POLS 60837: Maximum Likelihood Estimation and Limited Dependent Variables (Fall 2016)

Course Information
Class Time: Mon., 6:30pm–9:15pm
Classroom: DeBartolo Hall 108

Professor:
Gary E. Hollibaugh, Jr.
Email: gholliba@nd.edu
Office: Decio Hall 409
Office Hours: Tues., 3:30pm–5:30pm (DeBartolo Hall 331 Computer Lab)
Wed., 3:30pm–5:30pm
By Appointment

Course Description

This course presents an overview of some regression-based methods widely used in political science today. The emphasis of the course is on models where the traditional assumptions of ordinary least-squares regression are violated, primarily in a cross-sectional context and because the dependent variable is non-continuous. The course will focus on maximum likelihood estimation of models of various kinds of limited-dependent and qualitative response variables. Specific topics covered will include binary response models, ordered response models, multinomial response models, event count models, duration models, censored/truncated regression models, and selection models. Throughout, we will focus on understanding the theoretical underpinnings of the various models and developing and evaluating applications of them to substantive problems in political science. Students will be asked to do data analysis exercises, to evaluate published research relying on quantitative techniques, and to do a research project on a topic of their own choosing.

The models discussed in this course are among the most widely used in political science today. It is not possible to function as an empirical political researcher without at least a passing familiarity with them; moreover, given the rapid and increasing rate at which more advanced models are being adopted in political science, these techniques increasingly represent a minimal level of statistical competence necessary to do publishable-quality quantitative work. Knowing these models, and using them appropriately and well, can increase your odds of landing a job, publishing books and articles, being granted tenure, and generally leading a happier and more fulfilling professional life.

Much of the material in this course is fairly technical and most of the readings will still require several readings to fully comprehend. Students are expected to have a basic acquaintance with differential and integral calculus, linear algebra, and probability/distribution theory. At the same time, it is impossible to learn statistics simply by reading books or articles and attending lectures.

Because of this incontrovertible fact, students will be required to complete problem sets regularly. Most of these exercises will be computer-based and use R and data I provide; some will replicate published work. The course will end with the submission of a piece of original research that applies the techniques used in class to a question of empirical importance.

Upon successfully completing the course, students will be able to:

- Understand and explain the logic behind maximum likelihood estimation and apply it in a variety of settings.
- Understand and explain the logic behind limited-dependent variable models and evaluate the validity of the assumptions in applied settings.
- Read, interpret, and evaluate statistical writing that uses limited-dependent variable models.
- 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.
- Manage data and conduct analyses R.
- Apply the tools learned to a question of your own choosing.
- Apply the theoretical foundations learned here in order to use (and develop) other models not covered in class.

The R Statistical Computing Environment

In analyzing and examining data during this course, we will use make use of the R language. 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. We will talk more about R, its benefits, and its potential problems during the course. Because we will be using R in class, those of you with laptops available should bring them. Of course, I will not tolerate the use of computers for anything other than class related work during active class time. **Note that all assignments, as well as your final papers, should include complete and well-commented R code. Otherwise they will not be accepted**.

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.

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

Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models by Julian J. Faraway.
· Errata

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

Unifying Political Methodology: The Likelihood Theory of Statistical Inference by Gary King.
· Errata

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!

Analysis of Ordinal Categorical Data by Alan Agresti.

Applied Regression Analysis and Generalized Linear Models by John Fox.

Categorical Data Analysis by Alan Agresti.

Econometric Analysis of Cross Section and Panel Data by Jeffrey M. Wooldridge.

Estimation and Inference in Econometrics by Russell Davidson and James G. McKinnon.

Event History Modeling: A Guide for Social Scientists by Janet M. Box-Steffensmeier and Bradford S. Jones.

Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models by Tim Futing Liao.

Limited-Dependent and Qualitative Variables in Econometrics by G. S. Maddala.

Linear Probability, Logit, and Probit Models by John H. Aldrich and Forrest D. Nelson.

Maximum Likelihood Estimation: Logic and Practice by Scott R. Eliason.

Microeconometrics: Methods and Applications by A. Colin Cameron and Pravin K. Trivedi.

Regression Analysis of Count Data by A. Colin Cameron and Pravin K. Trivedi.

Additionally, you may want to acquire some sort 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.

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

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.

If you want to use Stata for your own (*i.e.*, not for class) purposes, there are some excellent resources as well:

Generalized Linear Models: An Applied Approach by John P. Hoffmann

Regression Models for Categorical Dependent Variables Using Stata by J. Scott Long and Jeremy Freese

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

And finally, the Econometrics Academy is a useful resource for lots of models we will cover in class, as well as code for both R and Stata.

