Time Series Analysis

ARIMA II: Mo Parameters, Mo Problems

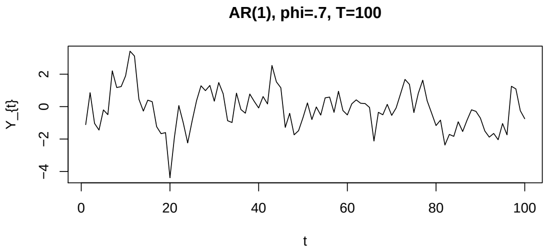
Michael Colaresi
colaresi@msu.edu

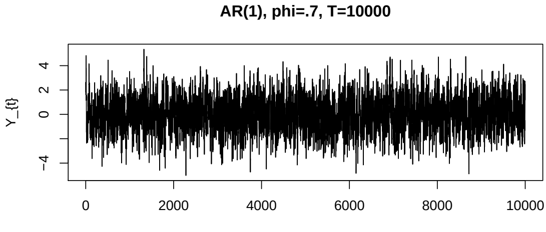
September 18, 2014

Outline

Autoregressive Process

- Autoregressive process of order p: Suppose ϵ_t is a purely random process with mean $\mu = 0$ and variance σ^2_ϵ. A process Y_t is said to be an *autoregressive process of order* p, $(AR\,(p))$, if
  $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t$.
- An $AR\,(\infty) = \phi(B) = 1 - \sum_{j=1}^{\infty} \phi_j B^j$ (written this way because we usually place it on the left hand side: $\phi(B) y_t = \epsilon_t$)


AR(1), phi=.7, T=100
Y_{t}
t
0
20
40
60
80
100
−4
−2
0
2


AR(1), phi=.7, T=10000
Y_{t}
0
2000
4000
6000
8000
10000
−4
0
2
4

AR Properties

- AR process does not depend on time
  1. $\epsilon_t s$ are independent
  2. Mean is constant
  3. Roots are stable $\therefore$ weakly stationary process
  4. If $\epsilon_t \sim N\left(\mu, \sigma^2\right) \Rightarrow$ strictly stationary process.
- ACF "dampens"

Moving Average

- Moving Average Process of order q: Suppose ϵ_t is a purely random process with mean $\mu = 0$ and variance σ^2_ϵ. A process Y_t is said to be a *moving average process of order q* ($MA(q)$), if

$$\begin{aligned} Y_t &= \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \ldots - \theta_q\epsilon_{t-q} \\ &= \left(1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \ldots - \theta_q B^q\right)\epsilon_t \end{aligned}$$

- An $MA(\infty) = \theta(B) = 1 + \sum_{j=1}^{\infty} -\theta_j B^j$.

MA Properties

- The MA process does not depend on time.
  1. $\epsilon_t s$ are conditionally independent.
  2. The mean is constant $\therefore$weakly stationary process.
  3. If $\epsilon_t \sim N\left(\mu, \sigma^2\right) \Rightarrow$ strictly stationary process.
  4. Roots are stable and process is "invertible".
- ACF "cuts off" or "ends" abruptly.

ARIMA(p,d,q)

It is sometimes the case that some data series exhibit mixed AR/MA behavior. As a generic example

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \ldots - \theta_q \epsilon_{t-q} \tag{1}$$

- Using the lag operator (B), we can rewrite (1) as

$$\phi(B)\, Y_t = \theta(B)\, \epsilon_t \tag{2}$$

- where $\phi(B), \theta(B)$ are polynomials of order p and q respectively. Now components in (2) can be written as:

$$\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p \tag{3}$$

$$\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q \tag{4}$$

AR to MA and Back Again

One of the most important things to remember about ARIMA models is that AR and MA process are mathematically equivalent to each other at the limit.

- You can turn an AR into an MA
- You can turn an MA into an MA
- *Without loss of information (just paper)*

To Infinity and Beyond

Both AR and MA processes can be represented as infinite order representations of the other.

The implication:

- AR process helps "identify" an MA process (more on this below).
- The MA process tells us whether an AR process is stationary.
- Using both allows us to better understand how to use ACF and PACFs.

