

02417: Time Series Analysis

Week 5 - AR, MA and ARMA processes

Peder Bacher

DTU Compute

Based on material previous material from the course

March 7, 2025

Week 5: Outline of the lecture

- ▶ Stochastic processes - 2nd part:
 - MA, AR, and ARMA-processes, Sec. 5.5
 - Non-stationary models, Sec. 5.6
 - Seasonal ARIMA models
 - Optimal Prediction, Sec. 5.7
- ▶ Estimation of parameters in linear dynamic models, Sec. 6.4

Linear process as a statistical model?

$$Y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \psi_3 \varepsilon_{t-3} + \dots$$

- ▶ Observations: $Y_1, Y_2, Y_3, \dots, Y_N$
- ▶ Task: Find an infinite number of parameters from N observations!
- ▶ Solution:

Linear process as a statistical model?

$$Y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \psi_3 \varepsilon_{t-3} + \dots$$

- ▶ Observations: $Y_1, Y_2, Y_3, \dots, Y_N$
- ▶ Task: Find an infinite number of parameters from N observations!
- ▶ Solution: Restrict the sequence $1, \psi_1, \psi_2, \psi_3, \dots$

MA(q), AR(p), and ARMA(p, q) processes

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

$\{\varepsilon_t\}$ is white noise

MA(q), AR(p), and ARMA(p, q) processes

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

$$Y_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} = \varepsilon_t$$

$\{\varepsilon_t\}$ is white noise

MA(q), AR(p), and ARMA(p, q) processes

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

$$Y_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} = \varepsilon_t$$

$$Y_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

$\{\varepsilon_t\}$ is white noise

MA(q), AR(p), and ARMA(p, q) processes

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

$$Y_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} = \varepsilon_t$$

$$Y_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

$\{\varepsilon_t\}$ is white noise

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

$$\phi(B)Y_t = \varepsilon_t$$

$$\phi(B)Y_t = \theta(B)\varepsilon_t$$

where

$$\phi(B) = (1 + \phi_1 B + \phi_1 B^2 + \dots + \phi_p B^p)$$

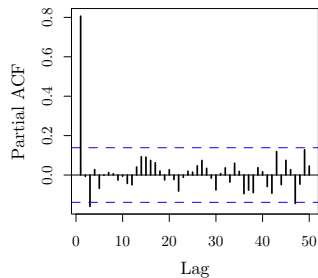
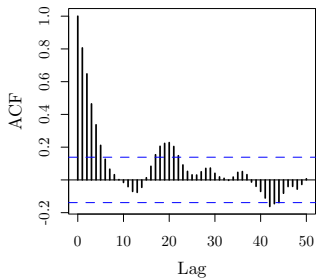
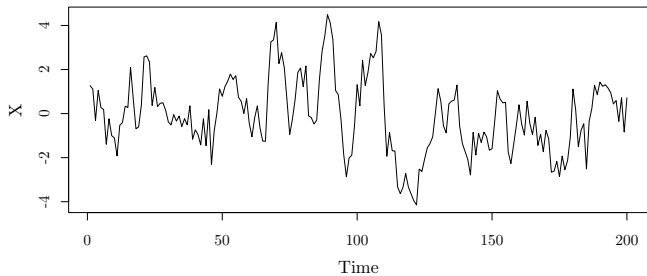
$$\theta(B) = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)$$

are polynomials in the backward shift operator B , ($BX_t = X_{t-1}$, $B^2 X_t = X_{t-2}$)

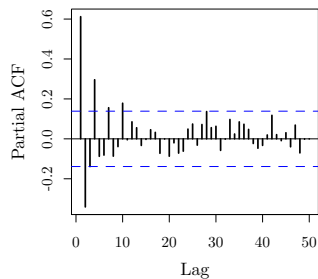
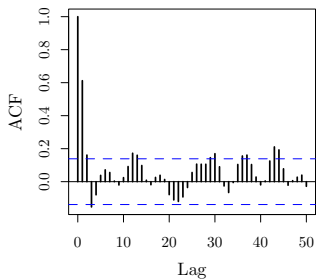
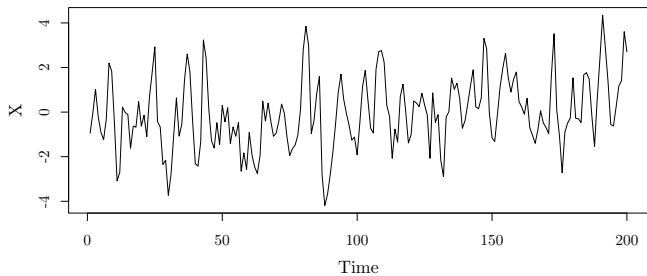

```
# Write an ARMA simulation  
n <- 200  
y <- numeric(n)  
# AR1:  
# (1 - phi1 * B) * Y_t = eps_t  
#
```

```
# Function for plot
plotit <- function(x){
  layout(rbind(1,2:3))
  par(mar=c(3,3,1,1), mgp=c(2, 0.7,0))
  plot(x, ylab="X")
  acf(x, lag.max=50, lwd=2)
  pacf(x, lag.max=50, lwd=2)
}
```

