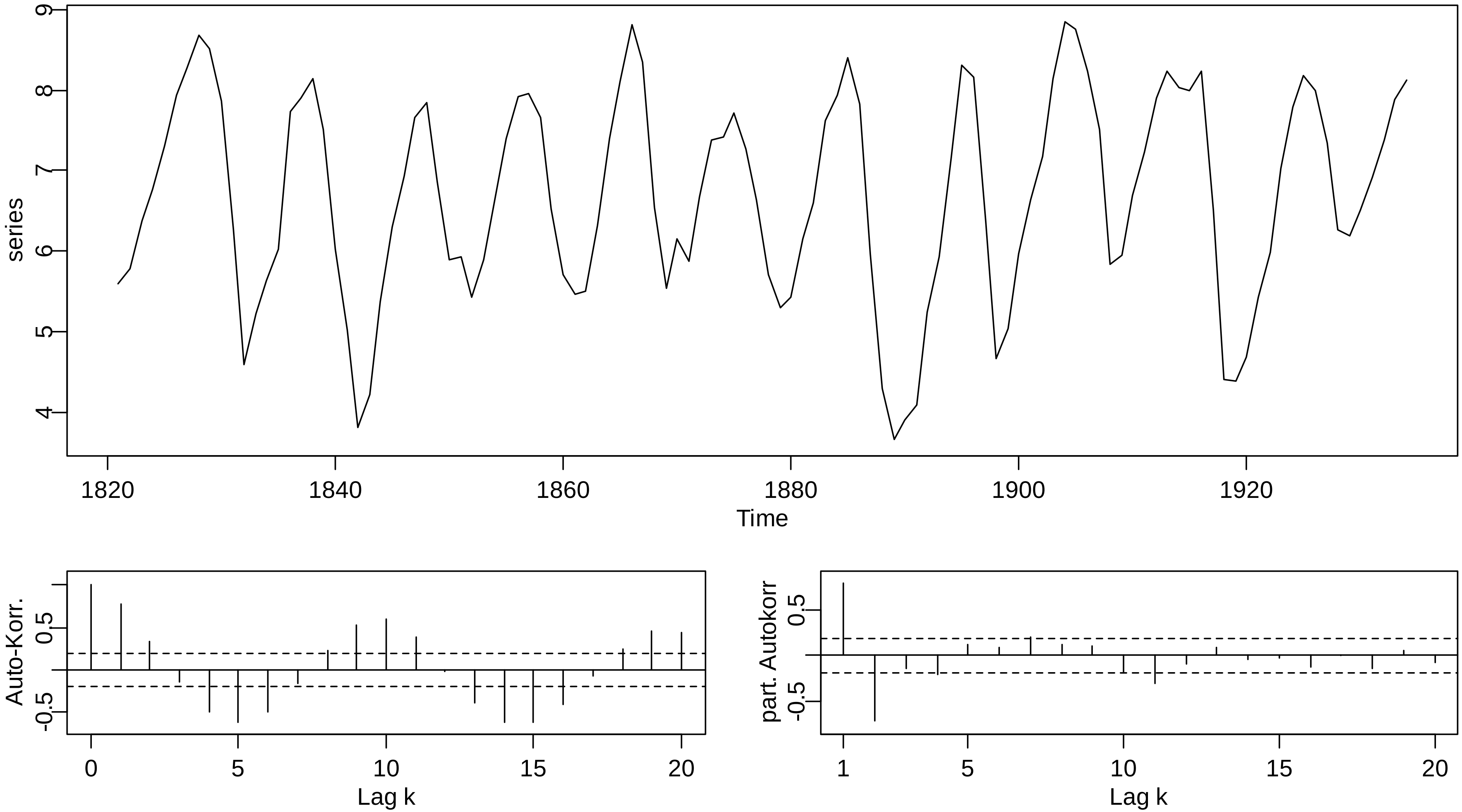
# *Model Order for sqrt(purses)*



# *Model Order for log(lynx)*



# Basic Idea for Parameter Estimation

We consider the stationary AR(p)

$$(X_t - \mu) = \alpha_1 (X_{t-1} - \mu) + \ldots + \alpha_p (X_{t-p} - \mu) + E_t$$

where we need to estimate

| | |
|---|---|
| $\alpha_1, \ldots, \alpha_p$ | model parameters |
| $\sigma_E^2$ | innovation variance |
| $\mu$ | general mean |

# *Approach 1: Regression*

Response variable: $X_t$, t = 1,…,n-p

Explanatory variables: $X_{t-1}$, t = 2,…,n-p+1

$X_{t-2}$, t = 3,…,n-p+2

…

$X_{t-p}$, t = p+1,…,n

We can now use the regular LS framework. The coefficient estimates then are the estimates for $\alpha_1,...,\alpha_p$. Moreover, we have

$$\sigma_E^2 = \frac{1}{n-2p-1}\sum_{i=1}^{n-p} r_i^2 \quad \text{and} \quad \hat{\mu} = \frac{\hat{\alpha}_0}{1-\hat{\alpha}_1-\hat{\alpha}_2-...-\hat{\alpha}_p}$$

# *Approach 1: Regression*

Preparing the design matrix

```
> d.Psqrt <- sqrt(Purses)
> d.Psqrt.mat <- ts.union(Y=d.Psqrt,X1=lag(d.Psqrt,-1),X2=lag(d.Psqrt,-2))
> d.Psqrt.mat[1:5,]
         Y    X1    X2
[1,] 3.162    NA    NA
[2,] 3.873 3.162    NA
[3,] 3.162 3.873 3.162
[4,] 3.162 3.162 3.873
[5,] 3.464 3.162 3.162
```

# *Approach 1: Regression*

Fitting the LS model

```
> r.Pfit <- lm(Y ~ .,data=data.frame(d.Psqrt.mat))

> summary(r.Pfit)

Call: lm(formula = Y ~ ., data = data.frame(d.Psqrt.mat))

Residuals:    Min      1Q  Median      3Q     Max

          -2.0925 -0.4088 -0.0536  0.4286  1.9774

Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept)    1.117      0.448    2.49  0.01513 *

X1             0.283      0.113    2.50  0.01474 *

X2             0.403      0.114    3.53  0.00077 ***
```

# Applied Time Series Analysis
## FS 2012 – Week 05

# *Approach 1: Regression*

Output from the LS model

```
Residual standard error: 0.8 on 66 degrees of freedom
Multiple R-Squared: 0.332,      Adjusted R-squared: 0.312
F-statistic: 16.4 on 2 and 66 DF,  p-value: 1.64e-006
```

Thus we have:

$$\hat{\alpha}_1 = 0.283, \hat{\alpha}_2 = 0.403$$

$$\hat{\mu} = \frac{1.117}{1 - 0.283 - 0.403} = 3.56$$

$$\hat{\sigma}_E^2 = (0.8004)^2 = 0.64$$

# *Overview of the Estimates*

| | Regression | Yule-Walker | MLE | Burg |
|---|---|---|---|---|
| $\hat{\alpha}_1$ | 0.283 | - | - | - |
| $\hat{\alpha}_2$ | 0.403 | - | - | - |
| $\hat{\mu}$ | 3.56 | - | - | - |
| $\hat{\sigma}_E^2$ | 0.64 | - | - | - |

