STATS 207: Time Series Analysis
Autumn 2020

Lectures 5-6: ARMA/ARIMA Modelling

Dr. Alon Kipnis

October 2nd 2020

Genera Info

- Homework assignment due Monday 10/5/2020.
- Office hours 11:20-12:20 after the lecture.

Outline

ACF of ARMA Models

Forecasting

Estimating **ARMA** parameters

Midway Summary

- **AR**(p) :

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_{t-p} x_p + w_t \qquad \text{or} \qquad \phi(B) x_t = w_t,$$

where $\phi(z)$ is the **autoregressive polynomial**

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p.$$

- **MA**(q):

$$x_t = w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}, \qquad \text{or} \qquad x_t = \theta(B) w_t,$$

where $\theta(z)$ is the **moving average polynomial**

$$\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^p.$$

- **ARMA**(p, q)

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_{t-p} x_p + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}$$

or

$$\phi(B) x_t = \theta(B) w_t.$$

Causality and MA Representation

- Definition: **ARMA**(p, q) model is called **causal** if

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B) w_t,$$

  where:

  - (w_t) is white noise,
  - $\sum_{j=0}^{\infty} |\psi_j| < \infty$,
  - $\psi_0 = 1$,
  - $\psi(B) \equiv \sum_{j=0}^{\infty} \psi_j B^j$.

- $\psi(B)$ gives an **infinite-order MA representation**.
- Example: The **AR**(1) process is causal iff $|\phi| < 1$. Equivalently, it is causal iff the magnitude of the root of $\phi(z) = 1 - \phi z$ is bigger than 1. More generally...

Causality of ARMA

- Property: An **ARMA**(p, q) model is causal iff $\phi(z) \neq 0$ for $|z| \leq 1$. In this case,

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \qquad |z| \leq 1.$$

"ARMA is causal if $\phi(z)$ has no roots inside the unit circle."

- Example 3.8:

$$x_t = 0.9x_{t-1} + w_t + 0.5w_{t-1}$$

$$\phi(z) = 1 - 0.9z$$
$$\theta(z) = 1 + 0.5z$$

The MA representation of x_t is obtained from $\phi(z)\psi(z) = \theta(z)$:

$$(1 - 0.9z)(1 + \psi_1 z + \ldots + \psi_j z^j + \ldots) = 1 + 0.5z$$
$$1 + (\psi_1 - 0.9)z + (\psi_2 - 0.9\psi_1)z^2 + \ldots + (\psi_j - 0.9\psi_{j-1})z^j + \ldots = 1 + 0.5z$$

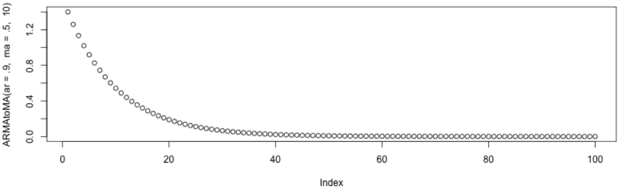
By matching coefficients:

$$x_t = w_t + 1.4 \sum_{j=1}^{\infty} 0.9^{j-1} w_{t-j}.$$

Using R (Example 3.8)

```
ARMAtoMA(ar = .9,  ma = .5,  10)   # first 10 psi-weights
```

```
1.4 1.26 1.134 1.0206 0.91854 0.826686 0.7440174 ...
```



Example: Textbook conditions for causality of **AR**(2) (Example 3.9):

$$\phi_1 + \phi_2 < 1, \quad \phi_2 - \phi_1 < 1, \quad \text{and} \quad |\phi_2| < 1.$$

No **simple** conditions for causality in the general case.

Example 3.11: AR with complex roots

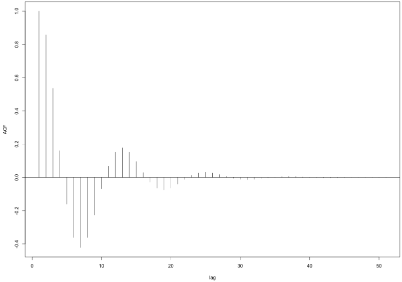
$$x_t = 1.5x_{t-1} - 0.75x_{t-2} + w_t, \qquad \sigma_w^2 = 1.$$

$$\phi(z) = 1 - 1.5z + 0.75z^2 = (1 - \sqrt{-1/3} - z)(1 + \sqrt{-1/3} - z).$$

```
z = c(1,-1.5,.75) # coefficients of the polynomial
polyroot(z) # print roots
```

```
1+0.57735026918963i 1-0.57735026918963i
```

```
ACF = ARMAacf(ar=c(1.5,-.75), ma=0, 50)
plot(ACF, type="h", xlab="lag")
abline(h=0)
```



Invertibility and AR Representation

- Definition: An **ARMA**(p, q) model is said to be **invertible** if (x_t) can be written as

$$\pi(B)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} = w_t,$$

where

  - $\pi(b) = \sum_{j=0}^{\infty} \pi_j B^j$
  - $\sum_{j=0}^{\infty} |\pi_j| < \infty$.
  - $\pi_0 = 1$

- Property: An **ARMA**(p, q) model is invertible iff $\theta(z) \neq 0$ for $|z| \leq 1$. In this case,

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \qquad |z| \leq 1.$$

"ARMA is invertible iff $\theta(z)$ has no roots inside the unit circle."

