Week 7:

ARMA

Fitting Models to Data

Last time

- stationarity, white noise WN
- linear process MA(∞)

$$Y_t = \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i} + \varepsilon_t = \psi(B)\varepsilon_t$$

- MA(q):

$$Y_t = \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t = \theta(B)\varepsilon_t$$

  ACF: $k_0 = q$, PACF: no k_0
- AR(p):

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t,$$

  ACF: no k_0, PACF: $k_0 = p$

ARMA

ARMA(p, q):

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

ARMA

ARMA(p, q):

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

Define

$$\varphi(z) = 1 - \varphi_1 z - \ldots \varphi_p z^p, \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$

then

$$\varphi(B) Y_t = \theta(B) \varepsilon_t.$$

ARMA

ARMA(p, q):

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

Define

$$\varphi(z) = 1 - \varphi_1 z - \ldots \varphi_p z^p, \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$

then

$$\varphi(B) Y_t = \theta(B) \varepsilon_t.$$

Properties:

- stationarity: if roots of $\varphi(z) = 0$ lie outside the unit circle

ARMA

ARMA(p, q):

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

Define

$$\varphi(z) = 1 - \varphi_1 z - \ldots \varphi_p z^p, \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$

then

$$\varphi(B) Y_t = \theta(B) \varepsilon_t.$$

Properties:

- stationarity: if roots of $\varphi(z) = 0$ lie outside the unit circle
- invertibility: if roots of $\theta(z) = 0$ lie outside the unit circle

ARMA

ARMA(p, q):

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

Define

$$\varphi(z) = 1 - \varphi_1 z - \ldots \varphi_p z^p, \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$

then

$$\varphi(B) Y_t = \theta(B) \varepsilon_t.$$

Properties:

- stationarity: if roots of $\varphi(z) = 0$ lie outside the unit circle
- invertibility: if roots of $\theta(z) = 0$ lie outside the unit circle
- ACF ρ_k
  - $\hookrightarrow$ Yule Walker equations $\rightsquigarrow$ difference equation for ρ_k
  - $\hookrightarrow$ no truncation point k_0

ARMA

ARMA(p, q):

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

Define

$$\varphi(z) = 1 - \varphi_1 z - \ldots \varphi_p z^p, \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$

then

$$\varphi(B) Y_t = \theta(B) \varepsilon_t.$$

Properties:

- stationarity: if roots of $\varphi(z) = 0$ lie outside the unit circle
- invertibility: if roots of $\theta(z) = 0$ lie outside the unit circle
- ACF ρ_k
  - $\hookrightarrow$ Yule Walker equations $\rightsquigarrow$ difference equation for ρ_k
  - $\hookrightarrow$ no truncation point k_0
- PACF ρ_{kk}: no truncation point k_0

Exercise

Consider ARMA(1,1)

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

- Express the stationarity condition.
- For a stationary process compute $\mathrm{E}\, Y_t$ and $\mathrm{Var}\, Y_t$.
- Compute ACF ρ_k.

Non-centered ARMA processes

- we have seen: $\{Y_t\}$ stationary ARMA(p, q) $\rightsquigarrow$ $\mathrm{E}\,Y_t = 0$

Non-centered ARMA processes

- we have seen: $\{Y_t\}$ stationary ARMA$(p, q) \rightsquigarrow \mathrm{E}\,Y_t = 0$
- If we want $\mathrm{E}\,Y_t = \mu \in \mathbb{R} \rightsquigarrow \mathrm{E}(Y_t - \mu) = 0$ and use

$$Y_t - \mu = \varphi_1(Y_{t-1} - \mu) + \cdots + \varphi_p(Y_{t-p} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

Non-centered ARMA processes

- we have seen: $\{Y_t\}$ stationary ARMA(p, q) $\rightsquigarrow$ $\mathrm{E}\,Y_t = 0$
- If we want $\mathrm{E}\,Y_t = \mu \in \mathbb{R}$ $\rightsquigarrow$ $\mathrm{E}(Y_t - \mu) = 0$ and use

$$Y_t - \mu = \varphi_1(Y_{t-1} - \mu) + \cdots + \varphi_p(Y_{t-p} - \mu) + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$$