# Approach 2: Yule-Walker

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

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

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

$\hat{\mu}$ is the arithmetic mean of the time series
$\hat{\sigma}_E^2$ is the estimated variance of the residuals

→ **see example on the blackboard for an AR(2)-model**

# *Approach 2: Yule-Walker*

The Yule-Walker-Estimation is implemented in R

```
> ar.yw(sqrt(purses))

Call:
     ar.yw.default(x = sqrt(purses))

Coefficients:
     1       2
0.2766  0.3817

Order selected 2  sigma^2 estimated as  0.639
```

# *Overview of the Estimates*

| | Regression | Yule-Walker | MLE | Burg |
|---|---|---|---|---|
| $\hat{\alpha}_1$ | 0.283 | 0.277 | - | - |
| $\hat{\alpha}_2$ | 0.403 | 0.382 | - | - |
| $\hat{\mu}$ | 3.56 | 3.61 | - | - |
| $\hat{\sigma}_E^2$ | 0.64 | 0.64 | - | - |

# *Approach 3: Maximum-Likelihood-Estimation*

**Idea**: Determine the parameters such that, given the observed time series $x_1,\ldots,x_n$, 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 $X = (X_1, \ldots, X_n)$**

# Approach 3: Maximum-Likelihood-Estimation

If we assume the AR(p)-model

$$(X_t - \mu) = \alpha_1 (X_{t-1} - \mu) + ... + \alpha_p (X_{t-p} - \mu) + E_t$$

and i.i.d. normally distributed innovations

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

the time series vector has a multivariate normal distribution

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

with covariance matrix V that depends on the model parameters $\alpha$ and $\hat{\sigma}_E^2$.

# *Approach 3: Maximum-Likelihood-Estimation*

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

$\alpha$ , $\mu$ and $\hat{\sigma}_E^2$.

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

→ **This is a highly complex non-linear optimization problem that requires sophisticated algorithms.**

