Time Series Analysis

Henrik Madsen

hm@imm.dtu.dk

Informatics and Mathematical Modelling
Technical University of Denmark
DK-2800 Kgs. Lyngby

Outline of the lecture

Identification of univariate time series models, cont.:

- Estimation of model parameters, Sec. 6.4 (cont.)
- Model order selection, Sec. 6.5
- Model validation, Sec. 6.6

Estimation – methods (from previous lecture)

- We have an appropriate model structure $AR(p)$, $MA(q)$, $ARMA(p,q)$, $ARIMA(p,d,q)$ with p, d, and q known
- **Task**: Based on the observations find appropriate values of the parameters
- The book describes many methods:
  - Moment estimates
  - LS-estimates
  - Prediction error estimates
    - Conditioned
    - Unconditioned
  - ML-estimates
    - Conditioned
    - Unconditioned (exact)

Maximum likelihood estimates

- $ARMA(p,q)$-process:

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

- Notation:

$$\begin{aligned} \boldsymbol{\theta}^T &= (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q) \\ \mathbf{Y}_t^T &= (Y_t, Y_{t-1}, \ldots, Y_1) \end{aligned}$$

- The Likelihood function is the joint probability distribution function for all observations for given values of $\boldsymbol{\theta}$ and σ_ε^2:

$$L(\mathbf{Y}_N; \boldsymbol{\theta}, \sigma_\varepsilon^2) = f(\mathbf{Y}_N | \boldsymbol{\theta}, \sigma_\varepsilon^2)$$

- Given the observations $\mathbf{Y}_N$ we estimate $\boldsymbol{\theta}$ and σ_ε^2 as the values for which the likelihood is maximized.

The likelihood function for $ARMA(p, q)$-models

- The random variable $Y_N|\mathbf{Y}_{N-1}$ only contains ε_N as a random component
- ε_N is a white noise process at time N and does therefore not depend on anything
- We therefore know that the random variables $Y_N|\mathbf{Y}_{N-1}$ and $\mathbf{Y}_{N-1}$ are independent, hence (see also page 3):

$$f(\mathbf{Y}_N|\boldsymbol{\theta}, \sigma_\varepsilon^2) = f(Y_N|\mathbf{Y}_{N-1}, \boldsymbol{\theta}, \sigma_\varepsilon^2) f(\mathbf{Y}_{N-1}|\boldsymbol{\theta}, \sigma_\varepsilon^2)$$

- Repeating these arguments:

$$L(\mathbf{Y}_N; \boldsymbol{\theta}, \sigma_\varepsilon^2) = \left(\prod_{t=p+1}^{N} f(Y_t|\mathbf{Y}_{t-1}, \boldsymbol{\theta}, \sigma_\varepsilon^2) \right) f(\mathbf{Y}_p|\boldsymbol{\theta}, \sigma_\varepsilon^2)$$

The conditional likelihood function

- Evaluation of $f(\mathbf{Y}_p|\boldsymbol{\theta}, \sigma_\varepsilon^2)$ requires special attention
- It turns out that the estimates obtained using the *conditional likelihood function*:

$$L(\mathbf{Y}_N; \boldsymbol{\theta}, \sigma_\varepsilon^2) = \prod_{t=p+1}^{N} f(Y_t|\mathbf{Y}_{t-1}, \boldsymbol{\theta}, \sigma_\varepsilon^2)$$

  results in the same estimates as the *exact likelihood function* when many observations are available
- For small samples there can be some difference
- Software:
  - The S-PLUS function `arima.mle` calculate conditional estimates
  - The R function `arima` calculate exact estimates

Evaluating the conditional likelihood function

- **Task**: Find the conditional densities given specified values of the parameters $\boldsymbol{\theta}$ and σ_ε^2
- The mean of the random variable $Y_t|\mathbf{Y}_{t-1}$ is the the 1-step forecast $\widehat{Y}_{t|t-1}$
- The prediction error $\varepsilon_t = Y_t - \widehat{Y}_{t|t-1}$ has variance σ_ε^2
- We assume that the process is Gaussian:

$$f(Y_t|\mathbf{Y}_{t-1}, \boldsymbol{\theta}, \sigma_\varepsilon^2) = \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \mathrm{e}^{-(Y_t - \widehat{Y}_{t|t-1}(\boldsymbol{\theta}))^2/2\sigma_\varepsilon^2}$$

