Empirical methodology in strategy and IB research
BA 590 GH

Instructor: Glenn Hoetker
BIF 4019
ghoetker@illinois.edu
265-4081

Class: M & W 1-2:30 in BIF 4001
Office hours: By appointment

Contents

Course outline and grades.................................................................1
Description of assignments...........................................................1
  Applied exercises .........................................................................1
  Discussion....................................................................................2
Textbooks—philosophy and recommendations.................................2
Schedule by session.........................................................................3

Course outline and grades
This class is idiosyncratic, broad and very applied. I assume that you have already taken as basic
statistics course and are comfortable with linear regression. The goal of this class is to make you
a well-informed consumer and user of econometric techniques. I want you to develop an
intuition based on both the underlying mathematics and practical experience.

Grades will be determined as follows
  • Homework 50%
  • Discussion: 50% for your contribution to classroom discussion

Description of assignments

Applied exercises
Statistical methodology is best learned through practice. Therefore, I have designed several
assignments over the seven weeks. These will give you a chance to apply what you’ve learned to
real data. Unlike most classes, you are welcome, nay, encouraged to work with others on these.
You can even turn in assignments representing the work of the group as a whole—just be sure
everyone’s name is on it. You are on your honor to only associate your name with assignments
on which you have made some contribution.
Discussion

Most of the learning in our class will occur through discussion. Therefore, we need everyone’s active participation. Meaningful participation obviously requires having carefully read and thought about the day’s material, but that is not enough. You also need to speak up. Brilliant insights are, of course, always appreciated, but you can also contribute by asking questions, politely disagreeing with me or another student, or following up on a fellow student’s comments with further insights or evidence. There are usually many valid ways of looking at any issue we study, so don’t worry about being wrong. In fact, in this class, if you aren’t confused at least on occasion, you probably aren’t paying enough attention.

When preparing for discussion of an application paper, remember that, unlike most seminars, we really aren’t interested in the theory or hypotheses. Rather, please focus on

- What method or methods did they apply?
- Did they apply them properly?
- If not, how did they fail?
- Are there other methods that they should have used? If so, why?

Textbooks—philosophy and recommendations

There is one required textbook for this class:


There are many, many good statistical textbooks available. There are also some pretty bad ones. Text vary along multiple dimensions.

- **Mathematical sophistication** How much mathematical sophistication do they assume?
- **Focus** Are they general or focused on a specific class of problem? The more focused texts are, of course, likely to provide deeper coverage of that class of problem, but how many textbooks do you really want to keep around?
- **Theory versus application**
- **Package specificity**

I find it useful to have a variety of texts available, ranging from fairly basic to fairly advanced. When I encounter a new technique/model/issue, I’ll start with a less mathematically sophisticated, more applied textbook to get a background and then move up the gradient to acquire a deeper understanding.

I’ll recommend focused textbooks under their appropriate topics. Here are some general texts that I highly recommend, along with my rationale for recommending them.


A good starting point if you hit a concept that is new to you. Good as a starting point and as a lead to other sources. A bit dated.
Greene, W. H. *Econometric Analysis*.  
“The” ultimate graduate level econometrics textbook. Very thorough. If you are comfortable with matrix algebra, you should be able to push your way through Greene. He is a very straight-forward presenter, but the material is very dense and primarily mathematical in its presentation. He doesn’t convey much of the intuition. Greene is particularly strong in limited dependent variables.


Collectively known as “Big Wooldridge” (2002) and “Little Wooldridge” (2003), these are my favorite econometrics texts. Wooldridge is a superb explainer. He conveys both mathematical intuition and practical concerns. I often start by looking something up in *Little W* in order to get an initial understanding and then move to *Big W* in order to get the underlying math.

**Schedule by session**

*Session 1 Introduction and thinking about data*
Mar 16, 2010

| Reading   | n/a                      |
| Application | n/a                      |
| Software  | n/a                      |
| Assignment | n/a                      |
| Comments  | If you need a basic resource for Stata, I strongly recommend http://www.ats.ucla.edu/stat/stata/. I also strongly recommend The Workflow of Data Analysis Using Stata by J. Scott Long |

*Session 2 Data organization and automation*
Mar 18, 2010

| Reading   | Hoetker G. 2010. A template for Stata .do files |
| Application | n/a                      |
| Software  | Stata: reshape, merge, mmerge*, esttab* |
| Assignment | n/a                      |
| Comments  | n/a                      |
Session 3 Review of OLS  
Mar 29, 2010

Reading     Kennedy, Ch. 1-4  
Application  n/a  
Software     Stata: regress, hettest, vif  
Assignment   n/a  
Comments     This seems like a lot of reading, but most of it should be quite familiar.  
             Plus, of course, you’ll have break to read it!  
             The goal is just to re-familiarize ourselves with some of the key concepts of 
             regression, before we take off on variations to them.

