POLS 60810: Quantitative Methods I (Spring 2015)

Course Information
Class Time: Thurs., 3:30pm–6:15pm
Classroom: DeBartolo Hall 332

Professor:
Gary E. Hollibaugh, Jr.
Email: gholliba@nd.edu
Office: Decio Hall 409
Office Hours: Mon., 3:30pm–5:30pm
Wed., 10:00am–12:00pm
By Appointment

Teaching Assistant:
Steven McDowell
Email: smcdowe2@nd.edu
Office: Hesburgh Library (First Floor)
Office Hours: Tues., 10:00am–12:00pm

Course Description

This course provides an introduction to quantitative research methods in political science. After a brief discussion of the basics of statistical analysis and hypothesis testing, the first part of the course will focus on ordinary least squares (OLS) regression, its assumptions, and its extensions. In the second part of the course, we will focus on widely-used methods that are appropriate when the assumptions of OLS are violated, and especially on limited dependent variable models. We will try to strike a balance between theory and mathematics on the one hand and the practical application and interpretation of statistics on the other hand. We will discuss the theoretical rationale behind and mathematical underpinnings of various statistical methods, how to apply those methods to real political questions, and how to conduct and interpret analyses using standard statistical packages.

By the end of the course, students should be able to:

- Identify the most appropriate methodological techniques for analysis given a research question and available data, as well as identify, understand the implications, and offer resolution to various problems encountered during quantitative analysis.
- Conduct data analyses using the methodologies covered in the course. In particular, students should be able to diagnose and test empirical models, and apply the techniques for correcting models that violate statistical assumptions.
- Manage data and conduct analyses using Stata and `R`.
- Apply the tools learned to a question of your own choosing.

Software

In analyzing and examining data during this course, we will primarily make use of the R language, though Stata will be used as well from time to time. Using R for data analysis has a number of advantages. First and foremost, the software is free! Students can download their own copy of R (for Windows, OS X, or Linux) by going to the Comprehensive R Network (CRAN) webpage

and selecting the appropriate version. **Please do this before the first class session.** Second, writing your own statistical code forces you to think carefully about the statistical assumptions that underlie your modeling decisions, in a way that using a point-and-click interface would not. Finally, if you find yourself needing to change software in the future, it is much easier to transition from R to more user-friendly statistical software (such as Stata or SPSS) than the other way around. Of course, if you prefer, you can also conduct your analyses using Stata unless I specify otherwise; I will require some assignments be conducted entirely in R because (a) some things are much easier to accomplish in R; and (b) past experience has taught me that some students will refrain from using R unless forced. Regardless, I **STRONGLY** recommend becoming comfortable with both, and it will be a good idea to use both interchangeably throughout the course of the semester. **Note that all assignments, as well as your final papers, must include complete and well-commented code. Otherwise they will not be accepted.** Finally, because we will be making use of software during the semester, I strongly recommend bringing your laptops to class.

Books

There are no required books (or articles) for this course; instead, you will be required to seek out materials that are best suited for **you** and **your learning style**. That said, I will structure my lectures around the following books and the articles listed later in the syllabus. I will try to make most of them available to you as we go if you cannot find them easily online yourselves. Again, this setup is largely because no book is perfect for all students. I suggest you ask around, look at other syllabi online, and just browse the shelves at the library and used bookstores to find books that make things clear to you.

Applied Regression Analysis and Generalized Linear Models (2nd edition) by John Fox.
- Errata
- Stata walkthrough

Econometric Analysis by William H. Greene (any edition will do, but the chapter numbers in the syllabus are from the 7th edition).
- Errata

Interpreting and Using Regression by Christopher H. Achen.

Introduction to Probability by Charles M. Grinstead and J. Laurie Snell.

Regression Models for Categorical and Limited Dependent Variables by J. Scott Long.

You may want to acquire some sort of reference material for R and/or Stata. Below are two that I have found quite useful:

An R Companion to Applied Regression (2nd edition) by John Fox and Sanford Weisberg.

Microeconometrics Using Stata (Revised edition) by A. Colin Cameron and Pravin K. Trivedi.

Additionally, there are multitudes of other books that may be of use in understanding the models and theoretical foundations described in this course. A sampling is described below. This list is not exhaustive, and if you find something that works for you but is not listed, please email me!

