Master of Science (MSc)

Applied Information and Data Science

Institute of Natural Sciences and
Humanities ING
Prof. Dr. Philipp Schütz
Lecturer

T direct +41 76 510 21 60
philipp.schuetz@hslu.ch

Horw 26.03.2020

RTP02 Discrete Response, Time Series and Panel Data

Where are we? What is up next?

RECAP AND PREVIEW

What did we do last time?

- **Diagnostic tools:**
  - Features of the autocorrelation function:

Property	ACF characteristic
Stationary	Fast (often exponential) decay
Trend	Slow decay
Seasonality	Oscillatory behaviour
Seasonality and trend	Oscillatory pattern and slow decay
Outliers	Small perturbations on whole ACF values

  - **Partial Autocorrelation function:** *Theoretical approach:*

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

    Practical realisation:
    Built-in function (pacf)

What did we do last time?

- **Models for stationary time series** (e.g. remainder)
  - White noise (iid distributed random variables)
  - AR(p) models
  - MA(q) models
  - ARMA(p,q) models
  - etc.

  Big question: How to find the right model?

  First partial answer: For white noise, all time steps are independent.

Recap: AR(p) models

- *AR(p) model* is a model of the form:

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + e_t$$

  i.e. the current value depends on the previous p values plus some innovation e_t

- Reminder: Back-shift operator B is defined as $\mathrm{B}\, X_t = X_{t-1}$

- To the model:

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_p X_{t-p} = e_t$$

  we define the *characteristic polynomial*:

$$\Phi(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$$

  such that

$$\Phi(B)\, X_t = e_t$$

- *AR(p) model* is stationary if all roots of the characteristic polynomial

$$\Phi(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$$

  has an absolute value larger than 1.

AIMS FOR TODAY

Guiding questions for today

Key question:

When to use which model?

Leading questions:

- What are the key properties of AR(p), MA(q) and ARMA(p,q) models?
- How to simulate these processes?
- How to fit the parameters of the models?

STATIONARITY OF AR(P) MODELS

When are AR(p) model stationary?

AR(p) models are supposed to fit stationary time series

⇒ When are these models stationary?

Theoretical result: AR(p) models are stationary when

- Mean is 0, i.e $E[X_t] = 0$

- The absolute value of the roots of the characteristic polynomial

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

  are all larger than 1.

The first condition may also be removed by defining a shifted AR(p) process:

$$X_t = m + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + e\_t$$

How to check this practically? (cf. exercise 2.7 and 3.1)

Let us check the AR(2) model:

$$X_t = 0.8\, X_{t-1} + 0.4\, X_{t-2} + e_t$$

It may be rewritten as:

$$X_t - 0.8\, X_{t-1} - 0.4\, X_{t-2} = e\_t$$

Therefore, the characteristic polynomial is:

$$\Phi(z) = 1 - 0.8\, z - 0.4\, z^2$$

The theoretical roots ($\Phi(z) = 0$) are:

$$z_{1,2} = \frac{0.8 \pm \sqrt{0.8^2 - 4 \cdot 1 \cdot (-0.4)}}{2 \cdot (-0.4)} = -1 \pm \frac{1}{0.8}\, \sqrt{2.24}$$

The roots are:

```
polyroot(c(1,-0.8,-0.4))
0.8708287+0i    -2.8708287+0i
```

As the first value has an absolute value below 1, the time series is not stationary.

When could an AR(p) model be suitable?

KEY PROPERTIES OF AR(P) MODELS

Example

The AR(1) process:

$$X_t = 0.75 \cdot X_{t-1} + e_t$$

Is stationary, because the characteristic polynomial:

$$\Phi(z) = 1 - 0.75\, z$$

Has the root $(\Phi(z_1) = 0)$ $z_1 = \frac{1}{0.75} \approx \frac{4}{3} > 1$ i.e. it is stationary.

