

PPA 810-1
Quantitative Methods II
Spring 2012

David Popp
426 Eggers Hall
office ph: 443-2482
dcpopp@maxwell.syr.edu

Office Hours:
Tuesday 10:00-noon
Wednesday 10:00-11:30
or by appointment

Course Description: This course is the second course in the methodology sequence for Public Administration PhD students. The course focuses on the use of regression analysis for social science research. We will begin with a description of the properties and assumptions of the basic multivariate linear regression model using ordinary least squares (OLS), along with the statistical inference tools necessary for hypothesis testing. The course continues by examining the consequences of violating the assumptions of the OLS model. Techniques for dealing with such cases are at the heart of empirical research. We discuss several such techniques, including adjustments for heteroskedasticity and autocorrelation, and the instrumental variables technique.

Computer Resources: This course is designed to prepare you to use regression analysis in your own research, and to make you an educated consumer of the work of others. While the course includes several examples of such analysis, regression analysis is best learned through practice, rather than through a textbook and lectures. Homework assignments will include computer exercises that make use of the techniques covered in class. All such exercises are designed to be completed using Stata, which has become the most commonly used software package regression analysis. Copies of Stata are available on computers in CPR, in CEPA/CTIP, the Social Sciences grad bay, Eggers 040 and on one computer in both the PA grad bay on the 2nd floor and in Moynihan. It is also available in various computer labs throughout campus (<http://its.syr.edu/labs/>). In addition, Stata is available for purchase at a discounted rate for students through the following link:

<http://www.stata.com/order/new/edu/gradplan.html>

Note that, as Syracuse University is already a participating campus for the GradPlan, I have not registered this course separately. If you plan on purchasing a perpetual license, so that you can use Stata in your own research, I highly recommend purchasing Stata SE, rather than Stata IC, as Stata SE is able to handle larger data sets. For the purposes of this class, Stata IC should be sufficient for most exercises (although it may be necessary to find a computer with Stata SE for some exercises in which you use matrix algebra, rather than Stata's own commands, to analyze data). You should not purchase Small Stata, as it is limited to data sets with 1,000 observations or less. That will not be sufficient for many of the data sets we will use.

For those unfamiliar with Stata, it is worth your effort to learn, as you are likely to continue to use Stata for your own research. The Stata website includes several links to introductory tutorials, including one at the Syracuse University library. These can be found at:

<http://www.stata.com/links/resources1.html>

I have found the introduction provided by the London School of Economics to be particularly helpful.

Prerequisites: Students should have successfully completed PPA 810, Quantitative Methods I, or an equivalent course that discusses research design. Students should have basic familiarity with probability and statistical inference. The class will use both matrix algebra and differential calculus. I will provide a brief introduction to matrix algebra, focusing on the tools to be used in class. For those needing review, the text provides a brief introduction to differential calculus. Useful web sites include:

<http://www.intmath.com/Calculus/Calculus-intro.php>
<http://www.sosmath.com/calculus/calculus.html>.

If you need other resources, please see me.

E-mail: All students in the class are required to have an e-mail account and to check e-mail regularly. I will use e-mail to make announcements about class, and may distribute supplementary materials for assignments via e-mail. Not checking e-mail is not an appropriate excuse for missing these announcements.

Reading: The required textbook for this class is:

Introductory Econometrics: A Modern Approach, 4th edition, by Jeffrey M. Wooldridge, South-western Cengage Learning, 2009.

In addition, the following text is optional:

Mostly Harmless Econometrics: An Empiricist's Companion, by Joshua D. Angrist and Jörn-Steffen Pischke, Princeton University Press, 2009.

Both texts are available at the Orange Bookstore. Older editions are fine. The level of material covered in class will fall somewhere in between these two books. Wooldridge's *Introductory Econometrics* provides an intuitive introduction to econometric methods, but makes little use of calculus and matrix algebra. However, most advanced textbooks using such tools provide more detail than necessary for this course.

As such, I will occasionally supplement the readings from Wooldridge with material from Angrist and Pischke's book, which introduces econometric techniques in a narrative, rather than textbook, form. At times the mathematics in this book become advanced, so I will use it sporadically throughout the semester. While it is therefore optional, those of you who plan on making extensive use of regression analysis in your research would benefit from purchasing and reading the book. I will include the relevant chapters in the syllabus for those who would like to read them. In addition, I will also make use of journal articles to illustrate the concepts discussed in class. While some are listed on the syllabus below, I may add additional articles as the semester progresses. These articles are all available electronically through the Syracuse University Library.

