POLS 503 ADVANCED QUANTITATIVE POLITICAL METHODOLOGY

MICHAEL D. WARD

Abstract. This class meets weekly in MGH 254 on Wednesdays from 13:30-16:20, with a break around 15:00.

Introduction

We will focus on applied regression analysis, spanning the linear model as well as introducing models of categorial variables and maximum likelihood methods. Work will concentrate on practical exercises utilizing data from political science, as well as the presentation of quantitative material for political science audiences.

There are two basic sets of assignments. The first consists of weekly assignments in which replications of published articles using maximum likelihood methods are employed. The second is a project in which data from a published article in political science, of the student's choosing, is analyzed and interpreted, with an eye toward eventual publication.

There will be a weekly lecture. Most weeks will involve a homework assignment in which students will undertake analysis of a database from a published article in the social sciences. We will develop the necessary code and explore the results in class.

Tentative Schedule, i.e., subject to revision

Week 1: March 29th, 2006 Review of Regression. Readings: Chapter 5 & 6, Fox 1997 Chapter 4, Fox 2002

Instructor's Note: This is a draft, under revision. It may be subject to revision as 503 evolves, as well.

Assignment for Now and Later. Do this today, if at all possible. You should find an article in political science that was published in the last five years that uses either ordinary least squares or utilized a logit model (probit is fine, too). You need to make a PDF of the article and find a copy of the data employed. I'll talk more about this in class. By next week you need to have a) the PDF, b) the data, and c) a summary that you have written about the specific Table or Graphic you will be replicating.

Assignment for next week. Some of you will remember the U.S. Presidential Election in 2000, and the controversy that surrounded the outcome. One important article about this in political science is: Wand, Jonathan N. A., Kenneth Shotts, Walter R. Mebane, Jasjeet S. Sekhon, Michael Herron, and Henry E. Brady (2001) "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida" *American Political Science Review,*, 95(4):793–810. Available electromajically via JSTOR and other electronic vendors. The data from the Wand et alia study is available at `http://elections.berkeley.edu/data/apsr_distribution/`, in case you are interested. But you won't need these, this week.

Instead, your assignment this week is to use the county level data (available on the class web site) to run a plain vanilla regression (OLS) and report the results.

There are two parts to the assignment.

(1) Run the plain vanilla regression of the Buchanan vote by county in Florida for the 2000 Presidential Election as a function of the population, percentage of whites, percentage of blacks, percentage of hispanics and other races, percentage of population over 65 years of age, the percentage completing high school, and the percentage in each county with a college education. Display and interpret the results, as you would in an article or book chapter. This means you need (at a minimum) to a) present the results, b) interpret the results, and c) tell the reader what the results mean and why they are important. I expect two pages of material from you on this aspect of the assignment.

(2) Using what you have done in the first part, I want you to present empirical and textual material that you alone have created that will make the case that Buchanan voters were really Gore voters. Again make a convincing case using the empirical materials available to you.

Week 2: April 5th, 2006 Sampling Variation and Cross Validation. Readings: Chapter 16, Fox 1997.

Assignment Handed Out. Replication, Phase I, due. Identification, PDF, Selection of Tables/Graphs.

Week 3: April 12th, 2006 Unusual and Influential Data. Readings: Chapter 11, Fox 1997 Chapter 6, Fox 2002.

Assignment Handed Out.

Week 4: April 19th, 2006 Diagnosing Problems. Readings: Chapters 12 & 13, Fox 1997 Chapter 7, Fox 2002

Assignment Handed Out. Replication, Phase II, due. Basic Exploratory Data Analysis of replication data.

Week 5: April 26th, 2006 Beyond Linear Least Squares. Readings: Chapter 14, Fox 1997. Luke Keele, "3-D Graphics in R," *The Political Methodologist*, Volume 13, Number 2, Fall 2005.

Assignment Handed Out.

Week 6: May 3rd, 2006 Logit and Probit Models. Readings: Chapter 15, Fox 1997.

Assignment Handed Out. Replication, Phase III, due. Replication of table/graphic accompanied by stylish write-up.

