PUBPOL 890.06 Advanced Statistics and Econometrics for Public Policy

Duke University, Sanford School of Public Policy

Fall 2015

Lecture: M/W 10:05 – 11:20 AM in Sanford 04

Lab/Recitation: F 1:25 - 2:40 in Sanford 04

Professor: Matt Harding

Office: Rubenstein 198

Office Hours: Wed. by apt. (see page 3)

Email: matthew.harding@duke.edu

TA: Danielle Vance-McMullen

Office: Rubenstein 283

Office Hours: Fri. 3:30-4:30 PM or by appt.

Email: danielle.vance@duke.edu

Overview:

This course develops the theory and applications of regression analysis, which is the primary tool for empirical work in public policy and economics. Emphasis is placed on techniques for testing hypotheses and estimating causal relationships, especially from non-experimental data. The course briefly reviews probability, statistics and the foundations of causal inference, including randomized control trials. We move on to focus on methods of drawing inference from non-experimental, cross-sectional data. The foremost among these is the basic linear regression model (OLS). We will learn how to estimate, test and predict single variable and multivariate relationships. We also examine common issues in implementing OLS regression, including measurement error, collinearity, functional form assumptions, heteroskedasticity, and omitted variable bias. We will then discuss estimation methods for when some of the assumptions of OLS do not hold, including instrumental variables regression, maximum likelihood estimation, and probit and logit models.

The course will cover both theoretical and practical issues and problems sets will contain extensive applications to real data and require the use of statistical software (Stata). The goal of the course is to provide students with a toolkit of methods for drawing causal inferences from cross-sectional data and an understanding of the common issues that can occur when applying these tools. It is especially targeted to the techniques that students at the Master's and Ph.D. level need to help them conduct policy-relevant empirical research as part of their degree-related academic interests or in their future careers.

Prerequisites:

This class is designed to be taken by MPP and MIDP students. Some MA Economics students or PhD students whose previous coursework has not covered these topics may also take the course. The course assumes a basic understanding of probability, statistics, and simple regression analysis, although these topics will be reviewed during the course. The course also assumes proficiency in mathematics up through calculus. Calculus is not a requirement for this course, although students with this background will likely identify how calculus techniques make solving the problems presented easier.

Course Expectations:

Preparation: Class preparation is important to success in this course. Prior to each class, students should complete the assigned readings from the book or another source. If you find the material difficult to grasp during your pre-class reading, you may prefer to skim this material before class and more thoroughly study it after class discussion.

Attendance and Participation: Students are expected to attend and actively participate in each class and recitation/lab meeting. For this course, participation in lecture will include "understanding check" questions that students should respond to using the classroom response system. Missing class is permitted for illness, family emergencies, religious holidays, and other similar important and unavoidable reasons. Please let the TA know before class if you are missing class for a permitted reason.

Assignments/Grading:

Problem Sets: There are 9 problem sets due throughout the semester. The problem sets will ask you to demonstrate your understanding of statistical concepts and require you to complete hands-on data analysis using instructor-provided datasets. These problem sets are graded. Half of the grade is based on a "good faith" effort to solve each of the problems; the other half of the grade is based on your answers to selected problems. Reviewing problem set answer keys should help you identify areas to improve before the exam.

Due dates for problem sets are listed on the syllabus course schedule, but may be extended and will be finalized by the instructor (or TA) as the course progresses. Due dates roughly align with the week a chapter has been covered in class. The time of submission will **always be 5:00 PM.**

Class Participation: The in-class response system will be used to track your class participation over the course of the semester. To earn full participation points, you must respond to each "understanding check" question posed in class or lab (unless your absence was excused).

Exams: There are two 75-minute, in-class midterm tests in this course. Each will be graded anonymously. All exams will be closed book. However, students will be allowed to carry a letter sized (8.5x11 inch) page "cheat-sheet" with them to the exams. You can write anything you want on the cheat sheet. Only one cheat sheet is allowed in each exam.

Replication Project: For your final project, you will replicate the analysis of a published paper using the original data and Stata. You will produce a memo describing your work. The replication project will be due during finals week. Additional details on the project will be available on Sakai.