Example 3.8 (cont'd)

- Example (Example 3.8):

$$x_t = 0.9x_{t-1} + 0.5w_{t-1} + w_t.$$

- We have: $\phi(z) = 1 - 0.9z$, $\theta(z) = 1 + 0.5z$.
- Invertible representation:

$$\theta(z)\pi(z) = \phi(z)$$
$$(1 + 0.5z)(1 + \pi_1 z + \pi_2 z^2 + \dots) = 1 - 0.9z.$$

We get: $\pi_j = (-1)^j 1.4(0.5)^{j-1}$, $j \geq 1$, hence

$$x_t = 1.4 \sum_{j=1}^{\infty} (-0.5)^{j-1} x_{t-j} + w_t.$$

- Using R:

```
ARMAtoAR(ar = .9, ma = .5, 8) # first 8 pi-weights
```

```
-1.4 0.7 -0.35 0.175 -0.0875 0.04375 -0.021875 0.0109375
```

π and Prediction

- **Prediction is easy** for processes with AR representation

$$x_t = \sum_{j=1}^{\infty} a_j x_{t-j} + w_t, \qquad \sum_{j=0}^{\infty} |a_j| < \infty.$$

- Indeed, if (w_t) is iid noise,

$$\mathbb{E}\left[x_{t+1} | x_t, x_{t-1}, \ldots\right] = \sum_{j=1}^{\infty} a_j x_{t-j-1} = \sum_{j=0}^{\infty} a_{j+1} x_{t-j}.$$

- Property: Existence of such representation for **stationary ARMA processes** is equivalent to **invertibility**.

π and Prediction (cont'd)

- Property: Existence of such representation for **stationary ARMA processes** is equivalent to **invertibility**.

  Proof: Rewrite

  $$x_t - \sum_{j=1}^{\infty} a_j x_{t-j} = \pi(B) x_t = w_t,$$

  where $\pi_0 = 1$, $\pi_j = -a_j$, $j \geq 1$.

- **Consequence**: **Given** the invertibility relation

  $$\pi(B) x_t = w_t,$$

  we can predict using the **weights embedded** in π:

  $$x_{t+1}^t = -\sum_{j=0} \pi_{j+1} x_{t-j}$$

AR and MA Polynomials – Recap

Assume **causality** and **invertibility** of **ARMA**(p, q)

- At the **process** level:

$$w_t = \pi(B)x_t \qquad \text{and} \qquad \psi(B)w_t = x_t.$$

- At **operator** level

$$\pi(B)\psi(B) = I.$$

- At level of **power/Laurent series**

$$\pi(z)\psi(z) = 1, \quad |z| \leq 1.$$

ACF of ARMA

- Causal linear process representation

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$$

- **Autocovariance**

$$\gamma(h) = \sigma_w^2 \sum_{t=0}^{\infty} \psi_t \psi_{t+h}, \quad h \geq 0,$$

(sliding window/convolution interpretation)

- **Special case**: **MA**(q) process ($\psi_j = \theta_j$, for $j = 0, \ldots, h$)

$$\gamma(h) = \sigma_w^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h}, \quad 0 \leq h \leq q.$$

- **Very special case**: **MA**(1), $\sigma_w^2 = 1$,

$$\gamma(h) = \begin{cases} 1 + \theta^2 & h = 0, \\ \theta & h = \pm 1, \\ 0 & \text{otherwise.} \end{cases}$$

ACF of ARMA (cont'd)

- Linear **Difference Equation** for ACF

$$\gamma(h) = \sum_{i=1} \phi_i \gamma(h - i), \qquad h \geq \max(p, q + 1),$$

with **initial conditions**

$$\gamma(h) = \sum_{j=1}^{p} \phi_j \gamma(h - j) + \sigma_w^2 \sum_{j=h}^{q} \theta_j \psi_{j-h}, \quad 0 \leq h < \max(p, q + 1).$$

- See Chapter 3.2 for solving linear difference equations.

Example 3.11:

$$x_t = \phi x_{t-1} + \theta w_{t-1} + w_t, \quad |\phi| < 1.$$

The autocovariance satisfies

$$\gamma_x(h) = \phi \gamma_x(h-1), \quad h \geq 2,$$

with initial condition:

$$\gamma_x(1) = \phi \gamma_x(0) + \sigma_w^2 \theta, \quad \text{and} \quad \gamma_x(0) = \phi \gamma_x(-1) + \sigma_w^2 (1 + \theta\phi + \theta^2).$$

Solving for $\gamma_x(0)$ and $\gamma_x(1) = \gamma_x(-1)$:

$$\gamma_x(0) = \sigma_w^2 \frac{1 + 2\theta\phi + \theta^2}{1 - \phi^2}, \qquad \gamma_x(1) = \sigma_w^2 \frac{(1 + \theta\phi)(\phi + \theta)}{1 - \phi^2}.$$

From here,

$$\gamma_x(h) = \sigma_w^2 \frac{(1 + \theta\phi)(\phi + \theta)}{1 - \phi^2} \phi^{h-1}, \quad h \geq 1.$$

Note: ACF of **ARMA**(1,1) is of the **same form as the ACF of AR(1)**.
To distinguish the two, we introduce Partial ACF.

