Research in Mathematical Political Science: Crafting Problem-Specific Estimators

Political Science 867
Winter 2002
Dr. Sanford Gordon

I. Introduction

The purpose of this course is to give students the tools to develop their own methodological techniques to address the substantive political problems in which they are interested. Too often, political science falls victim to "flavor of the month syndrome." Practitioners with a good statistical package and passing familiarity with the "next big thing" attempt to cram a political science dataset into a statistical model that someone else has developed for a different problem in a different field.[1] Too often, it is either ill-suited or methodological overkill.[2]

We will adopt an approach in which the methods one employs act in service to the theory that one has deduced, and not the other way around. Now that you have completed the introductory sequence at Ohio State, you are in a position to think more carefully about data generating processes and relationships between variables, and to derive *problem-specific* methodological approaches. During most of the course, we will get "under the hood" of different techniques to assess how theories about politics might suggest appropriate estimators.

Most of the techniques we will discuss are maximum likelihood estimators, though we will discuss others as well, including non-parametric and Bayesian approaches. Throughout the course and in a variety of contexts, we will rely heavily on Monte Carlo simulation. Simulation will pervade all aspects of the course: as a means of solving complicated probability problems, of evaluating the properties of estimators, of estimating parameters, and of presenting results.

One might say that good empirical social research unfolds as a series of four questions.

1) What is the *underlying social process* responsible for generating my data?
2) How can I operationalize this process with the simplest possible model?
3) What constitutes an appropriate critical test of my hypotheses regarding the underlying social process?
4) How can I present my findings in a way that both methodologists and non-methodologists can easily understand?

It is my hope that this class will give you the tools to facilitate answering 2-4. Answering the first requires imagination, intelligence, and patience, three things that cannot easily be taught.

---

[1] For purposes of disclosure, I count myself among the guilty.
[2] Usually, our econometric theory skills are sufficiently poor that our ability to design ever-more complex statistical models far outstrip our ability to understand the fundamentals of those models -- questions like large sample properties and whether their parameters are even identified.

II. Contact Information

Professor:	Sanford Gordon	2114 Derby Hall Phone: 292-7839 E-mail: gordon.256@osu.edu Web: http://psweb.sbs.ohio-state.edu/faculty/sgordon/ Office Hours: W 3:30-5:30, or by appointment

Class Meetings
Monday/Wednesday 9:30-11:18am, 29 Derby Hall

III. Plan of the Course

Class time will be split between lecture and applications, called workshops. During lecture, I will discuss some general principles of inference and simulation, as well as use specific estimators as examples of more general approaches.

The core of the class, however, is the hands-on "workshop," in which we experiment with programming algorithms, estimators, and simulations. As with many areas of life, statistical methodology is best learned by doing, and not solely by reading. Doing in this case does not mean typing commands in STATA or SPSS, but programming estimators more or less from scratch. As such, we will make extensive use throughout the course of a very flexible programming environment called GAUSS. Be warned: Learning GAUSS is not easy. But GAUSS has a number of distinct advantages over competitors: It can handle matrix operations very quickly, its fundamentals resemble those of other computing languages (GAUSS resembles FORTRAN, which is relatively straightforward), and it contains a module of routines appropriate for doing things like maximizing likelihood functions. *NOTE: Previous experience with computer programming is NOT a prerequisite for this course.*

Once you get the hang of it, you will be surprised to learn just how useful GAUSS is; for example, given a matrix of X variables and a vector of *Y* variables, acquiring the OLS estimates for a coefficient vector is as easy as typing inv(x'x)*x'y, which should (I hope) looks familiar to you.[3] Later in the course, we will also play with another program, called WinBUGS, a version of a Bayesian statistical package (BUGS stands for Bayesian Inference using Gibbs Sampling).

IV. The Philosophy of the Course is Simplicity and Elegance

We are going to be doing some complicated stuff, so always keep this rule in mind: Never use an estimator more complicated than the one you need. If you believe that the assumptions of OLS hold for your data, then for heaven's sake, use OLS!

There are two cardinal sins, put here in the form of questions you should NEVER ask yourself:

1) How complicated must my model be for people to notice my work?
2) Gee, the Illudium Q-36 space modulator regression sure sounds like a cool statistical model, and I'm sure to land a methods job if I use it in my dissertation. I wonder what dataset would lend itself to the Q-36?

