# Time Series Analysis

## Lecture 5: ARIMA models-2
## Estimation, PACF, Forecasting

**Tohid Ardeshiri**

Linköping University
Division of Statistics and Machine Learning

September 16, 2019

# Maximum likelihood estimation: reminder

**Time series data are NOT independent**

- Likelihood of $\alpha$ given observations $x_1, \cdots, x_t$ is

$$L(\alpha) = f(x_1, \cdots, x_t | \alpha)$$

- Maximum likelihood: Optimal $\alpha$

$$\widehat{\alpha} = \arg\max_{\alpha} L(\alpha)$$

- Dependent data (time series): chain rule

$$L(\alpha) = f(x_1|\alpha) f(x_2|\alpha, x_1) f(x_3|\alpha, x_2, x_1)...$$

- Negative log-likelihood $l(\alpha) = -\sum_i \log(f(x_i|\alpha, x_{i-1}, \cdots))$
- Maximum likelihood: Optimal $\alpha$

$$\max_{\alpha} L(\alpha) = \min_{\alpha} l(\alpha)$$

# Maximum likelihood estimation: reminder

- Normal distributions: if $x_i \sim N(\mu, \sigma^2)$ , iid.

$$L(\theta) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{\sum_i (x_i - \mu)^2}{2\sigma^2}}$$

- Maximum likelihood

$$\widehat{\mu} = \bar{x}$$
$$\widehat{\sigma}^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2$$

# ML for AR(1) **Whiteboard**

- For ARMA models, assume normality of $w_t$!
- Negative log-likelihood

$$l(\mu, \phi, \sigma_w^2) = \frac{S(\mu, \phi)}{2\sigma_w^2} + \frac{n}{2}\log(2\pi\sigma_w^2) - \frac{1}{2}\log(1-\phi^2)$$

$$S(\mu, \phi) = (1-\phi^2)(x_1-\mu)^2 + \sum_{t=2}^{n}[(x_t-\mu) - \phi(x_{t-1}-\mu)]^2$$

- How to find optimum?
  - For $\sigma^2$ explicit

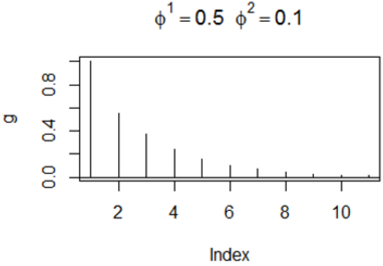
$$\widehat{\sigma}_w^2 = \frac{1}{n} S(\widehat{\mu}, \widehat{\phi})$$

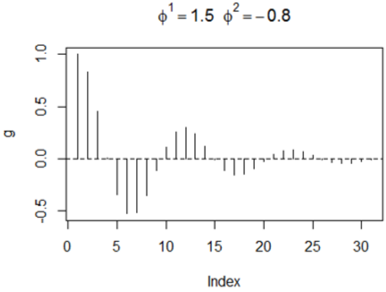
  - Otherwise numerical optimization (unconstrained optimization)

# ARMA

- **Autoregressive moving average ARMA($p, q$)**
  $x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}$
  - $\phi_p \neq 0, \theta_q \neq 0$
  - Is stationary
  - $Ex_t = 0$
- ACF for AR(1), MA(1), MA(2)
  $\rightarrow$ how to compute ACF for a general ARMA?