- And therefore:

$$L(\mathbf{Y}_N; \boldsymbol{\theta}, \sigma_\varepsilon^2) = (\sigma_\varepsilon^2 2\pi)^{-\frac{N-p}{2}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=p+1}^{N} \varepsilon_t^2(\boldsymbol{\theta})\right)$$

ML-estimates

- The (conditional) ML-estimate $\widehat{\boldsymbol{\theta}}$ is a prediction error estimate since it is obtained by minimizing

$$S(\boldsymbol{\theta}) = \sum_{t=p+1}^{N} \varepsilon_t^2(\boldsymbol{\theta})$$

- By differentiating w.r.t. σ_ε^2 it can be shown that the ML-estimate of σ_ε^2 is

$$\widehat{\sigma}_\varepsilon^2 = S(\widehat{\boldsymbol{\theta}})/(N-p)$$

- The estimate $\widehat{\boldsymbol{\theta}}$ is asymptoticly “good” and the variance-covariance matrix is approximately $2\sigma_\varepsilon^2 \boldsymbol{H}^{-1}$ where $\boldsymbol{H}$ contains the 2nd order partial derivatives of $S(\boldsymbol{\theta})$ at the minimum

Finding the ML-estimates using the PE-method

- 1-step predictions:

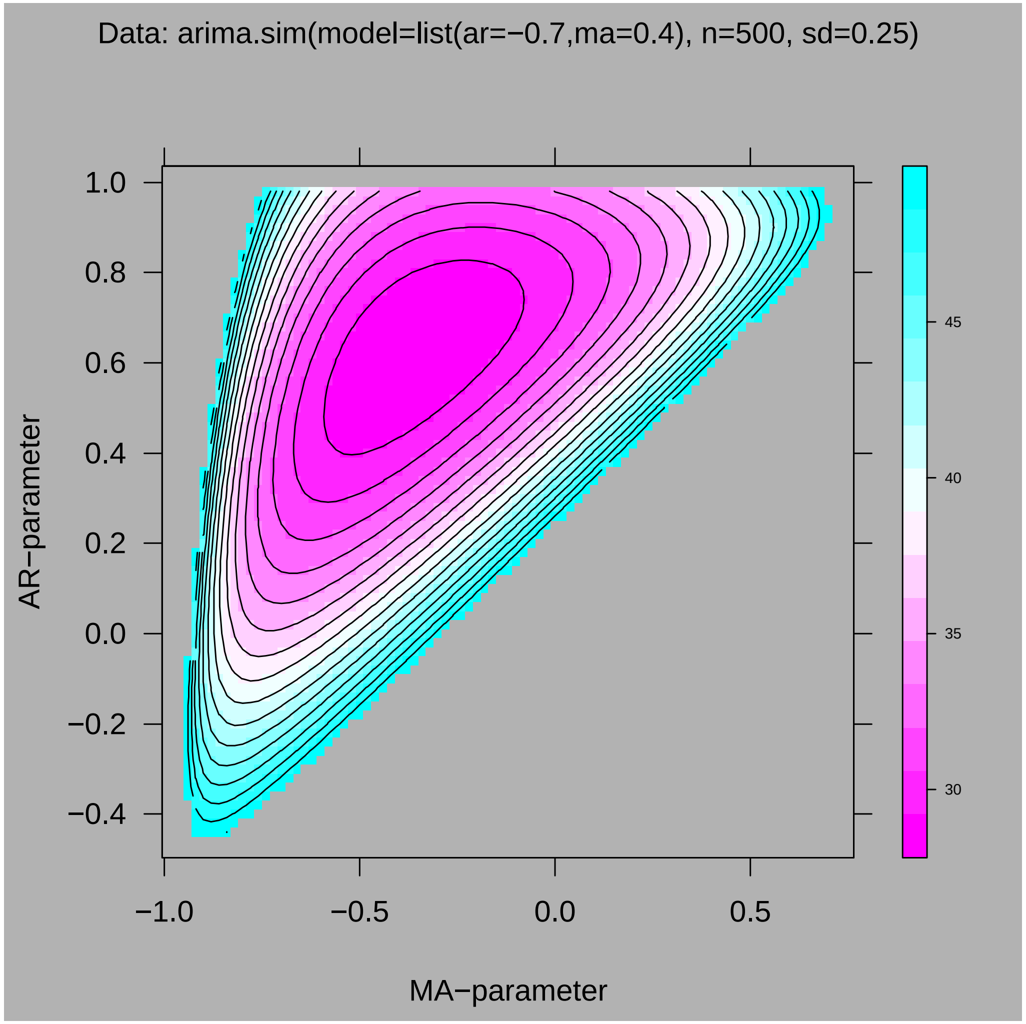
$$\widehat{Y}_{t|t-1} = -\phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} + 0 + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

