Course description and goals:
The purpose of this course is to expose students to econometric techniques frequently used by economists working with cross-sectional and/or panel data. I will cover ten broad topics: i) panel data, ii) correlated errors, clustering and the “Moulton” problem, iii) instrumental variables estimation, iv) regression discontinuity, v) Non-linear optimization and maximum likelihood models, vi) discrete data, vii) limited dependent variables, viii) duration data, iv) matching estimators, and v) quantile regressions. For each topic, course time will be split evenly between theory and examples. I will first present a standard textbook treatment of the topic. Next, we will discuss a number of papers that have used the techniques outlined in class. Students will be expected to read the assigned papers and be able to discuss not only the econometric techniques used but also the basic economic issues as well. I also have sample STATA and/or Matlab programs that estimate each of the techniques we consider. On the class web page is a 20 page handout that outlines Stata. From your other first-year classes you should be intimately familiar with Matlab.

Textbook and course readings:
Required:


Recommended:


I have selected Greene for the class textbook, primarily because you used it the first semester. It is a solid cookbook that has extensive coverage of topics and a nice bibliography. The book is a little dated in that it covers lots of time series econometrics (that nobody uses anymore) and it does not cover more recent cross-sectional topics such as quantile regressions, propensity score matching, etc. Despite being
rather broad, the book is lacking on intuition. Therefore, the other required book is Angrist and Pischke’s *Mostly Harmless Econometrics*, a fantastic book that should be read cover-to-cover by any young applied micro economist. The book provides an excellent mix of statistical detail, econometric intuition and practical instruction. The topic coverage is limited but covers about 60 percent of the topics in this course, which are also the bulk of econometric tools used in the vast majority of applied micro economics. I wish there was an econometrics textbook this well done when I was in graduate school.

I have also listed two recommended textbooks. Wooldridge is a classic econometrics textbook that examines many cross-sectional topics in detail. He derives estimators, talks about their finite sample and asymptotic properties and provides some nice examples. The book has a great coverage of topics for applied micro and is a standard reference for applied micro economists. Cameron and Trivedi is a at reference for model applied micro econometrics. It covers both old (panel data, IV) and the new topics (bootstrapping, quantile regressions, clustered standard errors, etc.) and it has a nice blend of statistical intuition and mathematical rigor.

Electronic versions of the class readings are available on the course web page. Some of the links take you directly to JSTOR which is only accessible through a university IP address. Others are downloadable from my web page. To comply with federal copy write laws, these files are password protected with the login/password being your NetId/password combination.

**Prerequisites:**
A course comparable to a first-year graduate econometrics sequence or permission of the instructor.

**Expectations:**
Students are expected to attend class, to read the reading prior to class, to NOT be late to class, to participate in classroom discussions, to hand in assignments when due, and to NOT engage in academic dishonesty.

**Grading:**
Grades for the course will be based on weekly problem sets (25 percent of the course grade), a midterm (25 percent) and a cumulative final exam (25 percent) and a replication exercise (25 percent of the course grade).

*Exams:* The midterm exam is tentatively scheduled for Wednesday, March 7th, at the regular class time, while the final exam is scheduled for Thursday, May 10, from 10:30am - 12:30pm. I am going to allow some wiggle room on these dates and times. I need to arrange the first year test times with the other first year classes.

**Problem sets**
The problem sets will have two types of questions. The first will be exam-type questions that ask you to “prove” or “show.” For the second type of question, you will be given data and asked to estimate an econometric model. To assist you in your programming and to provide you with a better understanding of the techniques we will be discussing, I will distribute sample programs for each topic. The programs will be written in either Matlab or STATA and the programs are available on the class web page. You will be given anywhere from one to two weeks to complete each problem set, depending on the difficulty of the problem set. You may work on the problem sets in a group but each student is required to hand in their own answers.
Replication Project

Students will be required to replicate, or attempt to replicate, results from a paper published in a refereed journal. For this exercise, I want you to select a paper where the original data is available on the Web. So for example, any paper that uses data from the Current Population Survey, National Health Interview Survey, or the Census Public Use Micro Samples, would be great since this data is downloadable from various web pages. You are forbidden to email a researcher and ask for their data or their code. I want you to start with original sources and construct the sample, then try to reproduce the results. This assignment is due by noon on the final day of class. You cannot work on the replication exercise in a group but you can ask your colleagues for advice/suggestions. The grades for the assignment will be based not only on the clarity of your analysis but the quality of your writing. Please see the complete description of the project on the class web page.

Reading List
ECOE 60303
Spring 2012

I. A Brief Review of Linear Models

Greene, Chapters 2-4
Angrist and Pischke, Chapters 2-3
Wooldridge, Chapters 2-4

II. Panel Data: Fixed and Random Effects, Difference-in-difference models

Greene, Chapter 8 (The Generalized Regression Model)
Greene, Chapter 9 (Models for Panel Data)
Angrist and Pischke, Chapter 5
Wooldridge, Chapter 7 (Generalized Least Squares) and Chapter10 (Panel Data)


### III. Correlated Errors

Greene, Section 8.8

Angrist and Pischke, Chapter 8


IV. Instrumental Variables Models

Greene, Chapters 12 and 13 (IV and Systems of equations)

Greene, Chapter 15 (Generalized Method of Moments)

Angrist and Pischke, Chapter 4

Wooldridge, Chapters 5 and 6.


V. Regression Discontinuity Design

Angrist and Pischke, Chapter 7


**VI. Maximum Likelihood Estimation and Nonlinear Optimization**

Greene, Chapter 16 (Maximum Likelihood Estimation)

Greene, Appendix E (Computation and Optimization)

A simple example we will follow: Count data models: Greene, Chapter 25

Wooldridge, Chapter 12, Section 12.7 (Non-linear optimization)

Wooldridge, Chapter 13 (Maximum likelihood models)


**VII. Models with Discrete Dependent Variables**

Greene, Chapter 23 (Discrete Choice Models)

Greene, Chapter 25 (Count data models)

Wooldridge, Chapter 15 (Discrete Data), Chapter 19 (Count Data)


VIII. Truncated and Censored Data, Sample Selection

Greene, Chapter 24.

Wooldridge, Chapter 16 and 17.


IX. Duration Data

Greene, Chapter 25.

Wooldridge, Chapter 20.


X. Matching methods

Angrist and Pischke (2009), Chapter 3


**XI. Quantile regressions**

Angrist and Pischke, Chapter 7