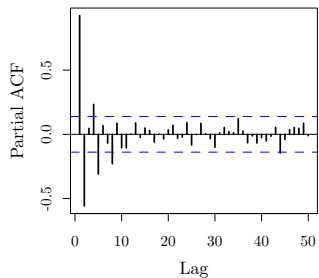
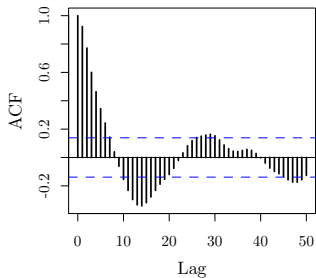
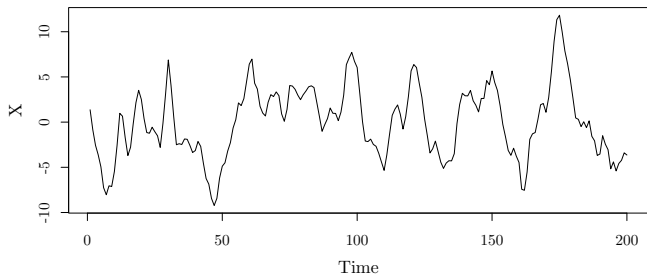
```
# AR(1)
n <- 200
plotit( arima.sim(model=list(ar=c(0.8)), n=n) )
```



```
# MA(2)
plotit( arima.sim(list(ma=c(0.9, 0.8)), n) )
```



```
# arma(1,2)
plotit( arima.sim(list(ar=c(0.8),ma=c(0.9, 0.8))), n )
```



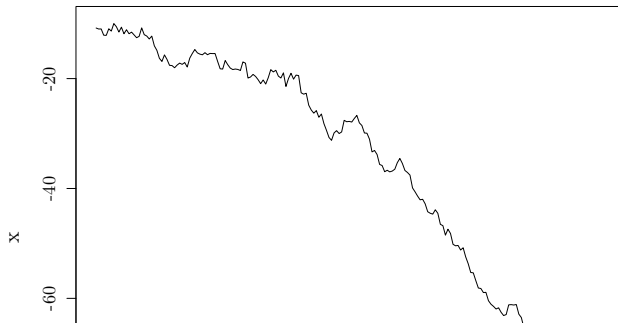
```

# A simulation function for ARMA simulation, use model as arima.sim, i.e. flip sign of phi (into ar) coefficients
sim <- function(model, n, nburnin=100){
  n <- n + nburnin
  # Take the ar and ma part
  ar <- model$ar
  ma <- model$ma
  # The order (i.e. the number of lags)
  p <- length(ar)
  q <- length(ma)
  # The vector for the simulation result
  y <- numeric(n)
  # Generate the random normal values
  eps <- rnorm(n)
  # Run the simulation
  for(i in (max(p,q)+1):n){
    y[i] <- eps[i] + sum(y[i-(1:p)] * ar) + sum(eps[i-(1:q)] * ma)
  }
  # Return without the burn-in period
  return(y[(nburnin+1):n])
}

```

```
# Test it by comparing
model <- list(ar=c(0.4), ma=c(0.2))
set.seed(12)
sim(model, 10, nburnin=100)
set.seed(12)
arima.sim(model, 10, n.start=100)

# Non-stationary process
# Do the simulation and plot
n <- 200
model <- list(ar=c(1.01))
#arima.sim(model, n)
x <- sim(model, n)
plot(x, type="l", ylab="x", xlab="t")
```



Invertibility and Stationarity

- ▶ A stochastic process is said to be *invertible* if a finite amount of observations can determine its state.

Invertibility and Stationarity

- ▶ A stochastic process is said to be *invertible* if a finite amount of observations can determine its state.

▶ A stochastic process is said to be *stationary* if?

Invertibility and Stationarity

- ▶ A stochastic process is said to be *invertible* if a finite amount of observations can determine its state.

- ▶ A stochastic process is said to be *stationary* if its distribution does not change over time.

Invertibility and Stationarity of ARMA models

► $MA(q) : Y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q}$

Invertibility and Stationarity of ARMA models

► $MA(q) : Y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q}$

Always stationary

Invertibility and Stationarity of ARMA models

► $MA(q) : Y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q}$

Always stationary

Invertible if the roots in $\theta(z^{-1}) = 0$ with respect to z all are within the unit circle

Invertibility and Stationarity of ARMA models

▶ $MA(q) : Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$

Always stationary

Invertible if the roots in $\theta(z^{-1}) = 0$ with respect to z all are within the unit circle

▶ $AR(p) : Y_t + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} = \varepsilon_t$

Invertibility and Stationarity of ARMA models

▶ $MA(q) : Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$

Always stationary

Invertible if the roots in $\theta(z^{-1}) = 0$ with respect to z all are within the unit circle

▶ $AR(p) : Y_t + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} = \varepsilon_t$

Always invertible

Invertibility and Stationarity of ARMA models

▶ $MA(q) : Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$

Always stationary

Invertible if the roots in $\theta(z^{-1}) = 0$ with respect to z all are within the unit circle

▶ $AR(p) : Y_t + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} = \varepsilon_t$

Always invertible

Stationary if the roots of $\phi(z^{-1})$ with respect to z all lie within the unit circle

Invertibility and Stationarity of ARMA models

▶ $MA(q) : Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$

Always stationary

Invertible if the roots in $\theta(z^{-1}) = 0$ with respect to z all are within the unit circle

▶ $AR(p) : Y_t + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} = \varepsilon_t$

Always invertible

Stationary if the roots of $\phi(z^{-1})$ with respect to z all lie within the unit circle

▶ $ARMA(p, q)$

Stationary if the roots of $\phi(z^{-1})$ with respect to z all lie within the unit circle

Invertible if the roots in $\theta(z^{-1})$ with respect to z all are within the unit circle

Autocorrelations

MA(2):

$$Y_t = (1 + 0.9B + 0.8B^2)\epsilon_t$$

Autocorrelations

MA(2):

$$Y_t = (1 + 0.9B + 0.8B^2)\epsilon_t$$

zero after lag 2

Autocorrelations

MA(2):

$$Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

zero after lag 2

AR(1):

$$(1 - 0.8B)Y_t = \varepsilon_t$$

Autocorrelations

MA(2):

$$Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

zero after lag 2

AR(1):

$$(1 - 0.8B)Y_t = \varepsilon_t$$

exponential decay (damped sine in case of complex roots)

Autocorrelations

MA(2):

$$Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

zero after lag 2

AR(1):

$$(1 - 0.8B)Y_t = \varepsilon_t$$

exponential decay (damped sine in case of complex roots)