Partial ACF

- Definition: **Partial Correlation**

$$\begin{aligned}&\mathrm{ParCorr}(X, Y|Z_1, \ldots, Z_h)\\&= \mathrm{Corr}\left(X - \hat{X}(Z_1, \ldots, Z_h), Y - \hat{Y}(Z_1, \ldots, Z_h)\right)\end{aligned}$$

  Meaning: Linear relationship between X and Y that **cannot be attributed** to $Z_1, \ldots, Z_h$.

- Definition: The **partial ACF** (PACF) of a stationary process x_t

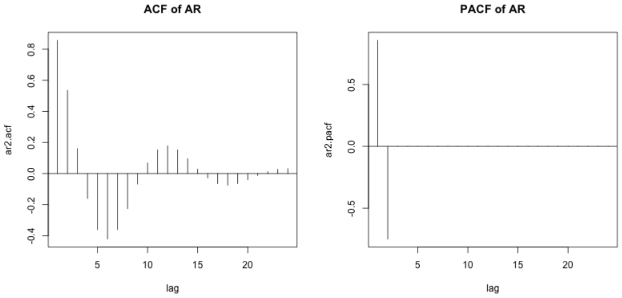
$$\phi_{hh} \equiv \mathrm{ParCorr}(x_{t+h}, x_t | x_{t+1}, \ldots, x_{t+h-1}),$$

  where $\hat{X}(Z_1, \ldots, Z_h)$ is the regression of X on $Z_1, \ldots, Z_h$.
  Meaning: Linear relationship between two lagged values that **cannot be attributed** to shorter lags.

Example – Partial ACF of AR

Example 3.16: $x_t = 1.5x_{t-1} - 0.75x_{t-2} + w_t$

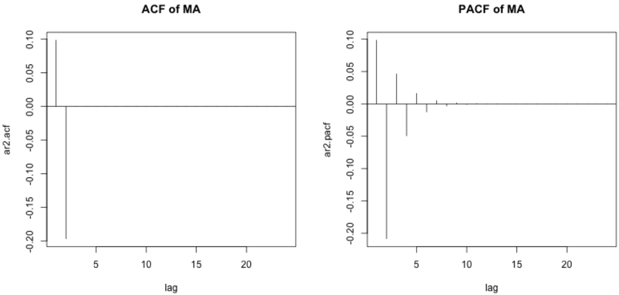
```
ar2.acf = ARMAacf(ar=c(1.5,-.75), ma=0, 24)[-1]
ar2.pacf = ARMAacf(ar=c(1.5,-.75), ma=0, 24, pacf=TRUE)
par(mfrow=c(1,2))
plot(ar2.acf, type="h", xlab="lag", main="ACF of AR")
abline(h=0)
plot(ar2.pacf, type="h", xlab="lag", main="PACF of AR")
abline(h=0)
```



Example – Partial ACF of MA

Example: $x_t = w_t + 1.5w_{t-1} - 0.75w_{t-2}$

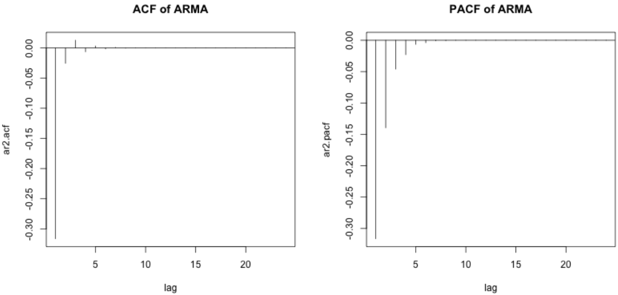
```
ar2.acf  =  ARMAacf(ar=0,ma=c(1.5,-.75),  24)[-1]
ar2.pacf  =  ARMAacf(ar=0, ma=c(1.5,-.75), 24, pacf=TRUE)
par(mfrow=c(1,2))
plot(ar2.acf, type="h", xlab="lag", main="ACF of MA")
abline(h=0)
plot(ar2.pacf, type="h", xlab="lag", main="PACF of MA")
abline(h=0)
```