[3] And if it doesn't, this may not be the course for you.

V. Evaluation

- Four Problem sets (15% each). Problem sets for the most part will include a computer portion and a written portion. For the computer portion, students should e-mail me the computer program that produced the results.
- Paper assignment (40%). You will write a twenty-page paper applying the techniques from the course to a topic of interest to you. In the space of only twenty pages, you will be unlikely to produce a publishable document. You should, however, make a convincing case for the link between theory and an appropriate model and statistical test. You will also provide the computer code that produced your results. ALL PAPER TOPICS MUST BE APPROVED BY ME IN ADVANCE. Students are encouraged to choose topics that might lead to broader dissertation study. The paper is not due until the final exam period; however, in week 10, you will present your preliminary findings to the class.

VI. Readings

The following will be available for purchase at the University and local bookstores (You may have some of these already):

- William Greene, *Econometric Analysis, 4th edition.*
- Gary King, *Unifying Political Methodology*
- J. Scott Long, *Regression models for categorical and Limited Dependent Variables*
- Christopher Mooney, *Monte Carlo Simulation*

Virtually all articles that follow are available online, either at JSTOR or where otherwise indicated. In addition, articles and selected chapters will be available for reading and photocopying in the Department library (2174 Derby).

V. Academic Integrity

All of the work you do in this course is expected to be your own. *Absolutely no cheating or plagiarism* (using someone else's words or ideas without proper attribution) *will be tolerated*. Any cases of cheating or plagiarism will be handled according to university policy and reported to the University Committee on Academic Misconduct. For more on university policy, see http://www.osu.edu/offices/oaa/procedures.

VI. Students with Disabilities

If you have any condition, such as a physical, psychiatric/emotional, medical or learning disability, which will make it difficult for you to carry out the work as outlined in this syllabus, or which will require extra time for exams, please notify me and the Office for Disability Services in the first two weeks of the course so that we may make appropriate arrangements. All information and documentation of disability is confidential. Course materials are available in alternative formats upon request. For such materials, please contact Wayne DeYoung, 2140 Derby Hall, 154 North Oval Mall, 292-2880.

VII. Weekly Schedule

NOTE: Whether readings are assigned for the entire week or for specific class meetings varies from week to week.

Week 1. Preliminaries

Monday, January 7

a. Orientation

b. Workshop: Introduction to the GAUSS programming language, part I: scalars, vectors, and matrices; file I/O; programming etiquette

Wednesday, January 9

a. Methods of Inference: Likelihood and Bayesian approaches

b. Workshop: Introduction to the GAUSS programming language, part II: procedures, loops, and branches; vectorizing

Reading for Week 1:

- King, chs. 1-3 (review)
- Achen, Christopher H. "An Agenda for the New Political Methodology: Microfoundations and ART." Paper presented at the 2001 Annual Meeting of the American Political Science Association. Available at pro.harvard.edu.
- Nagler, Jonathan. 1995. "Coding Style and Good Computing Practices." *The Political Methodologist* 6. Available at http://www.nyu.edu/classes/nagler/quant2/coding_style.html.

Week 2. Simulation

Monday, January 14

a. The Monte Carlo Principle

b. Workshop: Monte Carlo Simulation -- OLS and spherical errors

Wednesday, January 16

a. Techniques for Number Generation: Discrete and Continuous

b. Workshop: Random Number Generators

Readings for Week 2:

- Greene, 5.1-5.3
- Mooney, *Monte Carlo Simulation,* entire

Week 3. Maximization

Monday, January 21

NO CLASS: MARTIN LUTHER KING JR.'S BIRTHDAY

<u>Wednesday, January 23</u>

a. *Numerical maximization*
b. *Workshop: Write your own maximization procedures: Newton-Raphson maximization of a deterministic function and a simple log-likelihood.*

Readings for Week 3:

- Greene, 5.5

Week 4. A Review of Some Basic Stochastic Processes

<u>Monday, January 28</u>

a. *Some important discrete stochastic processes*
b. *Workshop: Poisson regression, down and dirty*

Readings:

- Long, ch. 8
- King, Gary. 1989. "Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator." *American Journal of Political Science* 33, 762-784.