ARMA(1,2):

$$(1 - 0.8B)Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

Autocorrelations

MA(2):

$$Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

zero after lag 2

AR(1):

$$(1 - 0.8B)Y_t = \varepsilon_t$$

exponential decay (damped sine in case of complex roots)

ARMA(1,2):

$$(1 - 0.8B)Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

exponential decay from lag $q + 1 - p = 2 + 1 - 1 = 2$ (damped sine in case of complex roots)

Autocorrelations

MA(2):

$$Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

zero after lag 2

AR(1):

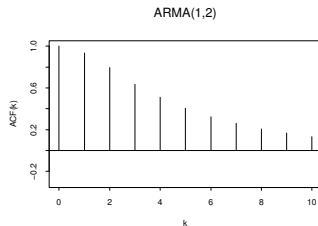
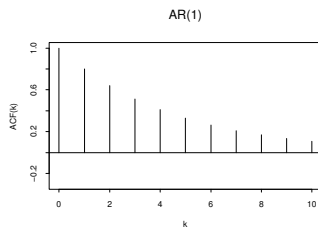
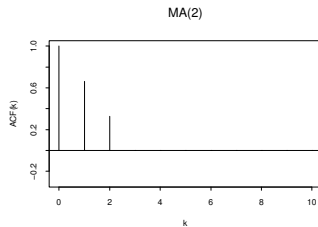
$$(1 - 0.8B)Y_t = \varepsilon_t$$

exponential decay (damped sine in case of complex roots)

ARMA(1,2):

$$(1 - 0.8B)Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

exponential decay from lag $q + 1 - p = 2 + 1 - 1 = 2$ (damped sine in case of complex roots)



Partial autocorrelations

MA(2):

$$Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

Partial autocorrelations

MA(2):

$$Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

AR(1):

$$(1 - 0.8B)Y_t = \varepsilon_t$$

Partial autocorrelations

MA(2):

$$Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

AR(1):

$$(1 - 0.8B)Y_t = \varepsilon_t$$

zero after lag 1

ARMA(1,2):

$$(1 - 0.8B)Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

Partial autocorrelations

MA(2):

$$Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

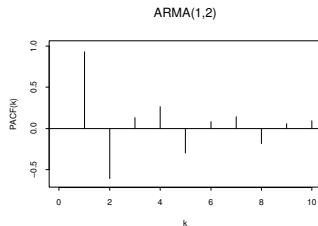
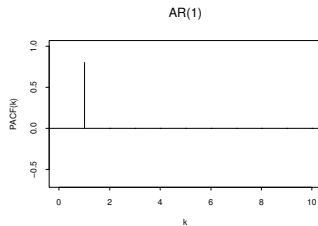
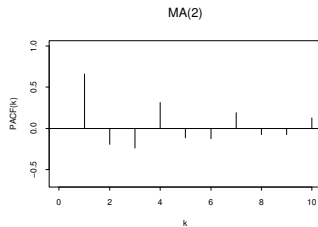
AR(1):

$$(1 - 0.8B)Y_t = \varepsilon_t$$

zero after lag 1

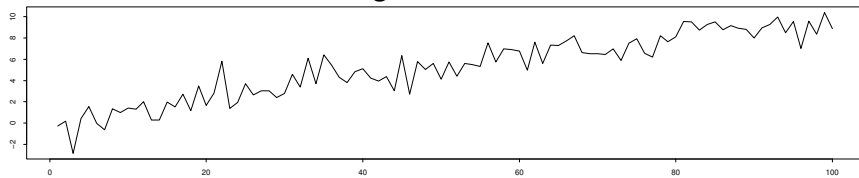
ARMA(1,2):