or equivalently

$$Y_t = \alpha + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$$

where

$$\alpha = \mu(1 - \varphi_1 - \ldots - \varphi_p)$$

Non-centered ARMA processes

- we have seen: $\{Y_t\}$ stationary ARMA(p, q) $\leadsto$ $\operatorname{E} Y_t = 0$
- If we want $\operatorname{E} Y_t = \mu \in \mathbb{R}$ $\leadsto$ $\operatorname{E}(Y_t - \mu) = 0$ and use

$$Y_t - \mu = \varphi_1(Y_{t-1} - \mu) + \cdots + \varphi_p(Y_{t-p} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

or equivalently

$$Y_t = \alpha + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

where

$$\alpha = \mu(1 - \varphi_1 - \ldots - \varphi_p)$$

Be careful with R functions: if it estimates μ or α

Fitting models to data

Construction of Models by Box–Jenkins Methodology

Data: $Y_1, \ldots, Y_n$ observations from some stationary $\{Y_t\}$

1. Identification of model.
   - $\hookrightarrow$ Based on ACF and PACF
   - $\hookrightarrow$ Based on information criteria
2. Estimation of model.
3. Verification of model.

Construction of Models by Box–Jenkins Methodology

Data: $Y_1, \ldots, Y_n$ observations from some stationary $\{Y_t\}$

1. Identification of model.
   - $\hookrightarrow$ Based on ACF and PACF
   - $\hookrightarrow$ Based on information criteria
2. Estimation of model.
3. Verification of model.

There is no true model. *All models are wrong, but some of them are useful.*

Construction of Models by Box–Jenkins Methodology

Data: $Y_1, \ldots, Y_n$ observations from some stationary $\{Y_t\}$

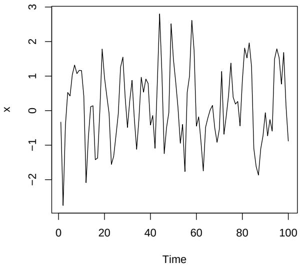
1. Identification of model.
   - ↪ Based on ACF and PACF
   - ↪ Based on information criteria
2. Estimation of model.
3. Verification of model.

There is no true model. *All models are wrong, but some of them are useful.*

⇝ We look for a parsimonious model which is able to describe our data well enough

Example

Data: $Y_1, \ldots, Y_{100}$



1. Based on some criteria $\rightsquigarrow$ choose AR(2) model
2. Estimation $\rightsquigarrow$ $\widehat{\varphi}_1 = 0.6634$, $\widehat{\varphi}_2 = -0.3137$
3. Model diagnostics (do residuals $\widehat{\varepsilon}_t$ look like white noise?), stationarity, evaluation of predictive ability etc.

1A. Identification based on ACF and PACF

Truncation point k_0: such that $\rho_k = 0$ whenever $k > k_0$

	MA(q)	AR(p)	ARMA(p, q)
ACF ρ_k	$k_0 = q$	no k_0	no k_0
PACF ρ_{kk}	no k_0	$k_0 = p$	no k_0

$\rightsquigarrow$ identification of the model using sample ACF and PACF

1A. Identification based on ACF and PACF

Truncation point k_0: such that $\rho_k = 0$ whenever $k > k_0$

	MA(q)	AR(p)	ARMA(p, q)
ACF ρ_k	$k_0 = q$	no k_0	no k_0
PACF ρ_{kk}	no k_0	$k_0 = p$	no k_0

$\rightsquigarrow$ identification of the model using sample ACF and PACF

Procedure:

1. plot ACF and PACF $\rightsquigarrow$ correlograms
2. try to look for truncation points

$\rightsquigarrow$ we need appropriate bounds for r_k under non-iid case to be able to decide if $\rho_k = 0$ or $\rho_k \neq 0$

⋆ Bartlett's formula for a linear process

Theorem. Let $\{Y_t\}$ is a stationary linear process with iid white noise $\{Z_k\}$ with $\mathrm{E}Z_k^4 < \infty$. Then for $k > 1$

$$\sqrt{n}(r_k - \rho_k) \xrightarrow{D} \mathrm{N}(0, w_{kk})$$

where

$$w_{kk} = \sum_{i=1}^{\infty} (\rho_{i+k} + \rho_{i-k} - 2\rho_k \rho_i)^2,$$

i.e. for n large $r_k \dot{\sim} \mathrm{N}(\rho_k, w_{kk}/n)$.