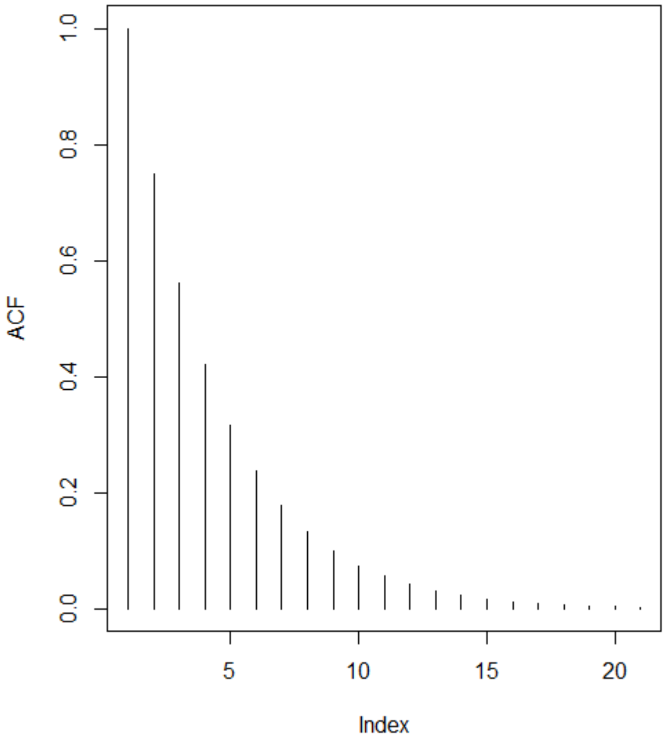
What are the peculiarities of its theoretical (partial) autocorrelation function?

The theoretical (partial) autocorrelation function may be calculated in R by:

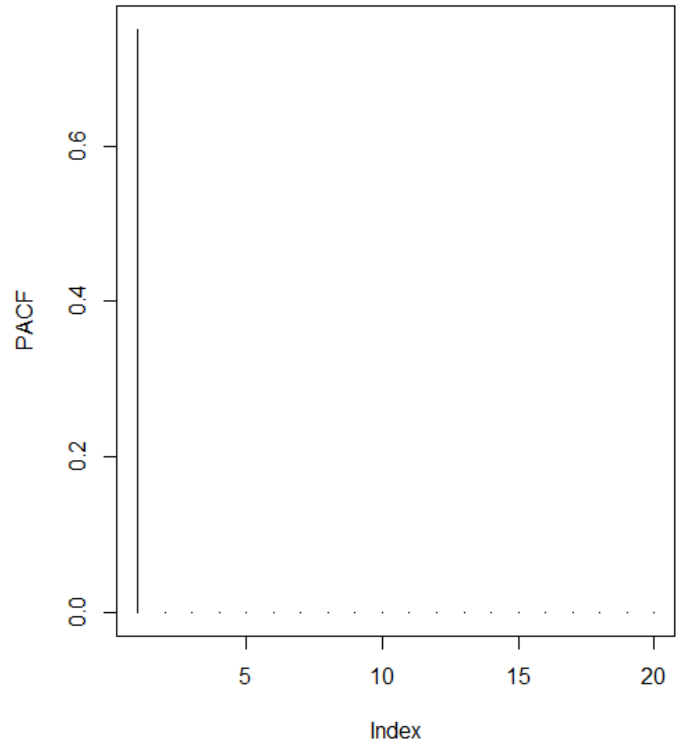
```
ARMAacf(ar=(alpha_1,alpha_2,…))

ARMAacf(ar=(alpha_1,alpha_2,…),pacf=TRUE)
```

Example

```
ta<-ARMAacf(ar=(0.75),lag.max=22)
plot(ta,type=«h»)
```




```
tp<-ARMAacf(ar=(0.75),pacf=TRUE)
plot(tp,type=«h»)
```



Observation: ACF decays quickly, PACF has a clear cutoff (at 1 for AR(1) model)