MA process to infinite-order AR process

- Consider the following $MA(1)$ process:

$$Y_t = \epsilon_t - \theta\epsilon_{t-1} \tag{5}$$

- Shift (5) back one period:

$$\begin{aligned} Y_{t-1} &= \epsilon_{t-1} - \theta\epsilon_{t-2} && (6) \\ \epsilon_{t-1} &= Y_{t-1} + \theta\epsilon_{t-2} && (7) \end{aligned}$$

- Substitute (7) into (5) to get a new expression for ϵ_t :

$$\begin{aligned} Y_t &= \epsilon_t - \theta\left(Y_{t-1} + \theta\epsilon_{t-2}\right) && (8) \\ Y_t &= \epsilon_t - \theta Y_{t-1} - \theta^2\epsilon_{t-2} \end{aligned}$$

MA to AR II

- Repeat this process for two lags:

$$\epsilon_{t-2} = Y_{t-2} - \theta\epsilon_{t-3} \tag{9}$$

- Substitute (9) into (8):

$$\begin{aligned} Y_t &= \epsilon_t - \theta Y_{t-1} - \theta^2 (Y_{t-2} - \theta\epsilon_{t-3}) \\ &= \epsilon_t - \theta Y_{t-1} - \theta^2 Y_{t-2} + \theta^3 \epsilon_{t-3} \end{aligned}$$

- and so on. If we ignore $\epsilon_{t-\infty}$ (since $\theta^{\infty} \to 0$), will get us:

$$Y_t = \epsilon_t - \sum_{i=1}^{\infty} \theta^i Y_{t-i} \tag{10}$$

- which is an infinite-order AR process.

AR process to infinite-order MA process

- Alternatively, we can express an AR process as a infinite order MA process. Consider the following $AR(1)$ process:

$$Y_t = \phi Y_{t-1} + \epsilon_t \tag{11}$$

- Again, shift back (11) one period:

$$Y_{t-1} = \phi Y_{t-2} + \epsilon_{t-1} \tag{12}$$

- Plugging (12) into equation (11):

$$\begin{aligned} Y_t &= \phi(\phi Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \phi^2 Y_{t-2} + \phi\epsilon_{t-1} + \epsilon_t \end{aligned} \tag{13}$$

AR to MA II

- Repeat the process for two lags:

$$Y_{t-2} = \phi Y_{t-3} + \epsilon_{t-2} \tag{14}$$

- Substituting equation (14) into equation (13):

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

- Like (10), one process disappears (this time it is the AR process), and another assumes an infinite-order MA process:

$$Y_t = \sum_{i=1}^{\infty} \phi^i \epsilon_{t-i} \tag{15}$$

A Higher Order

- There is no reason AR process would just be of order 1.
  - What would ACF of an AR(2) look like?
  - AR(3)?
- Simulate the process: $y_t = .5y_{t-1} + .3y_{t-2} + \epsilon_t$

Time 0: $0 + 1 \leftarrow OneUnitShock$
Time 1: $.5(1) + .3(0) = .5$
Time 2: $.5(.5) + .3(1)) = .55$
Time 3: $.5(.55) + .3(.5) = .425$
Time 4: $.5(.425) + .3(.55) = .3775$

PACF

PACF The PACF has the pattern of the AR(k) form:

$$Y_t = \phi_{k_1} Y_{t-1} + \phi_{k_2} Y_{t-2} + \phi_{k_3} Y_{t-3} + \ldots + \phi_{k_k} Y_{t-k} + \epsilon_t$$

An AR(1) provides a single spike such that $Y_t = \phi Y_{t-1}$ and an AR(2) provides $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2}$. The PACF of MA processes will be infinitely lagged and decaying because the inverted MA is an infinitely lagged AR.

Rinse and Repeat

Box-Jenkins Methods

1. Plot the data
2. Difference/Transform to Stationarity
3. Check ACF/PACF and Diagnose
4. Filter and get residuals
5. Test residuals for patterns
6. Repeat until you have white noise.

Danger Will Robinson

!!!!!!!!!!!Overfitting the sample!!!!!!!!!!!

More on that later...

Independence Tests: Mind Your Q's

- The Q-**statistic** is given by

$$Q = T\sum_{k=1}^{m} r_k^2 \sim \chi^2_{(m-p-q)}$$

where T = total number of time point; m = the number of "events" in the correlogram (max lag); k = the index for each autocorrelation parameter; p = the order of the AR component; and q = the order of the MA component. This test is asymptotically valid for the null hypothesis that errors are white noise. Remember that if the null is true, the expected value is $(m-p-q)$.

Also called the Portmaneau test.

Box-Ljung

- The **Box-Ljung statistic** is given by

$$T\,(T+2)\sum_{k=1}^{m}\frac{r_k^2}{(T-k)} \sim \chi^2_{(m-p-q)}$$

  This statistic represents an attempt to obtain better small sample properties than the Q. Note it is asymptotically valid.