Application

⋆ Bartlett's formula for a linear process

Theorem. Let $\{Y_t\}$ is a stationary linear process with iid white noise $\{Z_k\}$ with $\mathrm{E}Z_k^4 < \infty$. Then for $k > 1$

$$\sqrt{n}(r_k - \rho_k) \xrightarrow{D} \mathrm{N}(0, w_{kk})$$

where

$$w_{kk} = \sum_{i=1}^{\infty}(\rho_{i+k} + \rho_{i-k} - 2\rho_k\rho_i)^2,$$

i.e. for n large $r_k \dot{\sim} \mathrm{N}(\rho_k, w_{kk}/n)$.

Application

$\hookrightarrow$ If $\{Y_t\}$ iid $\rightsquigarrow$ $w_{kk} = 1$

⋆ Bartlett's formula for a linear process

Theorem. Let $\{Y_t\}$ is a stationary linear process with iid white noise $\{Z_k\}$ with $\mathrm{E}Z_k^4 < \infty$. Then for $k > 1$

$$\sqrt{n}(r_k - \rho_k) \xrightarrow{D} \mathrm{N}(0, w_{kk})$$

where

$$w_{kk} = \sum_{i=1}^{\infty}(\rho_{i+k} + \rho_{i-k} - 2\rho_k\rho_i)^2,$$

i.e. for n large $r_k \,\dot{\sim}\, \mathrm{N}(\rho_k, w_{kk}/n)$.

Application

↪ If $\{Y_t\}$ iid $\rightsquigarrow$ $w_{kk} = 1$

↪ if k_0 if truncation point, then for $k > k_0$

$$w_{kk} = \sum_{i=1}^{\infty} \rho_{i-k}^2 = \sum_{i=1}^{\infty} \rho_{k-i}^2 = \sum_{i=-k_0}^{k_0} \rho_i^2 = 1 + 2\sum_{i=1}^{k_0} \rho_i^2$$

so

$$\sqrt{n} \cdot r_k \xrightarrow{D} \mathrm{N}\left(0, 1 + 2\sum_{i=1}^{k_0} \rho_i^2\right) \quad \text{for } k > k_0$$

Example

MA(1):

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Then

$$\rho_1 = \frac{\theta_1}{1+\theta_1^2}, \quad \rho_k = 0, \quad k > 1,$$

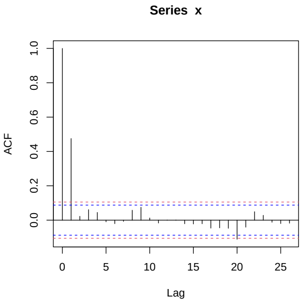
so

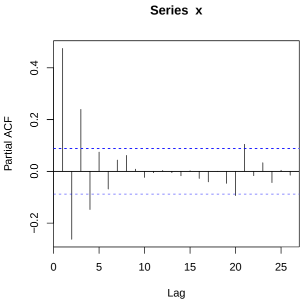
$$1 + 2\sum_{i=1}^{k_0} \rho_i^2 = 1 + 2\rho_1^2 = 1 + \frac{2\theta_1^2}{(1+\theta_1^2)^2}$$

for large *n*

$$r_k \dot{\sim} \mathrm{N}\left(0, \frac{1}{n}\left[1 + \frac{2\theta_1^2}{(1+\theta_1^2)^2}\right]\right), \quad k > 1$$

Example: MA(1)


Series x
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
Lag


Series x
Partial ACF
0.4
0.2
0.0
−0.2
0
5
10
15
20
25
Lag

Asymptotics for sample PACF

Quenouille's approximation. If $\{Y_t\}$ is an AR(p) process, then for $k > p$

$$\sqrt{n} r_{kk} \xrightarrow{D} \mathrm{N}(0, 1), \quad n \to \infty$$

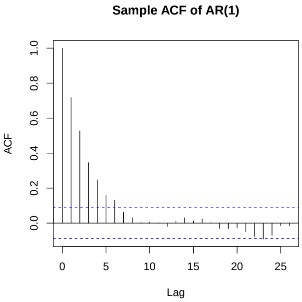
and r_{kk} and r_{jj} are asymptotically independent for $k > j > p$.

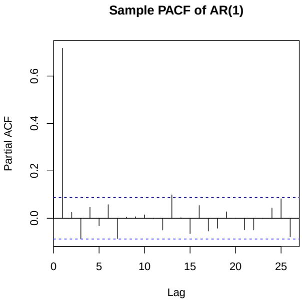
Asymptotics for sample PACF

Quenouille's approximation. If $\{Y_t\}$ is an AR(p) process, then for $k > p$

$$\sqrt{n} r_{kk} \xrightarrow{D} \mathrm{N}(0,1), \quad n \to \infty$$

and r_{kk} and r_{jj} are asymptotically independent for $k > j > p$.