# ACF for AR(2)


$\phi^1 = 0.5$ $\phi^2 = 0.1$
g
0.0
0.4
0.8
2
4
6
8
10
Index


$\phi^1 = 1.5$ $\phi^2 = -0.8$
g
-0.5
0.0
0.5
1.0
0
5
10
15
20
25
30
Index

# ACF for AR(p), MA(p)

- AR(p): using difference equations
- MA(q): using difference equations

$$\rho(h) = \begin{cases} \frac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{1+\theta^2+\ldots+\theta_q^2} & 0 \le h \le q \\ 0 & h > q \end{cases}$$

# ACF for ARMA(p,q)

- ARMA(p,q):

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

- Causal ARMA: $x_t = \phi^{-1}(B)\theta(B)w_t = \psi(B)w_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$
  - How to find $\psi_j$ in practice? Coefficient matching
- Theorem: ACF of ARMA(p,q) can be found by solving general homogeneous equations:

$$\gamma(h) - \phi_1\gamma(h-1) - \ldots - \phi_p\gamma(h-p) = 0, \qquad h \geq max(p, q+1)$$

  - Initial conditions

$$\gamma(h) - \phi_1\gamma(h-1) - \ldots - \phi_p\gamma(h-p) = \sigma_w^2 \sum_{j=h}^{q} \theta_j \psi_{j-h}, 0 \leq h < \max(p, q+1)$$

# ACF for ARMA(1,1)

- Show for ARMA(1,1)

$$\rho(h) = \frac{(1+\theta\phi)(\phi+\theta)}{1+2\theta\phi+\theta^2}\phi^{h-1}, h \geq 1$$

- Note: pattern similar to AR(1) $\rightarrow$ hard to differentiate
- Note: ACF is 0 for $h > q$ from MA(q) $\rightarrow$ MA(q) is identifiable from ACF
- How to differentiate between AR(p)? ARMA(p)?

# Partial correlation

**A Gaussian intuition:**

- Conditional density: $f(x, y|z) = \frac{f(x,y,z)}{f(z)}$
- if $x$, $y$ and $z$ were jointly normal then

$$f(x, y|z) = N\left(\begin{bmatrix} x \\ y \end{bmatrix}; \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

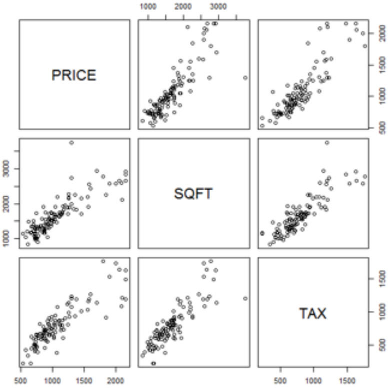
- Also,

$$\rho_{xy|z} = \frac{\text{cov}(x, y|z)}{\sqrt{\text{var}(x|z)\,\text{var}(y|z)}} = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}$$

- **What if $\Sigma_{12} = 0$ ?**

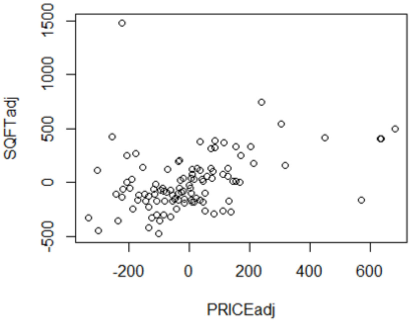
# Partial autocorrelation

- Example: Albuquerque home prices
  - What if we remove information stored in TAX from PRICE and SQFT?



# Partial autocorrelation

- $\hat{y} = \hat{\alpha}_0 + \hat{\alpha}_1 z$
- $\hat{x} = \hat{\beta}_0 + \hat{\beta}_1 z$
- $x' = x - \hat{x}$
- $y' = y - \hat{y}$
- **Partial autocorrelation**




  - If we know $\alpha$, $\beta$ and $z$, we can reduce connection between x and y

$>$ corr(cbind(PRICEadj,SQFTadj))
[1] 0.3675204

# PACF

- Partial autocorrelation function (PACF) for a stationary process

$$\phi_{11} = \text{corr}(x_{t+1}, x_t)$$

$$\phi_{hh} = \text{corr}(x'_{t+h}, x''_t), \quad h > 1$$

  - where $x'_{t+h} = x_{t+h} - \sum_{j=1}^{h-1} \hat{\beta}_j x_{t+h-j}$
  - and $x''_t = x_t - \sum_{j=1}^{h-1} \hat{\beta}_j x_{t+j}$
  - Note: coefficients in $x''_{t+h}$ and $x'_{t+h}$ are the same (stationarity)
- Example: AR(1) $\phi_{11} = \phi, \phi_{22} = 0$