- If we use $\varepsilon_p = \varepsilon_{p-1} = \cdots = \varepsilon_{p+1-q} = 0$ we can find:

$$\widehat{Y}_{p+1|p} = -\phi_1 Y_p - \cdots - \phi_p Y_1 + 0 + \theta_1 \varepsilon_p + \cdots + \theta_q \varepsilon_{p+1-q}$$

- Which will give us $\varepsilon_{p+1} = Y_{p+1} - \widehat{Y}_{p+1|p}$ and we can then calculate $\widehat{Y}_{p+2|p+1}$ and ε_{p+2} ... and so on until we have all the 1-step prediction errors we need.
- We use numerical optimization to find the parameters which minimize the sum of squared prediction errors

$S(\boldsymbol{\theta})$ for $(1 + 0.7B)Y_t = (1 - 0.4B)\varepsilon_t$ with $\sigma_\varepsilon^2 = 0.25^2$



Moment estimates

- Given the model structure: Find formulas for the theoretical autocorrelation or autocovariance as function of the parameters in the model
- Estimate, e.g. calculate the SACF
- Solve the equations by using the lowest lags necessary
- **Complicated!**
- **General properties of the estimator unknown!**

Moment estimates for $AR(p)$-processes

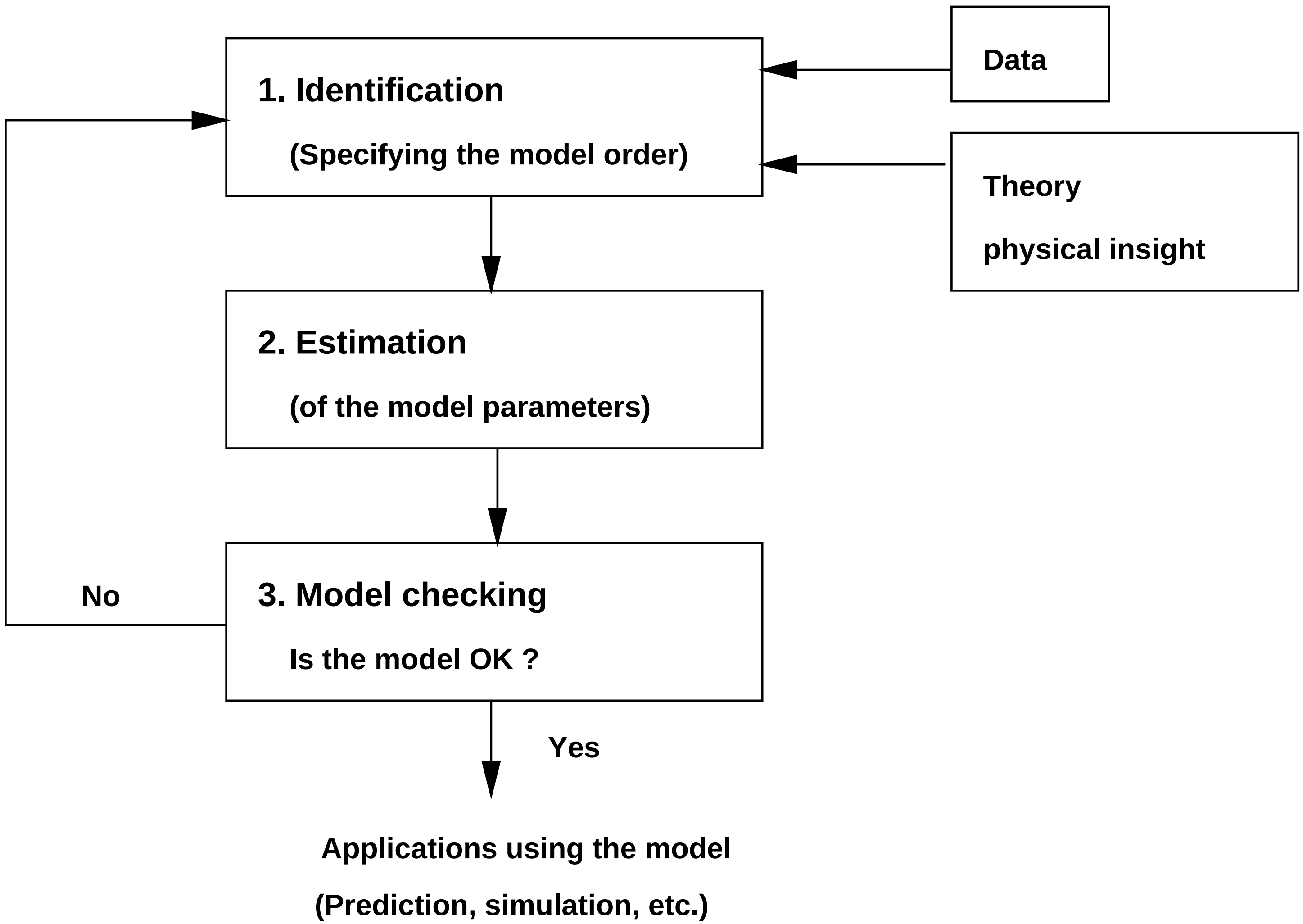
In this case moment estimates are simple to find due to the Yule-Walker equations (page 104). We simply plug in the estimated autocorrelation function in lags 1 to p:

$$\begin{bmatrix} \widehat{\rho}(1) \\ \widehat{\rho}(2) \\ \vdots \\ \widehat{\rho}(p) \end{bmatrix} = \begin{bmatrix} 1 & \widehat{\rho}(1) & \cdots & \widehat{\rho}(p-1) \\ \widehat{\rho}(1) & 1 & \cdots & \widehat{\rho}(p-2) \\ \vdots & \vdots & & \vdots \\ \widehat{\rho}(p-1) & \widehat{\rho}(p-2) & \cdots & 1 \end{bmatrix} \begin{bmatrix} -\phi_1 \\ -\phi_2 \\ \vdots \\ -\phi_p \end{bmatrix}$$

and solve w.r.t. the ϕ's

The function `ar` in S-PLUS or R use this approach as default

Model building



Validation of the model and extensions / reductions

- Residual analysis (Sec. 6.6.2): Is it possible to detect problems with residuals? (the 1-step prediction errors using the estimates, i.e. $\{\varepsilon_t(\widehat{\boldsymbol{\theta}})\}$, should be white noise)
- If the SACF or the SPACF of $\{\varepsilon_t(\widehat{\boldsymbol{\theta}})\}$ points towards a particular ARMA-structure we can derive how the original model should be extended (Sec. 6.5.1)
- If the model pass the residual analysis it makes sense to test null hypotheses about the parameters (Sec. 6.5.2)

