Time Series Analysis

Lasse Engbo Christiansen

Department of Applied Mathematics and Computer Science
Technical University of Denmark

October 27, 2017

Outline of the lecture

- Estimation of parameters in linear dynamic models, Sec. 6.4
- Example in R

Identification of the ARMA-part

Characteristics for the autocorrelation functions:

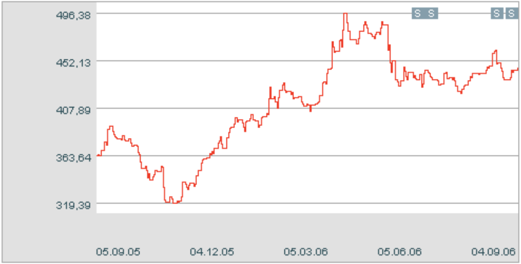
	ACF $\rho(k)$	PACF ϕ_{kk}
AR(p)	Damped exponential and/or sine functions	$\phi_{kk} = 0$ for $k > p$
MA(q)	$\rho(k) = 0$ for $k > q$	Dominated by damped exponential and or/sine functions
ARMA(p, q)	Damped exponential and/or sine functions after lag $\max(0, q - p)$	Dominated by damped exponential and/or sine functions after lag $\max(0, p - q)$

The IACF is similar to the PACF; see the book page 133

Estimation

- 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)

Example



Using the autocorrelation functions we agreed that

$$\hat{y}_{t+1|t} = a_1 y_t + a_2 y_{t-1}$$

and we should select a_1 and a_2 so that the sum of the squared prediction errors is minimized.

To comply with the notation of the book we will write the 1-step forecasts as $\hat{y}_{t+1|t} = -\phi_1 y_t - \phi_2 y_{t-1}$

The errors given the parameters (ϕ_1 and ϕ_2)

- Observations: $y_1, \; y_2, \; \ldots, \; y_N$
- Errors: $e_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t} = y_{t+1} - (-\phi_1 y_t - \phi_2 y_{t-1})$

$$
\begin{aligned}
e_{3|2} &= y_3 + \phi_1 y_2 + \phi_2 y_1 \\
e_{4|3} &= y_4 + \phi_1 y_3 + \phi_2 y_2 \\
e_{5|4} &= y_5 + \phi_1 y_4 + \phi_2 y_3 \\
&\vdots \\
e_{N|N-1} &= y_N + \phi_1 y_{N-1} + \phi_2 y_{N-2}
\end{aligned}
$$

$$
\begin{bmatrix} y_3 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} -y_2 & -y_1 \\ \vdots & \vdots \\ -y_{N-1} & -y_{N-2} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} + \begin{bmatrix} e_{3|2} \\ \vdots \\ e_{N|N-1} \end{bmatrix}
$$

Or just:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

Solution

To minimize the sum of the squared 1-step prediction errors $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ we use the result for the General Linear Model from Chapter 3:

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

With

$$\boldsymbol{X} = \begin{bmatrix} -y_2 & -y_1 \\ \vdots & \vdots \\ -y_{N-1} & -y_{N-2} \end{bmatrix} \quad \text{and} \quad \boldsymbol{Y} = \begin{bmatrix} y_3 \\ \vdots \\ y_N \end{bmatrix}$$

- Asymptotically: $V(\hat{\theta}) = \sigma_\epsilon^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$
- The method is called the LS-estimator for dynamical systems
- The method is also in the class of prediction error methods since it minimize the sum of the squared 1-step prediction errors
- **How does it generalize to AR(p)-models?**

Small illustrative example using R

```
obs <- c(-3.51, -3.81 ,-1.85, -2.02, -1.91, -0.88) ;
N <- length(obs)
(Y <- obs[3:N])

## [1] -1.85 -2.02 -1.91 -0.88

(X <- cbind(-obs[2:(N-1)], -obs[1:(N-2)]))

##      [,1] [,2]
## [1,] 3.81 3.51
## [2,] 1.85 3.81
## [3,] 2.02 1.85
## [4,] 1.91 2.02

solve(t(X) %*% X, t(X) %*% Y) # Estimates

##            [,1]
## [1,] -0.1474288
## [2,] -0.4476040
```

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:

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

- 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

- 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 (almost) the same estimates as the *exact likelihood function* when many observations are available
- For small samples there can be some differences
- Software:
  - The R function `arima` calculate exact estimates per default

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:

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

- 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 (remember that p is the order of the AR part):

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

- The estimate $\widehat{\boldsymbol{\theta}}$ is asymptotically unbiased and efficient, 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:

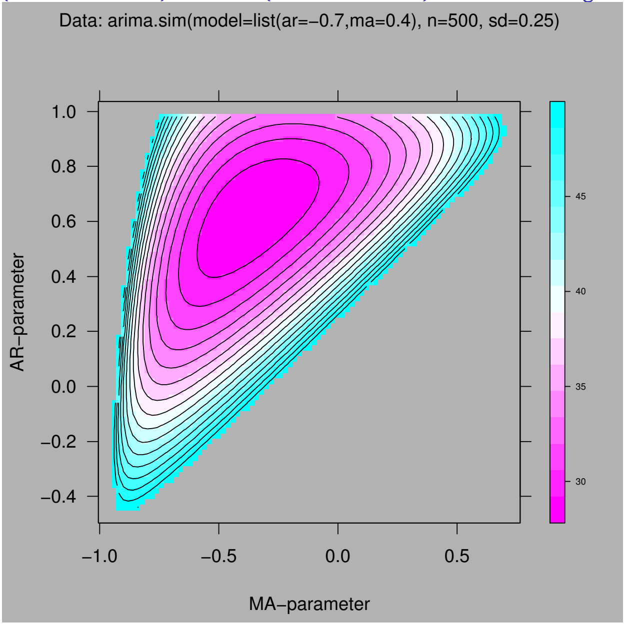
$$\widehat{Y}_{t|t-1} = -\phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

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

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

- 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 are unknown**

Moment estimates for $AR(p)$-processes

In this case moment estimates are simple to find due to the Yule-Walker equations. 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 R does this

Highlights

- Maximum likelihood estimation by looking at independent one step prediction errors.
- "Essentially, all models are wrong, but some are useful."
  (George E. P. Box)