Chapter 1

Installing R and RStudio

The Nature of Econometrics and Econometric Data

Jim Bang

Syllabus Information

What is Econometrics?

- "How much?"
  - The quantitative analysis of economic data
  - Helps to quantify relationships

General Information

- Office Hours: MWF 11am-12:30pm; TR 8:30-10am; liberally by appointment: book me here!
- Research Paper Deadlines:
- Class Preparation, Participation, and Discussion
  - The class participant's engagement actively contributes to general student learning – particularly important with a very small class.

Resources

- Blackboard
- Software
  - R (Free Download): http://cloud.r-project.org
  - RStudio (Free Download – Recommended): https://www.rstudio.com/

Getting Started with R

Why R?

R helps you talk to a computer and help solve big problems. Employers use R (or other similar languages like Python or Julia) for statistics and data analysis. R is not usually associated with calculus or mathematics, but can do anything that expensive mathematical-analysis programs like Matlab or Mathematica can do - and possibly more! Also, R is free!

We will use R in this class for two reasons:

1. As a practical alternative to pencil-and-paper calculus;
2. To develop basic proficiency with a computing language.

Installing R

Updating R

RStudio

What is RStudio?

Why RStudio?

RStudio is an Integrated Development Environment (IDE) for R. What does that mean? Well, if you think of R as a language, which it is, you can think of RStudio as a program that helps you write and work in the language. RStudio makes programming in R much easier and I suggest that you use it!

Where do I get RStudio?

Do I still need R?

R Packages

How do I install packages?

How often do I install a package?

Economics and the Scientific Method

1. Ask a Question
2. Form a Hypothesis
3. Review Existing Research
4. Collect Data
5. Analyze the Data

- Test *Causal* Hypotheses
- Forecast/Predict

6. Form a Conclusion

Example 1: Effects of Training on Productivity

$$wage = f(educ, exper, training; u), \text{ where}$$

- *wage* is the worker's hourly wage, - *educ* is a measure of schooling, - *exper* is a measure of workforce experience, - *training* is a measure of time spent training, and - u is random noise.

Assuming we can approximate $f(\cdot)$ with a linear function, we would

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 training + u$$

An econometric model consists of a systematic part – $f(*)$ – and a random error – u

Example 2: Effects of Sentencing on Crime

$$crime = f(crimewage, marketwage, otherincome, p_{caught}, p_{convicted|caught}, \hat{sentence}, age; u), \text{ where}$$

- *crime* is the number of hours spent in illegal activities,
- *crimewage* is the "wage" per hour for illegal work,
- *marketwage* is the wage for legal employment,
- *otherincome* is income from other sources,
- p_{caught} is the probability of getting caught,
- $p_{convicted|caught}$ is the probability of conviction if caught,
- *sentence* is the expected sentence if convicted, and
- *age* is the person's age.

Some variables might not be observed (crime wage); others might only be observe by proxy or as an average (p_{caught}, $p_{convicted|caught}$, *sentence*)

$$crime = \beta_0 + \beta_1 marketwage + \beta_2 otherinc + \beta_3 freqarr + \beta_4 freqconv + \beta_5 avesen + \beta_6 age + u$$

Data: Sources and Types

- Experimental versus Observational Data
- Cross-section
- Time-series data
- Pooled cross sections
- Panel/longitudinal data

Causality

How could we determine whether spending on education *causes* better performance?

Hold all else equal: *ceteris paribus*

Problems:

- Confounding (omitted) variables
  - Solution: Add those variables to the model.
- Non-random sample selection
  - Solution: Tobit (Chapter 17; Tobin, 1958) Heckit (Chapter 17; Heckman, 1976)
- Non-random treatment assignment
  - Solution: Instrumental Variables (Chapter 15) Matching treatment and controls (Rosenbaum & Rubin, 1983); Covariate balancing (Horvitz & Thompson, 1952; Robins et al., 2000)