ARMA: Model Fitting

This Lecture

- ACF and PACF rules of thumb
- Information Criteria: AIC and BIC
- A rough guide to model selection
- Checking coefficients and refining

Later lectures:

1. Properties of stationary and nonstationary ARMA processes
2. Forecasting using ARMA

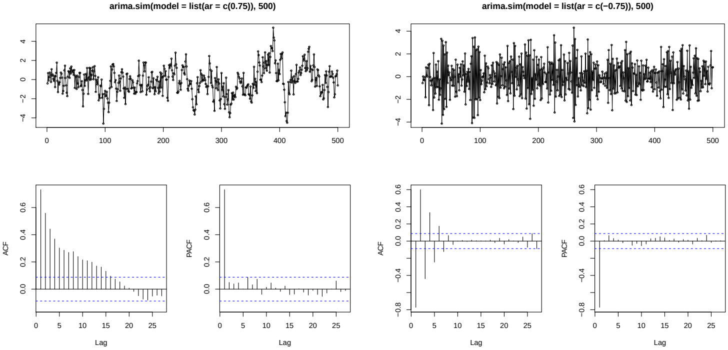
Important

- **If true model is ARMA(p,q), ACF and PACF are not THAT helpful for choosing p and q**
- **Just use them to get a sense of the extent of autocorrelation and a rough place to start**

AR(1)

For stationary AR(1)

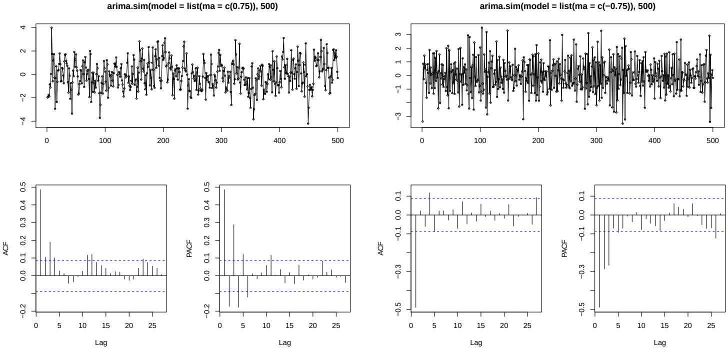
- ACF has exponential decay if $1 > \phi_1 > 0$ and oscillating decay if $0 > \phi_1 > -1$.
- PACF cuts off at one lag.



MA(1)

For MA(1), opposite of AR(1)

- ACF cuts off at one lag.
- PACF has exponential decay if $0 > \theta_1 > -1$ and oscillating decay if $1 > \theta_1 > 0$.



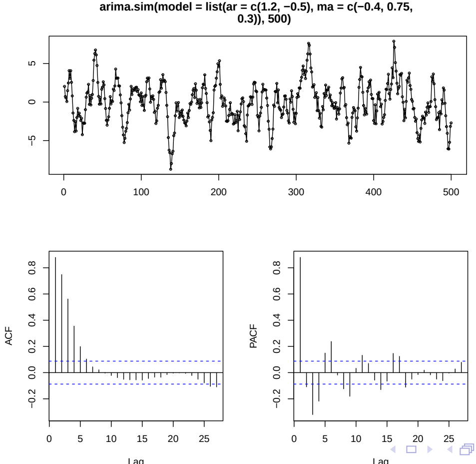
More general ARMA

- For AR(p):
  - ACF has exponential or oscillating decay depending on ϕ coefficient signs
  - PACF will cut off after p lags
- For MA(q):
  - ACF cuts off after q lags
  - PACF has exponential or oscillating decay depending on θ coefficient signs
- For ARMA(p,q):
  - It's a mix – not informative for selecting p and q
  - WE ARE OFTEN IN THIS SCENARIO!