Key properties of AR(p) processes

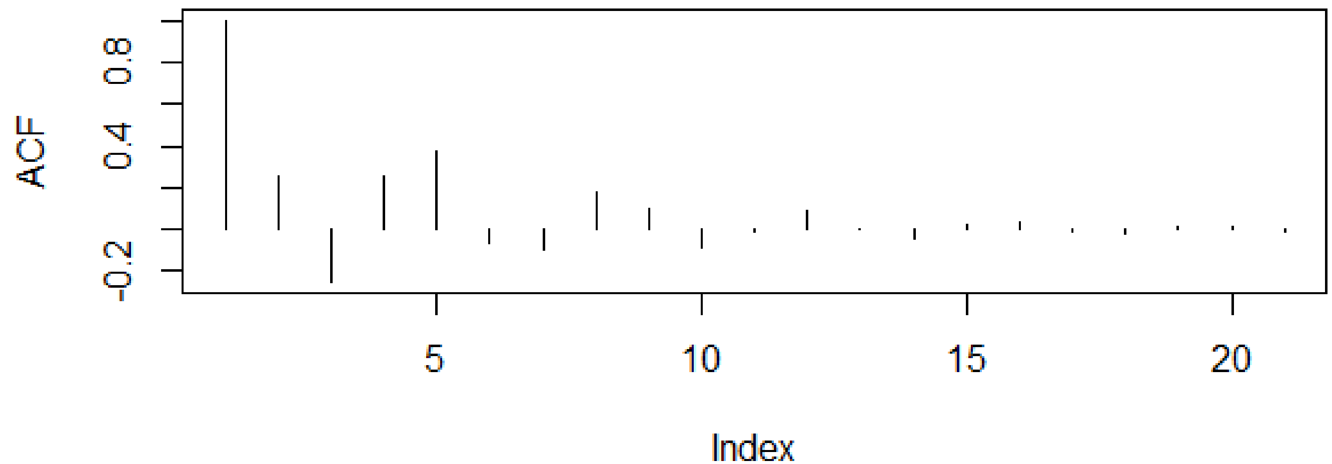
For the model identification of AR(p) processes two properties are key:

a) Autocorrelation decays quickly or displays a damped sinusoid.

b) Partial autocorrelation function has a clear cut-off at p.

Example for a): Consider the AR(3) process

$$X_t = 0.5\,X_{t-1} - 0.5\,X_{t-2} + 0.5\,X_{t-3} + e_t$$



Quick check: Why is it stationary?

Which AR-model to choose?

MODEL ORDER IDENTIFICATION

How to fit an AR(p) model?

- Check whether (partial) autocorrelation function fulfils the required specification?

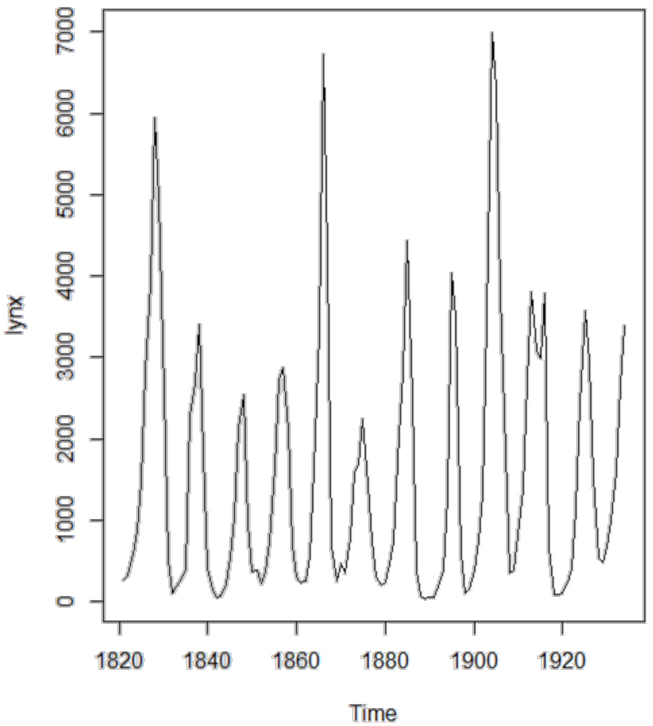
  WARNING: Fulfilling the above stated properties does not imply that the AR(p) model is the correct model.

- Let us try this with an example. The lynx data set in R describes the number of lynx trappings in Canada from 1821 to 1934.