# PACF for AR(p)

$$x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + w_t$$

- It can be shown:
  - $\phi_{pp} = \phi_p$
  - $\hat{\beta}_1 = \phi_1, ...\hat{\beta}_p = \phi_p, \hat{\beta}_{p+1} = 0, ..., \hat{\beta}_h = 0$ for $h > p$

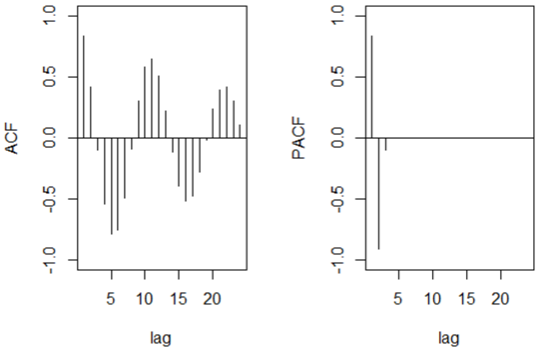
- It means

$$\phi_{hh} = \text{cov}(x_{t+h} - \sum_{j=1}^{p} \phi_j x_{t+h-j}, x_{-} \sum_{j=1}^{p} \phi_j x_{t+j})$$

$$= \text{cov}(w_{t+h}, x_t - \sum_{j=1}^{p} \phi_j x_{t+j}) = 0, \quad \text{when h} > p$$

# PACF for AR(p)

- Example: AR(3) $\phi_1 = 1.5$, $\phi_2 = -0.75, \phi_3 = -0.1$

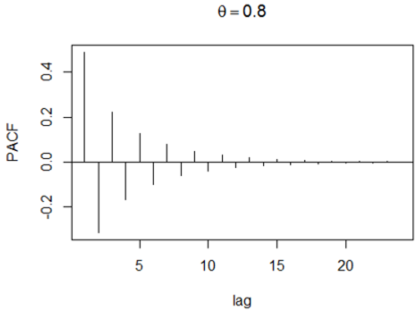


# PACF for MA(1)

- If invertible,

$$\phi_{hh} = -\frac{(-\theta)^h(1-\theta^2)}{1-\theta^{2h+2}}, \quad h \geq 1$$

Decreases exponentially with h



# ACF and PACF

| | AR($p$) | MA($q$) | ARMA($p, q$) |
|---|---|---|---|
| ACF | Tails off | Cuts off after lag $q$ | Tails off |
| PACF | Cuts off after lag $p$ | Tails off | Tails off |

How to differentiate between ARMA(p,q)?

# Empirical ACF (EACF)

Idea:

- ARMA(p,q): $x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + \sum_{j=1}^{q} \theta_j w_{t-j} + w_t$
- If we can estimate $\phi_j \rightarrow x'_t = x_t - \sum_{j=1}^{p} \phi_j x_{t-j}$ is linear function in $w_t, \ldots, w_{t-q}$
- If we run regression $x'_t$ against $w_t \ldots w_{t-j}$:
  - Residuals are white noise, $j \geq q \rightarrow$ ACFs not significant
    - ⋆ Some of the coefficients will be 0
  - Residuals are not white noise, $j < q \rightarrow$ ACFs significant
  - Note: $w_t s$ substituted by lagged residuals from a series of regressions
- If $x'_t = x_t - \sum_{j=1}^{k} \phi_j x_{t-j}, k < p \rightarrow$ white noise will never be achieved $\rightarrow$ ACFs are not zero