<u>Wednesday, January 30</u>

a. *Some important continuous stochastic processes*
b. *Workshop: Modeling ancillary parameters as functions of regressors*

Readings:

- Greene, chs. 3 (This should be review for most of you) and 19.9.1-19.9.4
- Franklin, Charles H. 1991. "Eschewing Obfuscation? Campaigns and the Perception of U.S. Senate Incumbents." *American Political Science Review* 85, 1193-1214.
- Zorn, Christopher J.W. 2000. "Modeling Duration Dependence." *Political Analysis* 8, 367-380. Available at http://web.polmeth.ufl.edu/pa/vol8no4.html.

Week 5. Special Topics: Latent variables, Threshold models, and Sample Selection

<u>Monday, February 4</u>

a. *Continuous and Discrete Latent Variables models*
b. *Workshop: Cutpoint estimation in an ordered probit*

Readings:

- Greene, 19.1-19.4, 19.8, 20.3, 20.5
- Long, chs. 5,7
- Canes-Wrone, Brandice. 2001. "Bureaucratic Decisions and the Composition of the Lower Courts: An Analysis of Wetlands Permitting." Typescript.
- 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, 1151-1174.

<u>Wednesday, February 6</u>

Threshold and Sample Selection Models: Hurdle, ZIP, ZINB, and Heckit

Readings:

- Greene, 19.9.6, 20.4
- Second paper TBA

Week 6. Special Topics: Relaxing Parametric and Distributional Assumptions

Monday, February 11

a. *Relaxing Distributional Assumptions: MAD and Cox Regression*
b. *Workshop: Comparing the Cox model with fully parametric approaches*

Readings:

- Western, Bruce. 1995. "Concepts and Suggestions for Robust Regression Analysis." *American Journal of Political Science* 39, 786-817.
- Bienen, Henry, and Nicolas Van de Walle, 1989. "Time and Power in Africa." *American Political Science Review* 83, 19-34.

Wednesday, February 13

Relaxing linearity assumptions: GAM and Neural Networks

Readings:

- Beck, Nathaniel, and Simon Jackman. 1998. "Beyond Linearity by Default: Generalized Additive Models." *American Journal of Political Science* 42, 596-627.
- Beck, Nathaniel, Gary King, and Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94, 21-35.

Week 7. Statistical Tests of Complicated Hypotheses

Monday, February 18

Hypothesis testing using unrestricted models: Wald, F, and LR tests

Wednesday, February 20

Hypothesis testing using restricted models: the LM test

Readings:

- Long, ch. 4
- Greene, ch. 4.9, 7.1-7.5

Week 8. Interpreting and Presenting Results

Monday, February 25

Forecasts, first differences, and marginal effects: linear models and linear approximations

Wednesday, February 27

a. *Simulation-based approaches*
b. *Presentation*

Readings:

- King, Gary, Michael Tomz, and Jason Wittenberg. 2000. “Making the Most of Statistical Analyses: Improving Interpretation and Presentation.” *American Journal of Political Science* 44(April): 347-361.
- Tufte, Edward. 1983. *The Visual Display of Quantitative Information*, chapters 1,4,6

Week 9. “Hard” integrals and an introduction to MCMC

Monday, March 4

a. Random effects in nonlinear models: Alternative approaches
b. Workshop: Comparing numerical and Monte Carlo integration

Reading:

- Greene, 5.4
- Gordon, Sanford C. 2002. “Stochastic Dependence in Competing Risks.” *American Journal of Political Science* 46, 200-217.

Wednesday, March 6

a. A (too) brief introduction to Markov Chain Monte Carlo
b. Workshop: Applications of MCMC: MNP in WinBUGS

Reading:

- Jackman, Simon. 2000a. “‘Estimation and Inference via Bayesian Simulation: an introduction to Markov Chain Monte Carlo.” *American Journal of Political Scienc*e 44, 375-404.
- Jackman, Simon. 2000b. ‘‘Estimation and Inference are ‘Missing Data’ Problems: Unifying Social Science Statistics via Bayesian Simulation.’’*Political Analysi*s. 8, 307--332. Available at http://web.polmeth.ufl.edu/pa/vol8no4.html.
- Quinn, Kevin M., Andrew D. Martin, and Andrew B. Whitford. 1999. “Voter Choice in Multi-Party Democracies: A Test of Competing Theories and Models.” *American Journal of Political Science* 43, 1231-1247.

Week 10. Wrapping Up and Student Presentations