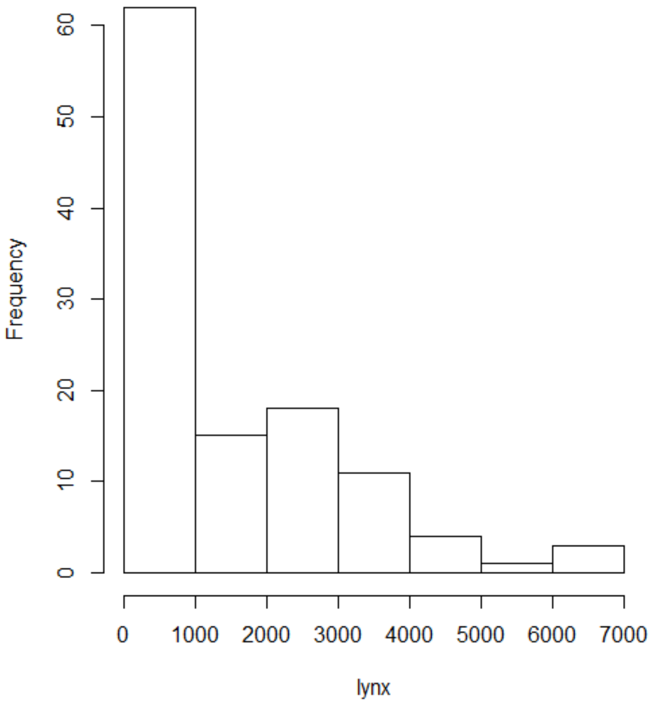
Source: www.wikipedia.org

Walk-through lynx example

Plot of lynx data set



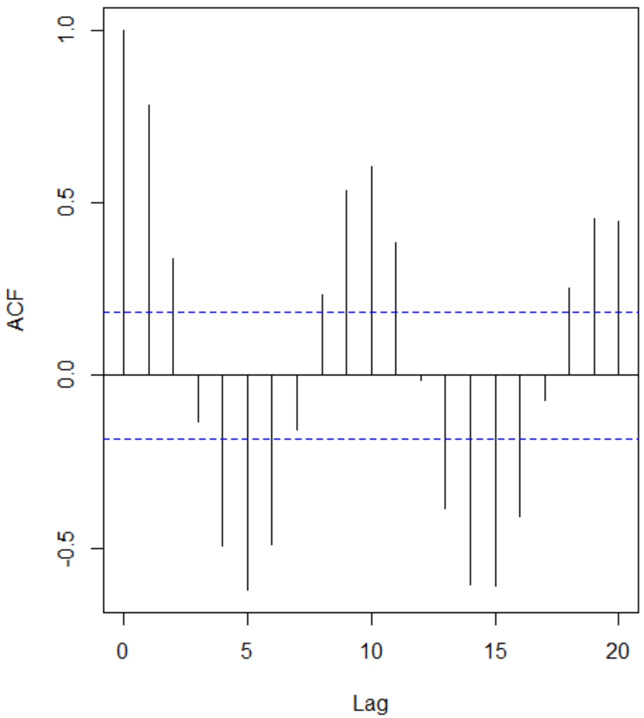
Histogram of lynx dataset



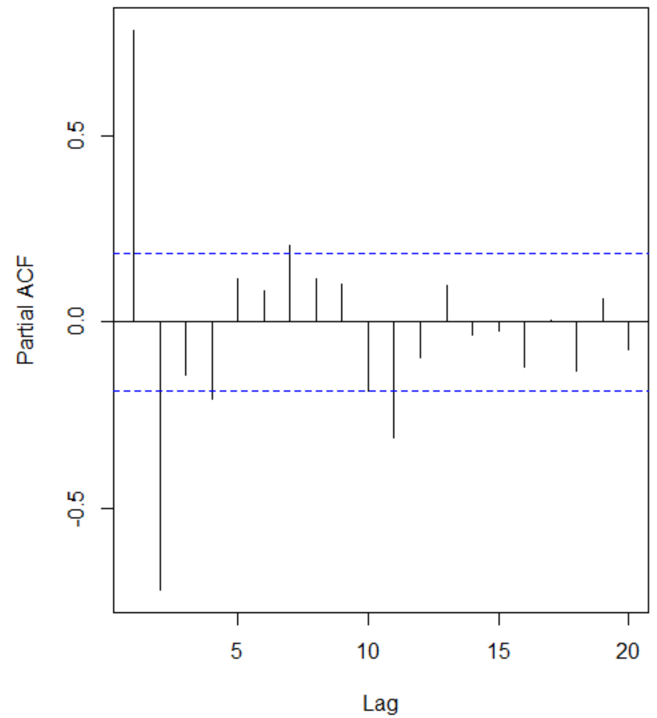
The data is considerably right-skewed. Therefore, we consider the log-transformed data.

Walk-through lynx example

Plug-in estimated ACF of log(lynx)



Partial ACF



ACF displays damped sinusoid

PACF has clear cutoff after lag 11, 7, 4 or 2 (depending on threshold).

⇒ Try fitting an AR(2), AR(4), AR(7), AR(11)

How to find the AR(p) coefficients?

MODEL PARAMETER IDENTIFICATION

Motivation Parameter identification

Often the time series are not centred, i.e. instead of a classical AR(p) process a centred version:

$$Y_t = m + X_t = m + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + e_t$$

has to be fitted.