Week 7: May 10th, 2006 Logit and Probit Models, Ordered.

Assignment Handed Out. Replication, Phase IV, due.

Week 8: May 17th, 2006 Maximum Likelihood Approaches. Readings: Chapters 1-4, King (1989/1998)

Assignment Handed Out. Replication, Phase V is due.

Week 9: May 24th, 2006: More on Maximum Likelihood. Readings: Chapters 5 & 11, King (1989/1998)

No Assignment Handed Out.

Assignments

Detailed each week in class.

- Replication Phase I, due before class on April 5th: a labeled CD including a short PDF document indicating the article you are replicating, along with a descriptive summary of the data to be employed and and indication of the Table (one column is sufficient) and Graph (if appropriate) that you will replicate, along with a PDF of that article. Your note should not be longer than a page single spaced and it should indicate three possible improvements on the published analysis or write-up that you are contemplating.
- Replication Phase II, due before class on April 19th: An exploratory data analysis of the pertinent variables from your replication data set. Yes, this means that you have to get the data quickly; see last week's assignment. I want this written up as a data appendix for each variable, as well as some basic exploratory analysis of the interrelationships among the variables.
- Replication Phase III, due May 3rd: Basic replication of table and graphic proposed earlier, written up anew. This will be at most two pages single spaced, including the table and graphic, and will describe the results interpreting them anew, as if they were not a replication, but original research.
- Replication Phase IV, due May 10th: This week you will add your improvement, and describe it anew with a draft of up to 4 pages single spaced, including text, tables, and graphics. You will also deliver a CD including the data, code, and PDF of your effort. I will share that with one of your colleagues, who will replicate your work and return comments to you by next week.

- Replication Phase V, due May 17th: This week you will write a commentary based on your replication of your colleague's work and deliver it in the form of a CD including your PDF comments, as well as any code. Your assignment is to assume that you are a coauthor working together on this and provide a set of suggestions or rewrites for the draft pages you received.
- Replication Phase VI, due on June 2nd at 17:00pm: A CD with all of your work on the replication. I want you to take the comments you received and use and incorporate them as appropriate and presenting a shortish (10 pages single spaced max) PDF document that includes a) description of the model, b) the replication, c) your improvements, d) substantive interpretation of the results you found from your improvements, and e) the data appendix. This should be written in journal style (think APSR) with professional looking graphics (don't think APSR here, you can do better).

Books

- King, Gary. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* Cambridge, England and New York: Cambridge University Press, 1989. Reprinted, Ann Arbor: University of Michigan Press, 1998. Get the 1998 version.
- Fox, John. Applied Regression Analysis, Linear Models, and Related Methods. Thousand Oaks, California: Sage Publications, 1997. But note that a lot of collateral material is available online at `http://socserv.mcmaster.ca/jfox/Books/Applied-Regression/index.html`.
- Fox, John. *An R and S-PLUS Companion to Applied Regression.* Thousand Oaks, California: Sage Publications, 2002 `http://socserv.mcmaster.ca/jfox/Books/Companion/index.html`.

Office Hours

I'll be in my office most days. Monday afternoons, I teach another class from 16:30-19:20; Tuesdays, I'll hold formal office hours from 15:00-17:00 pm; On Wednesdays from 12:30-14:00, I attend the weekly CSSS seminar. Email is the best way to reach me reliably.

Xun Cao is the Teaching Assistant for this class. He will hold office hours in the Smith Computer Classroom from 15:00-17:00 on Mondays, and by appointment.

Grades

There are nine weekly homework assignments, six degrees of replication, and participation in class. You may drop one of the weekly homework assignments after the first week.

(1) 8 homework assignments, each worth 20 points (160 points total)
(2) 5 preliminary replication project deliverables, each worth 40 points (totalling 200 points)
(3) final replication delivery, worth (200 points)
(4) 40 points for staying awake in class, I mean class participation.
(5) Total is 600 points, divided by 150 should be your grade.

Late assignments will not be graded.

Michael D. Ward: Department of Political Science, University of Washington, Seattle, Washington, USA, 98195-3530

E-mail address: `mdw@u.washington.edu`