Statistical Models: Theory and Practice by David A. Freedman.

Basic Econometrics by Damodar N. Gujarati and Dawn C. Porter.

A Guide to Econometrics by Peter Kennedy.

Introductory Econometrics: A Modern Approach by Jeffrey M. Wooldridge.

And finally, there are other alternatives in terms of reference material for R:

A Beginner's Guide to R by Alain Zuur, Elena N. Ieno, and Erik Meesters.

A First Course in Statistical Programming with R by W. John Braun and Duncan J. Murdoch.

An Introduction to R by Peter Haschke.

An Introduction to R by W. N. Venables, D. M. Smith, and the R Core Team.

Getting Started in Data Analysis using Stata and R by Princeton University's Data and Statistical Services Lab.

ggplot2: Elegant Graphics for Data Analysis by Hadley Wickham.

Introducing R by the Statistical Consulting Group at the UCLA Institute for Digital Research and Education.

Introduction to R: A First Course in R by Michael Clark

Introduction to R: A Second Course by Michael Clark

Quick-R.

R for Beginners by Emmanuel Paradis.

R for Stata Users by Robert A. Muenchen.

R Page at the UCLA Institute for Digital Research and Education.
· Data Analysis Examples

R Reference Card by Matt Baggott and Tom Short.

The R Inferno by Patrick Burns.

Using R for Data Analysis and Graphics: Introduction, Code, and Commentary by J. H. Maindonald.

And Stata:

Getting Started in Data Analysis using Stata and R by Princeton University's Data and Statistical Services Lab.

Stata Page at the UCLA Institute for Digital Research and Education.
· Data Analysis Examples

Grading

Class attendance is not explicitly required, though there is little prospect of success without it. That said, ability in quantitative methods varies and not everyone will be able to "master" the material. If you need a particular methodological skill to conduct your work, I do not want grades to dissuade someone from taking a course that can help them. You will receive feedback along the way on problem sets and know how you are doing in the class, but the final grade submitted in the books will be uninformative, so long as you make a good-faith effort. My goal is to give you the skills to conduct your own research, not weed people out of the program or send signals about hiring.

Problem Sets

There will be several problem sets. Absolutely no late assignments will be accepted. No assignment will be accepted without a code appendix or reproduction archive attached (or available to me online). No assignment will be accepted unless it is in Portable Document Format (PDF). No assignment will be accepted with cut and pasted computer output in the place of well presented and replicable figures and tables.

Original Research Paper

Conduct empirical research on the question of your choice using a coauthor (or coauthors; if there is an odd number of students in the class, I will allow for one—and only one—paper-writing group of three) and dataset of your choice. All students in the group will receive the same grade on the paper. Note that you must use some sort of linear model/generalized linear model in your analysis. The goal of the paper is to apply some method to, or develop one for, a substantive problem in your field of study. You should aim to produce a publishable article. You must send me your final paper groups by **February 5**. A research proposal that includes the dataset, the research question and details about the key variables of interest is due on **March 5**. The research paper is due on **April 23**. **Unless you receive prior authorization from the instructor in writing (which simply means you should email me ahead of time), late papers will be penalized ten percentage points for each day—or fraction thereof—that they are late.** This paper will include an introduction to the research question, a literature review, a description of the

data and measures, including descriptive statistics, an explanation of the empirical methodology, results, and the discussion/interpretation of the results. The style of the paper should be similar to that of published journal articles. If you are unsure of what this entails, consult recent issues of journals like the *American Political Science Review*, the *American Journal of Political Science*, the *Journal of Politics*, *World Politics*, *International Organization*, or others that publish your favorite quantitative research. (After all, this **is** a methods class!) If you need additional guidance, "Publication, Publication" by Gary King is useful reading; you should expect your final essay to comport with his guidelines (sans the discussion of replication, as you will be conducting your own original research).