Fitting parameters are
- the global mean m
- the AR-coefficients $\alpha_1, \ldots, \alpha_p$
- the parameters defining the distribution of the innovation e_t. Here we assume a normal distribution, i.e. σ_e^2 has to be fitted.

In the next slides we will discuss, four different approaches:
a) Ordinary least square (OLS)
b) Yule-Walker equations
c) Burg's algorithm
d) Maximum Likelihood Estimator (MLE)

How to find the AR(p) coefficients?

PARAMETERS FROM ORDINARY LEAST SQUARE

Parameter identification with ordinary least square (OLS)

We want to estimate the parameters:

$$m, \alpha_1, \dots, \alpha_p, \sigma_E^2$$

from the equation:

$$Y_t = m + X_t = m + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + e_t$$

We have a total of $n - p$ points for this because we have a total of n time slices and each equation links p time instances.

The OLS procedure is:

1. Estimate the global mean $\hat{m} = \bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$ and determine $x_t = y_t - \hat{m}$
2. Perform a regression on x_t without intercept. The resulting parameters are $\hat{\alpha}_1, \dots, \hat{\alpha}_p$.
3. Estimate the standard deviation of the innovation $\hat{\sigma}_E^2$ from the standard deviation of the residual.

Practical implementation: `ar.ols(data,order=p)`

How to prevent overfitting?

YULE WALKER EQUATIONS

Motivation Yule-Walker Equations

Basic idea: The model coefficients and the ACF values are entangled.

Example: AR(2) process: $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + e_t$

For the covariance in $\rho(2) = \frac{Cov(X_t, X_{t-2})}{\sqrt{Var(X_t) \cdot Var(X_{t-2})}}$ we obtain:

$$Cov(X_t, X_{t-2}) = \alpha_1 Cov(X_{t-1}, X_{t-2}) + \alpha_2 Cov(X_{t-2}, X_{t-2}) + Cov(e_2, X_{t-2})$$

The first term is the covariance in $\rho(1)$, the second term is the variance of X_{t-2} and the third term is zero (as e_t is an innovation).

In consequence, we have:

$$\rho(2) = \alpha_1 \rho(1) + \alpha_2 \rho(0)$$

i.e. the values of the auto-correlation function depend on the model coefficients.

Yule-Walker equations and YW fitting

A careful calculation yields for an AR(p) model the following equations, called *Yule-Walker* equations:

$$\sum_{i=1}^{p} \alpha_i \, \rho(k-i) = \rho(k) \quad for \; k > 0$$

$$\sum_{i=1}^{p} \alpha_i \; \rho(i) = \rho(0) - \sigma_E^2 \;\; for \;\; k = 0$$

(Note that the autocorrelation function is symmetric, i.e. $\rho(-k) = \rho(k)$ for all k)

To fit the AR(p) model parameters, the above equations are solved with the estimated values of the auto correlation function.

Practical implementation: `ar.yw(data,order=p)`

Further options for fitting

- **Burg’s algorithm:**
  Offers another fitting cost function that exploit also the first p function values.

  Practical implementation: `ar.burg(data,order.max=p)`

- **Maximum Likelihood Estimator (MLE):**
  Determines the parameter based on an optimisation of the maximum likelihood function.

  Practical implementation: `arima(Data,order=c(2,0,0))`

Please try it out yourself on the lynx data set

```
f.ols<-ar.ols(log(lynx),order=2,intercept=F)

f.yw<-ar.yw(log(lynx),order=2)

f.burg<-ar.burg(log(lynx),order=2)

f.mle<-arima(log(lynx),order=c(2,0,0))
```

What is the difference between the different fitted parameters?

Method	$\hat{m}$	α_1	α_2
OLS			
Yule-Walker			
Burg			
MLE			

Hints and results

Hints:
- Printing the variables displays the model parameters
- Mean can be visualised by variable$x.mean

Method	$\hat{m}$	α_1	α_2
OLS	6.6859	1.3844	-0.7479
Yule-Walker	6.6859	1.3504	-0.7200
Burg	6.6859	1.3831	-0.7461
MLE	6.6863	1.3776	-0.7399

Observation:
- Values are very similar.
- MLE also provides error estimates on parameters

Which method to use when?

- All methods are equivalent for theoretical models and yield similar results on realisations.

- Beware that the methods are susceptible to outliers and perform best on data following a normal distribution.