Example – Partial ACF of ARMA

Example: $x_t = -0.5x_{t-1} + w_t + 1.5w_{t-1} - 0.75w_{t-2}$

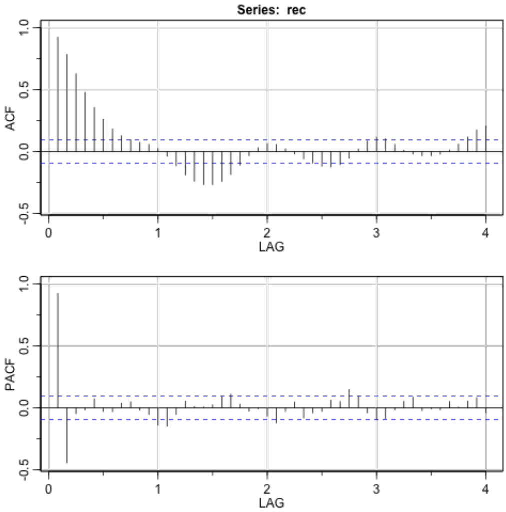
```
ar2.acf = ARMAacf(ar = -0.5, ma=c(1.5,-.75), 24)[-1]
ar2.pacf = ARMAacf(ar= -0.5, ma=c(1.5,-.75), 24, pacf=TRUE)
par(mfrow=c(1,2))
plot(ar2.acf, type="h", xlab="lag", main="ACF of ARMA")
abline(h=0)
plot(ar2.pacf, type="h", xlab="lag", main="PACF of ARMA")
abline(h=0)
```



Contrasting the PACF and ACF

	AR(p)	**MA**(q)	**ARMA**(p,q)
ACF	Decays	Cutoff q	Decays
PACF	Cutoff p	Decays	Decays

Example – ACF and PACF of Recruitment Data



- Looks like **AR**(2)!

Example – PACF and OLS

Fit an **AR**(2):

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t.$$

```
(regr = ar.ols(rec, order=2, demean=F, intercept=TRUE))  # regression
regr$asy.se.coef  # standard errors
```

(fitting using the OLS method, more details later)

```
Call:
ar.ols(x = rec, order.max = 2, demean = F, intercept = TRUE)

Coefficients:
      1        2
 1.3541  -0.4632

Intercept: 6.737 (1.111)

Order selected 2  sigma^2 estimated as  89.72
$x.mean
1.11059887598347
$ar
0.0417890066543078 0.0418794219793683
```

Forecasting I – Definitions

- The **best** m-step ahead predictor

$$x_{n+m}^{n,\dagger} = \mathbb{E}\left[x_{n+m} | x_1, \ldots, x_n\right]$$

minimizes the mean-squared error

$$P_{n+m}^{n} = \mathbb{E}\left[\left(x_{n+m} - x_{n+m}^{n}\right)^2\right].$$

- Definition: **Linear predictor**, m-step ahead, from data $x_1, \ldots, x_n$,

$$x_{n+m}^{n} \equiv x_{n+m}^{n}(\alpha_1, \ldots, \alpha_n) = \mu + \sum_{k=1}^{n} \alpha_k (x_k - \mu).$$

- **Simplification**: Assume $\mu = 0$.
- Definition: **Best** linear predictor:

$$x_{n+m}^{n,*} \equiv x_{n+m}^{n}(\alpha_{n+m,1}^{n,*}, \ldots, \alpha_{n+m,n}^{n,*}) = \sum_{k=1}^{n} \alpha_{n+m,k}^{n,*} x_k,$$

where

$$(\alpha_{n+m,1}^{n,*}, \alpha_{n+m,2}^{n,*}, \ldots, \alpha_{n+m,n}^{n,*}) \equiv \underset{(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n}{\operatorname{argmin}} P_{n+m}^{n}\left(x_{n+m}^{n}(\alpha_1, \ldots, \alpha_n)\right).$$

Forecasting II – Deriving the Best Linear Predictor

(1) Express mean-squared estimation error using variances (recall $\mathrm{Cov}(X, Y) = \mathbb{E}[XY]$ if $\mathbb{E}[X] = 0$):

$$\begin{aligned} P_{n+m}^n &= \mathrm{Var}(x_{n+m} - x_{n+m}^n) \\ &= \mathrm{Var}\left(x_{n+m}\right) - 2\mathrm{Cov}\left[x_{n+m}\sum_{k=1}^n \alpha_{n+m,k}^n x_k\right] + \mathrm{Var}\left(\sum_{k=1}^n \alpha_{n+k}^n x_k\right) \end{aligned}$$

(2) Evaluate covariances using matrices/vectors:

$$\mathrm{Var}\left(\sum_{k=1}^n \alpha_{n+k}^n x_k\right) = \boldsymbol{\alpha}'\boldsymbol{\Gamma}_n\boldsymbol{\alpha}, \qquad (\boldsymbol{\alpha})_k = \alpha_{n+m,k}^n, \qquad (\boldsymbol{\Gamma}_n)_{ij} = \mathbb{E}\left[x_i x_j\right]$$

$$\mathrm{Cov}\left[x_{n+m}\sum_{k=1}^n \alpha_{n+m,k}^n x_k\right] = \boldsymbol{\gamma}_{n+m}^{'n}\boldsymbol{\alpha}, \qquad (\boldsymbol{\gamma}_{n+m}^n)_k = \mathbb{E}\left[x_{n+m}x_k\right].$$