Identification of the model

- plot ACF and PACF $\rightsquigarrow$ correlogram
- try to look for truncation points

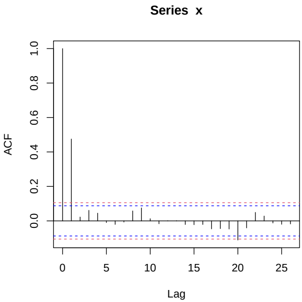
Identification of the model

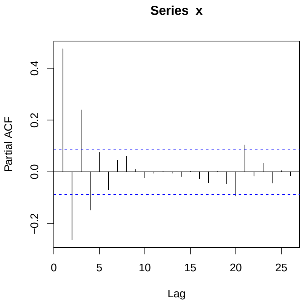
- plot ACF and PACF $\rightsquigarrow$ correlogram
- try to look for truncation points
- ACF
  - $\hookrightarrow$ bounds for iid:
  $$\left(-\frac{u_{0.975}}{n}, \frac{u_{0.975}}{n}\right)$$
  - $\hookrightarrow$ MA(q): the bounds for $k > q$ (Bartlett's approximation): Look if
  $$|r_k| < \frac{u_{0.975}}{n}\sqrt{1 + 2\sum_{j=1}^{q} r_j^2}, \quad k > q$$
  But one typically uses the default bounds

Identification of the model

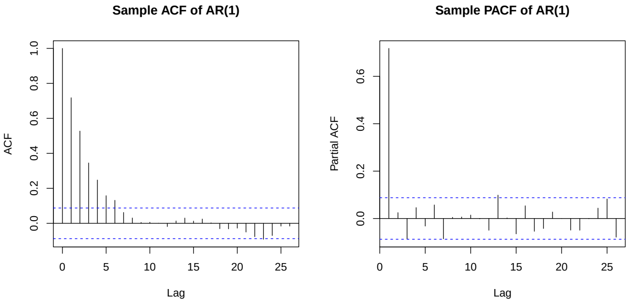
- plot ACF and PACF $\rightsquigarrow$ correlogram
- try to look for truncation points
- ACF
  - $\hookrightarrow$ bounds for iid:
  $$\left(-\frac{u_{0.975}}{n}, \frac{u_{0.975}}{n}\right)$$
  - $\hookrightarrow$ MA(q): the bounds for $k > q$ (Bartlett's approximation): Look if
  $$|r_k| < \frac{u_{0.975}}{n}\sqrt{1 + 2\sum_{j=1}^{q} r_j^2}, \quad k > q$$
  But one typically uses the default bounds
- PACF: AR(p) $\rightsquigarrow$ use the default bounds

Example: MA(1)

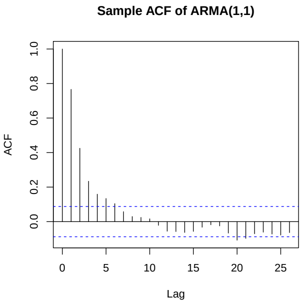

Series x
ACF
Lag
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25

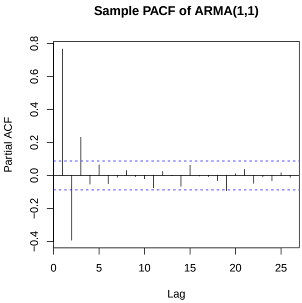

Series x
Partial ACF
Lag
−0.2
0.0
0.2
0.4
0
5
10
15
20
25

Example: AR(1)

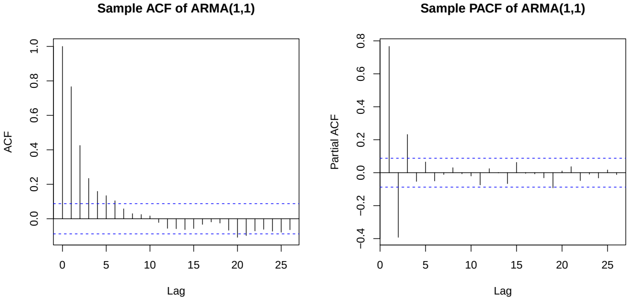

Sample ACF of AR(1)
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
Lag
Sample PACF of AR(1)
Partial ACF
0.6
0.4
0.2
0.0
0
5
10
15
20
25
Lag

Example: ARMA(1,1)





Example: ARMA(1,1)



data simulated from ARMA(1,1), but PACF looks like AR(3) . . .

