Model selection

Data to decisions

The problem

- Which model is best **<u>for the purpose of answering a particular question</u>**?
- Say, for concreteness, for forecasting purposes
- You could say that getting as close to the true DGP as possible is the obvious goal
- Yet:
  - There is no hope of knowing the true DGP in any interesting problem
  - Among misspecified models, models closer to the true DGP may forecast more poorly
  - Or, more generally, models closer to the true DGP can be worse for the purpose of answering the question at hand

Context matters

- “Models are rats”
- (See Kocherlakota’s ‘Model Fit and Model Selection’, St-Louis Fed review, April 2007)
- Models are wrong but, hopefully, useful for the specific purpose at hand
- But we should be able to rank alternatives: **model selection is model comparison**

Is fit the answer?

- Why not use the model that fits the evidence best for forecasting purposes?
- Fit can always be improved by adding more variables…
- … even variables that have nothing to do with the true DGP but happen to correlate with random draws in the sample at hand
- In fact, maximizing fit would invariably lead one to add variables that matter in sample through chance only and will hurt forecasting performance
- This is called *overfitting*

General principles

- There is a tradeoff between fit and parsimony
- **Principle #1:** among models that fit (=explain) about the same choose the more parsimonious one
- **Principle #2:** reward fit but penalize complexity (wiki information criterion)

Cross-validation and testing

- If our purpose is to forecast, forecasting ability is the right criterion to use
- How can one test forecasting ability since, by definition, we do not know what the right answer is?
- Split your sample into:
  1. *Training sample* (used to fit model)
  2. *Validation sample* (used to rank the performance of various models)
  3. *Testing sample* (used only after **it's all said and done** to gauge the generalizabality of the model to other data sets)
- Most common validation criterion: RMSE (root mean squared error)
- *k-fold validation*: use k random splits rather than just one

Warning

- Designing a model that fits historical evidence is trivial
- Forecasting is tough
- More complex models fit better, but forecast poorly (Wiki "overfitting")
- Only criterion that matters: out-of-sample forecasting fit
- In other words, how has your forecast performed?
- Truth: beating naïve models is tough, and naïve models are free

Information criteria

- Information criteria are quantitative measures of a model's performance in the fit/parsimony space

- Example: Akaike's information criterion (AIC)

$$AIC = 2k - 2\log(L)$$

where k is the number of parameters and L is the likelihood of the data under the model

- Some do model selection simply by ranking models according to information criteria (= picking the model with the lowest AIC)

- Meh…

But enough chit-chat, it’s time to looks at some examples