Residual analysis

- Plot $\{\varepsilon_t(\widehat{\boldsymbol{\theta}})\}$; do the residuals look stationary?
- Tests in the autocorrelation. If $\{\varepsilon_t(\widehat{\boldsymbol{\theta}})\}$ is white noise then $\hat{\rho}_\varepsilon(k)$ is approximately Gaussian distributed with mean 0 and variance $1/N$.
  If the model fails calculate SPACF also and see if an ARMA-structure for the residuals can be derived (Sec. 6.5.1)
- Since $\hat{\rho}_\varepsilon(k_1)$ and $\hat{\rho}_\varepsilon(k_2)$ are independent (Eq. 6.4) the test statistic $Q^2 = \sum_{k=1}^{m} \left(\sqrt{N}\hat{\rho}_{\varepsilon_t(\widehat{\boldsymbol{\theta}})}(k)\right)^2$ is approximately distributed as $\chi^2(m-n)$, where n is the number of parameters.

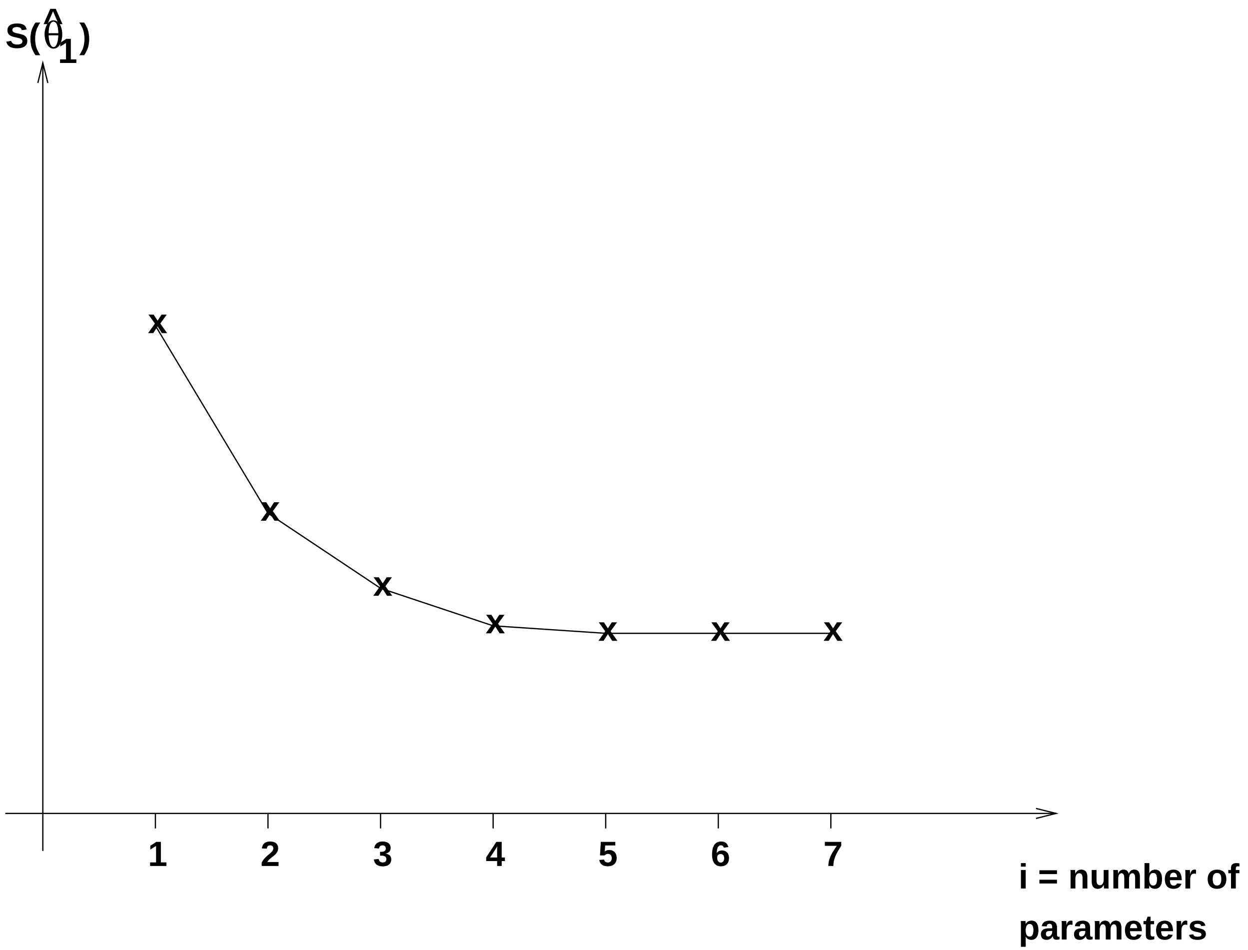
S-PLUS: `arima.diag('output from arima.mle')`