1B. Identification based on information criteria

- popular approach in practice
- typically based on normal likelihood

Procedure:

1. Choose P, Q and fit all models

$$\text{ARMA}(p, q) \text{ for } 0 \leq p \leq P,\ 0 \leq q \leq Q.$$

2. For each model compute a chosen criterion: AIC, BIC.
3. Take

$$(\widehat{p}, \widehat{q}) = \mathrm{argmin}_{p,q}\, \text{criterion}(p, q).$$

Information criteria

$Z_1, \ldots, Z_n$ data: Model with log likelihood $l(\boldsymbol{\theta})$ depending on parameter $\boldsymbol{\theta} \in \mathbb{R}^k$:

$$\text{criterion } = -2l(\widehat{\boldsymbol{\theta}}) + p(k),$$

Namely,

$$AIC = -2l(\widehat{\boldsymbol{\theta}}) + 2k,$$
$$BIC = -2l(\widehat{\boldsymbol{\theta}}) + \log(n) \cdot k.$$

ARMA(p, q):

- $p + q + 1$ parameters if intercept not included

Information criteria

$Z_1, \ldots, Z_n$ data: Model with log likelihood $l(\boldsymbol{\theta})$ depending on parameter $\boldsymbol{\theta} \in \mathbb{R}^k$:

$$\text{criterion } = -2l(\widehat{\boldsymbol{\theta}}) + p(k),$$

Namely,

$$AIC = -2l(\widehat{\boldsymbol{\theta}}) + 2k,$$
$$BIC = -2l(\widehat{\boldsymbol{\theta}}) + \log(n) \cdot k.$$

ARMA(p, q):

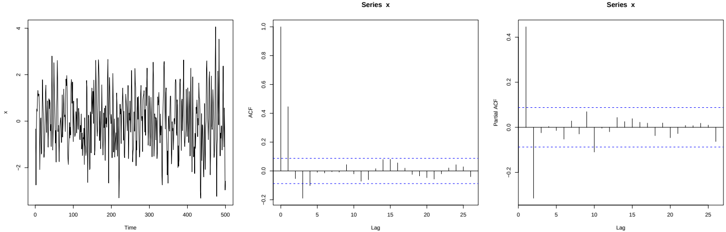
- $p + q + 1$ parameters if intercept not included

If l is normal regression log likelihood

$$-2l(\widehat{\boldsymbol{\theta}}) = n + n\log(2\pi) + n\log\left(\widehat{\sigma}^2\right)$$

Example

Data simulated from AR(2) with $\varphi_1 = 0.6$ and $\varphi_2 = -0.3$



p, q	AIC 0	AIC 1	AIC 2	AIC 3	AIC 4	BIC 0	BIC 1	BIC 2	BIC 3	BIC 4
0	1605	1461	1458	1455	1447	1610	1470	1471	1472	1468
1	1496	1460	1459	1451	1446	1504	1473	1476	1472	1471
2	1445	1447	1449	1451	1445	1458	1464	1470	1476	1474
3	1447	1449	1451	1452	1447	1464	1470	1476	1482	1481
4	1449	1451	1447	1449	1449	1470	1476	1477	1483	1487

Model estimation

Various possible estimation methods

- AR models
  - Least squares conditional on first p values (CLS)
  - Moment methods (based on Yule Walker equations)
  - Maximum likelihood (MLE) or conditional maximum likelihood (CMLE) under normality
- MA models and ARMA models
  - recursive nonlinear least squares (NLS)
  - CMLE under normality

AR: CLS estimation

Data $Y_1, \dots, Y_n$

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t, \quad t = p+1, \dots, n,$$

$\rightsquigarrow$ a regression form $\boldsymbol{Y} = \mathbb{X}\boldsymbol{\varphi} + \boldsymbol{\varepsilon}$ $\rightsquigarrow$ least squares estimation

$$\widehat{\boldsymbol{\varphi}} = \mathrm{argmin}_{\boldsymbol{\varphi} \in \mathbb{R}^p} \sum_{t=p+1}^{n} (Y_t - \varphi_1 Y_{t-1} - \ldots - \varphi_p Y_{t-p})^2$$

with

$$\widehat{\sigma}^2 = \frac{1}{n-p} \sum_{t=p+1}^{n} (Y_t - \widehat{\varphi}_1 Y_{t-1} - \ldots - \widehat{\varphi}_p Y_{t-p})^2.$$