(3) LS solution:

$$\boldsymbol{\alpha}_{n+m}^{n,*} \equiv \boldsymbol{\alpha}^* \equiv \mathrm{argmin}_{\boldsymbol{\alpha}}\left(\boldsymbol{\alpha}'\boldsymbol{\Gamma}_n\boldsymbol{\alpha} - 2\boldsymbol{\gamma}_{n+m}^{'n}\boldsymbol{\alpha}\right)$$

$$\boldsymbol{\Gamma}_n\boldsymbol{\alpha}_{n+m}^{n,*} = \boldsymbol{\gamma}_{n+m}^n.$$

Forecasting II' - Alternative Derivation

(3') Alternative derivation:

- Property 3.3 ("orthogonality principle") x_{n+m}^n is the best linear predictor based on $x_1, \ldots, x_n$ iff

$$\mathbb{E}\left[(x_{n+m} - x_{n+m}^n)x_k\right] = 0, \quad k = 1, \ldots, n$$

(proof by the Projection Theorem in B.1)

- Consequently, $\alpha_{n+m,1}^{n,*}, \ldots, \alpha_{n+m,n}^{n,*}$ must satisfy

$$\mathbb{E}\left[x_{n+m}x_k\right] = \mathbb{E}\left[x_{n+m}^n x_k\right]$$
$$(\boldsymbol{\gamma}_{n+m}^n)_k = \sum_{j=1}^{n} \alpha_{n+m,j}^{n,*} \mathbb{E}\left[x_j x_k\right]$$

for all $k = 1, \ldots, n$.

- In vector notation:

$$\boldsymbol{\gamma}_{n+m}^n = \boldsymbol{\Gamma}_n \boldsymbol{\alpha}_{n+m}^{n,*}.$$

- If x_t is stationary

$$(\mathbf{\Gamma}_n)_{ij} = \mathbb{E}\left[x_i x_j\right] = \gamma_x(i-j), \qquad \text{(Toeplitz, non-negative definite)}$$

$$(\boldsymbol{\gamma}^n_{n+m})_k = \mathbb{E}\left[x_{n+m}x_k\right] = \gamma_x(n+m-k).$$

- Example: One-step ahead prediction

$$\mathbf{\Gamma}_n \boldsymbol{\alpha}^{n,*}_{n+1} = \boldsymbol{\gamma}^n_{n+1}.$$

Prediction error satisfies

$$P^n_{n+1} \equiv \mathbb{E}\left[(x_{n+1} - x^1_{n+1})^2\right] = \gamma_x(0) - \boldsymbol{\gamma}'^{n}_{n+1}\mathbf{\Gamma}_n\boldsymbol{\gamma}^n_{n+1}$$

Proof:

$$\begin{aligned} P^n_{n+1} &= \mathbb{E}\left[(x_{n+1} - \boldsymbol{\alpha}'^{n,*}_{n+1}x)^2\right] = \mathbb{E}\left[(x_{n+1} - \boldsymbol{\gamma}'^{n}_{n+1}\mathbf{\Gamma}^{-1}_n x)^2\right] \\ &= \mathbb{E}\left[x^2_{n+1} - 2\boldsymbol{\gamma}'^{n}_{n+1}\mathbf{\Gamma}^{-1}_n x x_{n+1} + \boldsymbol{\gamma}'^{n}_{n+1}\mathbf{\Gamma}^{-1}_n x x' \mathbf{\Gamma}^{-1}_n \boldsymbol{\gamma}^n_{n+1}\right] \\ &= \gamma(0) - 2\boldsymbol{\gamma}'^{n}_{n+1}\mathbf{\Gamma}^{-1}_n\boldsymbol{\gamma}^n_{n+1} + \boldsymbol{\gamma}'^{n}_{n+1}\mathbf{\Gamma}^{-1}_n\mathbf{\Gamma}_n\mathbf{\Gamma}^{-1}_n\boldsymbol{\gamma}^n_{n+1} \\ &= \gamma(0) - \boldsymbol{\gamma}'^{n}_{n+1}\mathbf{\Gamma}^{-1}_n\boldsymbol{\gamma}^n_{n+1}. \end{aligned}$$

- The **Durbin-Levinson Algorithm** (Property 3.2) solves $\boldsymbol{\alpha}^{n,*}_{n+1}$ and P^n_{n+1} recursively.