Residual analysis (continued)

- Test for the number of changes in sign. In a series of length N there is $N-1$ possibilities for changes in sign. If the series is white noise (with mean zero) the probability of change is $1/2$ and the changes will be independent. Therefore the number of changes is distributed as $Bin(N-1, 1/2)$
  S-PLUS: `binom.test(N-1, 'No. of changes')`

- Test in the *scaled cumulated periodogram* of the residuals is done by plotting it and adding lines at $\pm K_\alpha/\sqrt{q}$, where $q = (N-2)/2$ for N even and $q = (N-1)/2$ for N odd. For $1-\alpha$ confidence limits K_α can be found in Table 6.2
  S-PLUS (95% confidence interval):
  ```
  library(MASS)
  cpgram('residuals')
  ```

Sum of squared residuals depend on the model size



(It is assumed that the models are nested)

Test is the model

- The test essentially checks if the reduction in SSE ($S_1 - S_2$) is large enough to justify the *extra* parameters in model 2 (n_2 parameters) as compared to model 1 (n_1 parameters). The number of observations used is called N.
- If vector $\boldsymbol{\theta}_{extra}$ is used to denote the extra parameters in model 2 as compared to model 1, then the test is formally:

$$H_0: \quad \boldsymbol{\theta}_{extra} = \mathbf{0} \;\; vs. \;\; H_0: \quad \boldsymbol{\theta}_{extra} \neq \mathbf{0}$$

- If H_0 is true it (approximately) hold that

$$\frac{(S_1 - S_2)/(n_2 - n_1)}{S_2/(N - n_2)} \sim \mathrm{F}(n_2 - n_1, N - n_2)$$

(The likelihood ratio test is also a possibility)

Testing one parameter for significance

$$H_0 : \theta_i = 0 \text{ against } H_1 : \theta_i \neq 0$$

- Can be done as described on the previous slide
- Alternatively we can use a t-test based on the estimate and its standard error: $\hat{\theta}_i / \sqrt{\hat{V}(\hat{\theta}_i)}$
- Under H_0 and for an $ARMA(p, q)$-model this follows a $t(N-p-q)$ distribution (or $t(N-1-p-q)$ if we estimated an overall mean of the series)
- Often N is so large compared to the number of parameters that we can just use the standard normal distribution

Information criteria

- Select the model which minimize some information criterion
- Akaike's Information Criterion

$$AIC = -2\log(L(\mathbf{Y}_N; \widehat{\boldsymbol{\theta}}, \hat{\sigma}_\varepsilon^2)) + 2\, n_{par}$$

- Bayesian Information Criterion

$$BIC = -2\log(L(\mathbf{Y}_N; \widehat{\boldsymbol{\theta}}, \hat{\sigma}_\varepsilon^2)) + \log N\, n_{par}$$

- Except for an additive constant this can also be expressed as

$$\begin{aligned} AIC &= N \log \hat{\sigma}_\varepsilon^2 + 2\, n_{par} \\ BIC &= N \log \hat{\sigma}_\varepsilon^2 + \log N\, n_{par} \end{aligned}$$

- BIC yields a consistent estimate of the model order

Example

A model for CO_2 ...