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, which will collectively count for 30% of the course grade. Absolutely no late assignments will be accepted, but your lowest grade will be dropped. **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), unless I instruct otherwise. 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

This paper will count for 50% of the course grade. 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 maximum likelihood model in your analysis. The goal of the paper is to apply some advanced 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 **September 26**. A research proposal that includes the dataset, the research question and details about the key vari-

ables of interest is due on **November 7**. The research paper is due on **December 5**. **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

This presentation will count for 10% of the course grade. On **December 5**, 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

The peer review will count for 10% of the course grade. 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 **December 11** by **11:59pm**. **The feedback you receive on your paper will not affect your final paper score. You will be evaluated based on how helpful, not how destructive, you are. If your paper is not turned in by December 5, 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 LaTeX allows 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 **September 25**.

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.

August 29: Introduction and Overview; R Programming and Monte Carlo Simulation

- **R Programming and Monte Carlo Simulation problem set distributed on August 29**
- Greene, Chapter 15
- Haschke, entire book
- *Introducing R*, full slideshow
- 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.
- 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.

September 5 and 12: Introduction to Maximum Likelihood Estimation

- **R Programming and Monte Carlo Simulation problem set due on September 12**
- **Maximum Likelihood Estimation problem set distributed on September 12**
- Faraway, Appendix A
- Green, Donald P. 2005. "Maximum Likelihood for the Masses." Unpublished Manuscript.
- Greene, Chapter 14
- King, Chapters 1–4
- Long, Chapter 2

September 19 and 26: Binary Response Models

- **FINAL PAPER GROUPS DUE ON SEPTEMBER 26 BY 6:30PM**
- **Maximum Likelihood Estimation problem set due on September 26**
- **Binary Response Models problem set distributed on September 26**
- Alvarez, R. Michael and John Brehm. 1995. "American Ambivalence Towards Abortion Policy: Development of a Heteroskedastic Probit Model of Competing Values." *American Journal of Political Science* 39(4): 1055–1082.

- Faraway, Chapter 2
- Greene, Chapter 17
- Hagle, Timothy M. and Glenn E. Mitchell II. 1992. "Goodness-of-Fit Measures for Probit and Logit." *American Journal of Political Science* 36(3): 762–784.
- Herron, Michael C. 1999. "Postestimation Uncertainty in Limited Dependent Variable Models." *Political Analysis* 8(1): 83–98.
- Huckfeldt, Robert and John Sprague. 1992. "Political Parties and Electoral Mobilization: Political Structure, Social Structure, and the Party Canvass." *American Political Science Review* 86(1): 70–86.
- King, Chapters 5.1, 5.2, 5.3, 5.5, and 5.6
- Long, Chapters 3 and 4
- Nagler, Jonathan. 1994. "Scobit: An Alternative Estimator to Logit and Probit." *American Journal of Political Science* 38(1): 230–255.
- Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25(1): 111–163.
- Vasquez, John A. 2004. "The Probability of War, 1816–1992." *International Studies Quarterly* 48(1): 1–27.

October 3 and 10: Ordered Response and Multinomial Choice Models

- **Binary Response Models problem set due on October 10**
- **Ordered Response and Multinomial Choice Models problem set distributed on October 10**
- 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.
- Faraway, Chapter 5
- Gelpi, Christopher F. and Michael Griesdorf. 2001. "Winners or Losers? Democracies in International Crisis, 1918–94." *American Political Science Review* 95(3): 633–647.
- Greene, Chapter 18
- King, Chapter 5.4
- Krehbiel, Keith and Douglas Rivers. 1988. "The Analysis of Committee Power: An Application to Senate Voting on the Minimum Wage." *American Journal of Political Science* 32(4): 1151–1174.
- Long, Chapters 5 and 6

- Long, J. Scott. 2012. "Regression Models for Nominal and Ordinal Outcomes." Unpublished Manuscript.
- Sanders, Mitchell S. 2001. "Uncertainty and Turnout." *Political Analysis* 9(1): 45–57.
- Winship, Christopher and Robert D. Mare. 1984. "Regression Models with Ordinal Variables." *American Sociological Review* 49(4): 512–525.