One-Step Ahead Forecasting for AR(p)

- Example 3.19: One-step ahead forecasting with causal **AR**(2) $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$:
  - Based on $x_1, x_2, \ldots$

$$\boldsymbol{\alpha}_{n+1,n}^{n,*} = \boldsymbol{\Gamma}_n^{-1} \boldsymbol{\gamma}_{n+1}^{n},$$

$$\alpha_{n+1,n}^{n,*} = \phi_1, \qquad \alpha_{n+1,n-1}^{n,*} = \phi_2, \qquad \alpha_{n+1,k}^{n,*} = 0, \quad k \geq n-2.$$

- Obvious generalization to **AR**(p).
- Asymptotic relation to π weights

$$\lim_{n \to \infty} \alpha_{n+1,n-k}^{n,*} = -\pi_k, \qquad k = 1, 2, \ldots, n$$

m-step Ahead Forecasting

- m-step ahead prediction error satisfies

$$P_{n+m}^n \equiv \mathbb{E}\left[(x_{n+m} - x_{n+m}^n)^2\right] = \gamma(0) - \boldsymbol{\gamma}_{n+m}^{\prime n} \boldsymbol{\Gamma}_n \boldsymbol{\gamma}_{n+m}^n$$

- **Prediction intervals**:

$$x_{n+m}^n \pm c_{\alpha/2}\sqrt{P_{n+m}^n}, \qquad \Pr(|w_t| \geq c_{\alpha/2}) \leq \alpha$$

  (need adjustment for multiple testing to get PI for more than one time period).

- Example: For **AR**(1) we have

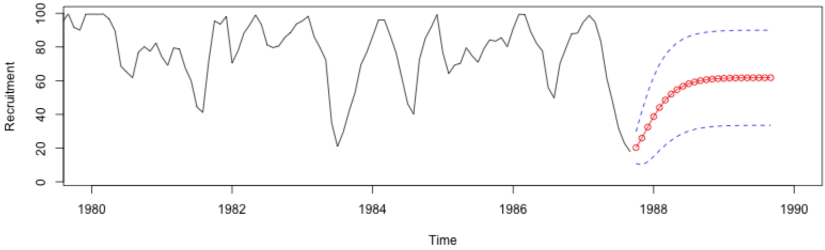
$$P_{n+1}^n = \sigma_w^2.$$

- Example: Long-range prediction

$$\lim_{m\to\infty} P_{n+m}^n = \gamma_x(0).$$

Forecasting – Example 3.26

```
regr = ar.ols(rec, order=2, demean=FALSE, intercept=TRUE)
fore = predict(regr, n.ahead=24)
ts.plot(rec, fore$pred, col=1:2, xlim=c(1980,1990), ylab="Recruitment")
lines(fore$pred, type="p", col=2)
lines(fore$pred+fore$se, lty="dashed", col=4)
lines(fore$pred-fore$se, lty="dashed", col=4)
```



Estimating ARMA parameters

Estimation – Motivation

- For an **ARMA**(p, q) model with **all parameters given**, we know:
  - Theoretical properties (invertibility, causality).
  - Description of second moments (ACF and PACF).
  - Predict future observations.

- Next:
  - **Estimate ARMA(p, q) parameters** given p and q under invertibility and causality assumptions.
  - Later: How to find p and q?
  - Later: Is **ARMA** even a good model?

Parameter Estimation in AR(p)

Three techniques:

1. Method of moments (MoM) and Yule-Walker estimates
2. Conditional Least Squares (C-LS)
3. Maximum Likelihood (ML)

MoM I: Estimating the Mean

- Use

$$\hat{\mu} = \bar{x} \equiv \frac{1}{n} \sum_{j=1}^{n} x_j.$$

- Theorem A.5: If x_t is a linear process and $\sum_j \psi_j \neq 0$, then

$$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} \mathcal{N}(0, V), \quad V = \sum_{h=-\infty}^{\infty} \gamma_x(h) = \sigma_w^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2.$$

- Henceforth assume $\mu = 0$.

MoM II: Estimating AR(p) Parameters Yule-Walker Equations

- Definition: The **Yule-Walker equations** are

$$\gamma(h) = \phi_1\gamma(h-1) + \ldots + \phi_p\gamma(h-p), \qquad h = 1, 2, \ldots, p,$$
$$\sigma_w^2 = \gamma(0) - \phi_1\gamma(1) - \ldots - \phi_p\gamma(p).$$

- In matrix notation

$$\boldsymbol{\Gamma}_p\boldsymbol{\phi} = \boldsymbol{\gamma}_p, \quad \sigma_w^2 = \gamma(0) - \boldsymbol{\phi}'\boldsymbol{\gamma}_p,$$

$$\boldsymbol{\Gamma}_p = \{\gamma(k-j)\}_{j,k=1}^p, \qquad \boldsymbol{\phi} = (\phi_1, \ldots, \phi_p), \qquad \boldsymbol{\gamma}_p = (\gamma(1), \ldots, \gamma(p))'$$

- **Method of moments** estimation (Yule-Walker estimators): Solve

$$\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\Gamma}}_p^{-1}\hat{\boldsymbol{\gamma}}_p, \qquad \hat{\sigma}_w^2 = \hat{\gamma}(0) - \hat{\boldsymbol{\gamma}}_p'\hat{\boldsymbol{\Gamma}}_p^{-1}\hat{\boldsymbol{\gamma}}_p.$$

where

$$\hat{\boldsymbol{\Gamma}}_p = \{\hat{\gamma}(k-j)\}_{j,k=1}^p, \qquad \hat{\boldsymbol{\gamma}}_p = (\hat{\gamma}(1), \ldots, \hat{\gamma}(p))'.$$