# *Approach 3: 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
```

# *Overview of the Estimates*

| | Regression | Yule-Walker | MLE | Burg |
|---|---|---|---|---|
| $\hat{\alpha}_1$ | 0.283 | 0.277 | 0.275 | - |
| $\hat{\alpha}_2$ | 0.403 | 0.382 | 0.395 | - |
| $\hat{\mu}$ | 3.56 | 3.61 | 3.55 | - |
| $\hat{\sigma}_E^2$ | 0.64 | 0.64 | 0.6 | - |

# *Approach 4: Burg‘s Algorithm*

**Idea**: Use non-linear optimization to minimize the in-sample forecasting error of a time-reversible stationary process.

→ **This estimation is distribution free!**

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

In R: > `ar.burg(d.Psqrt, order=2, demean=TRUE)`

# *Overview of the Estimates*

| | Regression | Yule-Walker | MLE | Burg |
|---|---|---|---|---|
| $\hat{\alpha}_1$ | 0.283 | 0.277 | 0.275 | 0.272 |
| $\hat{\alpha}_2$ | 0.403 | 0.382 | 0.395 | 0.397 |
| $\hat{\mu}$ | 3.56 | 3.61 | 3.55 | 3.61 |
| $\hat{\sigma}_E^2$ | 0.64 | 0.64 | 0.6 | 0.6 |

# *Summary of Estimation Methods*

## Regression:

+ simple, no specific procedures required