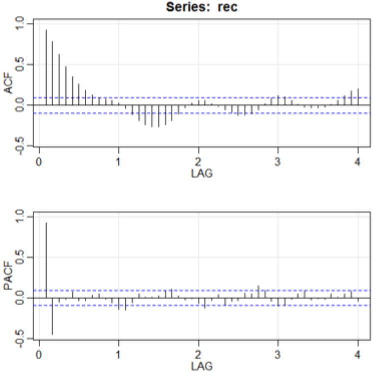
# Empirical ACF (EACF)

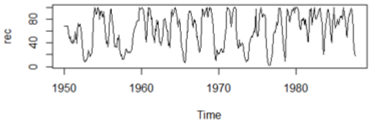
- $k > p$ General result: ACFs are 0 for $j > q + (k - p)$
  - Example: ARMA(0,1)
- General conclusion for AR,MA =(k,j):
  - This is theoretical one! $\rightarrow$ not exactly the same for the samples

| AR/MA | 0 | 1 | 2 | ... | ... | ... | ... |
|---|---|---|---|---|---|---|---|
| 0 | X | X | X | X | X | X | X |
| 1 | X | X | X | X | X | X | X |
| 2 | X | X | X | X | X | X | X |
| ... | X | X | X | X | X | X | X |
| ... | X | X | X | X | X | X | X |
| ... | X | X | 0 | 0 | 0 | 0 | 0 |
| ... | X | X | X | 0 | 0 | 0 | 0 |
| ... | X | X | X | X | 0 | 0 | 0 |
| ... | X | X | X | X | X | 0 | 0 |

# ARMA orders

- Recruitment series




Conclusion?

# ARMA orders

- EACF

```
> TSA::eacf(rec)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x x x x o o o  o  o  x
1 x x x o o o o o o o o  o  o  o
2 o o x x o o o o o o o  o  o  o
3 x o o x o o o o o o o  o  o  o
4 x x o o o o o o o o o  o  o  o
5 x x x o o o o o o o o  o  o  o
6 x x x o o o o o o o o  o  o  o
7 x x o o o o o o o o o  o  x  o
```

# ARMA orders

- AR(2) and ARMA(1,3)
  - Conclusions?

```
> arima(rec, order=c(2,0,0))

Call:
arima(x = rec, order = c(2, 0, 0))

Coefficients:
         ar1      ar2  intercept
      1.3512  -0.4612    61.8585
s.e.  0.0416   0.0417     4.0039

sigma^2 estimated as 89.33:  log likelihood = -1661.51,  aic = 3331.02
> arima(rec, order=c(1,0,3))

Call:
arima(x = rec, order = c(1, 0, 3))

Coefficients:
         ar1     ma1     ma2     ma3  intercept
      0.7826  0.5484  0.3239  0.2119    61.8609
s.e.  0.0390  0.0554  0.0621  0.0530     4.1953

sigma^2 estimated as 88.43:  log likelihood = -1659.24,  aic = 3330.48
>
```

# Model selection

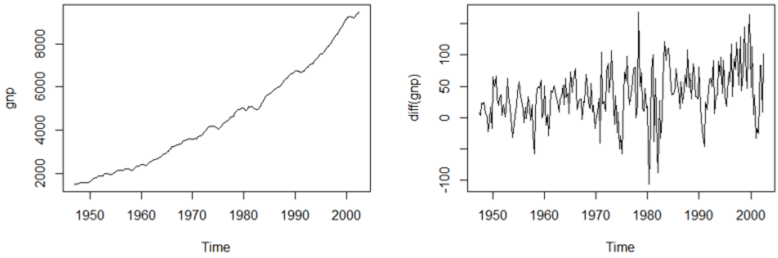
- Which model is suitable?
  - What is $p, d, q$ is ARIMA(p,d,q)?
  - $d$ is defined before!
  -
- Step 1: Check ACF, PACF and EACF to define a few tentative models

# Model selection

| | AR($p$) | MA($q$) | ARMA($p,q$) |
|---|---|---|---|
| ACF | Tails off | Cuts off after lag $q$ | Tails off |
| PACF | Cuts off after lag $p$ | Tails off | Tails off |