- We can calculate $\hat{\phi}$ without inverting $\hat{\boldsymbol{\Gamma}}_p$ using the **Durbin-Levinson algorithm.**

Yule-Walker Estimators – Large Sample Result

- Property 3.8: For a causal **AR**(p) process, as $n \to \infty$,

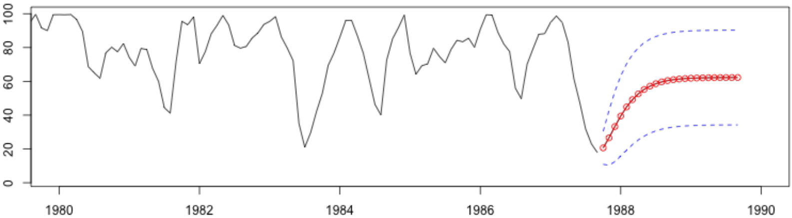
$$\sqrt{n}\left(\hat{\phi} - \phi\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_w^2 \boldsymbol{\Gamma}_p^{-1}\right), \qquad \hat{\sigma}_w^2 \xrightarrow{P} \sigma_w^2.$$

YW Estimation of the Recruitment Series

```
rec.yw = ar.yw(rec, order=2)
rec.yw$x.mean   # = 62.26278 (mean estimate)
rec.yw$ar       # = 1.3315874, -.4445447  (parameter estimates)
sqrt(diag(rec.yw$asy.var.coef))  # = .04222637, .04222637  (standard errors
rec.yw$var.pred  # = 94.79912 (error variance estimate)
```

Predicting using estimated **AR**(2) parameters:

```
rec.pr = predict(rec.yw, n.ahead=24)
U = rec.pr$pred + rec.pr$se
L = rec.pr$pred - rec.pr$se
minx = min(rec,L); maxx = max(rec,U)
ts.plot(rec, rec.pr$pred, xlim=c(1980,1990), ylim=c(minx,maxx))
lines(rec.pr$pred, col="red", type="o")
lines(U, col="blue", lty="dashed")
lines(L, col="blue", lty="dashed")
```



MoM III: Method of Moments Estimation for MA(1)

Example 3.29:

- Consider

$$x_t = w_t + \theta w_{t-1}, \quad |\theta| < 1.$$

- Recall $\gamma_x(0) = \sigma_w^2(1+\theta^2)$ and $\gamma_x(1) = \sigma_w^2\theta$. An estimator $\hat{\theta}$ for θ is obtained from

$$\hat{\rho}_x(1) = \frac{\hat{\gamma}_x(1)}{\hat{\gamma}_x(0)} = \frac{\hat{\theta}}{1+\hat{\theta}^2}.$$

  Two solutions exist. **We pick the invertible one**.

- If $|\hat{\rho}_x(1)| \le 1/2$,

$$\hat{\theta} = \frac{1-\sqrt{1-4\hat{\rho}(1)^2}}{2\hat{\rho}(1)}$$

  and (using the Delta Method)

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1+\theta^2+4\theta^4+\theta^6+\theta^8}{(1-\theta^2)^2}\right).$$

- If $|\hat{\rho}_x(1)| > 1/2$, **a real solution does not exist**.
- **Later**: The ML estimator has a better large sample behavior.

ML Estimation

- Definition: **Maximum Likelihood (ML) Estimator**
  1. Write likelihood function in terms of model parameters

     $$L(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_w^2) = L(x_1, \ldots, x_n; \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_w^2)$$

  2. Solve

     $$\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \hat{\theta}_q, \hat{\sigma}_w^2 \text{ argmin } L(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_w^2).$$

Example: (conditional) ML Estimation of AR(1)

$$x_t = \phi x_{t-1} + w_t, \qquad |\phi| < 1, \quad w_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_w^2).$$

- Likelihood function (conditioned on x_1)

$$L(\phi, \sigma_w^2) = f(x_1, \ldots, x_n; \phi, \sigma_w^2) = \prod_{j=2}^{n} f(x_j | x_{j-1}; \phi, \sigma_w^2),$$

where

$$f(x_j | x_{j-1}; \phi, \sigma_w^2) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left\{ -\left(x_j - \phi x_{j-1}\right)^2 / \sigma_w^2 \right\}, \qquad j = 2, \ldots, n$$

- ML estimate

$$\hat{\phi} = \text{argmin}_{\phi<1} S_x(\phi; x_1), \qquad \hat{\sigma}_w^2 = S(\hat{\phi})/(n-1),$$

where

$$S_c(\phi; x_1) \equiv \sum_{j=2}^{n} (x_j - \phi x_{j-1})^2.$$

Conditional LS Estimate

- Definition: **Conditional Least Squares Estimator for ARMA**(p, q)
  1. Set $\hat{w}_t = 0$ for $t \leq p$.
  2. Write

$$\hat{w}_t = x_t - \phi_1 x_{t-1} - \ldots - \phi_p x_{t-p} - \theta_1 \hat{w}_{t-1} - \ldots - \theta_q \hat{w}_{t-q}, \quad t = p+1, \ldots, n.$$

$$S_c(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q) \equiv \sum_{t=p+1}^{n} \hat{w}_t^2.$$

  3. Solve linear regression

$$\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q = \text{argmin}\, S_c(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q).$$

  4. Set

$$\hat{\sigma}_w^2 \equiv \frac{1}{n-p-1} S_c(\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q)$$

- For **AR**(1), C-LS is almost identical to ML.