- resulting AR(p) non-stationary, distribution assumption

## Yule-Walker:

+ easy to understand, no specific procedures required
- estimates will be biased, especially for short series

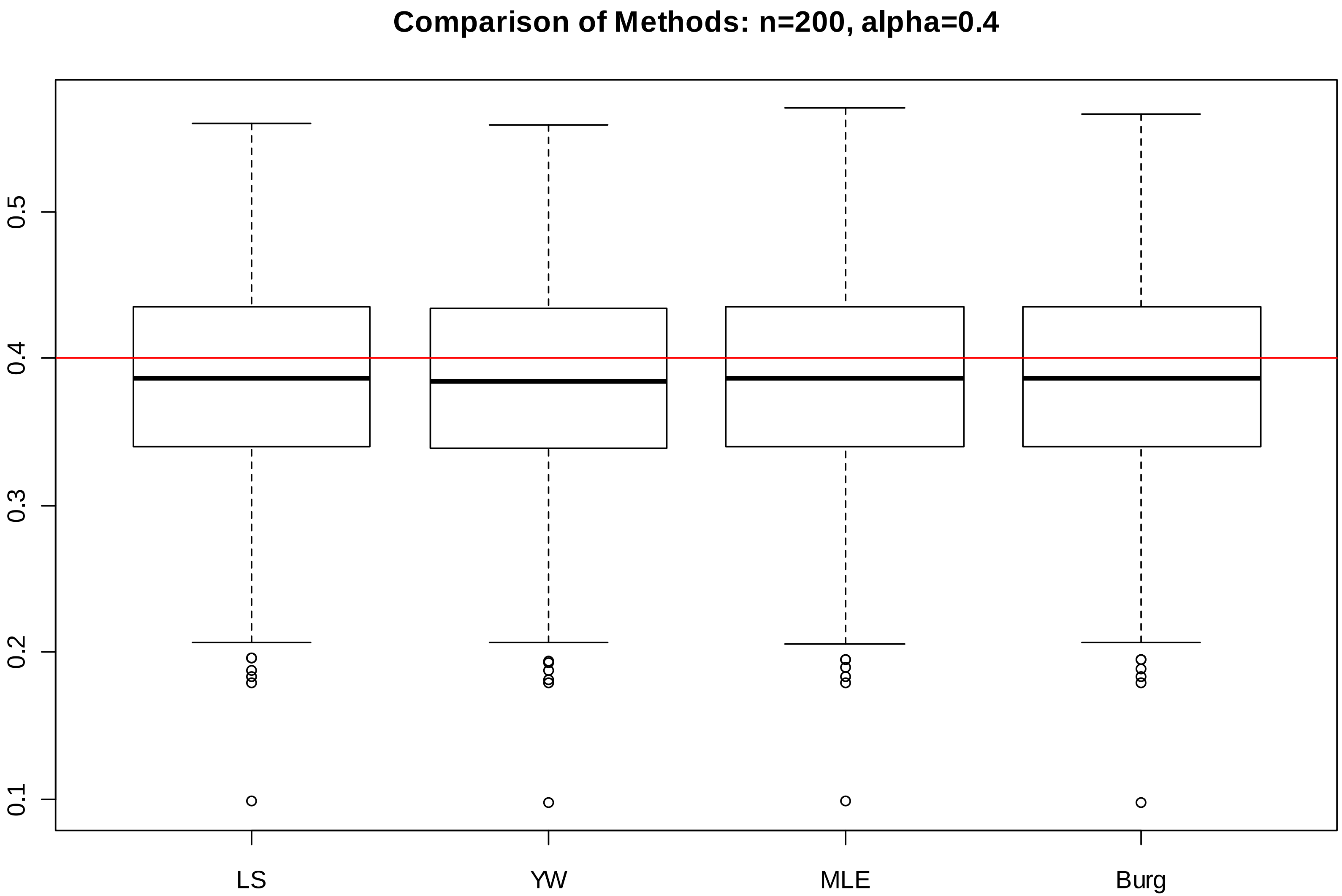
## MLE:

+ solves the problem „as a whole“, good theory behind
- heavy computation, convergence, distribution assumption

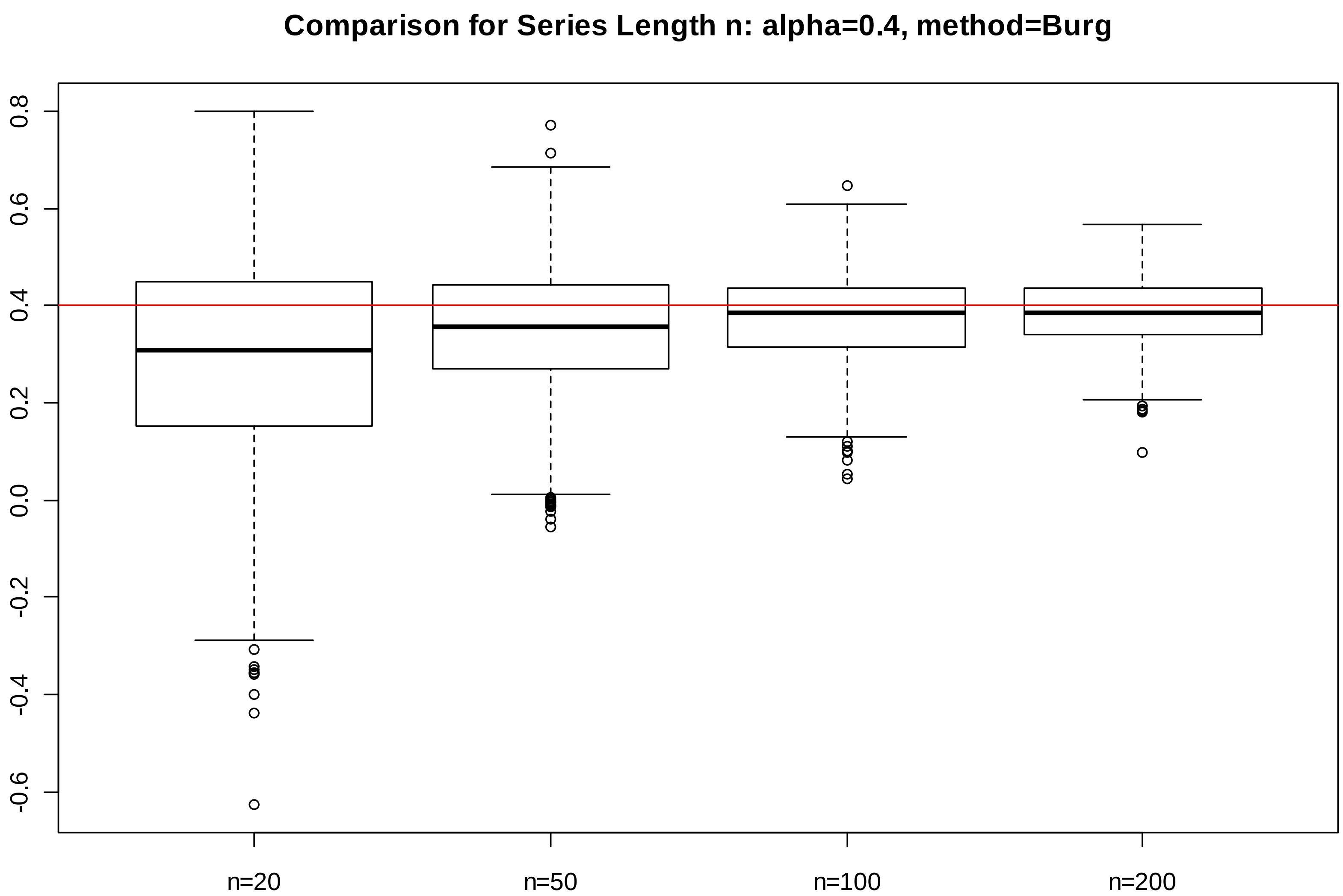
## Burg:

+ prediction oriented, no distribution assumption

# *Comparison: Alpha Estimation vs. Method*

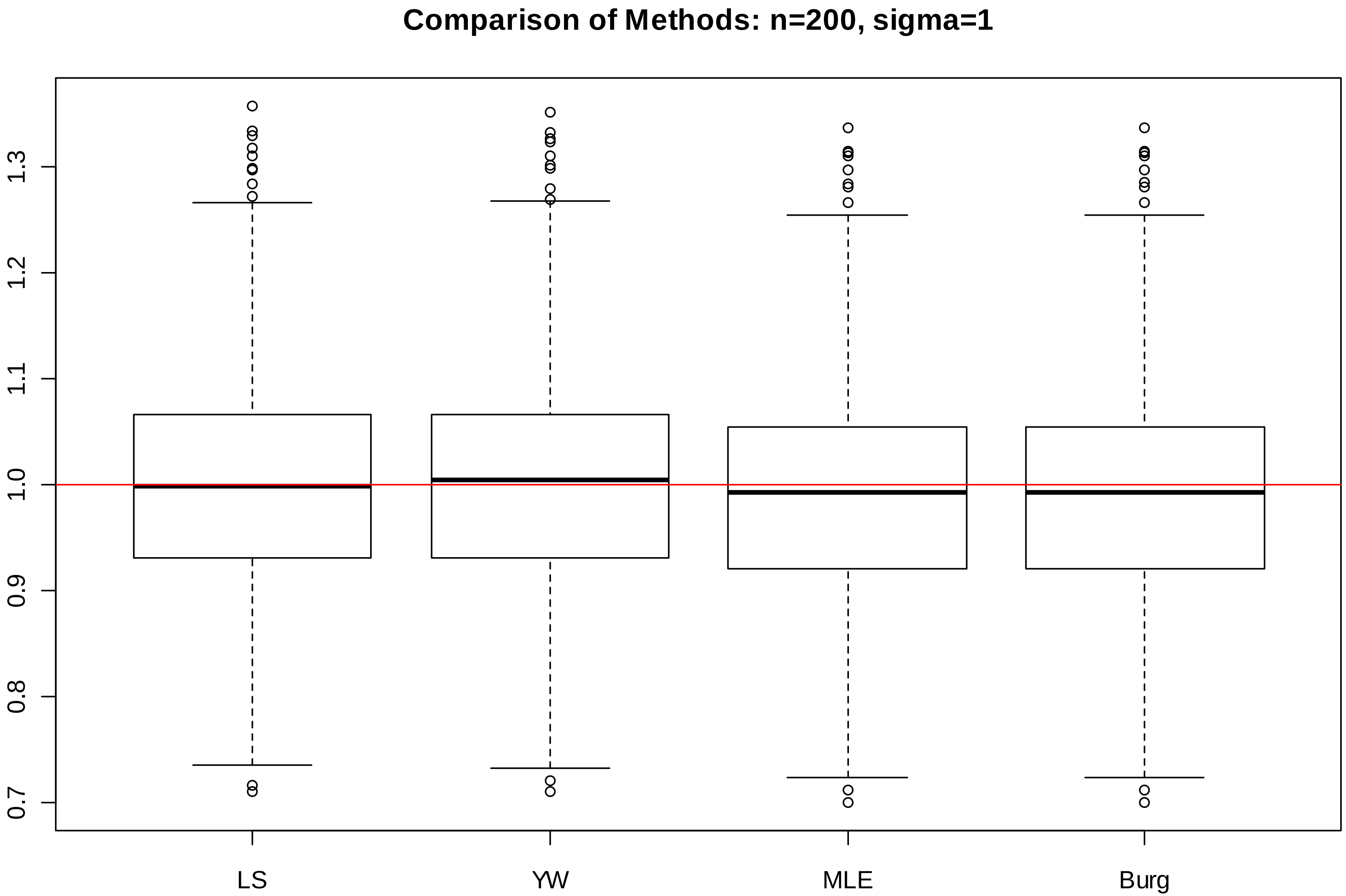


# *Comparison: Alpha Estimation vs. n*

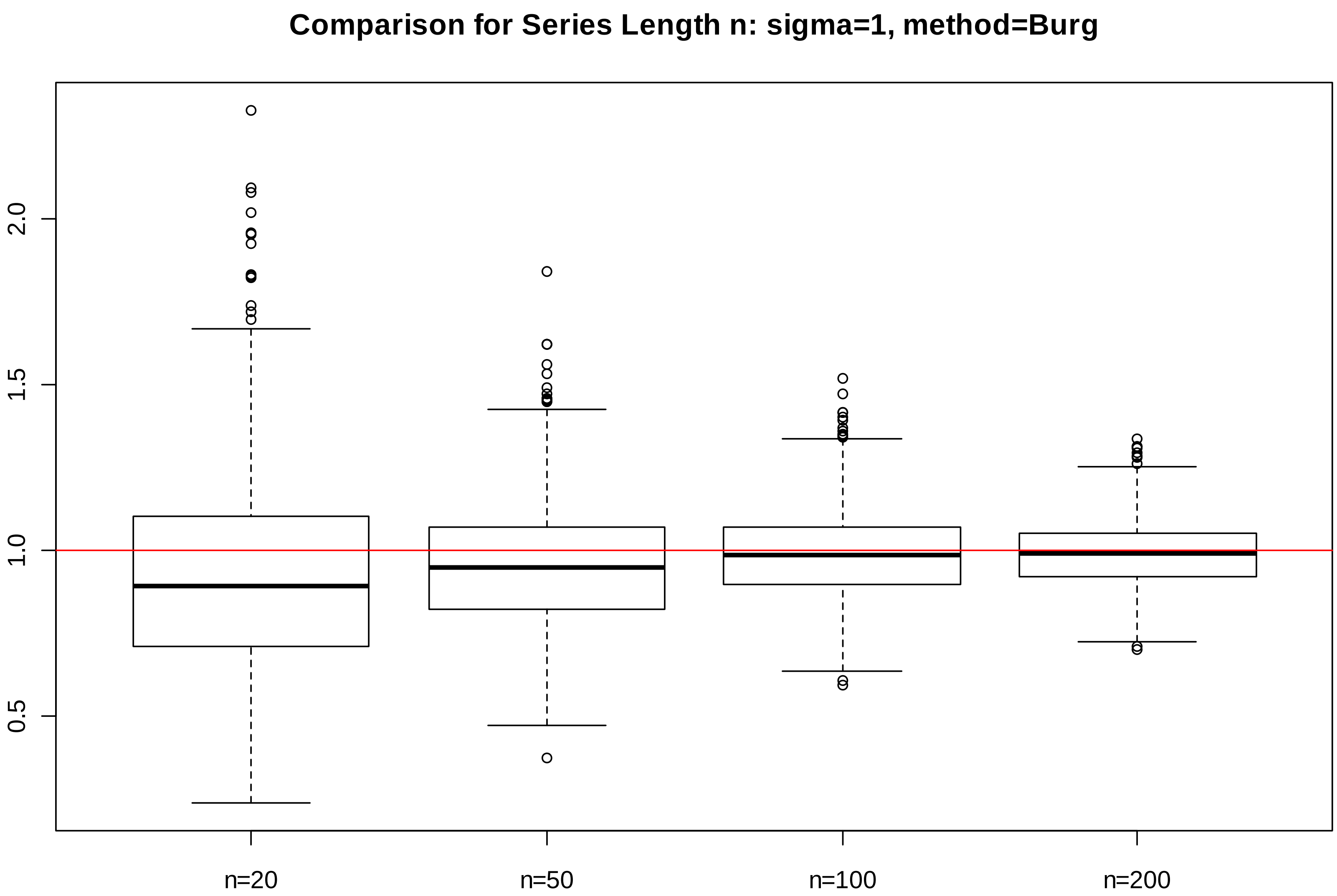


Applied Time Series Analysis
FS 2012 – Week 05

# *Comparison: Sigma Estimation vs. Method*



# *Comparison: Sigma Estimation vs. n*



# *Variance of the Arithmetic Mean*

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.

$$Var(\mu) = \frac{1}{n^2}\gamma(0)\left(n + 2 \cdot \sum_{k=1}^{n-1}(n-k)\cdot\gamma(k)\right)$$

## *Computation in Practice*

For adjusting the variance of the arithmetic mean do either:

1) **Estimate the theoretical ACF from the estimated AR-model**

