Econometric Theory and Methods II (90-907)
SPRING 2021

<u>Instructor</u>
Akshaya Jha
akshayaj@andrew.cmu.edu
Office: Hamburg Hall 2218
Phone: 608-347-2102
Office hours: by appointment

Lectures: Tuesday and Thursday, 1:30pm – 2:50pm, HBH 1006 (or remote)
Review sessions: Friday, 3:10pm-04:30pm, HBH 1006 (or remote)

Grading: 40% homework (approximately every two weeks)
                       30% final exam (take home and open book)
                       30% final project

Prerequisites: Statistical Theory for Social and Policy Research (90-905) and
                       Introduction to Econometric Theory (90-906), or equivalent PhD-level
                       econometrics with extensive coverage of linear regressions and sufficient
                       treatment of asymptotic theory. It is assumed that students are familiar
                       with <u>basic linear algebra, multivariate calculus, probability theory, and
                       statistical convergence concepts</u>.

Textbooks/Website

Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, 1st or 2nd edition. (required)

Cameron and Trivedi, *Microeconometrics: Methods and Applications,* 2005. (recommended)

Greene, *Econometric Analysis,* Sixth Edition, 2008. (additional supplement)

Course website is on Canvas. Readings and assignments will be posted there.

Computer Software

Many homework assignments will involve data analysis that requires specialized statistical
software. You are free to use any software that you like. We will provide instruction and
support for Stata, which is available on the Heinz PhD server and via the virtual lab. Some
assignments will require a symbolic programming language (i.e., not pre-packaged estimation
commands), and for this we will provide instruction and support for R. Some class examples
may be given in Matlab.

Course Description

This course covers a number of econometric models and techniques that are commonly used in applied microeconomics. The core topics include a general framework for estimators (which includes maximum likelihood and generalized method of moments), discrete outcome models, sample selection (and related limited dependent variable or switching models), duration and count models, time series models, panel data models, variance estimation (including clustering and the bootstrap), and non-parametric techniques. The course is designed for PhD students who have completed 90-906 (PhD Econometrics I) or an equivalent course.

Course Objectives

1. Understand the fundamental properties of M-estimators, both formally and intuitively.
2. Formulate appropriate econometric models for specific applications within the classes of data covered by the course, addressing issues of endogeneity and error correlation.
3. Derive properties of the estimators based on the above models.
4. Apply these estimators to data provided as part of course assignments.

Course Policies

Please turn in homework at the beginning of class when it is due; late assignments are not accepted. Students may work on the assignments in groups—in fact this is recommended—but each student must write their own submission and these cannot be direct copies of each other's work or be copied from any other source (e.g., something online). The final exam will be take home and open book. Any suspected incidents of cheating or plagiarism will be recorded with Heinz College administration at the same time the student is notified.

COURSE OUTLINE

1. Frameworks for Estimators (4 lectures)
   1. General theory for M-estimators
   2. Maximum Likelihood Estimation
   3. Generalized Method of Moments
   4. Numerical optimization methods
   5. Hypothesis testing
      1. (Non-linear) testing for parameters
      2. Specification Tests

2. Discrete Outcome Models (4)
   1. Binary Logit, probit, and linear probability models
   2. Specification issues: heteroskedasticity and distributional assumptions
   3. Endogenous explanatory variables (if time)
   4. Multinomial logit, nested logit, multinomial probit
   5. Ordered logit and probit (if time)

3. Selection Models (4)
   1. Censored and truncated dependent variables
   2. ML estimation of the above
   3. Motivation for selection – agent choices affect observations
   4. Selection model variants
   5. ML and two-step estimation
   6. Specification issues

4. Duration Models (1)
   1. Typical duration outcomes and data
   2. Hazard functions
   3. ML estimation with continuous or discrete time
   4. Unobserved heterogeneity and other specification issues

5. Time Series Models (3)
   1. Stationarity
   2. Autoregressive and Moving Average Models
   3. Model Selection and Fitting

6. Panel Data Models (2)
   1. Linear panel model framework and permanent unobserved heterogeneity
   2. Estimation with heterogeneity: fixed effects and first differencing
   3. Dynamic linear panel models: predetermined and endogenous explanatory variables
   4. Binary outcome panel models

7. Non-parametric and Semi-parametric Estimation Methods (3)
   1. Density Estimation
   2. Kernel Regression
   3. Robinson's Semi-parametric Estimator
   4. Non-parametric Hypothesis Tests

8. Bootstrap and Sub-sampling Methods (2)
   1. Bootstrap Methods for Independent and Dependent Data
   2. Sub-sampling Methods

9. Survey Weights and Variance Estimation (2)
   1. Weighted estimation: theory and practice
   2. Heteroskedasticity- and cluster-robust standard errors
   3. Bootstrapped standard errors
   4. Multiple hypothesis correction