ARMA(p,q)

ARMA(2,3)

$$y_t = 1.2y_{t-1} - 0.5y_{t-2} + a_t - 0.4a_{t-1} + 0.75a_{t-2} + 0.3a_{t-3}$$



Information Criteria

- We want a good fit, but not overfit. What is good fit? What is overfit?
- Good fit: greatest likelihood value
- Overfit: too many parameters (i.e., coefficients)
- MLE: maximize likelihood for given number of parameters
- One idea: choose p and q with highest, maximized likelihood.
- BUT... penalize the number of parameters

Akaike Information Criteria

- Akaike Information Criterion = (-2/T)*log likelihood + (2/T)*number parameters

$$AIC(l) = ln(\tilde{\sigma}_l^2) + \frac{2l}{T}$$

  - l is number of estimated parameters (e.g., $p + q$)
  - smaller is better
  - First term shrinks with higher l (better fit)
  - Second term grows with higher l (penalize l)
  - As T gets big, penalty for l declines

Schwarz-Bayesian Information Criterion

$$BIC(l) = ln(\tilde{\sigma}_l^2) + \frac{l \cdot ln(T)}{T}$$

- Sample size doesn't alleviate the penalty as much
- One selection rule: estimate models with lags from 0 to l. Calculate the AIC or BIC for each one. Pick the model with lowest AIC or BIC.
  - In R: armaselect() within caschrono package, auto.arima() within the forecast package.
  - auto.arima() is less accurate but has more options, e.g., seasonal terms, etc.
- Will this rule result in white noise residuals?

General "Rule"

- Find the lowest AIC or BIC model that eliminates residual autocorrelation.

Rough Guide

1. Use ACF/PACF/auto.arima/armaselect to get rough guess at p and q.
2. Estimate several models in the neighborhood of your guess.
3. For each, check ACF/PACF of residuals and test residuals with Ljung-Box
   - Make sure to adjust degrees of freedom for number of coefficients
4. Among models with no residual autocorrelation, pick the one with the smallest AIC or BIC
   - If AIC and BIC disagree, pick the smaller model

Takeaways and Opinions

- A very large model is rarely buying you anything in terms of
  - forecast accuracy, fit, bias reduction, or efficiency.
- Large model uses degrees of freedom (bad in small sample), takes longer to fit in large sample, and is harder to interpret.
- If there is still a bit of autocorrelation after fitting your best model, is it large enough to matter? Magnitude counts.

Ljung-Box for Residuals

- Ljung-Box

$$Q(m) = T(T+2)\sum_{s=1}^{m} \frac{\hat{\rho}_s^2}{T-s} \sim \chi^2(m)$$

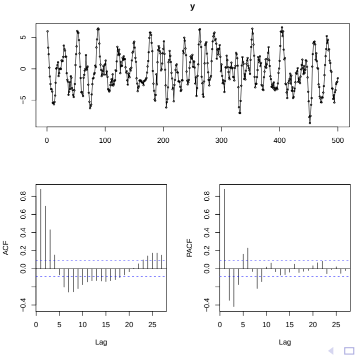
- m is whatever your hypothesis is, or $\approx ln(T)$ (unless you have seasonality). Use your knowledge.
- If you estimated $l = p + q$ coefficients, need to use $m - l$ degrees of freedom
- Test against a $\chi^2(m-l)$ distribution.
- Examples in R have been provided.

Example with ARMA(2,3)

I have no idea what p and q are from the figure, but know I need to worry about residual autocorrelation.

```
y <- arima.sim(model=list(ar=c(1.2,-0.5),ma=c(-0.4,0.75,0.3)),500)
```

```
tsdisplay(y)
```