$$(1 - 0.8B)Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$$

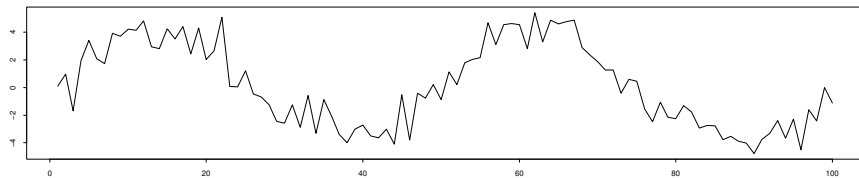


Non-stationary time series

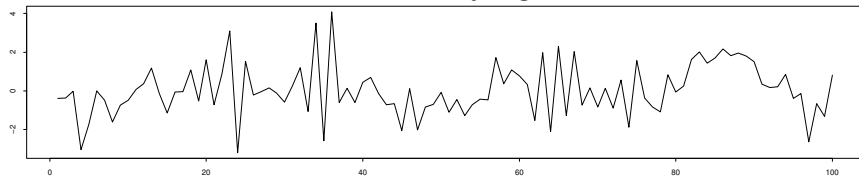
Long term trends



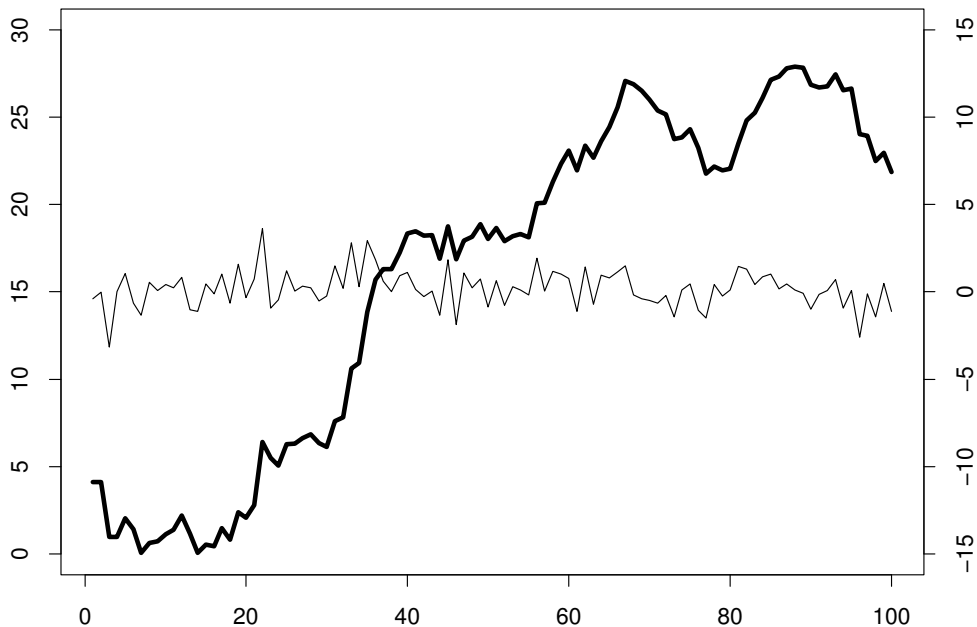
Periodic trends



General time varying behavior



Differencing



The $ARIMA(p, d, q)$ -process

- ▶ An $ARMA(p, q)$ model for:

$$W_t = \nabla^d Y_t = (1 - B)^d Y_t$$

where $\{Y_t\}$ is the series

The $ARIMA(p, d, q)$ -process

- ▶ An $ARMA(p, q)$ model for:

$$W_t = \nabla^d Y_t = (1 - B)^d Y_t$$

where $\{Y_t\}$ is the series

- ▶ That is:

$$\phi(B)\nabla^d Y_t = \theta(B)\varepsilon_t$$

The $ARIMA(p, d, q)$ -process

- ▶ An $ARMA(p, q)$ model for:

$$W_t = \nabla^d Y_t = (1 - B)^d Y_t$$

where $\{Y_t\}$ is the series

- ▶ That is:

$$\phi(B)\nabla^d Y_t = \theta(B)\varepsilon_t$$

- ▶ If we consider stationarity:

$$\phi(z^{-1})(1 - z^{-1})^d = 0$$

i.e. d roots in $z = 1 + 0i$, and the rest inside the unit circle

Seasonal Models

- ▶ In general, would you rather use new or old information in your models, for example would you prefer $Y_t = \theta Y_{t-1} + \epsilon_t$ or $Y_t = \theta Y_{t-2} + \epsilon_t$?

Seasonal Models

- ▶ In general, would you rather use new or old information in your models, for example would you prefer $Y_t = \theta Y_{t-1} + \epsilon_t$ or $Y_t = \theta Y_{t-2} + \epsilon_t$?
- ▶ When and why would it make sense to prefer older information over newer information?

The $(p, d, q) \times (P, D, Q)_s$ seasonal process

- ▶ A multiplicative (stationary) $ARMA(p, q)$ model for:

$$W_t = \nabla^d \nabla_s^D Y_t = (1 - B)^d (1 - B^s)^D Y_t$$

where $\{Y_t\}$ is the series

The $(p, d, q) \times (P, D, Q)_s$ seasonal process

- ▶ A multiplicative (stationary) $ARMA(p, q)$ model for:

$$W_t = \nabla^d \nabla_s^D Y_t = (1 - B)^d (1 - B^s)^D Y_t$$

where $\{Y_t\}$ is the series

- ▶ That is:

$$\phi(B)\phi(B^s)\nabla^d \nabla_s^D Y_t = \theta(B)\Theta(B^s)\varepsilon_t$$

The $(p, d, q) \times (P, D, Q)_s$ seasonal process

- ▶ A multiplicative (stationary) $ARMA(p, q)$ model for:

$$W_t = \nabla^d \nabla_s^D Y_t = (1 - B)^d (1 - B^s)^D Y_t$$

where $\{Y_t\}$ is the series

- ▶ That is:

$$\phi(B)\Phi(B^s)\nabla^d \nabla_s^D Y_t = \theta(B)\Theta(B^s)\varepsilon_t$$

- ▶ If we consider stationarity:

$$\phi(z^{-1})\Phi(z^{-s})(1 - z^{-1})^d(1 - z^{-s})^D = 0$$

i.e. d roots in $z = 1 + 0i$, $D \times s$ roots on the unit circle, and the rest inside the unit circle

The case $d = D = 0$; stationary seasonal process

► General:

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)\varepsilon_t$$

► Example:

$$(1 - \Phi B^{12})Y_t = \varepsilon_t$$

The case $d = D = 0$; stationary seasonal process

- ▶ General:

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)\varepsilon_t$$

- ▶ Example:

$$(1 - \Phi B^{12})Y_t = \varepsilon_t$$

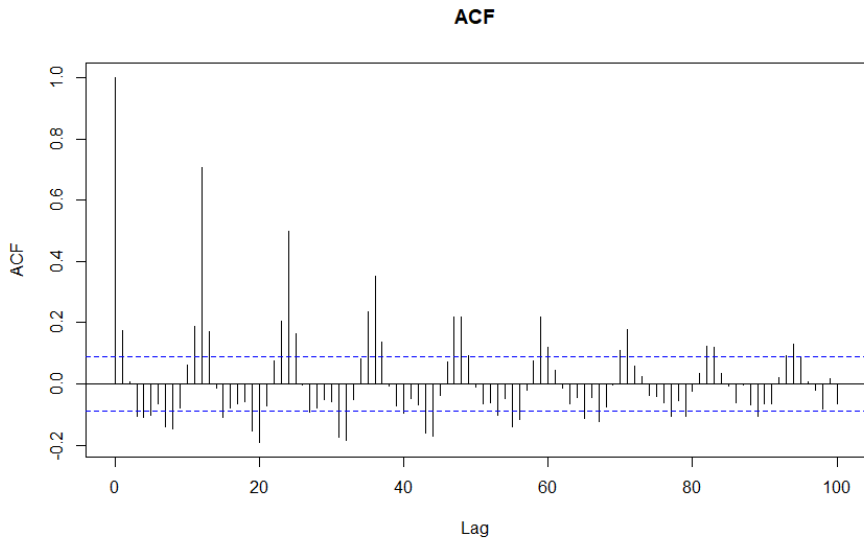
- ▶ Which can also be written:

$$Y_t = \Phi Y_{t-12} + \varepsilon_t$$

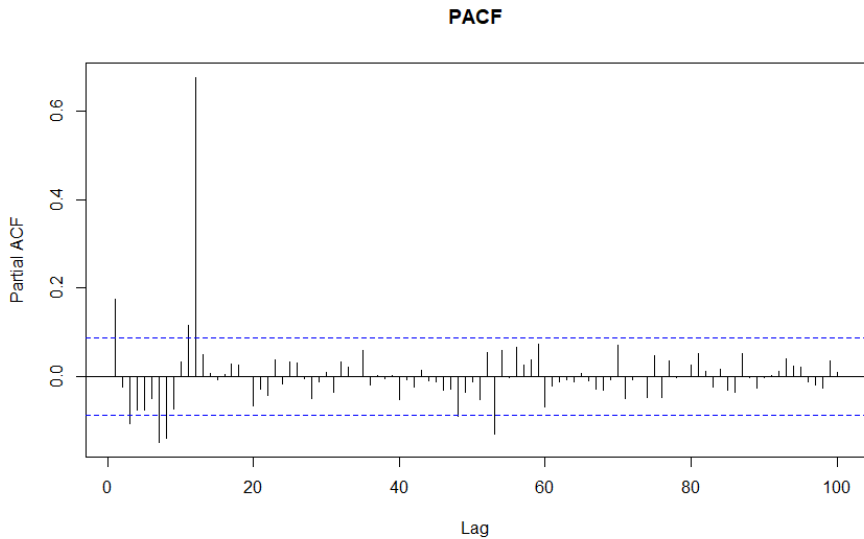
i.e. Y_t depend on Y_{t-12} , Y_{t-24} , ... (thereof the name)

- ▶ How would you think that the auto correlation function looks?

ACF and PACF of seasonal ARMA models



ACF and PACF of seasonal ARMA models



Prediction

- ▶ At time t we have observations $Y_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots$
- ▶ We want a prediction of Y_{t+k} , where $k \geq 1$

Prediction

- ▶ At time t we have observations $Y_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots$
- ▶ We want a prediction of Y_{t+k} , where $k \geq 1$
- ▶ Thus, we want the conditional expectation:

$$\hat{Y}_{t+k|t} = E[Y_{t+k} | Y_t, Y_{t-1}, Y_{t-2}, \dots]$$

Example – prediction in the $AR(1)$ model

- ▶ We write the model like $Y_{t+1} = \phi Y_t + \varepsilon_{t+1}$ (note the sign on ϕ)
- ▶ 1-step prediction:

$$\hat{Y}_{t+1|t} = E[Y_{t+1}|Y_t, Y_{t-1}, \dots]$$

Example – prediction in the $AR(1)$ model

- ▶ We write the model like $Y_{t+1} = \phi Y_t + \varepsilon_{t+1}$ (note the sign on ϕ)
- ▶ 1-step prediction:

$$\begin{aligned}\widehat{Y}_{t+1|t} &= E[Y_{t+1}|Y_t, Y_{t-1}, \dots] \\ &= E[\phi Y_t + \varepsilon_{t+1}|Y_t, Y_{t-1}, \dots]\end{aligned}$$

Example – prediction in the $AR(1)$ model

- ▶ We write the model like $Y_{t+1} = \phi Y_t + \varepsilon_{t+1}$ (note the sign on ϕ)
- ▶ 1-step prediction:

$$\begin{aligned}\widehat{Y}_{t+1|t} &= E[Y_{t+1}|Y_t, Y_{t-1}, \dots] \\ &= E[\phi Y_t + \varepsilon_{t+1}|Y_t, Y_{t-1}, \dots] \\ &= \phi Y_t + 0 = \phi Y_t\end{aligned}$$

Example – prediction in the $AR(1)$ model

- ▶ We write the model like $Y_{t+1} = \phi Y_t + \varepsilon_{t+1}$ (note the sign on ϕ)
- ▶ 1-step prediction:

$$\begin{aligned}\widehat{Y}_{t+1|t} &= E[Y_{t+1}|Y_t, Y_{t-1}, \dots] \\ &= E[\phi Y_t + \varepsilon_{t+1}|Y_t, Y_{t-1}, \dots] \\ &= \phi Y_t + 0 = \phi Y_t\end{aligned}$$

- ▶ 2-step prediction:

$$\widehat{Y}_{t+2|t} = E[Y_{t+2}|Y_t, Y_{t-1}, \dots]$$

Example – prediction in the $AR(1)$ model

- ▶ We write the model like $Y_{t+1} = \phi Y_t + \varepsilon_{t+1}$ (note the sign on ϕ)
- ▶ 1-step prediction:

$$\begin{aligned}\widehat{Y}_{t+1|t} &= E[Y_{t+1}|Y_t, Y_{t-1}, \dots] \\ &= E[\phi Y_t + \varepsilon_{t+1}|Y_t, Y_{t-1}, \dots] \\ &= \phi Y_t + 0 = \phi Y_t\end{aligned}$$

- ▶ 2-step prediction:

$$\begin{aligned}\widehat{Y}_{t+2|t} &= E[Y_{t+2}|Y_t, Y_{t-1}, \dots] \\ &= E[\phi Y_{t+1} + \varepsilon_{t+2}|Y_t, Y_{t-1}, \dots]\end{aligned}$$

Example – prediction in the $AR(1)$ model

► We write the model like $Y_{t+1} = \phi Y_t + \varepsilon_{t+1}$ (note the sign on ϕ)

