# PS271B: Quantitative Methods II

Prerequisites: PS204B, PS270

Spring 2010
Fridays 3:00-5:50, SSB 104

Langche Zeng
Zeng@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 that are not dealt with or dealt with adequately in the first course, including discrete choice data, ordinal and count data, duration/survival data, truncated/censored/sample selected data, and times series cross sectional 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), a midterm exam, and a group research project, the distribution of weights being approximately 30%/30%/40%. It is of vital importance that you make every effort to attend every class meeting, especially because none of the exsiting textbooks provides a perfect match to the materials we cover in class. You have the option of taking a final exam in place of the research project, in case the latter falls through for unforeseen reasons. 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 (6/4), when you will submit the paper and have an opportunity to 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 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 Univeristy 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.

Dedicated R users will find this classic of lasting value:

--Venables, William and Brian D. Ripley. *Modern Applied Statistics with S*. 4th Edition. Springer, 2002. website

If you use Stata a lot, 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, a few are here:

--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.

Some other useful references we draw on to various extent that are good resources for general consultation or further reading:

--Greene, William. *Econometric Analysis,* 6th Ed. Prentice Hall. 2008. website

A standard econometrics reference, comparable to Cameron and Trivedi (2005) in depth and is also quite comprehensive. Cameron and Trivedi (2005) has a broader and more up to date coverage of some practically important topics in microeconometrics (such as causal inference)

--Gelman, Andrew et al. *Bayesian Data Analysis*. 2nd ed. Chapman & Hall/CRC, 2004. website

--Gill, Jeff. *Bayesian Methods: A social and Behavioral Sciences Approach*. 2nd ed. Chapman & Hall/CRC, 2008. website

--Jackman, Simon. *Bayesian Analysis for the Social Sciences* . Wiley, 2009.

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

A good introduction to modern prediction and classification methods such as neural networks, support vector machines, classification trees and boosting.

--Pearl, Judea. *Causality: Models, Reasoning, and Inference*. Cambridge University Press. 2nd ed. 2009. website

A seminal work on causal inference.

## 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 (4/30), 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 (5/21), I expect a concise draft including key findings in tables/figures (or notification of your decision to take a final exam.) The "final" paper is due by the last class meeting (6/4). Given the nature of the course the term paper should focus on the methods part (data/coding/models/results..), with a few additional pages for providing the context and making sense of the hypotheses. We'll discuss the projects in class along the way.

*Supplementary papers for class use:* If you'd like to suggest a paper of special interest to use in class, please send me a pdf file. We'll use with other papers if possible. You will take turn presenting/leading discussions on the application papers (focusing on the methods).

*Topics:* The following is a tentative plan for the course content. 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. Any updates to the syllabus will be sent by email and/or posted here at http://dss.ucsd.edu/~lazeng/ps271.

1. Introduction and review/overview of fundamental concepts: Measurement, Modeling, Identification, Estimation, Inference; Prediction vs. Causal inference; Parametric, semi-parametric and nonparametric methods; Likelihood theory and Bayesian inference; Optimization vs. MCMC posterior sampling.

Cameron and Trivedi, chapters 1-3, 9, 13. King, 1 and 2.

2. Likelihood inference. Data generating processes. Specification of stochastic components and systematic components in a model. Bayesian methods.

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

3. R and Zelig; simulation; bootstrapping.

Faraway, chaps. 1, 6, appendix A. Cameron and Trivedi, 10, 11, 12. Imai et al. Zelig manual; Venables and Ripley, chapters 1-4.

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

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

5. Selection bias, censoring and truncation; Event count and duration/event history data

Cameron and Trivedi, chapters 16, 17 & 20; Faraway, chap.3. V & R 13, King, chapter 5 & 9; Long, chapters 7 & 8;

6. Midterm

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

Cameron and Trivedi, part V. Faraway, chaps 8-10. V&R chap. 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, Nonlinear models, Neural networks

Faraway, chaps. 11,12,14. Hastie et al, chapters 2, 7, 8, 11.

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 you cannot reach me by 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.

---