Political Science 794
Advanced Quantitative Analysis
Spring 2015

Professor Jonathan Hanson
johanson@syr.edu
324 Eggers Hall, 443-4032
Office Hours: Mondays & Thursdays 12:30-2:00 pm

This course has several goals. First, to provide you with the ability to use the classical linear regression model in a sophisticated and theoretically-informed way. Gaining both knowledge of the mathematical underpinnings of regression, and plenty of practice, is important in order to help you avoid being a naive user of this powerful tool. Second, to teach you extensions of the linear model to scenarios where the data violate its key assumptions, such as time-series and panel data. Critical to achieving this goal is the ability to recognize the nature of your data and choose the appropriate technique to analyze it. Third, to introduce you to a variety of estimation techniques that rely upon maximum likelihood estimation, such as situations in which the dependent variables of interest are categorical or count variables.

By nature of the material, this course is difficult. There will be times when you feel that you are just not getting it, and this is normal. Know this ahead of time, and keep working hard. There are no short cuts. Know also that I am on your side. My only goal is to help you learn, and I will push you because that is what I need to do in order for you to learn. The rewards will come if you persist!

Textbooks

The following book will be available for purchase at the Syracuse University Bookstore.

- Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 5th edition, (South-Western, Cengage Learning: 2013).

Older editions of this book may work. Although the will course will follow the presentation in Wooldridge, I strongly encourage you to make use of other resources. No single text will work well for all students, and access to multiple explanations with different emphases can be helpful. There are many good options to choose from, several of which are listed below:

- Richard A. Berk, *Regression Analysis: A Constructive Critique*, (Sage: 2004).

- Russell Davidson and James G. MacKinnon, *Econometric Theory and Methods*, (Oxford, 2003).
- Damodar Gujarati, *Basic Econometrics*, 4th edition, (McGraw-Hill/Irwin: 2002).
- William H. Greene, *Econometric Analysis*, 7th edition, (Prentice Hall: 2012).
- Peter Kennedy, *A Guide to Econometrics*, 6th edition, (Wiley-Blackwell: 2008).
- J. Scott Long, *Regression Models for Categorical and Limited Dependent Variables*, (Sage: 1997).

Other reading selections will be made available on the Blackboard site for the course.

A good math reference book may be useful at different points in the term. Here are some suggestions:

- Alpha C. Chiang, *Fundamental Methods of Mathematical Economics*, any edition, (McGraw-Hill).
- Will H. Moore and David A. Siegel, *A Mathematics Course for Political and Social Research*, (Princeton University Press: 2013).

Assignments and Grading

You will learn the material best when you use it. Problem sets will thus be assigned on a regular basis, accounting for 50% of the course grade. Group work on problem sets is permitted, and I encourage you to help each other.

The second portion of the course grade will come from two take-home exams worth 15% each. You are expected to work alone on these exams, with the aid of your notes and texts. The first exam will cover sections 1-6 of the course, and it will be due in early March, depending on our progress. The second exam will cover sections 7-12 and possibly some key material from the whole course. It will be due by May 5th.

The final 20% of the course grade comes from a paper that makes use of the statistical methods learned during the term. You have two options. First, you may write a "replication and extension" paper, in which you take a published work that interests you, attempt to replicate its statistical findings, and then extend the analysis in some way. Extensions include the use of additional data, different specifications, different estimation methods, etc. Second, you may write a research paper that involves an original statistical analysis. Ideally, this paper will be something that is useful for your research agenda. Additional details on this assignment, which is due on May 5th, will be provided as the semester gets underway.

Software

My recommendation is that you use either Stata or R for statistical work. Both are powerful and well-maintained. Learning a software package well requires an investment of time, and there is some path dependency once you become proficient in one package. Stata is an excellent program, and we will only scratch the surface of its capabilities in this course. It has very good documentation, a friendly user interface with both command line and menu options, and good tools for manipulating data. Since I trained using Stata, I am best able to help you with it. On the downside, Stata is expensive, and there a new version each year that requires more money if you want to stay up to date. Stata is available on campus lab machines, or you can purchase a GradPlan version at http://www.stata.com/order/new/edu/gradplans/campus-gradplan/. Note that there are different versions with different capabilities. For this course, Stata/IC should be fine. The Blackboard site has links to several Stata resources.