As you work on your papers, you will also learn to write about data analysis in a way that sounds and looks professional by using either a WYSIWYG system like Word, OpenOffice, or WordPerfect—or a typesetting system like LaTeX—to produce documents that are suitable for correspondence, collaboration, publication, and reproduction. **No paper will be accepted without a code appendix or reproduction archive attached (or available to me online). No paper will be accepted unless it is in Portable Document Format (PDF). No paper will be accepted with cut and pasted computer output in the place of well presented and replicable figures and tables.** Although good empirical work requires that the analyst understand her tools, she must also think about how to communicate effectively: ability to reproduce past analyses and clean and clear presentations of data summaries are almost as important as clear writing in this regard.

Presentation

On **April 23**, all groups will present their original research papers to the class. Each group will be allotted approximately 30 minutes for presenting and each presentation will be followed by approximately 10 minutes' worth of questions from the class. I will provide more concrete times in the weeks before the presentation. At a minimum, each group will present the following about their assigned paper: the research question, the data used, the empirical methodology, and the findings. The presentation will include an interpretation of the findings and how these results answer the original research question. **If your paper is not turned in by the start of class on the day of presentations, you will not be able to present and you will receive a zero for the presentation. No exceptions (unless you receive prior authorization from the instructor in writing).**

Peer Review

You are to read and comment on a different group's paper (the group you are to review will be assigned by me) and to grade this group's paper according to certain guidelines we will provide. Your main objective is to give the group feedback on what parts of the paper were done well and why, as well as any changes and improvements need to happen in order for the paper to be published. Your comments on your fellow student's paper are due on **April 30** by **11:59pm**. **You will be evaluated based on how helpful, not how destructive, you are. If your paper is not turned in by April 23, you will not be able to review a paper, nor will you receive feedback on your own. No exceptions (unless you receive prior authorization from the instructor in writing).**

Typing Mathematical Notation

The most convenient way to write math is to use LaTeX, a system for preparing documents with mathematical symbols. If you learn LaTeX, you will be able to write mathematical formulae on your iPad, iPhone or any computing device with at least a text editor. Then, all you need to do is upload your work to typesetting program and produce a camera ready PDF. Using LaTeXallows you to signal to your readers that you know the language of mathematics and that you take quantitative research seriously. Obviously, this is a good signal to send. (Nevermind the implications that follow from much of the discipline making judgments about the quality of one's work solely on one's choice of writing tool.)

If all this is not enough to convince you to invest in the time to learn LaTeX, keep in mind that it is very easy to prepare a presentation in Beamer out of a document written in LaTeX. You can literally copy and paste the mathematical formulas from the paper version into the presentation. **Finally, keep in mind that all material you turn in must be typed. There will be no exceptions.** You do not have to use LaTeX (Microsoft's Equation Editor is reasonable as well), but you need to make an investment in *something*.

Collaboration

One of the best ways that people learn is by teaching and collaborating with others. In this class we facilitate collaboration in two different ways:

1. On the homework assignments, you will be allowed to work in groups of no larger than three. That does not mean, however, that students are allowed to turn in identical assignments. Work out the solutions together and then write up the final answer separately. Identical assignments will result in failure of the assignment.

2. Your final paper will be coauthored and you will choose your co-collaborator. This will give you the chance to write a journal-quality research paper with the help of your peers. You must work in groups of two, unless there is an odd number of students in the class; in that case, I will allow for one (and only one) group of three. All students in the group will receive the same grade on the paper. You must send me your final paper groups by **February 5**.

Policy on Plagiarism

According to University of Notre Dame's Academic Code of Honor for Current Graduate Students, "any activity that compromises the pursuit of truth and the advancement of knowledge may undermine confidence in the academic enterprise. Violation of integrity in research includes, but is not limited to plagiarism; deliberate fabrication or fabrication in proposing, conducting, reporting, or reviewing research." To learn more about where to draw the line between misjudgment and academic misconduct, I encourage you to consult the website above. To avoid engaging in plagiarism make sure that you never use words that are not your own without proper attribution. According to our Academic Code of Honor for Current Graduate Students, "those who appropriate the words or ideas of another, and who attempt to present them as their own without proper acknowledgment of the source, whether intentional or not, are committing plagiarism or intellectual theft."

Policy on Disabilities

Any student who has a documented disability and is registered with Disability Services should speak with the professor as soon as possible regarding accommodations. Students who are not registered should contact the Office of Disability Services.

