PS271: Quantitative Methods II

Prerequisites: PS204B, PS270

Winter 2011
Fridays 3:00-5:50, SSB 353

Langche Zeng
lazeng@ucsd.edu
Office Hours: W 2-4, SSB399

Course Description

This course builds on PS204B (and PS270) and covers methods and models for a variety of data types frequently encountered in empirical work, including discrete choice data, ordinal and count data, duration/survival data, truncated/censored/sample selected data, and times series cross section or panel data. Likelihood and Bayesian inference and simulation methods are introduced. Issues and methods for causal inference are discussed, and selected topics such as machine learning methods are introduced as time permits. The course is application oriented, with emphasis on the understanding of the assumptions, the estimation and interpretation of various models, but maintains a level of analytic rigor necessary for a solid understanding of the inner working of the methods.

Course Requirements

Evaluation of course work will be based on exercises and in class participation (such as problem solving, paper presentation/discussion), peer feedback/evaluation, and a group research project, the distribution of weights being approximately 40%/15%/45%. It is of vital importance that you make every effort to attend every class meeting. The project should apply some methods/models relevant to the course to some substantive problem in the field of your interest (or develop some new methods.) Your team must consist of members of this class and all papers must be coauthored. The project should be finished, in the form of (mainly the data/methods/results/analysis part of) a potentially publishable research paper, by the last class meeting (3/11), when you will submit the paper and present your work to the class. All class work submissions should be in hard copies unless otherwise indicated.

Software

We will primarily use R for computation, and you are strongly encouraged to use LaTeX for typesetting and Emacs (or XEmacs ; also useful is ESS) for text editing. All are available on the web for free download.

Books and Internet Resources

You can find the following books in the UCSD bookstore:

--Cameron, C. and P.K. Trivedi, *Microeconometrics: Methods and Applications*. Cambridge University Press, 2005. website.

--Faraway, Julian J. Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Chapman and Hall/CRC, 2006. website

Cameron and Trivedi is quite comprehensive and can serve you not only during the course but also for years to come as a good econometrics reference book. Faraway provides detailed R examples for using a number of models covered in the course.

The following books are on reserve at the library:

--Angrist, J. and Pischke, JS. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2009

--King, Gary. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Ann Arbor: University of Michigan Press. 1998.

--Long, Scott. *Regression Models of Categorical and Limited Dependent Variables*. Sage Publications, 1997.

--Morgan, S. and Winship, C. *counterfactuals and Causal Inference*. Cambridge University Press, 2007.

Long (1997) is an exceptionally clear and easy to read exposition of models for nominal/ordinal/count/censored data. King (1998) is a political methodology classic covering a number of models we discuss. The other two books are primarily concerned with issues in causal inference.

Reference on machine learning:

--Hastie, Trevor, Robert Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer, 2009. website

A highly recommended R reference book is:

--Adler, J. *R in a Nutshell*. O'Reilly, 2009.

If you use Stata heavily, the companion book for the Cameron and Trivedi text is the best of its kind:

--Cameron and Trivedi. Microeconometrics Using Stata. 2010. Stata Press. website

There are extensive online resources for learning R, such as:

--Venables, William, D.M. Smith, and the R Development Core Team. An Introduction to R

An authoritative introduction to the language.

--JH Maindonald. Using R for Data Analysis and Graphics.

Helpful notes demonstrating essential data analysis and graphics using R.

--Imai, Kosuke, Gary King, and Olivia Lau. Zelig: Everyone's Statistical Software.

Zelig provides a unified syntax for using various R packages and for obtaining quantities of interest and uncertainty measures from various models.

Course Plan

The research project: You should find your coauthors and locate your data sets (if you don't already have one) as soon as possible. For comprehensive data resources, see for example the Social Science Data Services on campus, Micah Altman's The Impoverished Social Scientist's Guide to Data Resources , Paul Hensel's International Relations Data Site , and the Comparative Politics Data at U. Michigan. By the end of the 5th week (2/4), I expect a 2 page research proposal outlining the motivation of your project, the methods/models you plan to use, the status of data work, and expected contributions to the literature. Two weeks before the last class meeting (2/25), I expect a concise draft including key findings in tables/figures. We'll discuss the class projects along the way, but at these two points, each team will be paired with another for guaranteed peer feedback. The "final" paper is due by the last class meeting (3/11). Given the nature of the course the term paper should focus on the methods part (data/coding/models/results), with a couple of additional pages for providing the context. (At the last class meeting, you'll also submit a sheet of paper rating your experience with your teammates and the helpfulness of the peer feedback you have received. This is for my info only, as one factor for the course grades.)

Papers for class use: You are encouraged to suggest 1-2 papers to use in class. The papers should apply methods/models relevant to the course to some political science data. They can be of particular relevance to your project/field, or of general interest to the class. Please send the suggestions to me in pdf. You will take turn presenting/leading discussions on the application papers (focusing on the methods/models).

Topics: The following is a tentative plan for the course content. Some of the topics may take more or less than planned time, and we may adjust the plan according to the needs and pace of the class. Suggested readings from the books are noted below. Papers for class use will be sent to you a week in advance. Lecture notes will be made available along the way.

1. Introduction to the course and review/overview of fundamental concepts

Cameron and Trivedi, chapters 1-3; King, chapters 1&2.

2. Likelihood inference and Bayesian methods. Data generating processes.

Cameron and Trivedi, chapters 5&13; King, chapters 3&4; Long, chapters 1&2, 3.6, 4.1.

3. Simulation, Bootstrapping, R and Zelig

Cameron and Trivedi, 11, 12; Faraway, chaps. 1, 6, appendix A; Imai et al. Zelig manual. Venables and Ripley, 1-4.

4. R and Zelig (continued)

5. Binary and multiple choices; ordinal data; Interpretation.

Cameron and Trivedi, chapters 14 & 15. Faraway, chaps 2 and 6. Long, chapters 3-6; King, chapter 5

6. Selection bias, censoring and truncation; Event count data.

Cameron and Trivedi, chapters 16 & 20; Faraway, chap.3. King, chapter 5 & 9; Long, chapters 7 & 8.

7. Duration/event history data; Time series cross section/panel data; Multilevel modeling

Cameron and Trivedi, chap. 17, part V. Faraway, chaps 8-10.

8. Causal Inference: causal structure, nonparametric matching, model dependence. Treatment effects model, DID, regression discontinuity design, Instrumental variable methods.

Cameron and Trivedi, chapters 2 and 25; Morgan & Winship, Angrist & Pischke 4-6.

9. Causal Inference (continued). Prediction, machine learning, neural networks.

Hastie et al., 2, 11; Faraway, 14.

10. Group project presentations

Communication

I have a profound high frequency hearing loss that's not effectively helped by hearing aids, and I rely on speech reading and email for communication. This will certainly cause some inconvenience to you (e.g., I may ask you to repeat what you say, sometimes more than once; I cannot follow group discussion; In general I avoid the phone; etc.) I appreciate your understanding and accommodation. I will be compiling a class email list. Please send me your email address if it differs from the official UCSD address in the class roster.