```
> ARMAacf(ar = ar.coef, lag.max = r, pacf = FALSE)
```

and plug-in the result into the formula

2) **Work with function arima()**

```
> arima(sqrt(purses),order=c(2,0,0),include.mean=T)
         ar1     ar2  intercept
      0.2745  0.3947     3.5544
s.e.  0.1075  0.1089     0.2673
```

This directly gives the mean’s standard deviation.

## *Model Diagnostics*

**What we do here is Residual Analysis:**

„residuals“ = „estimated innovations“

$$= \hat{E}_t$$

$$= (x_t - \hat{\mu}) - \left(\hat{\alpha}_1 (x_{t-1} - \hat{\mu}) - \ldots - \hat{\alpha}_p (x_{t-p} - \hat{\mu})\right)$$

**Remember the assumptions we made:**

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

and probably

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

## *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 $\hat{E}_t$

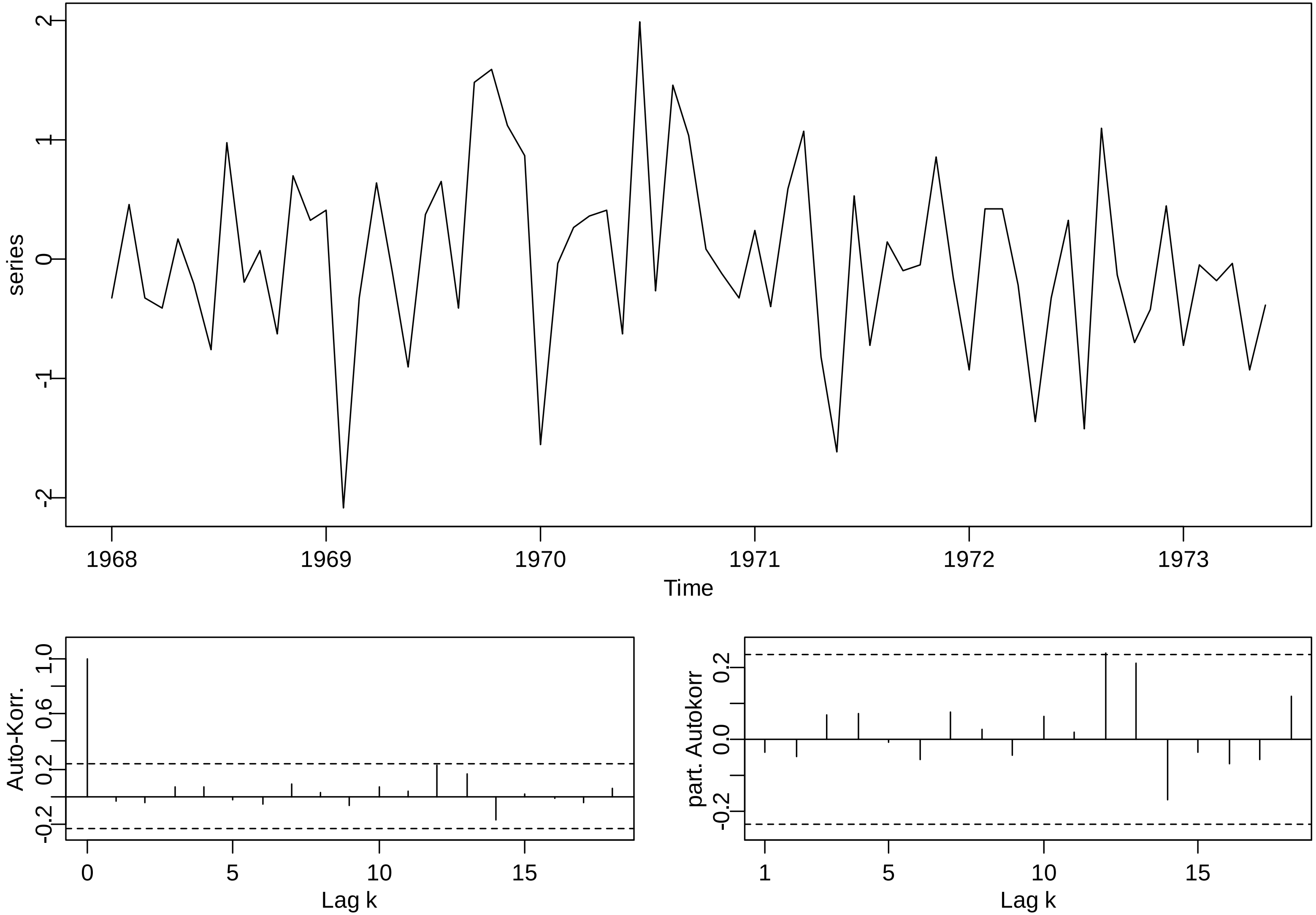
→ **The innovation time series $\hat{E}_t$ should look like white noise**

Purses example:

```
fit <- arima(sqrt(purses), order=c(2,0,0), include.mean=T)
acf(resid(fit)); pacf(resid(fit))
```

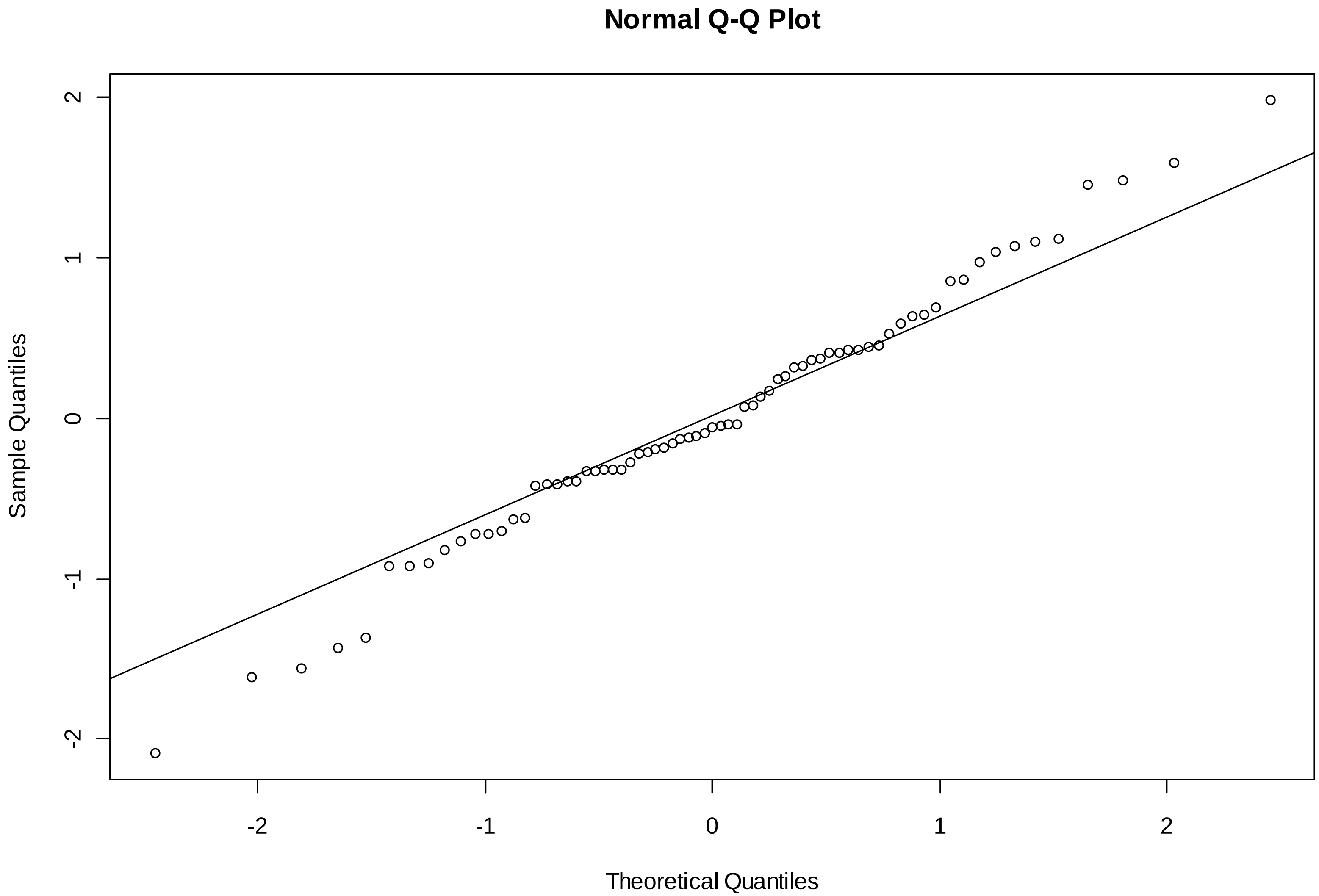
# Applied Time Series Analysis
## FS 2012 – Week 05

# *Model Diagnostics: sqrt(purses) data, AR(2)*



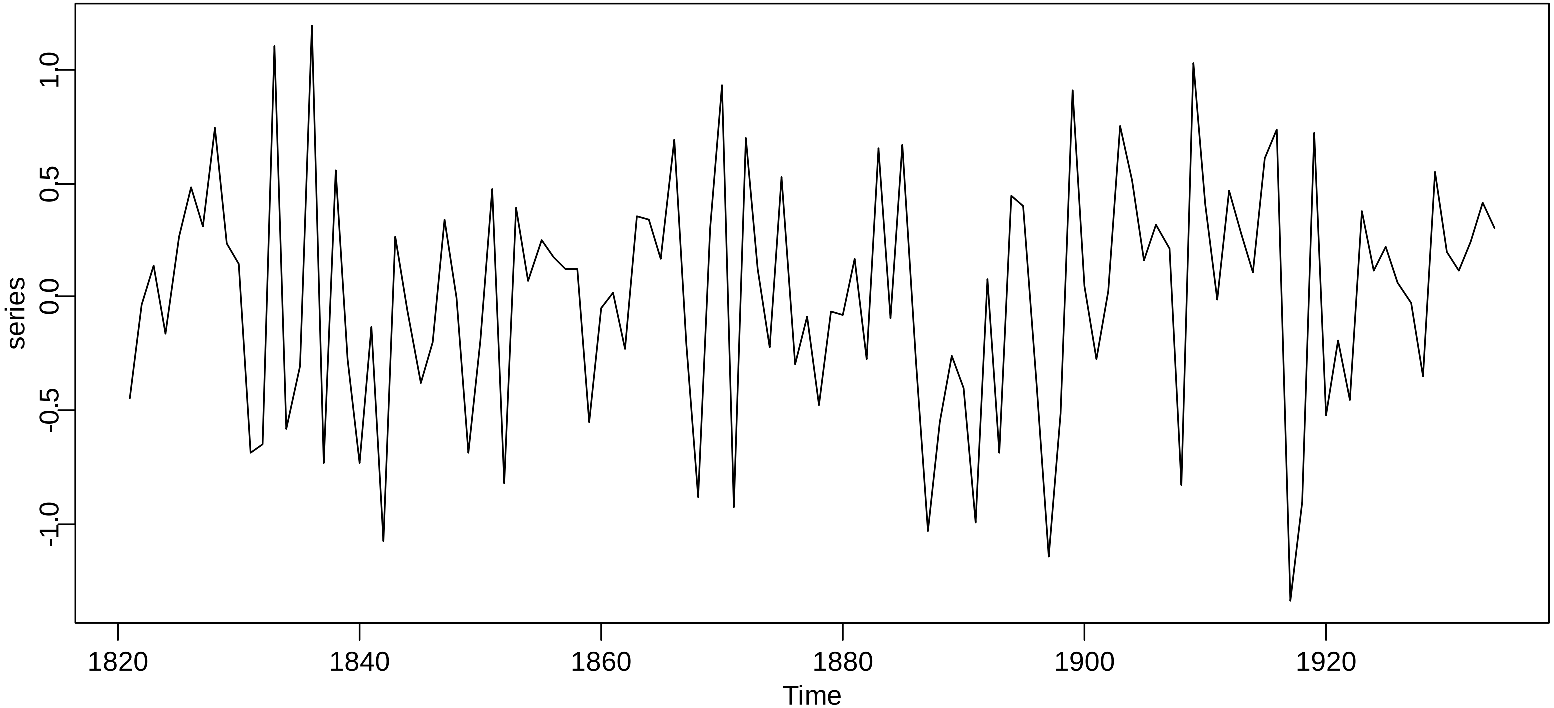
# Applied Time Series Analysis
## FS 2012 – Week 05

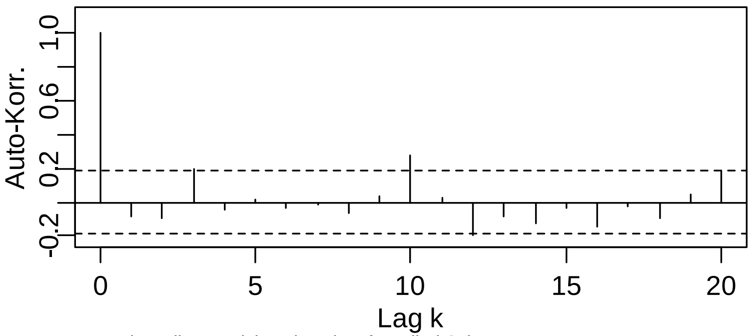
# *Model Diagnostics: sqrt(purses) data, AR(2)*

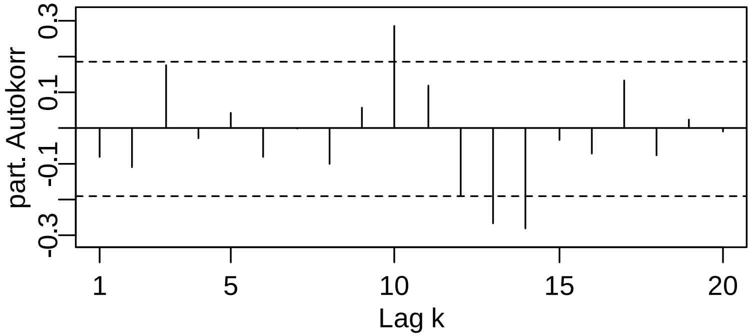


# Applied Time Series Analysis
## FS 2012 – Week 05

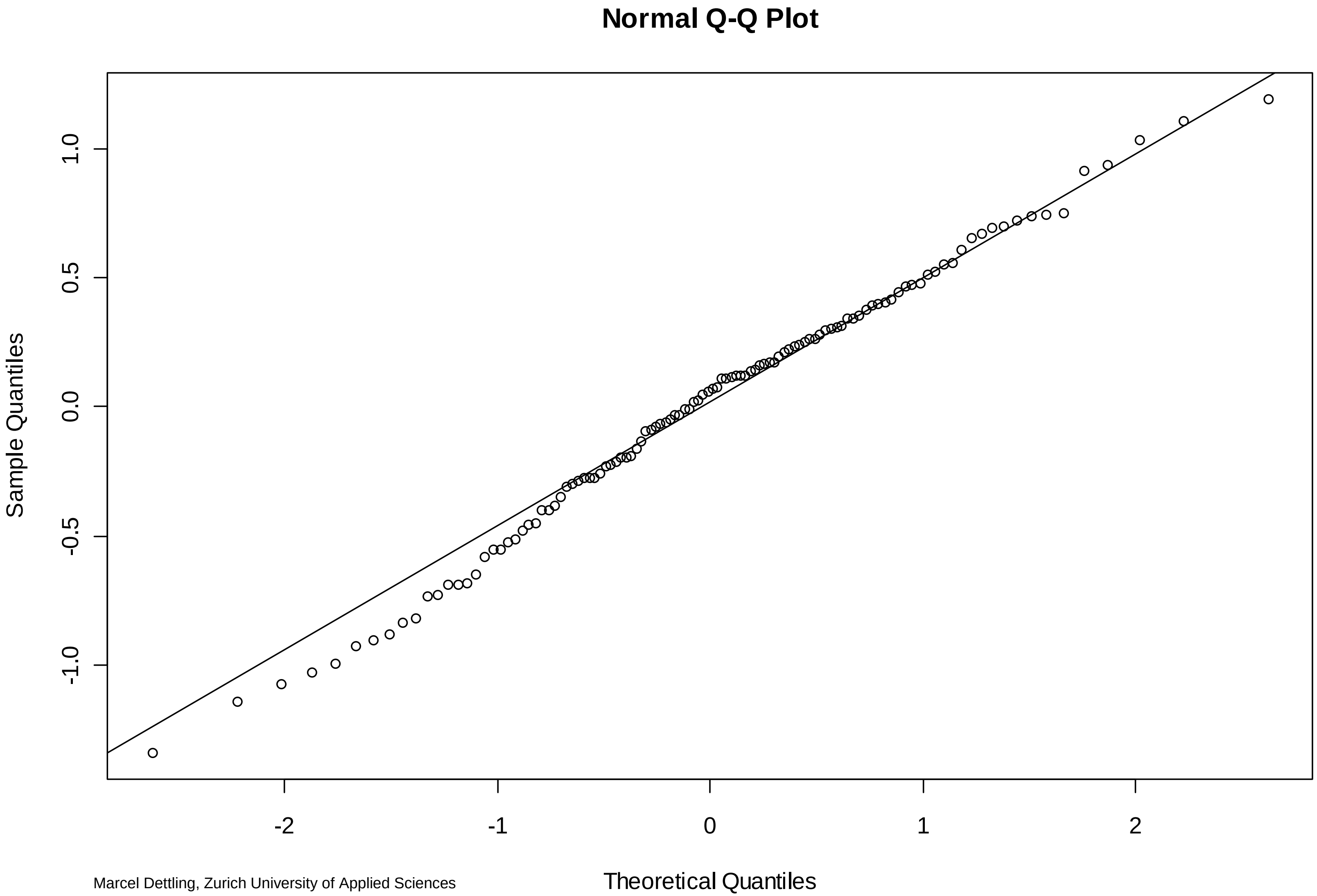