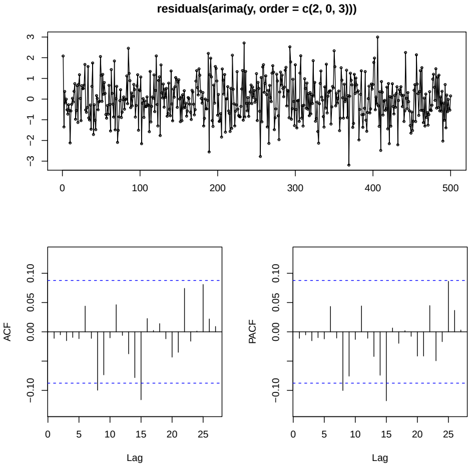


Example with ARMA(2,3)

```
# Use armaselect() from caschrono package, uses BIC only
armaselect(y,max.p=15,max.q=15,nbmod=5)

##      p q      sbc
## [1,] 2 3 54.31083
## [2,] 2 5 59.64113
## [3,] 3 3 59.81683
## [4,] 3 2 59.96569
## [5,] 2 4 60.50431

bt = Box.test(residuals(arima(y,order=c(2,0,3))),lag=20,type="Ljung-Box")
# uses 20 df, but should have 20-(2+3)=15.
1-pchisq(bt$statistic,15)

##   X-squared
## 0.09197976
```

Example with ARMA(2,3)

```
tsdisplay(residuals(arima(y,order=c(2,0,3))))
```

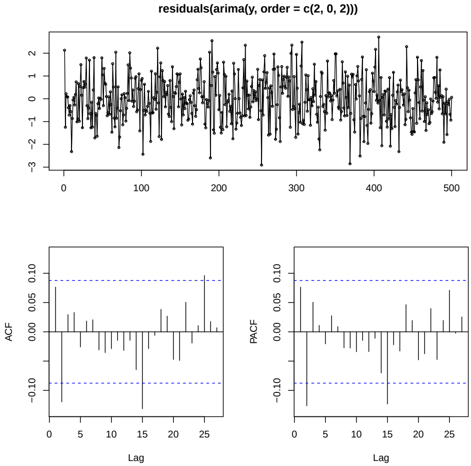


Example with ARMA(2,3)

```
# Use auto.arima
auto.arima(y)

## Series: y
## ARIMA(2,0,2) with non-zero mean
##
## Coefficients:
##          ar1      ar2      ma1     ma2     mean
##       1.4693  -0.6947  -0.6653  0.9055  -0.2690
## s.e.  0.0338   0.0341   0.0238  0.0251   0.2453
##
## sigma^2 estimated as 1.009:  log likelihood=-712.01
## AIC=1436.01   AICc=1436.19   BIC=1461.3

bt2 = Box.test(residuals(arima(y,order=c(2,0,2))),lag=20,type="Ljung-Box")
# uses 20 df, but should have 20-(2+2)=16.
1-pchisq(bt2$statistic,16)

##  X-squared
## 0.03122943
```

Example with ARMA(2,3)

```r
tsdisplay(residuals(arima(y,order=c(2,0,2))))
```



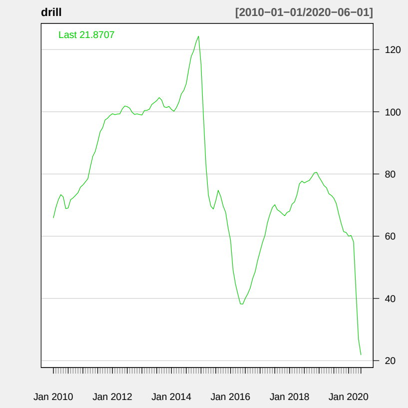
Example with Drilling Index

Ignore this warning message

```
## 'getSymbols' currently uses auto.assign=TRUE by default, but will
## use auto.assign=FALSE in 0.5-0.  You will still be able to use
## 'loadSymbols' to automatically load data.  getOption("getSymbols.env")
## and getOption("getSymbols.auto.assign") will still be checked for
## alternate defaults.
##
## This message is shown once per session and may be disabled by setting
## options("getSymbols.warning4.0"=FALSE). See ?getSymbols for details.

## [1] "IPN213111N"
```