Policy on Technology

This course relies heavily on access to computers, specific software, and the Internet. At some point during the semester you WILL have a problem with technology: your laptop will crash, a file will become corrupted, a server will go down, or something else will occur. These are facts of life, not emergencies. Technology problems will not normally be accepted as excuses for unfinished work. Count on "stuff" happening and protect yourself by doing the following:

- Plan ahead — start early, particularly if scarce resources are required.
- Save work often — at least every ten minutes.
- Make regular backups of files in a different location from the originals.
- Save drafts of work at multiple stages.
- When editing an image, set aside the original and work with a copy.
- Practice safe computing when surfing the web and checking email.
- On your personal computer, install and use software to control viruses and malware.

When submitting any assignment electronically in this course, you are responsible for any technological problems (*e.g.*, internet connection difficulties, corrupted files, etc.). To prevent problems along with the associated lateness penalties, you should submit papers well before the deadline and take proactive steps to ensure that the file was not corrupted (*e.g.*, check it after uploading to Sakai or copy yourself on emails and check the attached file). Again, please do not trust your computer to function correctly at the last minute.

Course Topics

Note: The course syllabus is a general plan for the course; deviations announced to the class by the instructor may be necessary. We may not cover all of these topics. Conversely, time permitting, other topics might be covered in this course.

Topic 1: The Central Limit Theorem and Statistical Inference

- Achen, Christopher H. 2002. "Advice for Students Taking a First Political Science Graduate Course in Statistical Methods." *The Political Methodologist* 10(2): 10–12.
- Gill, Jeff. 1999. "The Insignificance of Null Hypothesis Significance Testing." *Political Research Quarterly* 52(3): 647–674.
- Grinstead and Snell, Chapters 6, 7, 8 and 9

- Rainey, Carlisle. 2014. "Arguing for a Negligible Effect." *American Journal of Political Science* 58(4): 1083–1091.

Topic 2: Large Sample Theory, Regression, and the Linear Model

- Achen, Full Manuscript
- Fox, Chapters 5, 6, and 9
- Freedman, David A. 1991. "Statistical Models and Shoe Leather." *Sociological Methodology* 21(1): 291–313.
- Greene, Chapters 1, 2, 3, and 4
- King, Gary. 1986. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science* 30(3): 666–687.
- Leamer, Edward E. 1983. "Let's Take the Con Out of Econometrics." *American Economic Review* 73(1): 31–43.
- Long, Chapter 2
- Schrodt, Philip A. 2014. "Seven deadly sins of contemporary quantitative political analysis." *Journal of Peace Research* 51(2): 287–300.

Topic 3: Combinations of Predictors

- Berry, William D., Matt Golder, and Daniel Milton. 2012. "Improving Tests of Theories Positing Interaction." *American Journal of Political Science* 74(3): 653–671.
- Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14(1): 63–82.
- Braumoeller, Bear F. 2004. "Hypothesis Testing and Multiplicative Interaction Terms" *International Organization* 58(4): 807–820.
- Fox, Chapter 7

Topic 4: Problems with Data and Specification

- Achen, Christopher H. 2005. "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong." *Conflict Management and Peace Science* 22(4): 327–339.
- Clarke, Kevin A. 2005. "The Phantom Menace: Omitted Variable Bias in Econometric Research." *Conflict Management and Peace Science* 22(4): 341–352.
- Fox, Chapters 11, 12, 13, and 20
- Freedman, David A. 2006. "On the So-Called 'Huber Sandwich Estimator' and 'Robust Standard Errors.'" *The American Statistician* 60(4): 299–302.
- Greene, Chapters 6 and 9

- Harden, Jeffrey J. and Desmarais, Bruce A. 2011. "Linear Models with Outliers: Choosing between Conditional-Mean and Conditional-Median Methods." *State Politics & Policy Quarterly* 11(4): 371–389.

- King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95(1): 49–69.

- King, Gary and Margaret E. Roberts. Forthcoming. "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It." *Political Analysis*.

- Long, J. Scott and Laurie H. Ervin. 2000. "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model." *The American Statistician* 54(3): 217–224.

- Oneal, John R. and Bruce Russett. 2005. "Rule of Three, Let It Be? When More Really Is Better." *Conflict Management and Peace Science* 22(4): 293–310.