Session 4 Two OLS challenges: interaction terms & comparing coefficients  
Mar 31, 2010

Reading     Interaction terms  
             Brambor T, Clark WR, Golder M. 2006. Understanding Interaction Models: 
             Braumoeller BF. 2004. Hypothesis testing and multiplicative interaction 
             Comparing coefficients  
             Gelman A, Stern H. 2006. The difference between "significant" and "not 
             significant" is not itself statistically significant. American Statistician 60(4): 
             328-331  
             statistical test for the equality of regression coefficients. Criminology 36: 
             859-866  
             Wolfe R, Hanley J. 2002. If we're so different, why do we keep 
             overlapping? When 1 plus 1 doesn't make 2. Canadian Medical Association 
             Journal 166(1): 65-66  
             237-240  
Application   n/a  
Software     Stata: predict, mfx, suest, test  
Assignment   Begin Assignment 1: OLS. It is due session 7.  
Comments     Again, a lot of reading. The basic ideas are pretty straight-forward.
Session 5 Count models. Guest lecture by Deepak Somaya
Apr 5, 2010

Reading  n/a
Application n/a
Software  n/a
Assignment n/a
Comments

Session 6 Count models. Guest lecture by Deepak Somaya
Apr 7, 2010

Reading  n/a
Application n/a
Software  n/a
Assignment n/a
Comments

Session 7 Logit and probit models
Apr 12, 2010

Reading  Hoetker, 2007
        Long 1997, Ch. 3 (especially section 7 on interpretation)
        Train 2003, Ch 3
Application n/a
Software  Stata: Logit, probit, spost*
Assignment n/a
Comments  Two great sources here are Train (1986) and Train (2003). Train is an extremely clear writer. As a transportation economist, he comes at things a little bit differently than many other presentations, which I find helpful.

        Long’s book (there is now a Stata-specific version too) is a classic. Very strong on presentation. I don’t agree with everything he writes, but overall, great.
Session 8 Logit/probit II: Complications and challenges  
Apr 14, 2010

Reading

Buis M. 2008. Direct and indirect effects in a logit model. The Stata Journal


Application n/a
Software n/a
Assignment n/a
Comments I’d first read Allison, which sparked Hoetker 2007. Those two address the group comparison issue.

I’d then read Huang and Shields, followed by Norton, Wang and Ai.

I’d close with the Buis paper.

Session 9 Logit/probit III: Extensions  
Apr 19, 2010

Reading
Greene 2003, pgs. 736-740
Kennedy, Ch. 16
Train 2003, Ch. 3 again

Application n/a
Software Stata: mlogit, ologit, spost*
Assignment Assignment 2: Logitl ,due April 26
Comments I’d start with Kennedy and rereading Train. Note that Train doesn’t distinguish between logit and multinomial logit. To him, they are all logit and there is nothing magic about having just two categories. The Greene piece is most valuable for the two figures it includes.
Session 10 Hazard rate models I: Introduction, key concepts and data organization
Apr 21, 2010

Reading
Cleves MA. 2008. An introduction to survival analysis using Stata (2nd ed.). Stata Press: College Station, Tex. Ch. 1-3

Kennedy, section 17.4

Application n/a
Software stata: stset
Assignment n/a
Comments
Today we’ll focus on the basic concepts of hazard rate models, including:
Why OLS isn’t appropriate
Censoring and truncation
How to organize survival data in Stata

The two books from which the readings are drawn are excellent and deserve to be owned in their entirety.

Session 11 Hazard rate models II: Non-parametric approaches
Apr 26, 2010

Reading
Cleaves et al 2008, Ch. 8
Blossfeld et al 2007, Ch 3 (Optional and not on Compass)

Application n/a
Software n/a
Assignment Assignment 2 due
Comments
### Session 12 Hazard rate models III: Semiparametric models
Apr 28, 2010

**Reading**  
Cleaves et al 2008, p. 129-145 (included in last session’s readings)

**Application**  
n/a

**Software**  
n/a

**Assignment**  
n/a

**Comments**

### Session 13 Hazard rate models IV: Parametric models
May 3, 2010

**Reading**  
Blossfeld et al, Ch. 7

**Application**  
n/a

**Software**  
n/a

**Assignment**  
n/a

**Comments**

### Session 14 Hazard rate models V: Extensions and complication
May 5, 2010

**Reading**  
Cleaves et al, Ch. 15

**Application**  
n/a

**Software**  
n/a

**Assignment**  
n/a

**Comments**