October 17: NO CLASS — FALL BREAK

October 24 and 31: Censored/Truncated/Selection Regression Models

- **Ordered Response and Multinomial Choice Models due on October 31**
- **Censored/Truncated/Selection Regression Models problem set distributed on October 31**
- Bernhard, William and Brian R. Sala. 2006. "The Remaking of an American Senate: The 17th Amendment and Ideological Responsiveness." *Journal of Politics* 68(2): 345–357.
- Dubin, Jeffrey A. and Douglas Rivers. 1989. "Selection Bias in Linear Regression, Logit, and Probit Models." *Sociological Methods and Research* 18(2–3): 360–390.
- Greene, Chapters 8 and 19
- King, Chapter 9
- Lemke, Douglas and William Reed. 2001. "War and Rivalry among Great Powers." *American Journal of Political Science* 45(2): 457–469.
- Lewis, David E. 2007. "Testing Pendleton's Premise: Do Political Appointees Make Worse Bureaucrats?." *Journal of Politics* 69(4): 1073–1088.
- Long, Chapter 7
- Sigelman, Lee and Langche Zeng. 1999. "Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models." *Political Analysis* 8(2): 167–182.

November 7: Final Paper Workshop

- **FINAL PAPER PROPOSALS DUE ON NOVEMBER 7 BY 6:30PM**

November 14 and November 21: Event Count and Duration Models

- **Censored/Truncated/Selection Regression Models problem set due on November 14**
- **Event Count and Duration Models problem set distributed on November 14**
- Alt, James E., Gary King, and Curtis S. Signorino. 2001. "Aggregation Among Binary, Count, and Duration Models: Estimating the Same Quantities from Different Levels of Data." *Political Analysis* 9(1): 21–44.

- Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(4): 1260–1288.
  - Erratum
- 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.
- Box-Steffensmeier, Janet M., Dan Reiter, and Christopher Zorn. 2003. "Nonproportional Hazards and Event History Analysis in International Relations." *Journal of Conflict Resolution* 47(1): 33–53.
- Carter, David B. and Curtis S. Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18(3): 271–292.
- Faraway, Chapter 3
- Fox, John. 2002. "Cox Proportional-Hazards Regression for Survival Data." Unpublished Manuscript.
- Greene, Chapters 18 and 19.4
- 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. "Event Count Models for International Relations: Generalizations and Applications." *International Studies Quarterly* 33(2): 123–147.
- 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.
- King, Chapters 5.7, 5.8, and 5.9
- Long, Chapter 8
- Teachman, Jay D. and Mark D. Hayward. 1993. "Interpreting Hazard Rate Models." *Sociological Methods & Research* 21(3): 340–371.

November 28: Writing Your Own Likelihood Functions: Linking Theory and Empirical Models

- **Event Count and Duration Models problem set due**
- Bas, Muhammet Ali, Curtis S. Signorino, and Robert W. Walker. 2008. "Statistical Backwards Induction: A Simple Method for Estimating Recursive Strategic Models." *Political Analysis* 16(1): 21–40.
- Bas, Muhammet A., Curtis S. Signorino, and Taehee Whang. 2013. "Knowing one's future preferences: A correlated agent model with Bayesian updating." *Journal of Theoretical Politics* 26(1): 3–34.

- Kenkel, Brenton and Curtis S. Signorino. 2014. "Estimating Extensive Form Games in R." *Journal of Statistical Software* 56(8): 1–27.
- Lewis, Jeffrey B. and Kenneth A. Schulz. 2002. "Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information." *Political Analysis* 11(4): 345–367.
- Nieman, Mark David. 2015. "Statistical Analysis of Strategic Interaction with Unobserved Player Actions: Introducing a Strategic Probit with Partial Observability." *Political Analysis* 23(3): 429–448.
- Poole, Keith T. and Howard Rosenthal. 1985. "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science* 29(2): 357–384.
- Ramsay, Kristopher W. and Curtis S. Signorino. 2009. "A Statistical Model of the Ultimatum Game." Manuscript.
- Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93(2): 279–98.
- Signorino, Curtis S. 2003. "Structure and Uncertainty in Discrete Choice Models." *Political Analysis* 11(4): 316–344.
- Signornio, Curtis S. and Kuzey Yilmaz. 2003. "Strategic Misspecification in Regression Models." *American Journal of Political Science* 47(3): 551–566.
- Whang, Taehee. 2010. "Empirical Implications of Signaling Models: Estimation of Belief Updating in International Crisis Bargaining." *Political Analysis* 18(3): 381–402.

December 5: Class Presentations

- **FINAL PAPERS DUE ON DECEMBER 5 BY 6:30PM VIA EMAIL OR SAKAI**
- **PEER REVIEWS DUE ON DECEMBER 11 BY 11:59PM VIA EMAIL OR SAKAI**