Other textbooks: As noted above, there is no one “perfect” textbook for this course. Most graduate level econometrics texts are aimed at Economics Ph.D. students. They make heavy use of mathematics, and focus on proof of concept as well as application. However, as you go forward in your own research, you may find such texts a useful reference. Below I provide an annotated bibliography of alternative textbooks that you may find useful:

Marno Verbeek, *A Guide to Modern Econometrics: 3rd Edition*, John Wiley & Sons, 2008.

The level of this text is closest to what we will use in class. It makes use of calculus and matrix algebra, but focuses on applications, rather than proof of concept. It moves rather quickly, as it includes fewer examples than other books. While the text includes 10 chapters, most of the topics covered in this course are covered in chapters 1-5. However, it is affordable, as it is available in paperback on Amazon.com for less than \$60.

Christopher F. Baum, *An Introduction to Modern Econometrics Using Stata*, Stata Press, 2006.

This book is an introduction to using Stata for econometrics. It reviews the theory (using matrix algebra) and demonstrates how many of the techniques that we cover in class can be done using Stata. While it is not as comprehensive as a traditional textbook, it is a nice supplement, and the level of math use is comparable to what we cover in class. In fact, I use this book as a guide for preparing lectures and examples. It can be ordered from Stata for \$54 (<http://www.statapress.com/books/imeus.html>)

A. Colin Cameron and Pravin K. Trivedi, *Microeconometrics Using Stata*, Stata Press, 210.

A more advanced book that can complement Baum’s text, particularly for those planning on doing extensive regression analysis in their research. Like Verbeek, it moves quickly and includes several chapters that are beyond what we cover in this course. While the level of mathematics is similar to the more advanced econometric textbooks described below, the combination of theory and Stata examples in the presentation makes this book more accessible than other advanced books. It can be ordered from Stata for \$65 (<http://www.statapress.com/books/imeus.html>)

Peter Kennedy, *A Guide to Econometrics: 6th Edition*, Blackwell Publishing, 2008

This book is a classic reference for econometrics. Its style is unique. Each chapter focuses on a specific issue, and provides a short, intuitive explanation of the topic. More detailed technical notes are provided at the end of each chapter. There are older editions, but the newest is much more comprehensive.

Jack Johnston and John DiNardo, *Econometric Methods: 4th Edition*, McGraw-Hill, 1997.

This book is smaller than the other texts listed here, as its primary focus is technique, rather than applications. It is a good reference for regression techniques, with formal derivations using calculus and matrix algebra. Because it focuses on the math, it is rather dense, and probably most useful as a reference to pull off your shelf when looking for something specific, rather than reading as a text.

William E. Griffiths, R. Carter Hill, and George G. Judge, *Learning and Practicing Econometrics*, John Wiley & Sons, 1993.

This book was used as the primary textbook for this course for many years. Compared to other econometric texts, the math is a bit lighter, although it still relies on calculus and matrix algebra. The main difference between this text and others is that its focus is on doing empirical work. Thus, the emphasis is on technique, rather than on proofs. Unfortunately, it has not been updated, and thus does not include some topics included in newer books.

William H. Green, *Econometrics: 6th Edition*, Prentice Hall, 2007.

This text is often used in Ph.D. classes in economic departments. It is a very thorough reference, but more advanced than Griffiths, Hill, and Judge. It includes both applications and formal proofs.

Donald J. Treiman, *Quantitative Data Analysis: Doing Social Research to Test Ideas*, Jossey-Bass, 2009.

This book is less rigorous than the other books listed here. The introduction notes that the only math required is high-school algebra and an introductory statistics course. Unlike the other books here, its focus is on social sciences more broadly, rather than economics. The author is a sociologist. It covers both descriptive data and regression techniques. It is not as comprehensive as other texts. For instance, instrumental variables are only discussed on two pages, and neither heteroskedasticity nor autocorrelation are included in the index.