These are tests for “white noise”

Jarque-Bera Test

The Jarque-Bera test makes use of the following properties of normal distributions:

1. All odd moments greater than 2 equal zero.
2. The 4th central moment (kurtosis) is equal to 3.

Jarque-Bera Strikes Back

- Consider the $r'th$ central moment:
  $\mu_r = T^{-1} \sum_t (X_t - \mu)^r \quad \forall\ r, t$
  - When $r = 3$, the ratio for the coefficient of skewness becomes, $\beta_1 = \frac{\mu_3}{(\mu_2)^{3/2}}$
  - and when $r = 4$, the coefficient of kurtosis becomes, $\beta_2 = \frac{\mu_4}{\mu_2}$.

  If a random variable is normally distributed, skewness equals zero and kurtosis equals 3. In addition, the standard errors of these two sample moments are equal to:

$$\sqrt{Var(\beta_1)} = \sqrt{\frac{6}{T}}$$
$$\sqrt{Var(\beta_2)} = \sqrt{\frac{24}{T}}.$$

The Return of the Jarque-Bera

The **Jarque-Bera test** is equivalent to:

$$\left(\frac{T}{6}\right)\beta_1 + \left(\frac{T}{24}\right)(\beta_2 - 3)^2 \sim \chi^2_2$$

where under normality, the expected value of the test statistic is two.

- The hypotheses are set up as follows:

$$\begin{aligned} H_0 &: \quad \text{Normality} \\ H_A &: \quad \text{Non-normality} \end{aligned}$$

Blood, Shattered Glass and Other Slight Exaggerations

Before you get the keys to the car (stata and R-code), you have to watch the cautionary (and metaphorical) public service video about driving, ur, time series gone wrong:

- Overfitting.
- Overdifferencing.
- Canceling Effects

Overfitting and Overconfidence

- Say you have an underlying data-generation process that is:
  $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$
  where $|\phi_i| < 1$ and $\epsilon_t \sim N(0, \Sigma)$
- What are the chances that when you estimate a AR(p) model, where $p > 2$, some ϕ_j, where $j > 2$ is "significant", even though in the DGP, it is 0.
- Say, for example, you run an AR(5) model.
  - independent random draws give you a .14 $(1 - (1 - .05)^3)$ chance of at least one one of the extra AR parameters being significant with a conventional .05 significance level.
  - Now, lets say it is ϕ_4 that is significant. You then drop the other 2 AR terms.
  - What is your confidence that $|\phi_4| > 0$?
    $< .05$?

Intuition

- NO! .14
- The intuition here is that it could have been any of the extra ϕ's that were falsely significant. You just decided to drop the ones that were not.
- Same calculation as before. The problem is not the math. It is the mechanical dropping/keeping of terms.

This is overfitting your sample data! It means you are over-confident in your inference relative to the population.

Over-differencing

- What if you have a stationary white noise model $y_t = \epsilon_t$, and difference it?
  - $y_t - y_{t-1} = \epsilon_t - y_t$
  - $y_{t-1} = \epsilon_{t-1}$, so
  - $\Delta y_t = \epsilon_t - \epsilon_{t-1}$
  - You get a non-invertable error process. This will look like a large MA(1) process in sample data. This is easy to check for.
- What if you under-difference?
  - $y_t = y_{t-1} + \epsilon_t$
  - What should the theoretical ACF look like? and the PACF?
  - In sample data, this will still be apparent, although biased downward.

Redundant Effects

- Think about this model $y_t = .5y_{t-1} - .5\epsilon_{t-1} + \epsilon$
- What are the simulated dynamics?
- $t_0 = 1$
- $t_1 = .5(1) - .5(1) = 0$
- $t_2 = ..5(0) + .5(0) = 0$
- ...
- Which looks like what?

Avoiding these problems

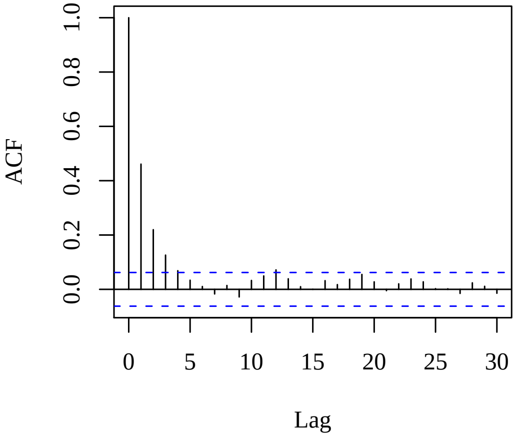
- Knowledge is power
- Parsimony helps
- Think about your model
- Split your data into a test and validation set.
  - Use Box-Jenkins methods on data from t_1 to t_i (test data).
  - Fit the "best" model for the test data to the data from t_{i+1} to t_T (validation data).
  - This is not a panacea. But it can help.