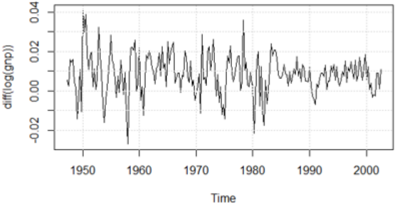
# Model selection

- Example: GNP data
  - Trying differencing $\rightarrow$ non-constant variance and maybe trend? $\rightarrow$ transformation



# Model selection

- Example: GNP data
  - Taking log and then differncing $\rightarrow$ still not perfect, but ... keep it as is.



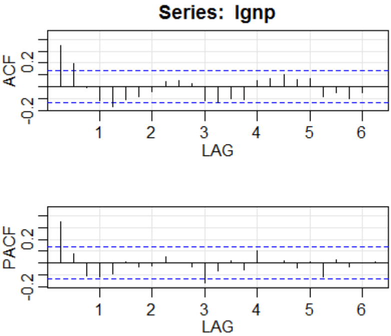

```
> adf.test(lgnp)

        Augmented Dickey-Fuller Test

data:  lgnp
Dickey-Fuller = -6.1756, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

# Model selection

- Example: GNP data
  - Testing ACF and PACF




Conclusion?

# Model selection

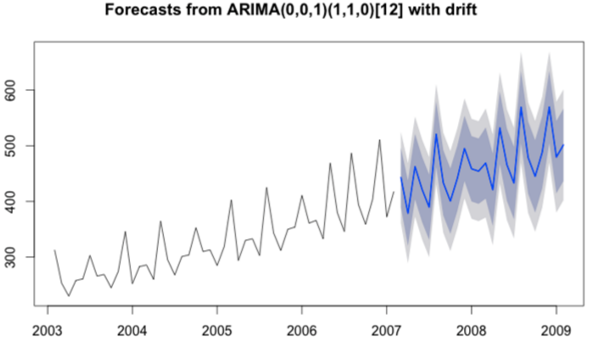
- Example: GMP data
  - Checking EACF