Academic Honesty: Students are expected to abide by the academic rules and regulations established by Syracuse University. These require students to “exhibit honesty in all academic endeavors. Cheating in any form is not tolerated, nor is assisting another person to cheat. The submission of any work by a student is taken as a guarantee that the thoughts and expressions in it are the student's own except when properly credited to another. Violations of this principle include giving or receiving aid in an exam or where otherwise prohibited, fraud, plagiarism, or any other deceptive act in connection with academic work. Plagiarism is the representation of another's words, ideas, programs, formulae, opinions, or other products of work as one's own, either overtly or by failing to attribute them to their true source” (*Syracuse University Bulletin* 2003-2004: p. 2). Of particular importance, while you are free to cite the views of others in your work, the final product must be *in your own words*, and any references to the works of others, whether directly quoted or merely paraphrased, must be cited. A good reference on the proper attribution of sources can be found at <http://www.dartmouth.edu/~sources/>. For more information on Syracuse University's academic integrity policies, see <http://academicintegrity.syr.edu>.

In compliance with section 504 and the Americans with Disabilities Act (ADA), Syracuse University is committed to ensure that “no otherwise qualified individual with a disability...shall, solely by reason of disability, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity...” If you feel that you are a student who may need academic accommodations due to a disability, then you should immediately register with the Office of Disability Services (ODS) at 804 University Ave., Room 309 3rd Floor, 315-443-4498 or 315-443-1371 (TDD only). ODS is the Syracuse University office that authorizes special accommodations for students with disabilities.

Grading: Your grade in this course will be based on participation in class (10%), homework exercises (25%), a referee report (25%) and a replication exercise (40%).

Homework assignments: Statistical analysis is best learned by doing. As such, there will be regular homework assignments throughout the semester. These will be a combination of written questions and data exercises using Stata. I will make any data sets needed for the assignments available on the G: drive. Homework assignments will be given in class. If you miss a class, it is your responsibility to find out if you missed any assignments or handouts. Not being present when an assignment was given out is **not** an acceptable excuse for missed or late work!

Replication Assignment: The major project for this class is a paper in which you replicate the results of a published research paper. You may choose any paper of interest to you, with the following guidelines: (1) the paper must make use of methods discussed in class, and (2) you may not replicate a paper using data sets from the Wooldridge text. Once you have chosen a paper, it must be approved by me.

To replicate the paper, you will need access to the paper's data. Some journals now require that authors make data available for replication purposes. These include the following:

- American Economic Review (<http://www.aeaweb.org/aer/index.php>, data sets included with articles)
- Applied AEA journals also make data available. Applied Economics and Economic Policy are good sources for relevant articles
- Journal of Peace Research (<http://jpr.sagepub.com/>, replication data at <http://www.prio.no/Research-and-Publications/Journal-of-Peace-Research/Replication-Data/>)
- Journal of Conflict Resolution (<http://jcr.sagepub.com/>; data sets included with the articles)
- Journal of Applied Econometrics (<http://qed.econ.queensu.ca/jae/>)

In addition, some journals have a policy requiring authors to make data available for replication, although the data is not readily available on the web site. Two examples are the Journal of Human Resources (<http://www.ssc.wisc.edu/jhr/>) and the Journal of Environmental Economics and Management (www.elsevier.com/locate/jee).

Finally, the following website offers links to articles and archives on replication in social science research:

<http://gking.harvard.edu/pages/data-sharing-and-replication>

Thus, you should first check the journal web site for any article that interests you. If that fails, you should contact the author directly to see if you can obtain the data (again, some journals require authors to do so if asked). I have been asked to do this myself, and it can be a good way for a young researcher to establish contact with a senior person in their field. In some cases, you may find that the data are publically available, although in those cases it may require some effort to clean the data. I would encourage each of you to talk with your faculty mentor for help finding an appropriate paper.

The write-up of your replication exercise should be similar to a research paper. Your introduction should state why the question is important. Based on the original paper, you should provide a brief literature review and describe the theory being tested. However, this can be brief, as I am *not* asking you to copy what has already been written. Paraphrasing is sufficient. You should describe the data and the methodology used to analyze the data. Please be certain to explain *why* the methodology used is appropriate (or, if you think it is not, please explain why, presenting evidence to support your claim). Your write-up of the results should be interpretive. You should consider any potential threats to the validity of the results, as well as how well the results are likely to apply outside of the sample in the paper. In your conclusion, you should consider any modifications and extensions of the work that may be justified.