Training

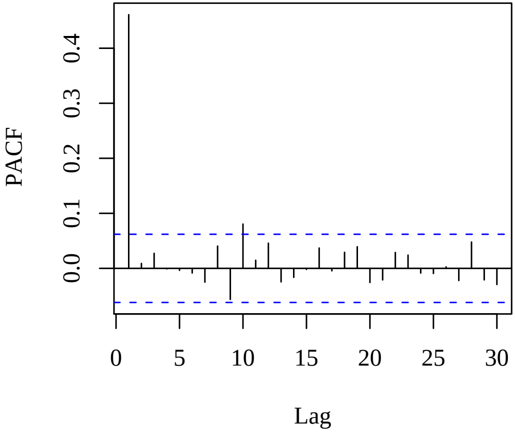
- In reality: We have a known sample and want to make inferences back to the unkown DGP.
- In class: We will have a known DGP, and learn what kind of samples that TYPICALLY creates.

Walk before you run.

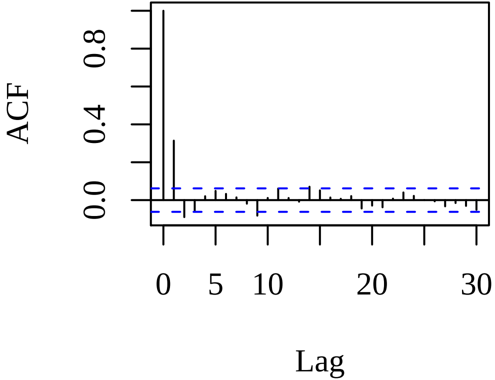
Example 1:ACF



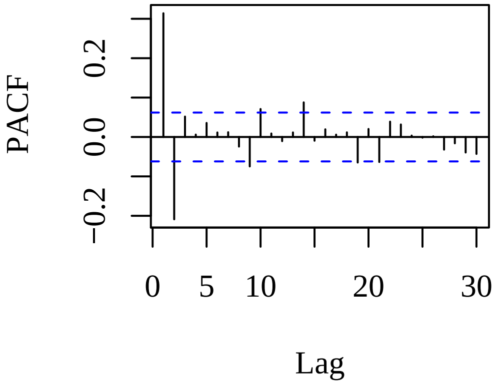
Example 1:PACF



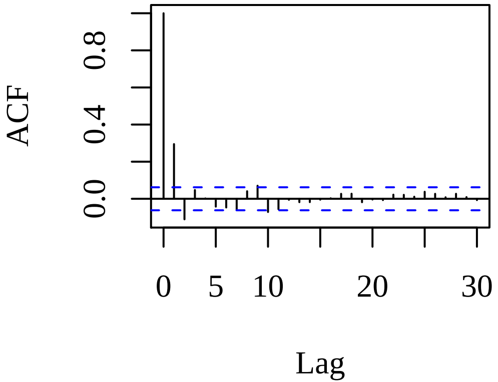
Example 2:ACF



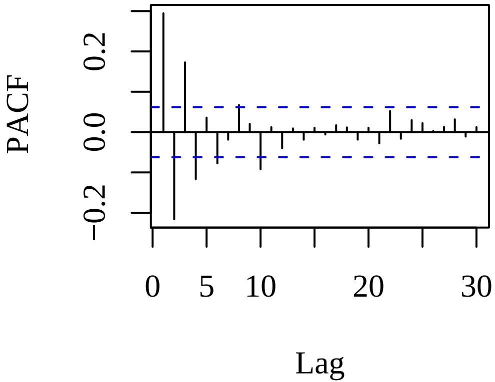