If I were starting from scratch right now, I would invest in learning R, which is open-source and downloadable at http://www.r-project.org. The use of R in political science has grown rapidly during the past decade, and it has become the statistical language of choice for many methodologists in the discipline. New packages are contributed by users regularly. The advantages are that it is free and very flexible. The disadvantages are that the learning curve is steeper, browsing and manipulating data is more difficult than in Stata, and the documentation is not as user-friendly as Stata's. The supplemental RStudio application, however, adds some features such as the ability to browse data more easily (http://www.rstudio.com/ide/). Here are some resources that may help you learn R:

- The R Manuals available online at http://www.r-project.org.
- User-contributed manuals available at http://cran.r-project.org/other-docs.html.
- Alain Zuur, Elena N. Ieno, and Erik Meesters, *A Beginner's Guide to R*, (Springer: 2009).
- John Fox, *An R and S-Plus Companion to Applied Regression*, (Sage: 2002).

When it comes to writing that uses math, such as your problem sets, learning LaTeX is clearly the best long-term option. LaTeX is an open-source typesetting program in which markup language that you enter into a text editor is compiled by a typesetting engine into a nicely-rendered pdf or postscript file. For those of you who are willing to invest time in learning LaTeX, I will gladly help you by supplying a template that you can use for your problem sets. Please see the documentation posted on the course Blackboard site.

- Windows users should start here http://miktex.org.
- Mac users should start here http://tug.org/mactex/.

Academic Integrity

The Syracuse University Academic Integrity Policy holds students accountable for the integrity of the work they submit. Students should be familiar with the Policy and know

that it is their responsibility to learn about instructor and general academic expectations with regard to proper citation of sources in written work. The policy also governs the integrity of work submitted in exams and assignments as well as the veracity of signatures on attendance sheets and other verifications of participation in class activities. Serious sanctions can result from academic dishonesty of any sort.

In particular, although I encourage you to work together on problem sets, the exams are a different story. I expect that you will not communicate with each other, or with any other individuals, about how to solve the problems on the exams.

Students with special needs

If you believe that you need accommodations for a disability, please contact the Office of Disability Services (ODS), http://disabilityservices.syr.edu, located at 804 University Avenue, room 309, or call 315-443-4498 for an appointment to discuss your needs and the process for requesting accommodations. ODS is responsible for coordinating disability-related accommodations and will issue students with documented disabilities "Accommodation Authorization Letters," as appropriate. Since accommodations may require early planning and generally are not provided retroactively, please contact ODS as soon as possible. You are also welcome to contact me privately to discuss your academic needs, although I cannot arrange for disability-related accommodations.

1 Introduction and Math Booster Shot

After a preview of the course and discussion of course mechanics, we will review some concepts in math and probability theory that will be useful throughout the course.

- Wooldridge, chapter 1, Appendix A-B and D.

2 The Classical Linear Regression Model

This section of the course, lasting about four sessions, will explore the mathematical underpinnings of ordinary least squares, using both single and multiple regression models. A key part of this section of the course is learning the assumptions under which ordinary least squares (OLS) is BLUE: the best linear unbiased estimator. Oftentimes one or more of these assumptions is not true in our data, and as we go on in the course it is important to understand the consequences of these violations and how we can address them.

- Wooldridge, chapters 2-3.

3 Inference and Asymptotic Properties

We work on interpreting regression results in order to draw inferences about the relationship between our variables of interest. Topics include: use of regression estimates for hypothesis testing, confidence intervals, testing restrictions on parameters, and the large-sample properties of OLS estimators, Roughly two sessions.

- Wooldridge, chapters 4-5.
- Richard A. Berk. 2004. "Causal Inference for the Simple Linear Model." In *Regression Analysis: A Constructive Critique*, chapter 5.

4 Functional Form and Specification Issues

The fact that the relationship between our independent and dependent variables is often non-linear does not mean we have to throw OLS out the window. Regression is an extremely flexible method, and we can make it work for us through various transformations of our data and changes in functional form.

- Wooldridge, chapters 6-7.
- Thomas Brambor, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14(1): 63–82.
- Achen, Christopher. 2005. "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where they Belong." *Conflict Management and Peace Science* 22: 327–339.

5 Data Issues

Many times our data do not cooperate, leading to phenomena with sometimes imposing names like multicollinearity, omitted variable bias, and measurement error. In this section of the course, our goal is to understand these phenomena and what to do (and what not to do) about them.