Grade Weights:

- Problems Sets 30%
- Participation 10%
- Midterm 1 20%
- Midterm 2 20%
- Final Project 20%

Assignment- and Grading-related Policies:

Makeup Exams: The exam dates listed on the syllabus are not flexible. By signing up for this class you agree to be available for all exams. If you miss a midterm due to any valid reason, the course grade will be based on the performance in the remaining exams and homework. Valid reasons for missing a midterm exam are limited to events beyond your control that make it impossible to attend the exam. Written documentation (doctor's note etc.) needs to be submitted. Make-up exams are *not* offered

Re-grade Requests: Any requests for problem set or exam regrading must be submitted within 48 hours of the problem set/exam return notice (posted to Sakai). You must submit an email to the instructor and TA stating the specific reasons that you believe your exam should be regraded *and* a hard copy of your problem set/exam (photocopy is okay) should be delivered to the TA's office along with a hard copy of the email. Take into account that if you ask for regrading, your whole problem set/exam may be checked again. Your point total may increase but may also decrease as the result of regrading.

Late Assignments: Students are expected to submit assignments (including the final project) on time. Late assignment will not be accepted. You are always welcome to submit your assignment early.

Working in Groups: You will get the most out of this course and the problem sets if you think about and try to answer each question yourself before working with others. Students may work together on problem sets, and in fact many students find that working in groups is a very effective way to lean and practice the concepts from the course. However, each student must hand in his or her own answers. This means that you should explain concepts in your own words, show your own work for math problems, and write and comment your own Stata code. For the objectives of the course (and for your future career), it is essential for you to be able to analyze data and interpret the results on your own.

Office Hours:

Your instructor and TA want you to succeed in this course! Both the instructor and TA will hold weekly office hours. TA office hours are on a first-come, first-served basis. Professor office hours are on an appointment system. You may sign up for an appointment using the scheduling website, http://mattharding.simplybook.me/sheduler/manage/event/1/unit/1/.

Occasionally our speaking obligations or other academic conflicts will force us to move office hours from the ones indicated on the first page of the syllabus. When this happens, the new office hours will be announced in class and on Sakai.

Other Policies:

University policies on cheating, plagiarism, students with disabilities, etc. apply to this course.

https://gradschool.duke.edu/academics/academic-policies-and-forms/standards-conduct

https://web.duke.edu/equity/resources.html

http://policies.duke.edu/students/universitywide/index.php

Textbooks/Materials:

Principles of Econometrics by R. Carter Hill, William E. Griffiths and Guay C. Lim (Wiley, 4th Edition, 2011).

Using Stata for Principles of Econometrics by Lee C. Adkins and R. Carter Hill (Wiley, 4th Edition, 2011).

OpenIntro Statistics, by Diez, Barr, and Cetinkaya-Rundel (2nd Edition, 2012). The book is available for free via the web, may be printed via a PDF or may be purchased for $8.69 in printed form through Amazon.com.

Stata (statistical software package). Course problem sets require you to use Stata. Stata is available on the computer labs in Sanford and Rubenstein, as well as all campus OIT labs. You may also purchase a Stata license for your personal computer through Duke OIT: https://software.oit.duke.edu/comp-print/software.

Socrative (class participation). This course will use the Socrative app/interface to facilitate and track class participation. More information on Socrative will be provided on the first day of class. A smartphone or laptop is required to use the Socrative system. Please see me after class if this is an issue for you. You may download Socrative's app or login to the computer-based system here: http://www.socrative.com/apps.php

Additional course materials (assignments, etc.) will be posted on the course's Sakai website. Students can log in with their netid and password at http://www.sakai.duke.edu/

Course Outline:

(Dates are subject to change if classes are cancelled due to weather, etc.)

August 24th	**Class Overview:** Introduction/Syllabus/Socrative Overview/What is Econometrics? Hill et al.: Ch. 1.1-1.5
August 26th	**Probability I:** Random variables, discrete vs. continuous, PDFs, computing probabilities, expectation, variance, joint and conditional probability, independence, Hill et al.: P.1-P.2, P.6 Supplemental: Diez Ch. 2 & 3
August 28th (Lab)	**Intro to Stata:** Managing your work in Stata (log files, do files, help files), opening data, describing data, summary statistics, finding joint and conditional probabilities in Stata Adkins & Hill: 1.1-1.13
August 31st	**Probability II:** Covariance and correlation, Z-scores, the normal distribution, other important distributions, interquantile range, Bayes rule Hill et al.: P.3-P.5 Supplemental: Diez Ch. 2 & 3
September 2nd	**Inference I:** Hypothesis testing, p-values, significance, rejection region Hill et al.: Appendix C.1-C.7 Supplemental: Diez Ch. 4-6