Throughout the semester, there will be three key deadlines for the replication assignment:

1. **March 21 – Approval of paper and data set:** To ensure that the methods you will be replicating apply the techniques that we'll cover in class, you must show me the paper that you intend to replicate and get my approval. You should meet with me as soon as possible to discuss possible papers. Once you have my approval and have obtained the data, the first step is to prepare a table of summary statistics. Most published papers will include such a table. This is an opportunity to verify that the data you are using matches that used in the published paper. For this deadline, you should submit a brief write-up of your data summary, including a description of the variables included.
2. **April 23 & 25 – Replication presentation:** During the last week of class, each student will present their replication to the class. This will take the same format as a conference presentation. Each student will have 15 minutes to present their findings. We will then have 5-10 minutes of questions from the audience. You should have completed your analysis by the presentation, but you do not need a finished paper.
3. **May 3 – Final paper due:** The final paper is due after classes end. This allows time to respond to any comments you may receive at your presentation.

Referee report: The other major assignment is to prepare a referee report on a paper of your choosing. For this assignment, it is best to choose a working paper, representing a work in progress, rather than a published paper. The referee report is an important task for academics. When you submit articles for publication, the work will be refereed, usually by 2-3 anonymous referees, who will provide the journal editor with guidelines as to the submission's quality for publication. I will provide more details on this assignment, including potential sources for working papers, in a separate document. As there is not a final exam for this course, you may use the exam period to complete this assignment if you wish. As such, the referee report will be due on the last day of the exam period, which is **Wednesday, May 9**.

Course Outline

I. Introduction

Wooldridge, chapter 1-2, Appendices C.1-C.3 & D
MHE, chapters 1-2

II. The Multiple Regression Model

Wooldridge, chapter 3 & Appendices C.4 & E
MHE, chapter 3.1-3.2 & 3.4.3 (optional)

III. Inference & Large Sample Theory

Wooldridge, Chapters 4 & 5
McCloskey, Deirdre N. and Stephen T. Ziliak (1996), "The Standard Error of Regressions," *Journal of Economic Literature*, 34(1), 97-114.
Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes and Andrei Shleifer (2010), "Disclosure by Politicians," *American Economic Journal: Applied Economics*, 2(2), 179-209.

IV. Model Specification

Functional Form

Wooldridge, Chapter 6

Dummy Variables & Regression Discontinuity

Wooldridge, Chapter 7

MHE, chapter 6.1

Krueger, Alan B. (1993), "How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989," *Quarterly Journal of Economics*, 108(1), 33-60.

Rockoff, Jonah and Lesley J. Turner (2010), "Short-Run Impacts of Accountability on School Quality," *American Economics Journal: Economic Policy*, 2(4), 119-147.

V. Relaxing the Assumptions of OLS: Heteroskedasticity

Wooldridge, Chapter 8

MLE, chapter 8 (optional – this chapter is very technical)

Moulton, Brent R. (1990), "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *Review of Economics and Statistics*, 72(2), 334-338.

VI. Relaxing the Assumptions of OLS: Autocorrelation

Wooldridge, Chapters 10 & 12

VII. Specification & Data Issues

Wooldridge, Chapter 9

VIII. Instrumental Variables

Wooldridge, Chapter 15

MHE, chapter 4.1-4.4 (optional)

Baum, Christopher F., Mark E. Schafer, and Steven Stillman (2003), "Instrumental Variables and GMM: Estimation and Testing," Boston College Department of Economics Working Paper No. 545.

Bound, John, David A. Jaeger, and Regina M. Baker (1995), "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90(430), 443-450.

Levitt, Steven D. (1997), "Using Electoral Cycles in Police Hiring to Estimate the Effect of Policy on Crime," *American Economic Review*, 87(3), 270-290.

Murray, Michael P. (2006), "Avoiding Invalid Instruments and Coping with Weak Instruments," *Journal of Economic Perspectives*, 20(4), 111-132

Wooldridge, Jeffrey M. (2001), "Applications of Generalized Method of Moments Estimation," *Journal of Economic Perspectives*, 15(4), 87-100.

IX. Panel Data

Wooldridge, Chapter 13 & 14.1

MHE, Chapter 5.1-5.2 (optional)

Cohen, Alma and Liran Einav (2003), "The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities," *Review of Economics and Statistics*, 85(4), 828-843.

X. Cause & Effect?

"Cause and defect," *The Economist*, August 15, 2009, 68.

Holland, Paul W., "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81(396) 945-960.

Greenstone, Michael and Ted Gayer (2009), "Quasi-experimental and Experimental Approaches to Environmental Economics," *Journal of Environmental Economics and Management*, 57(1), 21-44.

Smith, V. Kerry (2007), "Reflections on the Literature," *Review of Environmental Economics and Policy*, 1(2), 300-318.