# *Model Diagnostics: log(lynx) data, AR(2)*



# *Model Diagnostics: log(lynx) data, AR(2)*



# *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:

AIC = $-2\log(L) + 2p$
BIC = $-2\log(L) + 2\log(n)p$

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**

# *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:

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

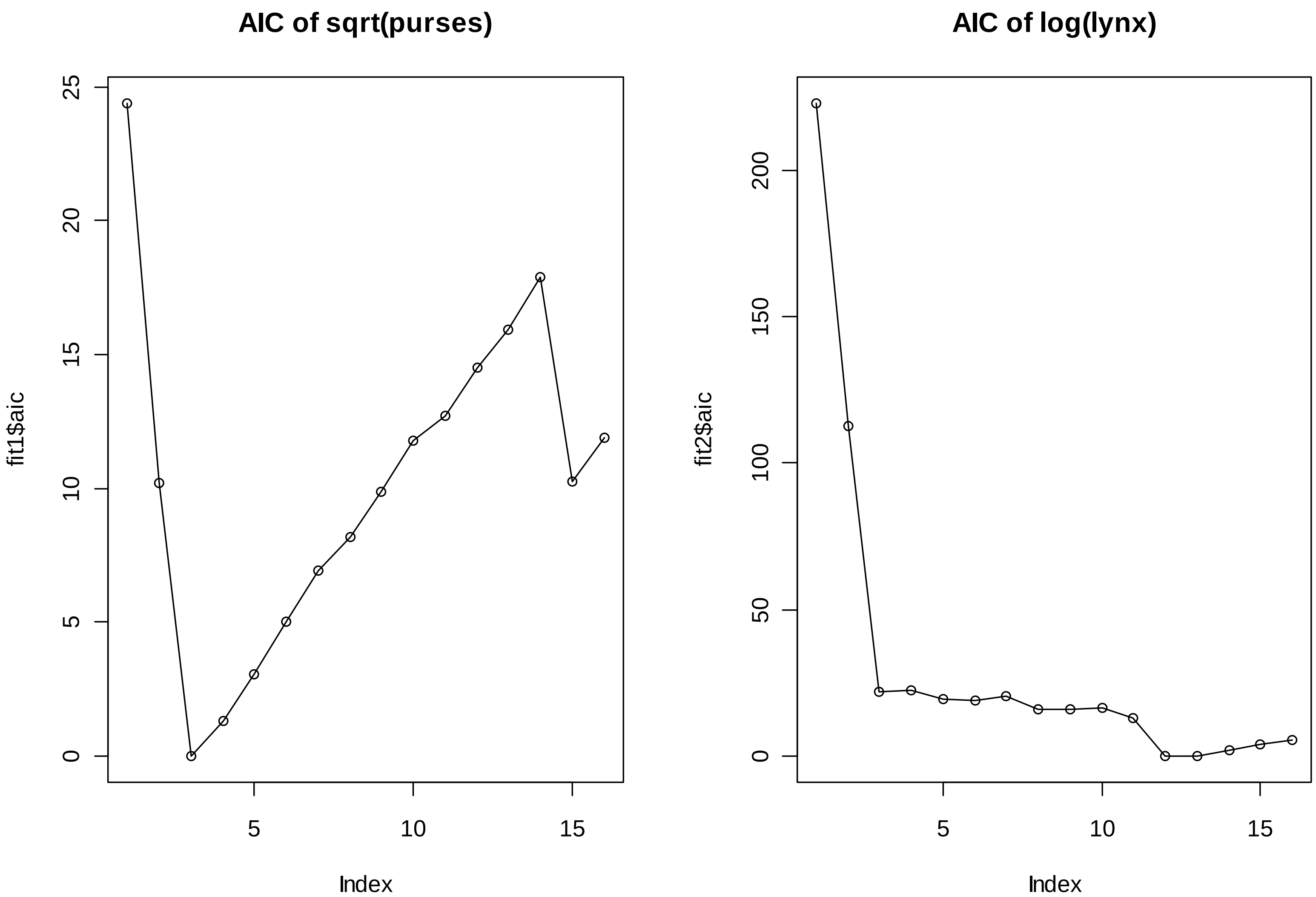
**Remarks**:

→ AIC tends to over-, BIC to underestimate the true p

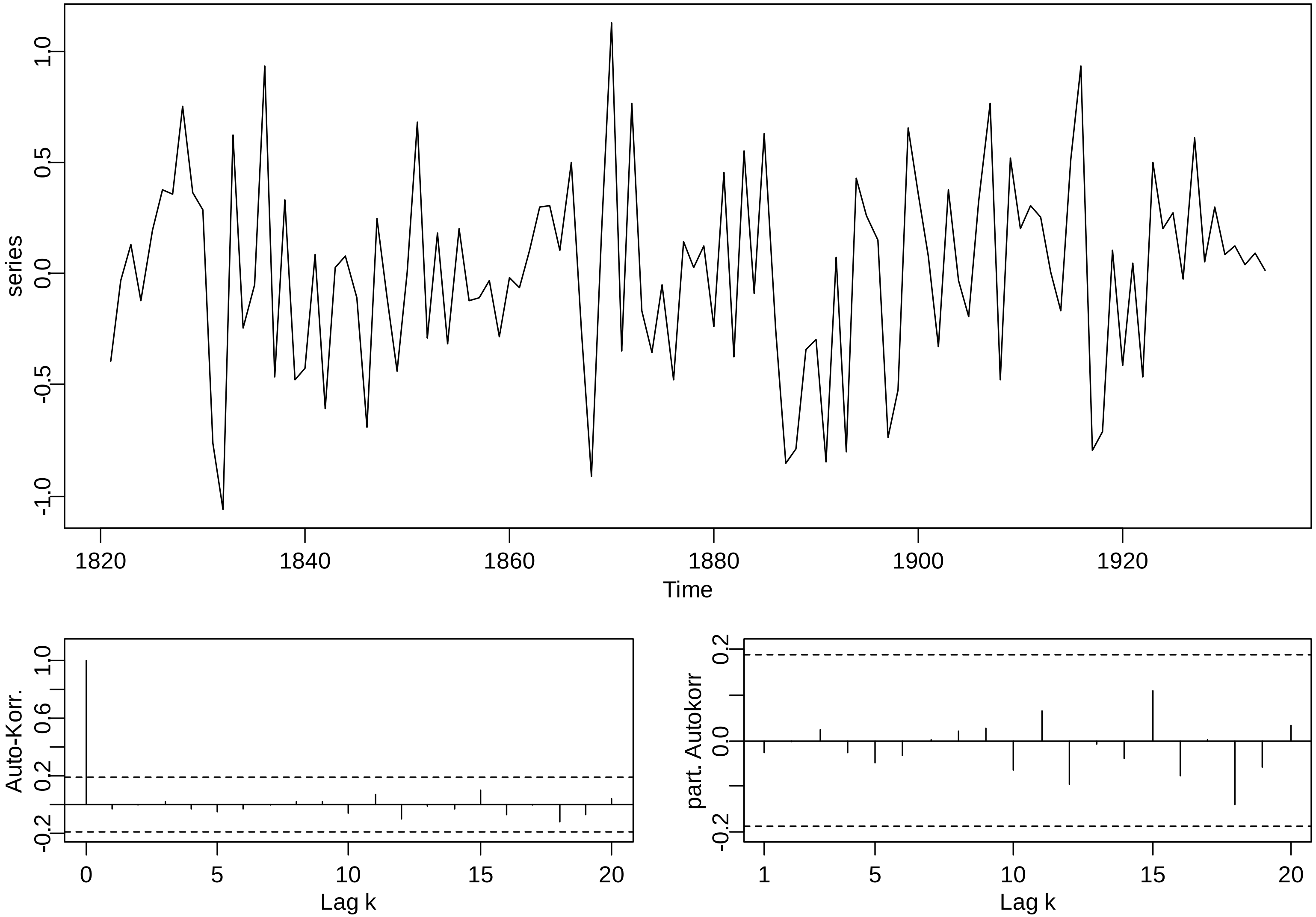