September 4th (Lab)	**Intro to Stata II:** Creating and managing variables, using Stata for visualization/graphs, coding loops and writing functions, using globals and locals, merging datasets Adkins & Hill: 1.14, 1.18
September 7th	**The SV Linear Model I:** Hill et al.: Ch. 2 Supplemental: Diez Ch. 7 ***Problem Set #1 Due***
September 9th	**The SV Linear Model II:** Hill et al.: Ch. 2 Supplemental: Diez Ch. 7
September 11th (Lab)	**Regression 1:** Simple Linear Regression Adkins & Hill: Ch. 2.1-2.5, 2A.1,
September 14th	**The SV Linear Model III:** Hill et al.: Ch. 2 Supplemental: Diez Ch. 7
September 16th	**Hypothesis Testing Using SV-OLS I:** Hill et al.: Ch. 3 Supplemental: TBD ***Problem Set #2 Due***
September 18th (Lab)	**Regression 2:** Hypothesis Testing Adkins & Hill: Ch. 3.1-3.3
September 21st	**Hypothesis Testing Using SV-OLS II:** Hill et al.: Ch. 3 Supplemental: TBD
September 23rd	**Goodness-of-Fit and Modeling Issues I:** *(Note this topic is not on Midterm I)* Hill et al.: Ch. 4 Supplemental: TBD ***Problem Set #3 Due***
September 25th (Lab)	**Review for Midterm 1** Review of Problem Sets 1, 2, and 3; Review of Key Class Concepts
September 28th	**MIDTERM I**
September 30th	**Goodness-of-Fit and Modeling Issues II:** Hill et al.: Ch. 4 Supplemental: TBD
October 2nd (Lab)	**Regression:** Prediction, Goodness of Fit, Outliers, Transforming the Data, Multicollinearity Adkins & Hill: Ch. 4
October 5th	**Multiple Regression:** Hill et al.: Ch. 5 Supplemental: TBD

October 7th	**Hypothesis Testing with MV-OLS** Hill et al.: Ch. 6 Supplemental: TBD
October 9th (Lab)	**Regression:** Multi-variable regression, Interaction Terms Adkins & Hill: Ch. 5
October 12th	**NO CLASS- FALL BREAK** ***Problem Set #4 Due***
October 14th	**Hypothesis Testing with MV-OLS** Hill et al.: Ch. 6 Supplemental: TBD
October 16th (Lab)	**Regression:** Hypothesis testing with multi-variable regression Adkins & Hill: Ch. 6
October 19th (Lab)	**Indicator Variables and Treatment Effects:** Adkins & Hill: Ch. 7
October 20th	***Problem Set #5 Due***
October 21st	**Indicator Variables and Treatment Effects:** Hill et al.: Ch. 7 Supplemental: TBD
October 23rd	**No Class**
October 26th	**Instrumental Variables I:** Hill et al.: Ch. 10 Supplemental: TBD ***Problem Set #6 Due***
October 28th	**Instrumental Variables II:** Hill et al.: Ch. 10 Supplemental: TBD
October 30st (Lab)	**Instrumental Variables:** Wald Estimator, Instrumental variable regression, Weak instruments Adkins & Hill: Ch. 10.1-10.6
November 2nd	**Maximum Likelihood Estimation:** Hill et al.: Appendix C.8 Supplemental: TBD ***Problem Set #7 Due***
November 4th	**Maximum Likelihood Estimation:** Hill et al.: Appendix C.8 Supplemental: TBD
November 6th (Lab)	**Maximum Likelihood Estimation, Probit and Logit:** Adkins & Hill: C.10, Adkins & Hill: Ch. 16

November 9th	**Qualitative and Limited Dependent Variable Models I:** Binary Choice Hill et al.: Ch. 16 Supplemental: TBD ***Problem Set #8 Due***
November 11th	**Qualitative and Limited Dependent Variable Models II:** Binary Choice Hill et al.: Ch. 16 Supplemental: TBD
November 13th (Lab)	**Review for Midterm II** Review of Problem Sets 4 - 8, Review of Key Class Concepts **Problem Set #9 Due**
November 16th	**Summary of Course Takeaways**
November 18th	**MIDTERM II**
November 20st (Lab)	**Skills for Final Projects/Wrapping it up:** Exporting regression results to Excel, Re-shape, Collapse, Class evaluations
November 23th	**Summary of Course Takeaways**
November 25th	**NO CLASS- THANKSGIVING BREAK**
November 27th	**NO LAB- THANKSGIVING BREAK**
November 30th to December 7th	**READING PERIOD**
December 10th	**FINAL PROJECT DUE: Thursday, December 10 at 5PM**