► 1-step prediction:

$$\begin{aligned}\widehat{Y}_{t+1|t} &= E[Y_{t+1}|Y_t, Y_{t-1}, \dots] \\ &= E[\phi Y_t + \varepsilon_{t+1}|Y_t, Y_{t-1}, \dots] \\ &= \phi Y_t + 0 = \phi Y_t\end{aligned}$$

► 2-step prediction:

$$\begin{aligned}\widehat{Y}_{t+2|t} &= E[Y_{t+2}|Y_t, Y_{t-1}, \dots] \\ &= E[\phi Y_{t+1} + \varepsilon_{t+2}|Y_t, Y_{t-1}, \dots] \\ &= \phi \widehat{Y}_{t+1|t} + 0\end{aligned}$$

Example – prediction in the $AR(1)$ model

► We write the model like $Y_{t+1} = \phi Y_t + \varepsilon_{t+1}$ (note the sign on ϕ)

► 1-step prediction:

$$\begin{aligned}\widehat{Y}_{t+1|t} &= E[Y_{t+1}|Y_t, Y_{t-1}, \dots] \\ &= E[\phi Y_t + \varepsilon_{t+1}|Y_t, Y_{t-1}, \dots] \\ &= \phi Y_t + 0 = \phi Y_t\end{aligned}$$

► 2-step prediction:

$$\begin{aligned}\widehat{Y}_{t+2|t} &= E[Y_{t+2}|Y_t, Y_{t-1}, \dots] \\ &= E[\phi Y_{t+1} + \varepsilon_{t+2}|Y_t, Y_{t-1}, \dots] \\ &= \phi \widehat{Y}_{t+1|t} + 0 \\ &= \phi^2 Y_t\end{aligned}$$

Example – prediction in the $AR(1)$ model

► We write the model like $Y_{t+1} = \phi Y_t + \varepsilon_{t+1}$ (note the sign on ϕ)

► 1-step prediction:

$$\begin{aligned}\widehat{Y}_{t+1|t} &= E[Y_{t+1}|Y_t, Y_{t-1}, \dots] \\ &= E[\phi Y_t + \varepsilon_{t+1}|Y_t, Y_{t-1}, \dots] \\ &= \phi Y_t + 0 = \phi Y_t\end{aligned}$$

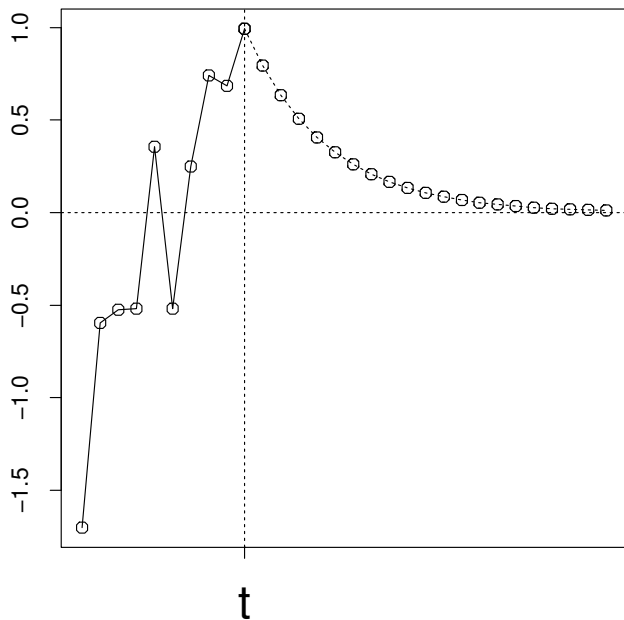
► 2-step prediction:

$$\begin{aligned}\widehat{Y}_{t+2|t} &= E[Y_{t+2}|Y_t, Y_{t-1}, \dots] \\ &= E[\phi Y_{t+1} + \varepsilon_{t+2}|Y_t, Y_{t-1}, \dots] \\ &= \phi \widehat{Y}_{t+1|t} + 0 \\ &= \phi^2 Y_t\end{aligned}$$

► k-step prediction:

$$\widehat{Y}_{t+k|t} = \phi^k Y_t$$

Example – prediction in $Y_t = 0.8 Y_{t-1} + \varepsilon_t$



Variance of prediction error for the $AR(1)$ -process

Prediction error:

$$e_{t+k|t} = Y_{t+k} - \hat{Y}_{t+k|t}$$

Variance of prediction error for the $AR(1)$ -process

Prediction error:

$$e_{t+k|t} = Y_{t+k} - \hat{Y}_{t+k|t} = Y_{t+k} - \phi^k Y_t$$

Variance of prediction error for the $AR(1)$ -process

Prediction error:

$$e_{t+k|t} = Y_{t+k} - \widehat{Y}_{t+k|t} = Y_{t+k} - \phi^k Y_t$$

Bring it on psi-form (MA-form):

$$Y_{t+k} = \phi Y_{t+k-1} + \varepsilon_{t+k}$$

Variance of prediction error for the $AR(1)$ -process

Prediction error:

$$e_{t+k|t} = Y_{t+k} - \widehat{Y}_{t+k|t} = Y_{t+k} - \phi^k Y_t$$

Bring it on psi-form (MA-form):

$$\begin{aligned} Y_{t+k} &= \phi Y_{t+k-1} + \varepsilon_{t+k} \\ &= \phi(\phi Y_{t+k-2} + \varepsilon_{t+k-1}) + \varepsilon_{t+k} \end{aligned}$$

Variance of prediction error for the $AR(1)$ -process

Prediction error:

$$e_{t+k|t} = Y_{t+k} - \widehat{Y}_{t+k|t} = Y_{t+k} - \phi^k Y_t$$

Bring it on psi-form (MA-form):

$$\begin{aligned} Y_{t+k} &= \phi Y_{t+k-1} + \varepsilon_{t+k} \\ &= \phi(\phi Y_{t+k-2} + \varepsilon_{t+k-1}) + \varepsilon_{t+k} \\ &= \phi^2 Y_{t+k-2} + \phi \varepsilon_{t+k-1} + \varepsilon_{t+k} \end{aligned}$$

