Maximum Likelihood Estimation for Generalized Linear Models

Professor Christopher Zorn

Course Description

This course is an overview of some regression-like statistical methods. The emphasis of the course is on likelihood-based models, and the bulk of our attention will be given to 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 include binary logit and probit, multinomial logit and probit, ordered logit and probit, and Poisson and other models for event counts, and (briefly) models for survival data.

The models discussed in this course are among the most widely used in the social sciences today. It is not possible to function as an empirical social scientist without at least a passing familiarity with these models; moreover, given the rapid and increasing rate at which more advanced models are being adopted in these fields, these techniques increasingly represent a minimal level of statistical competence necessary to do publishable-quality quantitative work. Put more briefly: knowing these models, and using them appropriately and well, can increase your odds of writing a strong (quantitative) dissertation, 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. While I have chosen readings that present the models as clearly and with as little jargon as possible, most of the material will still require several readings to fully comprehend. A solid understanding of scalar algebra is the only firm prerequisite for this class. Additionally, students are expected to have a nodding acquaintance with basic differential and integral calculus, linear algebra, and distribution theory. At the same time, it is impossible to learn statistics by reading books or articles and attending lectures. Because of this fact, students will be required to complete a series of lab exercises. Most of these exercises will be computer-based and make use of data I provide; some will replicate recent published work.

This syllabus is designed to provide an overview to the course. Clickable links are printed in green. Note that I am, by training, a political scientist, and many (read: most) of the examples we will use and discuss in class will come from that field. However, the methods covered here have wide applicability across the social sciences, including in sociology, economics, social psychology, anthropology, and other fields.
Course Readings

Required Text/Materials


Additional readings as necessary, all of which will be available on the course website and/or through JSTOR.

Strongly Recommended

Either:


or:


Other Useful/Recommended Readings


**Some Other Useful Resources**

The Political Methodology Section of the American Political Science Association was created to provide APSA members with an interest in political methodology with a forum in which to meet and discuss ideas. The section publishes a quarterly newsletter (*The Political Methodologist*), a quarterly journal on political methodology (*Political Analysis*), conducts a discussion list on topics relating to political methodology, and maintains an extensive electronic archive of papers, accessible via their homepage.

Also, just in case you weren’t aware of it, the Inter-University Consortium for Political and Social Research (ICPSR), at the University of Michigan, maintains an extensive archive of data in
the social and behavioral sciences. Much of it is accessible via their homepage.

Finally, a sample of courses similar to this one include those taught by:

- Fred Boehmke (University of Iowa).
- John Brehm (University of Chicago).
- Matt Golder (Florida State).
- James Honaker (Penn State).
- Simon Jackman (Stanford University).
- Gary King (Harvard University).
- Andrew Martin and Robert Walker (Washington University).
- Mike Ward (Duke University).
- Greg Wawro (Columbia University).

**Grading**

Grading will be based on six homework exercises. In most instances, exercises will be due 48 hours from being assigned (except when weekends intervene). Each exercise will be graded on the following scale:

- ✓+ indicates excellent,
- ✓ is good,
- ✓- is fair, and
- ✗ is unsatisfactory.

Homework exercises will generally involve estimation and interpretation of models on real/existing data, using statistical computer software (see below). Details for the homework assignments will be announced in class. Also, note that homework exercises should be submitted as hard (paper) copies. In the exceptional circumstance that you need to submit something electronically, **only PDF files will be accepted**, without exception. If you do not know how to create a PDF file, go learn. Now.
Software

You are welcome to make use of whatever statistical software you choose to complete the homework exercises, so long as the manner by which your results are generated and conclusions reached is transparent. However, due to the limits of instructor and TA time and patience, we will support only two software packages. Both are available on the machines in the Summer Program computing centers.

Stata

At the present time, Stata is probably the most widely-used statistical package in the social sciences. It is a powerful tool for data management, analysis, and display, and boasts some of the best manuals and on-line help of any existing software package. Stata is commercial software; the current version of Stata is 11.0, but previous versions (back to v. 8, at least) can also be used for the class. In the class notes, handouts, etc., Stata commands will appear in a fixed-width font and will be preceded by a period (“.”):

. regress Y X

Stata newbies may want to check out:

*Getting Started with Stata, Release 11*. 2009. College Station, TX: Stata Press.

