Section 16 Empirical Research Projects

Starting point: Question and data

- Starting point always must be "What question am I trying to answer?"
    - For thesis: something you can be interested in for a whole year
    - Something that can be answered
- Second consideration: "What data are available to help me find the answer?"
    - Macro data
    - Micro data from existing surveys
    - Collecting your own data from surveys

Methods

- Once you have the question and the data, you can carefully consider what method you should use
- Nature of dependent variable: continuous, limited?
    - Might need to consider LDV models
- What explanatory variables can you measure (and what is omitted)?
- Are there endogeneity concerns?
    - If yes, are appropriate instruments available to allow IV estimation?
- Are there other concerns about the error term?
    - Heteroskedasticity?
    - Autocorrelation?
- Are your data time series, cross section, pooled, or panel?
    - Appropriate models for each, including stationarity concerns
- What is the appropriate specification?
    - Functional form
    - Scaling and/or differencing to make the variables comparable

Estimation, diagnostic testing, re-estimation

- What did you learn from the first regression?
- Are there issues in the residuals or diagnostics based on the coefficients or residuals that suggest that your assumptions are incorrect?
    - Look for outliers and consider why they do not fit
    - (Errors in data)
- Can you test the underlying assumptions formally? Are they OK?

Writing the paper

- Introduction
    - What is the question?
    - How do you go about answering it?
    - What do you conclude?
- Theory section
    - What does economic theory tell us about the question?
    - What variables *should* be in the regression?
    - What considerations does theory suggest about functional form (e.g., CRTS)?
- Literature review
    - May come before theory section
    - Who else has explored this question and what did they find?
- Methods and data section
    - What estimation methods and tests are you proposing to use?
        - Why are these methods appropriate?
    - What data do you have (and not have)?
        - What issues of measurement might be important?
- Results section
    - Regression tables with basic description of results
    - Text must read as a narrative, referring to tables but not relying on them to tell the story.
- Analysis/interpretation/discussion section
    - What do the results mean?
    - Are there simulated experiments using your model that would help the reader understand your results?
    - How strong are the results?
    - Issues of internal and external validity: is it safe to draw conclusions based on your results?
- Conclusion
    - What do you conclude from your analysis?
    - What additional work remains to be done in future research?