Final question: How to choose the optimum model order p?

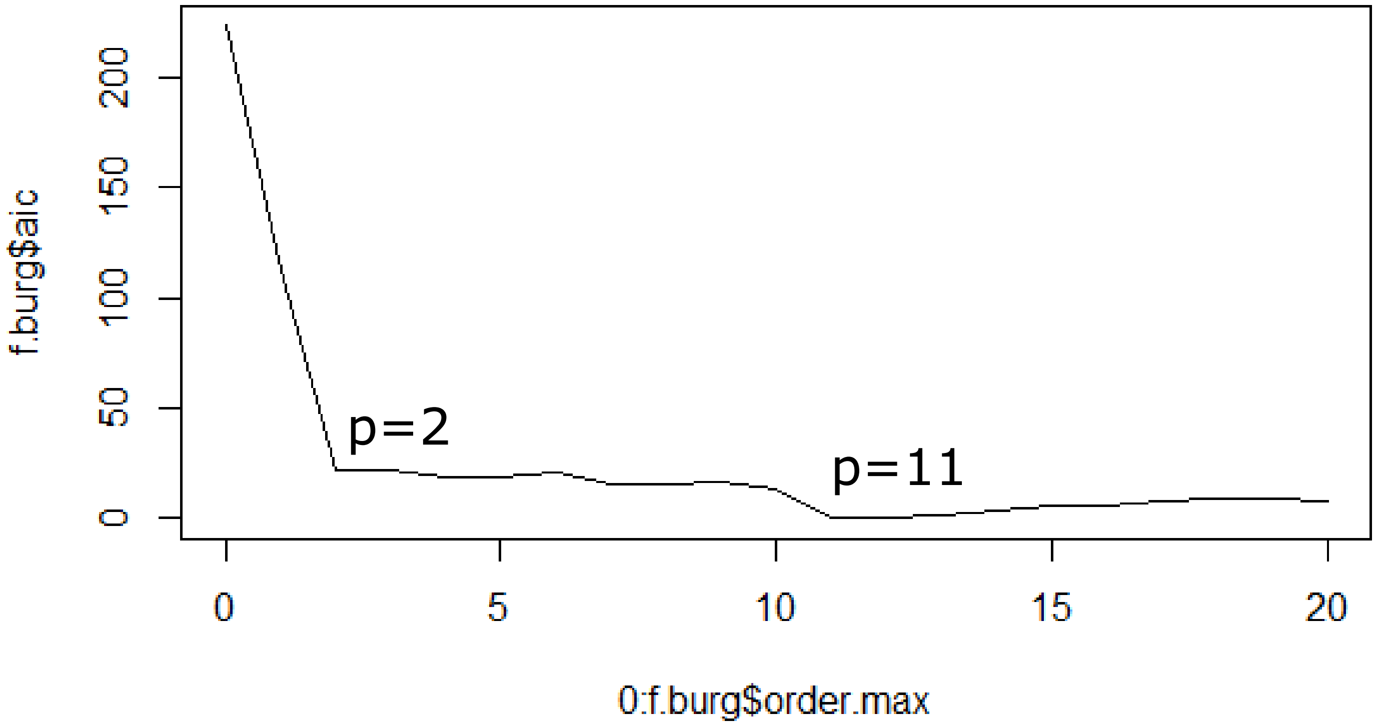
One option is to use the *Akaike information criterion (AIC).* AIC estimates the relative amount of information lost by the considered model. The less information a model loses, the higher is the quality of the model.

Practical example for order selection

```
f.burg<-ar.burg(log(lynx))
```

Note: no order parameter is given any more.

```
plot(0:f.burg$order.max,f.burg$aic)
```



Is it the right model?

MODEL DIAGNOSTICS

How to check whether the model fit was appropriate?