Beyond this, the Stata homepage is a valuable resource for questions about the Stata statistical software. There are a number of useful Stata references on the web, including Scott Long’s page at IU and an excellent Stata “help page” sponsored by UCLA.

R

R is a statistical environment and high-level programming language for data analysis and display. It is effectively the GNU version of the S language; as such, it is free (both as in speech and as in beer) and open source. The current (early-June 2010) version of R is 2.11.1. R is an object-oriented language; unlike Stata (and most other statistical packages), it operates by assigning values to objects in the workspace. In the notes, handouts, etc., R commands will be preceded by a caret (“>”):

> my.results<-lm(Y~X)

The Comprehensive R Archive Network (CRAN) is the go-to spot for all things R-related. I cannot begin to list all the R-related resources available on the web; for newbies, however, it might be useful to check out the Introduction to R, this page in getting data into R, and the various R “cheat sheets” here, here, and here. Stata users who are interested in learning R should check out the Moving from Stata to R page at the R Project’s wiki.
Course Schedule

Readings should be completed prior to coming to class on the assigned day. Note that we will not, in general, hew closely (or at all) to the readings themselves, other than topically. Links are to DOIs or to stable PDFs at JSTOR. References to Greene are to the Fifth Edition (2003), which your instructor slightly prefers. I won’t assign readings from either Long and Freese (2006) and to Faraway (2006), but students should consult the relevant parts of those texts for software guidance (depending on whether they are using Stata or R, respectively).

June 22: Overview, Notation, and an Introduction to Estimation

- Readings
  - Required:
    - None.
  - Recommended:
    - None. (Read Long, Chapter 1, for background).

June 23: Maximum Likelihood: Derivation and Properties

- Readings
  - Required:
  - Recommended:
    - Greene (2003), §17.4.
    - King (1989), Chapter 4.

June 24: Maximum Likelihood: Estimation and Inference

- Readings
  - Required:
    - Long, pp. 52-61.
  - Recommended:
    - Greene (2003), §E6.
June 25: Testing, Testing...

- **Readings**
  - **Required:**
    - Long, Chapter 4.
  - **Recommended:**
    - Greene (2003), pp. 484-496.

- **Exercise 1:** Use MLE to estimate the parameters of a linear regression model.

June 28: Binary Response Models, I

- **Readings**
  - **Required:**
    - Long, pp. 34-52.
  - **Recommended:**
    - Eliason, pp. 39-45.
June 29: Binary Response Models, II

- **Readings**
  - **Required:**
    - Long, pp. 61-112.
  - **Recommended:**

- **Exercise 2:** Estimate and interpret binary logit and probit models.

June 30: Binary Response Models, III

- **Readings**
  - **Required:**
  - **Recommended:**
July 1: Binary Response Models, IV

- **Readings**

  - **Required:**
  
  - **Recommended:**

- **Exercise 3:** Estimate and interpret heteroskedastic and bivariate probit models.

July 2: Ordered Response Models, I

- **Readings**

  - **Required:**
  
  - **Recommended:**

July 5: Ordered Response Models, II

- **Readings**

  - **Required:**
o **Recommended:**

- **Exercise 4:** Estimate and interpret ordered logit and probit models.

**July 6: Multinomial Choice Models, I**

- **Readings**
  - **Required:**
  - **Recommended:**

**July 7: Multinomial Choice Models, II**

- **Readings**
  - **Required:**
Exercise 5: Estimate and interpret some multinomial choice models.

July 8: Event Count Models, I

Readings

Required:

- Long, pp. 217-238.

Recommended:

- Cameron and Trivedi (1998), Chapter 3.
July 9: Event Count Models, II

- Readings
  - Required:
    - Long, pp. 239-250.
  - Recommended:

- Exercise 6: Estimate and interpret some models for event counts.

July 12: Bringing It All Together: Generalized Linear Models

- Readings
  - Required:
  - Recommended:
    - Gill (2000).

- Readings
  - **Required:**
  - **Recommended:**

July 14: Extensions: Models for Sample Selection

- Readings
  - **Required:**
  - **Recommended:**
July 15: Extensions: Item Response Models

- **Readings**
  - **Required:**
  - **Recommended:**

July 16: Wrap-Up

- **Readings**
  - **Required:**
    - None.
  - **Recommended:**