- Wooldridge, chapter 9.
- Peter Kennedy. 1998. "Violating Assumption Five: Multicollinearity." In *A Guide to Econometrics*, 4 ed., chapter 11.
- Kevin Clarke. 2005. "The Phantom Menace: Omitted Variable Bias in Econometric Research." *Conflict Management and Peace Science* 22: 341–352.

6 Non-Spherical Disturbances

In this section of the course, we learn that the formula for calculating the standard errors of regression coefficients is not valid when the variance of the disturbances (which come from that pesky stochastic part of the world) is not constant across the population after accounting for the independent variables. Additionally, OLS is no longer the BLUE. With appropriate techniques, we can detect and adjust for this phenomenon, which is known as heteroskedasticity.

- Wooldridge, chapter 8.

7 Regression with Time-Series Data

The standard assumption is that our data are independent and identically distributed. When we have time-series data, such as the results of monthly polls on presidential approval or some other case of repeated observations of the same object, this assumption is violated. The stochastic component of one observation may be correlated with the one preceding it and the one that follows. We spend about two sessions on this topic.

- Wooldridge, chapters 10-12.
- Jon C. Pevehouse and Jason D. Brozek. 2008. "Time-Series Analysis." In *The Oxford Handbook of Political Methodology*, chapter 19.

8 Regression with Panel Data

We have a panel when our data contain repeated observations of a sample of objects, such as a set of individuals who are surveyed periodically or a time series of cross-national data. In this scenario, we need to think about the non-independence of our observations both across time and space.

- Wooldridge, chapters 13-14.
- Nathaniel Beck and Jonathan N. Katz. 2011. "Modeling Dynamics in Time-Series-Cross-Section Political Economy Data." *Annual Review of Political Science* 14: 331–52.

9 Instrumental Variables Analysis

Oftentimes, there is strong reason to believe that the relationship between our dependent variable, and one or more of our independent variables, is endogenous. In such cases, our estimate of the coefficient on the independent variable will be biased. If we can find a third variable, however, which only has an effect on the dependent variable through the endogenous independent variable, we can identify the correct relationship.

- Wooldridge, chapter 15.
- Allison J. Sovey and Donald P. Green. "Instrumental Variables Estimation in Political Science: A Readers' Guide." *American Journal of Political Science* 55(1): 188–200.

10 Principles of Maximum Likelihood Estimation

We now move away from the classical linear regression model and its variants, all of which find the parameters of interest (i.e. the regression coefficients) though calculus. With Maximum Likelihood Estimation (MLE), these parameters are often found by a search algorithm which tours the parameter space and finds the values of the parameters that are "most likely" given the data. The power of MLE its flexibility. Many different functional forms are possible, facilitating analysis of many kinds of data for which linear regression is not suitable.

- Wooldridge, Appendix C.4
- Jae Myung. 2003. "Tutorial on Maximum Likelihood Estimation." *Journal of Mathematical Psychology* 47: 90–100.

11 Limited Dependent Variable Models

When our dependent variables come in the form of discrete categories, OLS can be problematic. For one thing, OLS produces predicted values of the dependent variable that fall into a continuous range rather than discrete values. Other problems arise with non-normality of standard errors and non-linearity of the effects of independent variables. With Maximum Likelihood Estimation, we can use functional forms that are designed for these data: probit analysis, logistic regression, and multinomial logit/probit. The purpose of this section is to become proficient in these techniques.

- Wooldridge, chapter 17.1-17.2
- J. Scott Long. 1997. "Binary Outcomes: The Linear Probability, Probit, and Logit Models" In *Regression Models for Categorical and Limited Dependent Variables*, chapter 3.
- Gary King, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses. Improving Interpretation and Presentation." *American Journal of Political Science* 44(2): 347–361.

12 Selection Models, Count Models, Duration Models, etc.

In this final section of the course, we will learn other advanced estimation techniques. The purpose of selection models is to address situations in which the cases that make it into the sample are different in important, unmeasured ways from those who do not, and these unmeasured factors are relevant for predicting the dependent variable. Count models are for cases in which the dependent variable is a count of the number of times something occurs. Duration models, also known as survival models, deal with situations in which we model the amount of time that some phenomenon lasts as a function of independent variables. If there is time, we may cover additional methods.

- Wooldridge, chapter 17.3-17.5.
- Gary King. 1989 "Event Count Models for International Relations: Generalizations and Applications." *International Studies Quarterly 33(2): 123–147.*