Example: Fitting AR(2) to Recruitment

Example 3.31

```
rec.mle = ar.mle(rec, order=2) rec.mle$x.mean
rec.mle$ar
sqrt(diag(rec.mle$asy.var.coef))
rec.mle$var.pred
```

```
62.2615261054892
1.35128085590502 -0.461273619173842
0.0409915944162095 0.0409915944162095
89.3359654276631
```

Compare with YW equations in Example 3.28:

```
rec.yw = ar.yw(rec, order=2)
rec.yw$x.mean
rec.yw$ar
sqrt(diag(rec.yw$asy.var.coef))
rec.yw$var.pred
```

```
62.2627816777042
1.33158738866791 -0.444544697634474
0.0422263743755033 0.0422263743755033
94.7991188417802
```

Fitting AR(2) to Recruitment (cont'd)

Compare with C-LS in Example 3.18:

```
fit <- ar.ols(rec , order=2, demean=TRUE , intercept=TRUE)
fit$x.mean
fit$ar
fit$asy.se.coef$ar
fit$var.pred
```

```
62.2627816777042
1.35406847266143 -0.46317843167489
0.041789006654309 0.0418794219793695
89.71705
```

Large Sample Distribution

Property 3.10: Consider a causal invertible **ARMA**(p, q) process. Denote

$$\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_p).$$

Under appropriate conditions, C-LS estimate and ML estimate, all provide **asymptotically optimal** estimates $\hat{\beta}$ of β and $\hat{\sigma}_w^2$ of σ_w^2. In particular

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_w^2 \mathbf{\Gamma}_{p,q}^{-1}\right), \qquad \mathbf{\Gamma}_{p,q} \equiv \begin{pmatrix} \mathbf{\Gamma}_{\phi\phi} & \mathbf{\Gamma}_{\phi\theta} \\ \mathbf{\Gamma}_{\theta\phi} & \mathbf{\Gamma}_{\theta\theta} \end{pmatrix},$$

Where:

- $\mathbf{\Gamma}_{\phi\phi} \equiv \{\gamma_x(i-j)\}_{i,j=1}^{p}$ where $\gamma_x(h)$ is the autocovariance of **AR**(p) $\phi(B)x_t = w_t$
- $\mathbf{\Gamma}_{\theta\theta} \equiv \{\gamma_y(i-j)\}_{i,j=1}^{q}$ where $\gamma_y(h)$ is the autocovariance of the **MA**(q) $y_t = \theta(B)w_t$.
- $\mathbf{\Gamma}_{\phi\theta} \equiv \{\gamma_{xy}(i-j)\}_{i,j=1}^{p,q}$ where $\gamma_{xy}(h)$ is the crosscovariance of x_t and y_t.
- $\mathbf{\Gamma}_{\theta\phi} = \mathbf{\Gamma}'_{\phi\theta}$.

Specific Asymptotic Distributions

- **AR**(1):

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} \mathcal{N}(0, 1 - \phi^2).$$

- **AR**(2):

$$\sqrt{n}\left(\begin{pmatrix}\hat{\phi}_1\\ \hat{\phi}_2\end{pmatrix} - \begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix}1-\phi_2^2 & -\phi_1(1+\phi_2)\\ -\phi_1(1+\phi_2) & 1-\phi_2^2\end{pmatrix}\right)$$

- **MA**(1):

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, 1 - \theta^2)$$

- **ARMA**(1, 1):

$$\sqrt{n}\left(\begin{pmatrix}\hat{\phi}\\ \hat{\theta}\end{pmatrix} - \begin{pmatrix}\phi\\ \theta\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix}(1-\phi_2^2)^{-1} & (1+\phi\theta)^{-1}\\ (1+\phi\theta)^{-1} & (1-\theta^2)^{-1}\end{pmatrix}^{-1}\right)$$

Note: Variance inflation due to fitting **AR**(2) to **AR**(1).

Recap

- **ARMA**(p, q) is a useful model for **stationary processes**.
- We can express **ACF** and **optimal linear** m**-step forecast** in terms of model's parameters.
- We can **fit ARMA**(p, q) **to data** using several techniques. Leading to asymptotically normal estimators.

Next:

- Extensions (ARIMA, SARIMA)
- How to select p and q?
- Is **ARMA** a good model for my data?