- Western, Bruce. 1995. "Concepts and Suggestions for Robust Regression Analysis." *American Journal of Political Science* 39(3): 786–817.

Topic 5: Instrumental Variables and Simultaneous Equations

- Bartels, Larry M. 1991. "Instrumental and 'Quasi-Instrumental' Variables." *American Journal of Political Science* 35(3): 777-800.

- Dunning, Thad. 2008. "Model Specification in Instrumental-Variables Regression." *Political Analysis* 16(3): 290–302.

- Gawande, Kishore and Hui Li. 2009. "Dealing with Weak Instruments: An Application to the Protection for Sale Model." *Political Analysis* 17(3): 236–260.

- Greene, Chapter 8

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

- Sovey, Allison J. and Donald P. Green. 2011. "Instrumental Variables Estimation in Political Science: A Readers' Guide." *American Journal of Political Science* 55(1): 188–200.

Topic 6: Longitudinal, Cross-sectional, and Panel Data

- Arceneaux, Kevin and David W. Nickerson. 2009. "Modeling Certainty with Clustered Data: A Comparison of Methods" *Political Analysis* 17(2): 177–190.

- Beck, Nathaniel. 2001. "Time-Series-Cross-Section Data: What Have We Learned in the Past Few Years?" *Annual Review of Political Science* 4(1): 271–293.

- Beck, Nathaniel and Jonathan N. Katz. 1995. "What To Do (and Not to Do) with Time-Series Cross-Section Data." *American Political Science Review* 89(3): 634–647.

- Clark, Tom and Drew A. Linzer. Forthcoming. "Should I Use Fixed or Random Effects?" *Political Science Research and Methods.*
- Fox, Chapter 16
- Greene, Chapters 11 and 20
- Honaker, James and Gary King. 2010. "What to Do about Missing Values in Time-Series Cross-Section Data." *American Journal of Political Science* 54(2): 561–581.
- Keele, Luke and Nathan J. Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14(2): 186–205.
- Steenbergen, Marco R. and Bradford S. Jones. 2002. "Modeling Multilevel Data Structures." *American Journal of Political Science* 46(1): 218–237.
- Wilson, Sven E. and Daniel M. Butler. 2007. "A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications." *Political Analysis* 15(2): 101–123.
- Zorn, Christopher J. W. 2001. "Generalized Estimating Equation Models for Correlated Data: A Review with Applications." *American Journal of Political Science* 45(2): 470–490.

Topic 7: Simulation Methods

- Efron, Bradley and Gail Gong. 1983. "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation." *The American Statistician* 37(1): 36–48.
- Fox, Chapter 21
- Greene, Chapter 15
- King, Gary, 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.
- MacKinnon, James G. 2002. "Bootstrap inference in econometrics." *Canadian Journal of Economics* 35(4): 615–645.
- Mooney, Christopher Z. 1996. "Bootstrap Statistical Inference: Examples and Evaluations for Political Science." *American Journal of Political Science* 40(2): 570–602.

Topic 8: Maximum Likelihood Estimation and Limited Dependent Variables

- Box-Steffensmeier, Janet M. and Bradford S. Jones. 1997. "Time is of the Essence: Event History Models in Political Science." *American Journal of Political Science* 41(4): 1414–1461."
- Dow, Jay K. and James W. Endersby. 2004. "Multinomial probit and multinomial logit: a comparison of choice models for voting research." *Electoral Studies* 23(1): 107–122.
- Fox, Chapter 14

- Green, Donald P. 2005. "Maximum Likelihood for the Masses." Unpublished Manuscript.
- Greene, Chapters 14, 17, 18, and 19
- King, Gary. 1988. "Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model." *American Journal of Political Science* 32(3): 838–863.
- King, Gary. 1989. "Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator." *American Journal of Political Science* 33(3): 762–784.
- Long, Chapters 3, 5, 6, 7, and 8
- Sigelman, Lee and Langche Zeng. 1999. "Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models." *Political Analysis* 8(2): 167–182.
- Winship, Christopher and Robert D. Mare. 1984. "Regression Models with Ordinal Variables." *American Sociological Review* 49(4): 512–525.