→ Plotting AIC/BIC values against p can give further insight. One then usually chooses the model where the last significant decrease of AIC/BIC was observed

# Applied Time Series Analysis
## FS 2012 – Week 05

# *AIC/BIC*



# *Model Diagnostics: log(lynx) data, AR(11)*



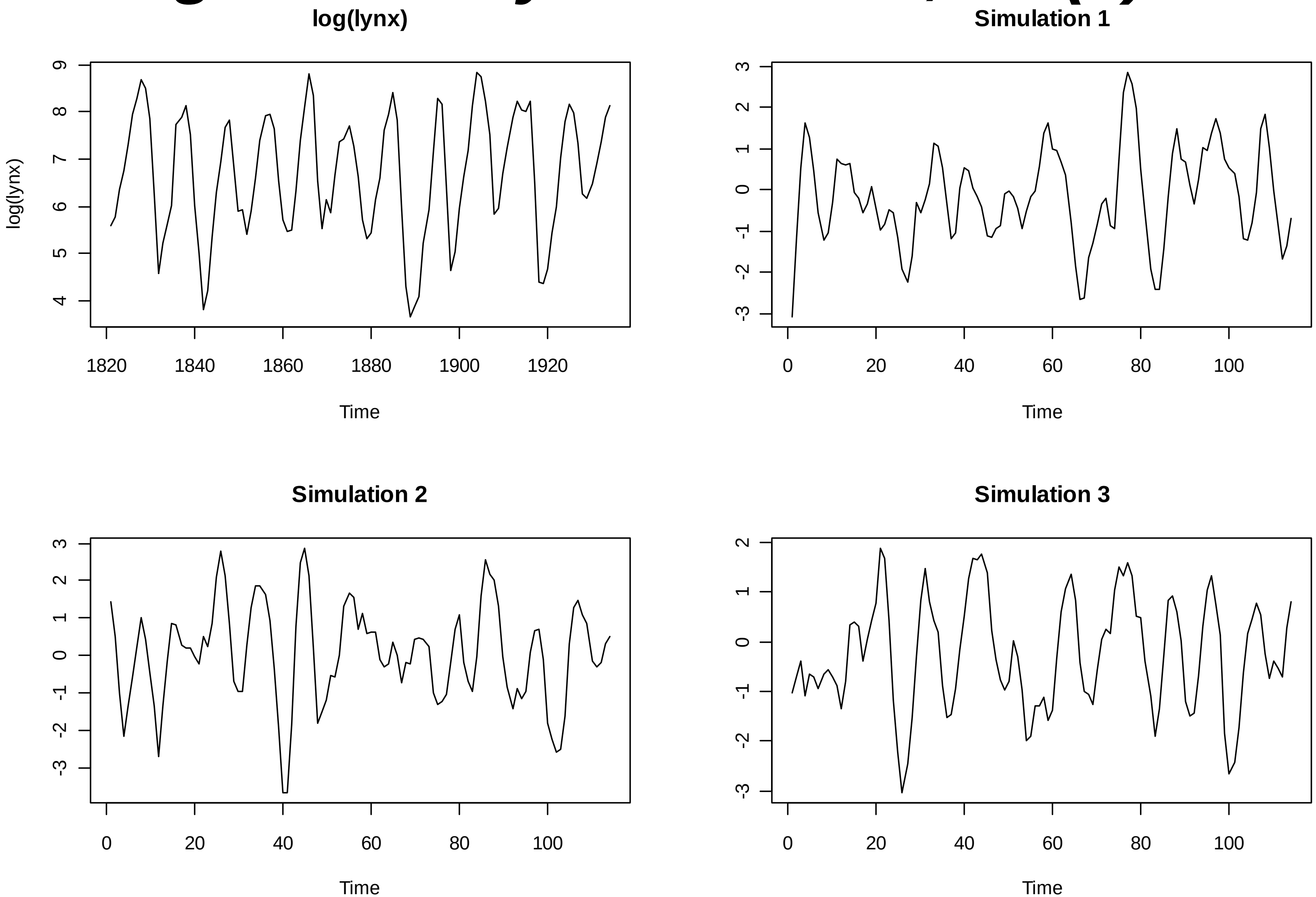
# *Diagnostics by Simulation*

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
## FS 2012 – Week 05

# *Diagnostics by Simulation, AR(2)*



# Applied Time Series Analysis
## FS 2012 – Week 05

# *Diagnostics by Simulation, AR(11)*