AR: CLS estimation

Data $Y_1, \ldots, Y_n$

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t, \quad t = p+1, \ldots, n,$$

$\rightsquigarrow$ a regression form $\boldsymbol{Y} = \mathbb{X}\boldsymbol{\varphi} + \boldsymbol{\varepsilon}$ $\rightsquigarrow$ least squares estimation

$$\widehat{\boldsymbol{\varphi}} = \mathrm{argmin}_{\boldsymbol{\varphi} \in \mathbb{R}^p} \sum_{t=p+1}^{n} (Y_t - \varphi_1 Y_{t-1} - \ldots - \varphi_p Y_{t-p})^2$$

with

$$\widehat{\sigma}^2 = \frac{1}{n-p} \sum_{t=p+1}^{n} (Y_t - \widehat{\varphi}_1 Y_{t-1} - \ldots - \widehat{\varphi}_p Y_{t-p})^2.$$

Properties: If the model is stationary and $\{\varepsilon_t\}$ white noise:

- For n large

$$\widehat{\boldsymbol{\varphi}} \dot{\sim} \mathrm{N}(\boldsymbol{\varphi}, \widehat{\sigma}^2 (\mathbb{X}^\top \mathbb{X})^{-1})$$

$\rightsquigarrow$ one can use `lm` and `summary` to get a valid inference

AR: Moment estimation

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t,$$

Yule Walker equations:

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \cdots + \varphi_p \rho_{k-p}, \quad k \geq 1.$$

AR: Moment estimation

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t,$$

Yule Walker equations:

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \cdots + \varphi_p \rho_{k-p}, \quad k \geq 1.$$

For $k = 1, \ldots, p \rightsquigarrow$ equation for $\boldsymbol{\varphi}$

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & & & & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \vdots & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{pmatrix}$$

$\rightsquigarrow$ replace ρ_k by r_k and solve it for $\boldsymbol{\varphi}$

AR: Conditional MLE

AR(1): Assume that ε_t are iid from N$(0, \sigma^2)$ and

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t,$$

AR: Conditional MLE

AR(1): Assume that ε_t are iid from $\mathrm{N}(0, \sigma^2)$ and

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t,$$

Conditional distribution

$$Y_t | Y_{t-1}, Y_{t-2}, \ldots \overset{D}{=} Y_t | Y_{t-1} \sim \mathrm{N}(\varphi_1 Y_{t-1}, \sigma^2)$$

AR: Conditional MLE

AR(1): Assume that ε_t are iid from $\mathrm{N}(0, \sigma^2)$ and

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t,$$

Conditional distribution

$$Y_t | Y_{t-1}, Y_{t-2}, \ldots \overset{D}{=} Y_t | Y_{t-1} \sim \mathrm{N}(\varphi_1 Y_{t-1}, \sigma^2)$$

$\rightsquigarrow$ it is easy to express conditional density $f_{t|t-1}$

AR: Conditional MLE

AR(1): Assume that ε_t are iid from $N(0, \sigma^2)$ and

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t,$$

Conditional distribution

$$Y_t | Y_{t-1}, Y_{t-2}, \ldots \stackrel{D}{=} Y_t | Y_{t-1} \sim N(\varphi_1 Y_{t-1}, \sigma^2)$$

$\rightsquigarrow$ it is easy to express conditional density $f_{t|t-1}$
$\rightsquigarrow$ joint distribution of $(Y_2, \ldots, Y_n)^\top$ given Y_1 has density

$$f(y_2, \ldots, y_n | y_1) = f(y_2|y_1) \cdot f(y_3|y_2) \cdot \ldots f(y_n|y_{n-1})$$

AR: Conditional MLE

AR(1): Assume that ε_t are iid from $\mathrm{N}(0, \sigma^2)$ and

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t,$$

Conditional distribution

$$Y_t | Y_{t-1}, Y_{t-2}, \ldots \overset{D}{=} Y_t | Y_{t-1} \sim \mathrm{N}(\varphi_1 Y_{t-1}, \sigma^2)$$