Variance of prediction error for the $AR(1)$ -process

Prediction error:

$$e_{t+k|t} = Y_{t+k} - \widehat{Y}_{t+k|t} = Y_{t+k} - \phi^k Y_t$$

Bring it on psi-form (MA-form):

$$\begin{aligned} Y_{t+k} &= \phi Y_{t+k-1} + \varepsilon_{t+k} \\ &= \phi(\phi Y_{t+k-2} + \varepsilon_{t+k-1}) + \varepsilon_{t+k} \\ &= \phi^2 Y_{t+k-2} + \phi \varepsilon_{t+k-1} + \varepsilon_{t+k} \\ &= \phi^2(\phi Y_{t+k-3} + \varepsilon_{t+k-2}) + \phi \varepsilon_{t+k-1} + \varepsilon_{t+k} \end{aligned}$$

Variance of prediction error for the $AR(1)$ -process

Prediction error:

$$e_{t+k|t} = Y_{t+k} - \widehat{Y}_{t+k|t} = Y_{t+k} - \phi^k Y_t$$

Bring it on psi-form (MA-form):

$$\begin{aligned} Y_{t+k} &= \phi Y_{t+k-1} + \varepsilon_{t+k} \\ &= \phi(\phi Y_{t+k-2} + \varepsilon_{t+k-1}) + \varepsilon_{t+k} \\ &= \phi^2 Y_{t+k-2} + \phi \varepsilon_{t+k-1} + \varepsilon_{t+k} \\ &= \phi^2(\phi Y_{t+k-3} + \varepsilon_{t+k-2}) + \phi \varepsilon_{t+k-1} + \varepsilon_{t+k} \\ &= \phi^3 Y_{t+k-3} + \phi^2 \varepsilon_{t+k-2} + \phi \varepsilon_{t+k-1} + \varepsilon_{t+k} \end{aligned}$$

Variance of prediction error for the $AR(1)$ -process

Prediction error:

$$e_{t+k|t} = Y_{t+k} - \widehat{Y}_{t+k|t} = Y_{t+k} - \phi^k Y_t$$

Bring it on psi-form (MA-form):

$$\begin{aligned} Y_{t+k} &= \phi Y_{t+k-1} + \varepsilon_{t+k} \\ &= \phi(\phi Y_{t+k-2} + \varepsilon_{t+k-1}) + \varepsilon_{t+k} \\ &= \phi^2 Y_{t+k-2} + \phi \varepsilon_{t+k-1} + \varepsilon_{t+k} \\ &= \phi^2(\phi Y_{t+k-3} + \varepsilon_{t+k-2}) + \phi \varepsilon_{t+k-1} + \varepsilon_{t+k} \\ &= \phi^3 Y_{t+k-3} + \phi^2 \varepsilon_{t+k-2} + \phi \varepsilon_{t+k-1} + \varepsilon_{t+k} \\ &\vdots \\ &= \phi^k Y_t + \phi^{k-1} \varepsilon_{t+1} + \phi^{k-2} \varepsilon_{t+2} + \dots + \phi \varepsilon_{t+k-1} + \varepsilon_{t+k} \end{aligned}$$

Variance of prediction error for the $AR(1)$ -process

Variance of prediction error:

$$V[e_{t+k|t}] = V[\phi^{k-1}\varepsilon_{t+1} + \phi^{k-2}\varepsilon_{t+2} + \dots + \phi\varepsilon_{t+k-1} + \varepsilon_{t+k}]$$

Variance of prediction error for the $AR(1)$ -process

Variance of prediction error:

$$\begin{aligned} V[e_{t+k}|t] &= V[\phi^{k-1}\varepsilon_{t+1} + \phi^{k-2}\varepsilon_{t+2} + \dots + \phi\varepsilon_{t+k-1} + \varepsilon_{t+k}] \\ &= (\phi^{2(k-1)} + \phi^{2(k-2)} + \dots + \phi^2 + 1)\sigma_\varepsilon^2 \end{aligned}$$

Variance of prediction error for the $AR(1)$ -process

Variance of prediction error:

$$\begin{aligned} V[e_{t+k|t}] &= V[\phi^{k-1}\varepsilon_{t+1} + \phi^{k-2}\varepsilon_{t+2} + \dots + \phi\varepsilon_{t+k-1} + \varepsilon_{t+k}] \\ &= (\phi^{2(k-1)} + \phi^{2(k-2)} + \dots + \phi^2 + 1)\sigma_\varepsilon^2 \end{aligned}$$

$(1 - \alpha) \times 100\%$ prediction interval:

$$\hat{Y}_{t+k|t} \pm u_{\alpha/2} \sqrt{V[e_{t+k|t}]}$$

$u_{\alpha/2}$ is the $\alpha/2$ -quantile in the standard normal distribution

Estimation

- ▶ Assume that 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

Estimation

- ▶ Assume that 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)

Estimation in AR(2) model

- ▶ Observations: y_1, y_2, \dots, y_N

Estimation in AR(2) model

- ▶ Observations: y_1, y_2, \dots, y_N
- ▶ Model: $y_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} = \epsilon_t$