Motivation:
Your aim was to fit the model

$$Y_t = m + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + e_t$$

to the data. Therefore, the new time series

$$E_t = Y_t - \hat{m} - \hat{\alpha}_1 X_{t-1} - \cdots - \hat{\alpha}_p X_{t-p}$$

should behave like the innovation, i.e.

$$E[E_t] = 0, Var(E_t) = \sigma_e^2$$

How to test this?
- Visually by plotting
- ACF/PACF of the new time series
- QQ-Plot of the new time series
- Simulating the time series models and qualitative comparison

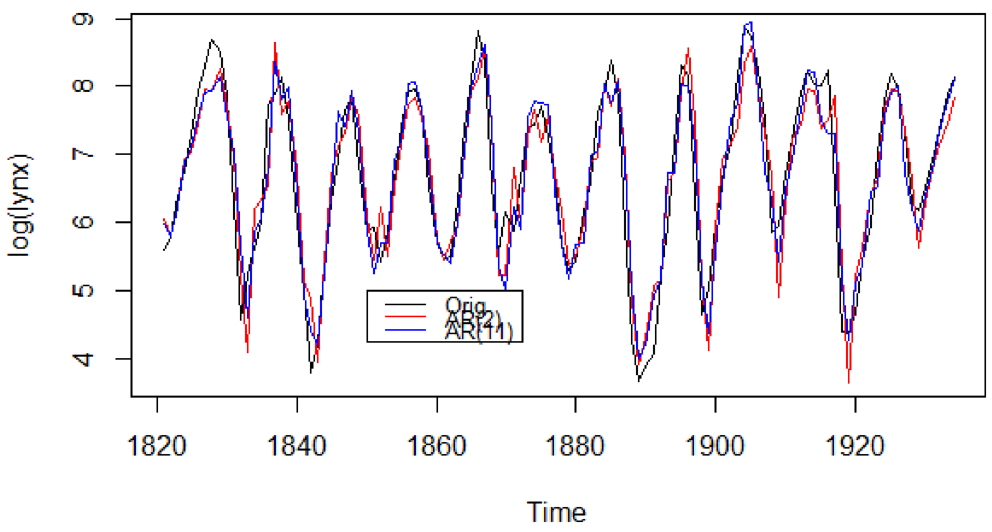
Walk-through example for lynx data

Fitting the two models:

```
fit.2<-arima(log(lynx),order=c(2,0,0))
fit.11<-arima(log(lynx),order=c(11,0,0))
```

Display the time series and model predictions:

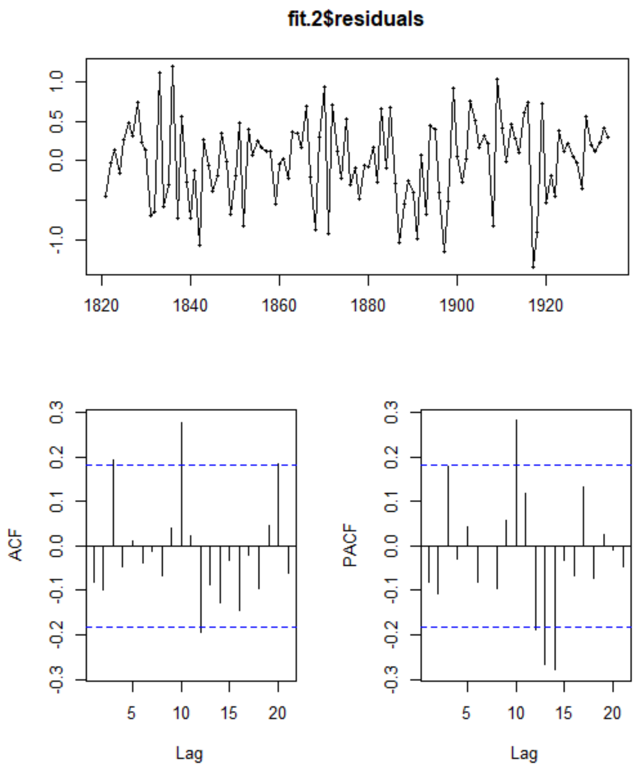
```
plot(log(lynx), main="Logged Lynx Data with ...")
lines(log(lynx)-fit.2$resid, col="red")
lines(log(lynx)-fit.11$resid, col="blue")
```