$\rightsquigarrow$ it is easy to express conditional density $f_{t|t-1}$
$\rightsquigarrow$ joint distribution of $(Y_2, \ldots, Y_n)^\top$ given Y_1 has density

$$f(y_2, \ldots, y_n | y_1) = f(y_2 | y_1) \cdot f(y_3 | y_2) \cdot \ldots f(y_n | y_{n-1})$$

$\rightsquigarrow$ likelihood conditional on Y_1:

$$L(\varphi_1, \sigma^2) = \prod_{t=2}^{n} f_{t|t-1}(Y_t | Y_{t-1}) = \prod_{t=2}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(Y_t - \varphi_1 Y_{t-1})^2}$$

obviously: $\widehat{\varphi}_1$ is the same as for CLS

ARMA: Conditional MLE

Consider ARMA(1,1):

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

ARMA: Conditional MLE

Consider ARMA(1,1):

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Conditional distribution

$$Y_t | Y_{t-1}, \varepsilon_{t-1}, Y_{t-2}, \varepsilon_{t-2} \ldots \stackrel{D}{=} Y_t | Y_{t-1}, \varepsilon_{t-1} \sim \mathrm{N}(\varphi_1 Y_{t-1} + \theta_1 \varepsilon_{t-1}, \sigma^2)$$

ARMA: Conditional MLE

Consider ARMA(1,1):

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Conditional distribution

$$Y_t | Y_{t-1}, \varepsilon_{t-1}, Y_{t-2}, \varepsilon_{t-2} \ldots \stackrel{D}{=} Y_t | Y_{t-1}, \varepsilon_{t-1} \sim \mathrm{N}(\varphi_1 Y_{t-1} + \theta_1 \varepsilon_{t-1}, \sigma^2)$$

If we knew ε_1 then for given $\varphi_1, \theta_1 \rightsquigarrow$

$$\varepsilon_2(\varphi_1, \theta_1) = Y_2 - \varphi_1 Y_1 - \theta_1 \varepsilon_1$$

and for $t > 2$crecursively

$$\varepsilon_t(\varphi_1, \theta_1) = Y_t - \varphi_1 Y_{t-1} - \theta_1 \varepsilon_{t-1}(\varphi_1).$$

ARMA: Conditional MLE

Consider ARMA(1,1):

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Conditional distribution

$$Y_t | Y_{t-1}, \varepsilon_{t-1}, Y_{t-2}, \varepsilon_{t-2} \ldots \stackrel{D}{=} Y_t | Y_{t-1}, \varepsilon_{t-1} \sim \mathrm{N}(\varphi_1 Y_{t-1} + \theta_1 \varepsilon_{t-1}, \sigma^2)$$

If we knew ε_1 then for given $\varphi_1, \theta_1 \rightsquigarrow$

$$\varepsilon_2(\varphi_1, \theta_1) = Y_2 - \varphi_1 Y_1 - \theta_1 \varepsilon_1$$

and for $t > 2$crecursively

$$\varepsilon_t(\varphi_1, \theta_1) = Y_t - \varphi_1 Y_{t-1} - \theta_1 \varepsilon_{t-1}(\varphi_1).$$

Likelihood:

$$L(\varphi_1, \theta_1, \sigma^2) = \prod_{t=2}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(Y_t - \varphi_1 Y_{t-1} - \theta_1 \varepsilon_{t-1}(\varphi_1, \theta_1))^2}$$

ARMA: NLS

$$L(\varphi_1, \theta_1, \sigma^2) = \prod_{t=2}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(Y_t - \varphi_1 Y_{t-1} - \theta_1 \varepsilon_{t-1}(\varphi_1, \theta_1))^2}$$

so

$$\max_{\phi_1, \theta_1, \sigma^2} \log L(\varphi_1, \theta_1, \sigma^2)$$

is equivalent to minimizing (non-linear) least squares

$$\min_{\phi_1, \theta_1} \sum_{t=2}^{n} (Y_t - \varphi_1 Y_{t-1} - \theta_1 \varepsilon_{t-1}(\varphi_1, \theta_1))^2$$

For ARMA(p, q) $\rightsquigarrow$ need to choose q starting values.

- ↪ set as 0
- ↪ more sophisticated methods (using preliminary estimation)
- ↪ possible to compute the unconditional likelihood (more complicated)

Practical estimation in R

Various functions in various packages

- `ar (stats)`
- `arima (stats)`
- `arma (tseries)`
- `auto.arima (forecast)`
- `Arima (forecast)`

Be careful:

- possibly different parametrizations
- different syntax, methods, initial values, . . .
- always read help