Estimation in AR(2) model

- ▶ Observations: y_1, y_2, \dots, y_N
- ▶ Model: $y_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} = \epsilon_t$

$$y_3 = \phi_1 y_2 + \phi_2 y_1 + e_3$$

$$y_4 = \phi_1 y_3 + \phi_2 y_2 + e_4$$

$$y_5 = \phi_1 y_4 + \phi_2 y_3 + e_5$$

$$\vdots$$

$$y_N = \phi_1 y_{N-1} + \phi_2 y_{N-2} + e_N$$

Estimation in AR(2) model

- ▶ Observations: y_1, y_2, \dots, y_N
- ▶ Model: $y_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} = \epsilon_t$

$$y_3 = \phi_1 y_2 + \phi_2 y_1 + e_3$$

$$y_4 = \phi_1 y_3 + \phi_2 y_2 + e_4$$

$$y_5 = \phi_1 y_4 + \phi_2 y_3 + e_5$$

$$\vdots$$

$$y_N = \phi_1 y_{N-1} + \phi_2 y_{N-2} + e_N$$

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

Estimation in AR(2) model

- ▶ Observations: y_1, y_2, \dots, y_N
- ▶ Model: $y_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} = \epsilon_t$

$$y_3 = \phi_1 y_2 + \phi_2 y_1 + e_3$$

$$y_4 = \phi_1 y_3 + \phi_2 y_2 + e_4$$

$$y_5 = \phi_1 y_4 + \phi_2 y_3 + e_5$$

$$\vdots$$

$$y_N = \phi_1 y_{N-1} + \phi_2 y_{N-2} + e_N$$

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

Or just:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$$

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:

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

With

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

► Asymptotically: $V(\hat{\boldsymbol{\theta}}) = \sigma_{\varepsilon}^2 (\mathbf{X}^T \mathbf{X})^{-1}$

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:

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

With

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

- ▶ Asymptotically: $V(\hat{\boldsymbol{\theta}}) = \sigma_{\varepsilon}^2 (\mathbf{X}^T \mathbf{X})^{-1}$
- ▶ How does it generalize to AR(p)-models?

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:

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

With

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

- ▶ Asymptotically: $V(\hat{\boldsymbol{\theta}}) = \sigma_{\varepsilon}^2 (\mathbf{X}^T \mathbf{X})^{-1}$
- ▶ How does it generalize to AR(p)-models?
- ▶ How about ARMA(p,q)-models?

Least squares for AR

```
# Test it by comparing
model <- list(ar=c(0.4))
set.seed(12)
sim(model, 10, nburnin=100)
set.seed(12)
x <- arima.sim(model, 100)

X <- lagdf(x, 0:3)
summary(lm(k0 ~ k1, X))
summary(lm(k0 ~ k1 + k2, X))
summary(lm(k0 ~ k1 + k2 + k3, X))
```

Maximum Likelihood estimates

- ▶ $ARMA(p, q)$ -process:

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

- ▶ Notation:

$$\boldsymbol{\theta}^T = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$$

$$\mathbf{Y}_t^T = (Y_t, Y_{t-1}, \dots, Y_1)$$

Maximum Likelihood estimates

- ▶ $ARMA(p, q)$ -process:

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

- ▶ Notation:

$$\boldsymbol{\theta}^T = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$$

$$\mathbf{Y}_t^T = (Y_t, Y_{t-1}, \dots, Y_1)$$

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

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

Maximum Likelihood estimates

- ▶ $ARMA(p, q)$ -process:

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

- ▶ Notation:

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

- ▶ 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

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
- ▶ $\{\varepsilon_t\}$ is a white noise process and therefore does not depend on anything

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
- ▶ $\{\varepsilon_t\}$ is a white noise process and therefore does not depend on anything
- ▶ Thus we know that the random variables $Y_N | \mathbf{Y}_{N-1}$ and \mathbf{Y}_{N-1} are independent,

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
- ▶ $\{\varepsilon_t\}$ is a white noise process and therefore does not depend on anything
- ▶ Thus we know that the random variables $Y_N | \mathbf{Y}_{N-1}$ and \mathbf{Y}_{N-1} are independent, hence:

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

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
- ▶ $\{\varepsilon_t\}$ is a white noise process and therefore does not depend on anything
- ▶ Thus we 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:

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

Evaluating the conditional likelihood function

- ▶ **Task:** Find the conditional 1-step densities, $f(Y_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}, \sigma_\varepsilon^2)$, given specified values of the parameters $\boldsymbol{\theta}$ and σ_ε^2

Evaluating the conditional likelihood function

- ▶ **Task:** Find the conditional 1-step densities, $f(Y_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}, \sigma_\varepsilon^2)$, 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 $\hat{Y}_{t|t-1}$

Evaluating the conditional likelihood function

- ▶ **Task:** Find the conditional 1-step densities, $f(Y_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}, \sigma_\varepsilon^2)$, 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

Evaluating the conditional likelihood function

- ▶ **Task:** Find the conditional 1-step densities, $f(Y_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}, \sigma_\varepsilon^2)$, 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:

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

Evaluating the conditional likelihood function

- ▶ **Task:** Find the conditional 1-step densities, $f(Y_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}, \sigma_\varepsilon^2)$, 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 $\hat{Y}_{t|t-1}$
- ▶ The prediction error $\varepsilon_t = Y_t - \hat{Y}_{t|t-1}$ has variance σ_ε^2
- ▶ We assume that the process is Gaussian:

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

- ▶ 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 $\hat{\theta}$ is a prediction error estimate since it is obtained by minimizing

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

ML-estimates

- ▶ The (conditional) ML-estimate $\hat{\boldsymbol{\theta}}$ 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):

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

ML-estimates

- ▶ The (conditional) ML-estimate $\hat{\boldsymbol{\theta}}$ 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):

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

- ▶ The estimate $\hat{\boldsymbol{\theta}}$ is asymptotically unbiased and efficient, and the variance-covariance matrix is approximately

$$2\hat{\sigma}_\varepsilon^2 \mathbf{H}^{-1}$$

where \mathbf{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:

$$\hat{Y}_{t+1|t} = -\phi_1 Y_t - \dots - \phi_p Y_{t-p+1} + \theta_1 \varepsilon_t + \dots + \theta_q \varepsilon_{t-q+1}$$

Finding the ML-estimates using the PE-method

- ▶ 1-step predictions:

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

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

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

Finding the ML-estimates using the PE-method

- ▶ 1-step predictions:

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

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

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

- ▶ 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.

Finding the ML-estimates using the PE-method

- ▶ 1-step predictions:

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

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

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

- ▶ 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