Example 2:PACF



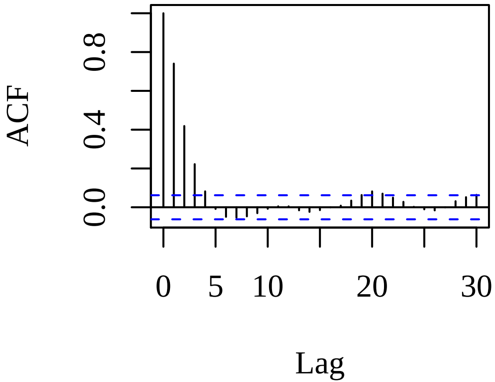
Example 3:ACF



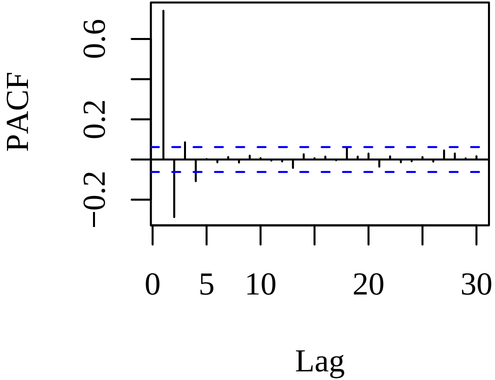
Example 3:PACF



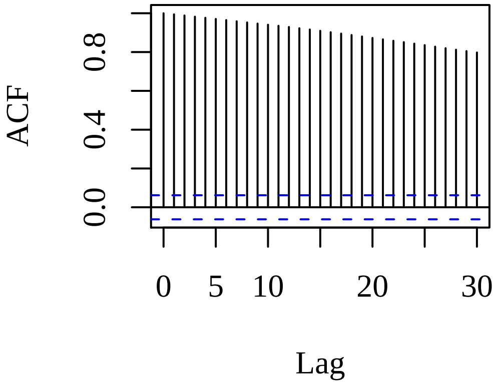
Example 4:ACF



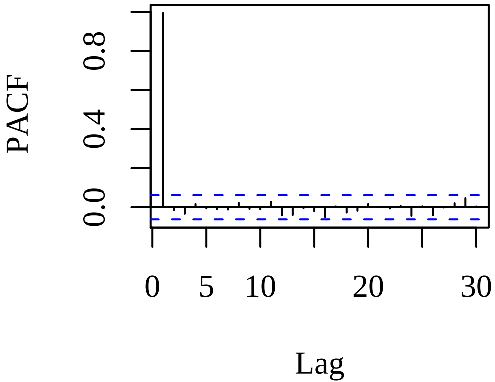
Example 4:PACF



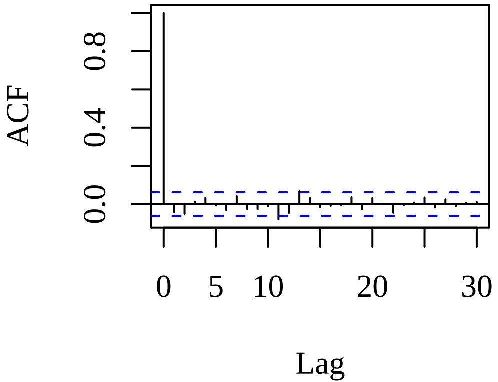
Example 5:ACF



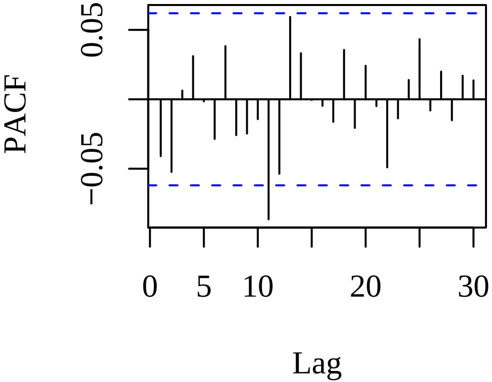
Example 5:PACF



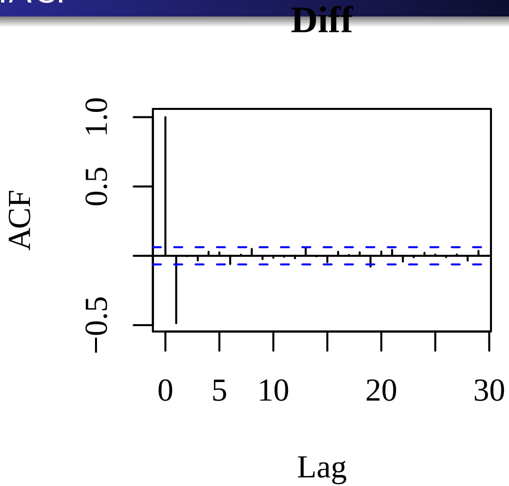
Example 6:ACF



Example 6:PACF



Example 7:ACF



Example 6:PACF