```
> TSA::eacf(lgnp)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x o o x o o o o o o  o  o  o
1 x x o o o o o o o o o  o  o  o
2 x x o o o o o o o o o  o  o  o
3 x o x o o o o o o o o  o  o  o
4 x o x o o o o o o o o  o  o  o
5 o x x o x o o o o o o  o  o  o
6 x x x x x o o o o o o  o  o  o
7 x x o x o o x o o o o  o  o  o
```

Conclusion?

# Forecasting

- We have our series $x_1 \ldots x_n$
- Use series to predict m steps ahead: $x^n_{n+m}$ should be based on our observed data $x^n_{n+m} = g(x_1, \ldots, x_n)$



# Forecasting

- Assume $g(x_1, ..., x_n) = \alpha_0 + \sum_{k=1}^{n} \alpha_k x_k$
  - Best linear predictors
- How to find $\alpha$'s?

$$minE\left[(x_{n+m} - g(x_1, ...x_n))^2\right]$$

- Prediction equations
  - Find $\alpha$'s by solving ($x_0 = 1$)

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

- Note: n+1 equations, n+1 unknowns

# One-step-ahead

- Denote $x_{n+1}^{n} = \phi_{n1}x_n + ...\phi_{nn}x_1$
- Prediction equations give

$$\Gamma_n \phi_n = \gamma_n$$

$$\Gamma_n = \begin{pmatrix} \gamma(1-1) & \gamma(2-1) & \dots & \gamma(n-1) \\ \gamma(2-1) & \gamma(2-2) & \dots & \gamma(n-2) \\ \dots & \dots & \dots & \dots \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(n-n) \end{pmatrix}$$

$$\phi_n = \begin{pmatrix} \phi_{n1} \\ \dots \\ \phi_{nn} \end{pmatrix} \qquad \gamma_n = \begin{pmatrix} \gamma_1 \\ \dots \\ \gamma_n \end{pmatrix}$$

- Note: for ARMA models $\Gamma_n$ is positive def $\rightarrow$ unique solution

# One-step-ahead

- Causal AR(p): for $n \geq p$ best linear prediction is

$$x_{n+1}^{n} = \phi_1 x_n + \ldots + \phi_p x_{n-p+1}$$

- In general, solve system of equations $\rightarrow O(n^3)$ operations

- Much faster algorithms exist
  - Durbin-Levinson algorithm
  - Innovations algorithm
- Property: PACF of a stationary process can be obtained as $\phi_{nn}$ by solving $\Gamma_n \phi_n = \gamma_n$

# One-step-ahead

- Mean square prediction error (MSPE)

$$P_{n+1}^{n} = E\left[(x_{n+1} - x_{n+1}^{n})^2\right] = \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n$$

- Confidence intervals for $x_{n+1}$

$$x_{n+1}^{n} \pm \alpha \sqrt{P_{n+1}^{n}}$$

- m-step ahead in general? Prediction equations
  - Difficult in general

# Read home

- Ch 3.2-3.4
- Paper "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation" by Tsay and Tiao
- R code: eacf in TSA package

# m-step-ahead for ARMA

- Assume causal and invertible ARMA(p,q)
- Finite past prediction

$$x_{n+1}^{n} = E(x_{n+1}|x_n, \ldots x_1)$$

- Infinite past prediction

$$\tilde{x}_{n+m}^{n} = E(x_{n+m}|x_n, \ldots x_1, x_0, x_{-1}, \ldots)$$

- m-step-ahead forecast for infinite past
  - Compute recursively

$$\tilde{x}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \hat{x}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j \tilde{x}_{n+m-j}, m = 1, 2, \ldots$$

- m-step ahead prediction error: $P_{n+m}^{n} = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2$

# Long-range forecasts

- What if $m \to \infty$?

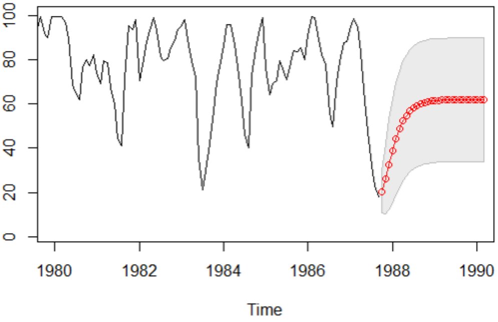
$$\tilde{x}_{n+m} \to 0(\mathrm{or}\mu)$$

$$P^n_{n+m} \to \sigma^2_x$$

# m-step-ahead

- Recruitment, AR(2)

$$x_{n+m}^{n} \pm 2\sqrt{P_{n+m}^{n}}$$



# Truncated prediction

- Ignore non-positive $j$ in $x_j$

$$\tilde{x}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{x}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j x_{n+m-j}, m = 1, 2, ...$$

- For ARMA, truncated prediction formula:
  - Recursive computation, explicit

$$\tilde{x}_{n+m}^n = \phi_1 \tilde{x}_{n+m-1}^n + ... \phi_p \tilde{x}_{n+m-p}^n + \theta_1 \tilde{w}_{n+m-1}^n + ... + \theta_q \tilde{w}_{n+m-q}^n$$
$$\hat{w}_t^n = \tilde{x}_t^n - \phi_1 \tilde{x}_{t-1}^n - ... - \phi_p \tilde{x}_{t-p}^n - \theta_1 \tilde{w}_{t-1}^n - ... - \theta_q \tilde{w}_{t-q}^n$$

- Boundary conditions: $\tilde{x}_t^n = x_n, 1 \leq t \leq n, \tilde{x}_t^n = 0, t \leq 0$

$$\tilde{w}_t^n = 0, t \leq 0 \quad \text{or } t > n$$