Financial Econometrics Econ 40357
ARIMA Model Selection
AIC, BIC, HQIC

N.C. Mark

University of Notre Dame and NBER

September 2, 2020

Model selection

- Ensure variable is stationary
- Estimate all the candidate models, keep variables with significant t-ratios. Not good
- Estimate candidate models, compare their forecasting accuracy. Doesn't work well either.
- Tradeoffs:
  - Underfitting: Omitted variables, produces bad forecasts
  - Overfitting: Additional sampling variability produces bad forecasts
  - Generally, lightly parameterized models produce better forecasts than heavily parameterized ones.
- **Information Criteria.** Let the data tell us how to specify the model. We use something called information criteria (IC).

A Little Background on Information Criteria

- Our friend **maximum likelihood estimation**
- Start with model
  - MA(1): $y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$
  - AR(1): $y_t = \mu(1-\rho) + \rho y_{t-1} + \epsilon_t$
- Let $f(\epsilon_t)$ be the pdf of ϵ_t. If ϵ_t is independent, then the **joint pdf** of $\epsilon_{t,}\epsilon_{t-1}$ is $f(\epsilon_t)\, f(\epsilon_{t-1})$.
- The product of the pdfs. Hence, the pdf of all the shocks is

$$f(\epsilon_{T,}\epsilon_{T-1}, ..., \epsilon_1) = f(\epsilon_T)\cdots f(\epsilon_1).$$

- Assume that ϵ_t is normally distributed. Independent and identically distributed with zero mean and variance σ_ϵ^2. Earlier, we showed the joint pdf to be,

$$f(\epsilon_T, ..., \epsilon_1) = \left(\frac{1}{\sqrt{\sigma_\epsilon^2 2\pi}}\right)^T e^{\frac{-1}{2\sigma_\epsilon^2}\sum_{t=1}^T \epsilon_t^2}$$

Background on Information Criteria

- Solve for ϵ_t in the model. Substitute these expressions into the joint pdf, gives the likelihood function.
- Searching for parameters $\theta, \sigma_\epsilon^2, \mu$ for the MA(1) or $\mu, \rho, \sigma_\epsilon^2$ for the AR(1) to maximize the likelihood function is equivalent to searching parameter values to maximize the logarithim of the likelihood function.
- We call it the log likelihood function.

$$LL = -T \ln\left(\sigma_\epsilon^2\right)^{\frac{1}{2}} - T \ln\left(2\pi\right)^{\frac{1}{2}} - \frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{T}\epsilon_t^2$$

$$\begin{aligned}
\text{Divide by } T \quad \frac{LL}{T} &= -\ln\left(\sigma_\epsilon^2\right)^{\frac{1}{2}} - \ln\left(2\pi\right)^{\frac{1}{2}} - \frac{1}{2\sigma_\epsilon^2}\underbrace{\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2}_{\hat{\sigma}_\epsilon^2} \\
&= -\frac{1}{2}\ln\left(\sigma_\epsilon^2\right)\underbrace{-\frac{1}{2}\ln\left(2\pi\right) - \frac{1}{2}}_{\text{constant}}
\end{aligned}$$

where it's understood that the ϵ_t represent the model.

Hence, the log likelihood function reduces to

$$LL = -\frac{1}{2} \ln \left(\sigma_{\epsilon}^{2} \right)$$

- Suppose we want to choose among ARMA(p,q), for $p = 0, ..., 5$, $q = 0, ..., 5$.
- Cannot use the highest likelihood across models for selection because it the maximized log likelihood (usually) continues to increase as you add parameters. Is like how R^2 keeps increasing when you add variables in regression.
- **Solution**: attach **penalty** for adding parameters. Different information criteria have different penalties.
- Maximizing the log likelihood, is to minimize $\ln(\sigma_{\epsilon}^{2})$. Information criteria: AIC, BIC, HPIC. The model that gives you the minimum IC is the one you want..
- First to do so was Akaike. False modesty to say A comes first in alphabet.

AIC, BIC, HPIC

Let k be number of parameters (count up the ρ_j, θ_j in ARMA model)

$$
\begin{aligned}
AIC &= \ln\left(\hat{\sigma}_{\epsilon}^{2}\right) + \frac{2k}{T} \\
BIC &= \ln\left(\hat{\sigma}_{\epsilon}^{2}\right) + \frac{k}{T}\ln\left(T\right) \\
HPIC &= \ln\left(\hat{\sigma}_{\epsilon}^{2}\right) + \frac{2k}{T}\ln\left(\ln\left(T\right)\right)
\end{aligned}
$$

In subsequent studies, AIC usualy chooses too many parameters, BIC, too few, HPIC is sort of just right.

How to do this on Eviews?

AIC, BIC, HPIC in Eviews

Dependent Variable: Y4
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 09/21/19 Time: 14:01
Sample: 4 200
Included observations: 197
Convergence achieved after 34 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.012659	0.077638	0.163050	0.8706
Y4(-1)	0.225327	0.090710	2.484035	0.0138
MA(1)	0.694689	0.062778	11.06577	0.0000
SIGMASQ	0.406187	0.038267	10.61443	0.0000
R-squared	0.470562	Mean dependent var		0.015227
Adjusted R-squared	0.462332	S.D. dependent var		0.878134
S.E. of regression	0.643899	Akaike info criterion		1.980890
Sum squared resid	80.01888	Schwarz criterion		2.047554
Log likelihood	-191.1177	Hannan-Quinn criter.		2.007876
F-statistic	57.17918	Durbin-Watson stat		1.828020
Prob(F-statistic)	0.000000			

Automatic ARIMA model selection

1. Open the series of interest
2. Click Proc, Automatic ARIMA Forecasting
3. In Options tab, choose Model Selection and the Information Criterion you want to use.