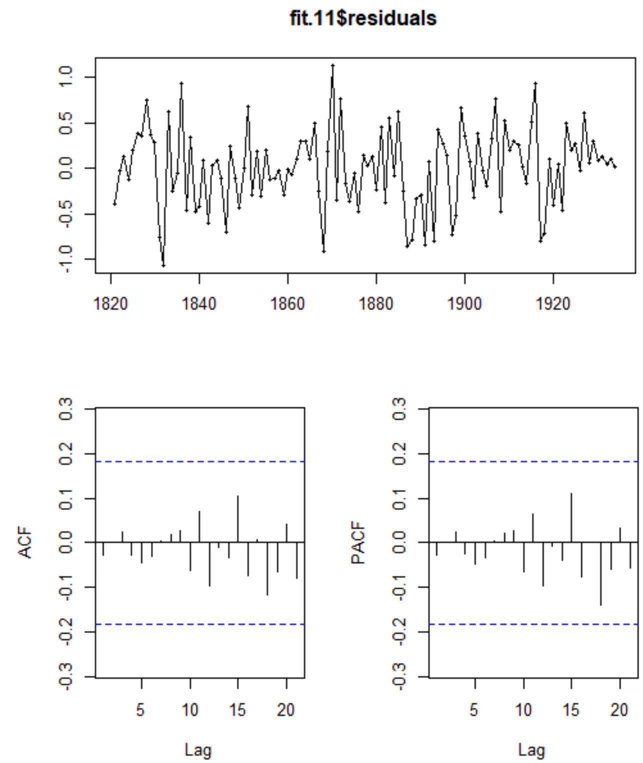


Walk-through example for lynx data

(P)ACFs of residuals:

```
tsdisplay(fit.2$residuals)
tsdisplay(fit.11$residuals)
```

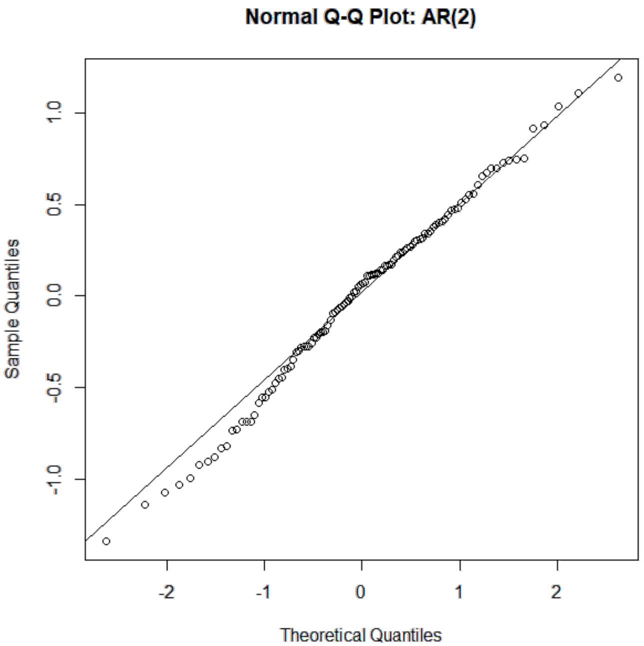


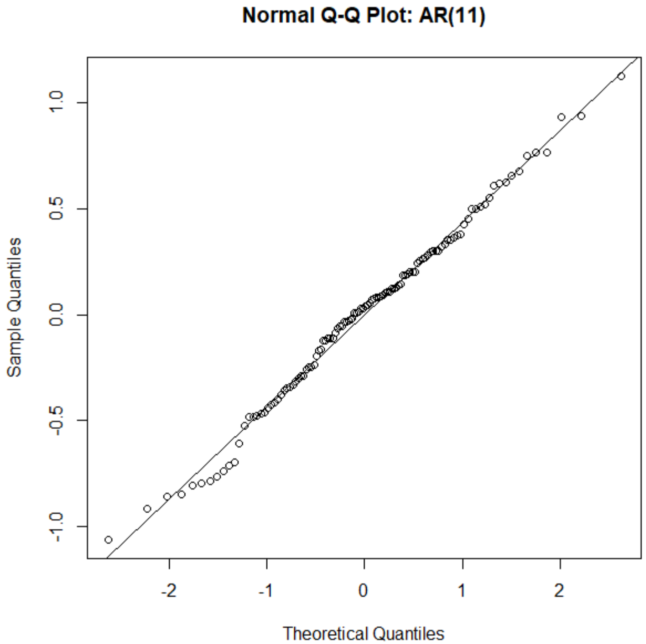


Walk-through example for lynx data

QQ-plots:

```
qqnorm(fit.2$residuals,main="Normal Q-Q Plot: AR(2)")
qqline(fit.2$residuals)

qqnorm(fit.11$residuals,main="Normal Q-Q Plot: AR(11)")
qqline(fit.11$residuals)
```





Final check of model by comparing data with simulation

```
par(mfrow=c(1,2))
plot(arima.sim(n=114, list(ar=fit.2$coef[1:2])))
plot(arima.sim(n=114, list(ar=fit.11$coef[1:11])))
```