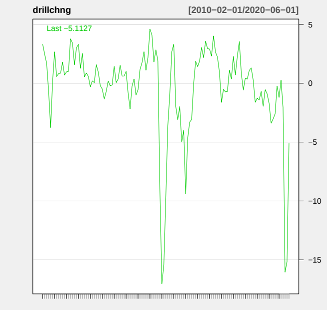
Example with Drilling Index

```
drill <- IPN213111N[paste("2010-01-01","2020-06-01",sep="/")]
chartSeries(drill,theme="white")
```



Example with Drilling Index

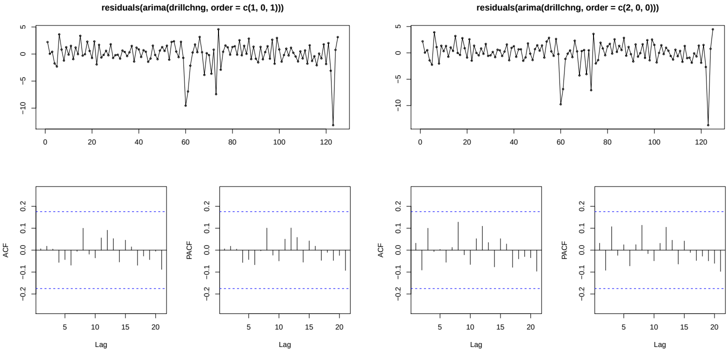
```
drillchng <- na.omit(diff(drill[,1]))
Box.test(drillchng,lag=10,type="Ljung-Box")

##
##  Box-Ljung test
##
## data:  drillchng
## X-squared = 97.72, df = 10, p-value < 2.2e-16

chartSeries(drillchng,theme="white")
```



Example with Drilling Index

```
armaselect(drillchng,nbmod=5)

##      p q      sbc
## [1,] 2 0 221.6247
## [2,] 3 0 226.5478
## [3,] 4 0 229.7136
## [4,] 1 0 232.7329
## [5,] 1 1 233.0572

auto.arima(drillchng)

## Series: drillchng
## ARIMA(1,0,1) with zero mean
##
## Coefficients:
##          ar1     ma1
##       0.5639  0.5158
## s.e.  0.0879  0.0929
##
## sigma^2 estimated as 5.361:  log likelihood=-281.91
## AIC=569.82   AICc=570.02   BIC=578.3
```

Example with Drilling Index



Check Coefficients

- The p'th lag might be important, but do we need lag $p-1$?
- Tsay: drop insignificant coefficients and reestimate
  - Does AIC/BIC go down? Are residuals still white noise?

Check Coefficients: Gold Example

- Stick with AR(p) model just for illustration
- AIC picks AR(12)

```
getSymbols("GOLDPMGBD228NLBM",src="FRED")

## [1] "GOLDPMGBD228NLBM"
```

```
gold <- na.omit(GOLDPMGBD228NLBM[paste("1968-04-01","2017-09-03",sep="/")])
goldrtn <- na.omit(diff(log(gold[,1])))
ar(goldrtn,order.max=15)

##
## Call:
## ar(x = goldrtn, order.max = 15)
##
## Coefficients:
##       1        2        3        4        5        6        7        8
## -0.0192  -0.0165   0.0213   0.0034  -0.0013  -0.0077   0.0032  -0.0029
##       9       10       11       12
##  0.0315   0.0009  -0.0014   0.0360
##
## Order selected 12  sigma^2 estimated as  0.000155
```

Check Coefficients: Gold Example

```
m1 <- arima(goldrtn,order=c(12,0,0))
m1


##
## Call:
## arima(x = goldrtn, order = c(12, 0, 0))
##
## Coefficients:
##             ar1      ar2     ar3     ar4      ar5      ar6     ar7      ar8
##         -0.0192  -0.0165  0.0213  0.0034  -0.0013  -0.0077  0.0032  -0.0029
## s.e.     0.0090   0.0090  0.0090  0.0090   0.0090   0.0090  0.0090   0.0090
##            ar9   ar10     ar11   ar12  intercept
##         0.0316  9e-04  -0.0014  0.036      3e-04
## s.e.    0.0090  9e-03   0.0090  0.009      1e-04
##
## sigma^2 estimated as 0.0001548:  log likelihood = 36844.9,  aic = -73661.8

# only lags 1, 2, 3, 9 and 12 are significant. AIC -73661.8
```

Check Coefficients: Gold Example

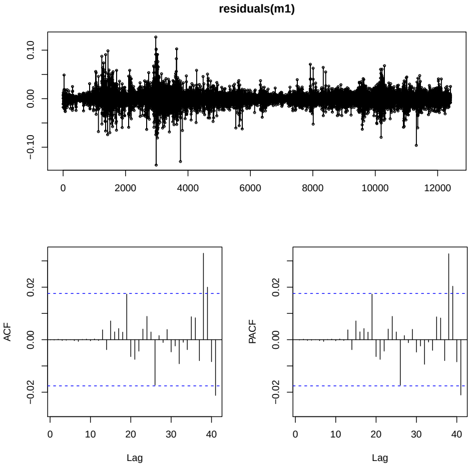
	Model 1
ar1	-0.02 $(0.01)^{*}$
ar2	-0.02 (0.01)
ar3	0.02 $(0.01)^{*}$
ar4	0.00 (0.01)
ar5	-0.00 (0.01)
ar6	-0.01 (0.01)
ar7	0.00 (0.01)
ar8	-0.00 (0.01)
ar9	0.03 $(0.01)^{***}$
ar10	0.00 (0.01)
ar11	-0.00 (0.01)
ar12	0.04 $(0.01)^{***}$
intercept	0.00 $(0.00)^{*}$
AIC	-73661.80
BIC	-73557.83
Num. obs.	12415

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table: Pretty Table Using TeXReG

Check Coefficients: Gold Example

```
tsdisplay(residuals(m1))
```



Check Coefficients: Gold Example

```
# create a model that constrains insignificant coefficients to 0.
c1 <- c(NA,NA,NA,0,0,0,0,0,NA,0,0,NA,NA) # last entry is for intercept
m2 <- arima(goldrtn,order=c(12,0,0),fixed=c1) # AIC -73674.61 an improvement

## Warning in arima(goldrtn, order = c(12, 0, 0), fixed = c1):  some AR
parameters were fixed:  setting transform.pars = FALSE

m2


##
## Call:
## arima(x = goldrtn, order = c(12, 0, 0), fixed = c1)
##
## Coefficients:
##            ar1      ar2     ar3  ar4  ar5  ar6  ar7  ar8     ar9  ar10  ar11
##        -0.0192  -0.0167  0.0211    0    0    0    0    0  0.0314     0     0
## s.e.    0.0090   0.0090  0.0090    0    0    0    0    0  0.0090     0     0
##          ar12  intercept
##        0.0361      3e-04
## s.e.   0.0090      1e-04
##
## sigma^2 estimated as 0.0001548:  log likelihood = 36844.3,  aic = -73674.61
```

Check Coefficients: Gold Example

	Model 1
ar1	$-0.02\ (0.01)^{*}$
ar2	$-0.02\ (0.01)$
ar3	$0.02\ (0.01)^{*}$
ar4	0.00
ar5	0.00
ar6	0.00
ar7	0.00
ar8	0.00
ar9	$0.03\ (0.01)^{***}$
ar10	0.00
ar11	0.00
ar12	$0.04\ (0.01)^{***}$
intercept	$0.00\ (0.00)^{*}$
AIC	-73674.61
BIC	-73622.62
Num. obs.	12415

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table: Pretty Table Using TeXReG

Check Coefficients: Gold Example

```
tsdisplay(residuals(m2),main='AR Model With Lags 1-3, 9, 12')
```