POLS W4912
Multivariate Political Analysis

Spring 2008

Prof. Gregory Wawro — gjw10@columbia.edu
741 International Affairs Bldg. — Office Hours: Tues. 2:30–4:30 and by appt.

TA: Matthew Winters — msw22@columbia.edu
270 International Affairs Bldg. — Office Hours: Thurs. 2:15-4:15

Course Goals

The motivation for this course is first and foremost to give you the ability to use the classical linear regression model with confidence—both in theory and practice—to test causal models. By the end of the course, you should be aware of when ordinary least squares (OLS) yields desirable properties, the cases when it does not, how to test for these instances, and what procedures you should use for correct estimation and inference. In addition, you will have been exposed to a number of variants on the classical linear regression model that are applicable for the many different types of data you might encounter.

Instructional Approach

The course will be taught using matrix notation and thus assumes some knowledge of matrix algebra. The absence of training in matrix algebra should not, however, constitute an obstacle to taking the course so long as you are prepared to put in some additional work early on learning a few, key matrix operations. The use of matrix notation allows for the derivation of central proofs of the properties of OLS (and other estimators) that would be cumbersome or impossible without it.

Assignments

There will be a mid-term and a final exam. The grades for these two exams will account for 60 percent of your final grade. In addition, there will be several problem sets along the way (approximately 5), and these will contribute the additional 40 percent of your grade. We may schedule weekly review session to discuss questions from the problem sets, computing issues, and other issues that arise in lecture.

Statistical Software

Homework assignments and class examples will be in `R`, which has become one of the most widely used software packages in the social sciences. While somewhat more difficult to learn than other packages, `R` is particularly useful for learning regression analysis in terms of matrix algebra. `R` is free and available for download from `http://www.r-project.org/`, and students should install this software on their own computers. There is a substantial network of `R` users that students can tap into for support. Students are welcome to use any software package they are familiar with, but we will provide support only for `R` and several assignments will require writing `R` code. We will hold lab sessions to get students acquainted with using `R`.

Students may want to supplement the use of `R` with `Stata`. It is assumed that you have the required Cunix ID to be able to use the Columbia PC labs, including that at 323 IAB, where Stata is available on-line. For a full list of AcIS computer lab locations please see the following link: `http://www.columbia.edu/acis/facilities/labs/locations/`. In addition, those of

you with networked PCs at home should be able to access `Stata` via Cunix. Finally, you can purchase the latest version of `Stata` for a much-reduced price ($145 without manuals for the "Intercooled" version) via Columbia's Gradplan. For information on buying the software, see `http://www.columbia.edu/acis/software/licenses/stata/`.

Lecture Notes
A PDF version of my lecture notes are available from the course web site (`http://www.columbia.edu/~gjw10/w4912.html`). Students should download and print up a copy of the notes for themselves so that they can follow along with the lectures. Students who do not have a copy of the notes will be at a severe disadvantage. The notes will appear in installments over the semester. Students should check the web site frequently for updates.

Books
I have ordered the following books for this course through bookculture:

William H. Greene, *Econometric Analysis*, 2003, Fifth Edition, Prentice-Hall.

Damodar N. Gujarati, *Basic Econometrics*, 2003, Fourth Edition, McGraw Hill.

Peter Kennedy, *A Guide to Econometrics*, 2003, Fifth Edition, MIT Press.

John Fox, *An R and S-Plus Companion to Applied Regression*, 2002, Sage Publications.

Greene will serve as the primary text, but in my experience it is extremely useful for students to read more than one treatment of the material. Gujarati serves as a more introductory and intuitive supplement to Greene, but should not be used exclusively. While these books can be expensive, they are useful references that you will want to have on your shelf if you are at all serious about conducting quantitative analysis beyond this course. New or like new copies of these books can often be purchased at a reduced price through the second-hand feature at Amazon.com. I have requested all of these texts to be placed on reserve at Lehman. Other texts you may find useful are

G. S. Maddala, *Introduction to Econometrics*, 2001, Third Edition, John Wiley and Sons.

Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 2002, Southwestern College Publishing.

Jeffrey M. Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, 2002, MIT Press.

Outline of Classes

Part I: The Classical Linear Regression Model

Week One: Review of Matrix Algebra and Statistical Inference January 22nd, 24th

- Greene, 5th or 6th ed.: Appendices A and B.
- Gujarati: Appendices A.1 to A.6 and B.1 to B.6.

- Kennedy: Appendix A.

Week Two: The Classical Linear Regression Model January 29th, 31st

Criteria for estimators—unbiasedness and efficiency. The derivation of the classical linear regression model or OLS.

- Greene, 5th or 6th ed.: Chapter 2, Chapter 3.1–3.2.
- Gujarati: Chapters 1, 2 and 3.1–3.2.
- Kennedy: Chapters 2–3.

Week Three: Properties of OLS in Finite Samples February 5th, 7th

The Gauss-Markov Assumptions. Why OLS is BLUE (unbiased and efficient). The variance of the OLS regression coefficients, hypothesis tests, and confidence intervals.

- Greene, 5th or 6th ed.: 4.3–4.10; for maximum likelihood 17.1–17.6 in 5th ed., 16.1–16.4, 16.6, 16.9 in 6th ed.
- Gujarati: 3.3 to 3.4, 4.1 to 4.3, 5.1 to 5.8 and 7.1 to 7.4.

Week Four: Inference and Hypothesis Tests February 12th, 14th

Hypotheses on a coefficient, goodness of fit, joint tests, confidence intervals, constructing hypothesis tests as linear restrictions, plus dummy variables and using joint hypothesis tests to perform tests for structural change.

- Greene, 5th ed.: 6.1–6.4, 7.1–7.5; 6th ed.: 5.1–5.4, Ch. 6.
- Gujarati: 5.1 to 5.8, 7.5 to 7.8, and 8.1 to 8.8.
- Kennedy: Chapter 4.

Week Five: Issues of Correct Specification and Bias February 19th, 21st

Omission of relevant variables, multicollinearity, measurement error.

- Greene, 5th ed.: 4.9.1, 5.6, 8.1–8.2; 6th ed.: 4.8.1, 12.5, 7.1–7.2.
- Gujarati: 10.1 to 10.5 and 13.3 to 13.5.
- Kennedy: Chapter 6, Chapter 9, Chapter 11,

Week Six: Remaining Features of OLS February 26th, 28th

Regression diagnostics, non-nested hypothesis tests, prediction, and a short digression on asymptotics.

- Greene, 5th ed.: 4.9.2, 4.9.3, 5.2–5.3, 6.6; 6th ed.: 4.8.2, 4.9, 5.6.

- Gujarati, 5.10, 8.9, 13.10, Appendix A.7
- Kennedy: Chapter 5, Appendix C,

Week Seven: OLS as Maximum Likelihood March 4th, 6th

- Greene, 5th ed., 17.1–17.6; 6th ed.: 16.1–16.4, 16.6, 16.9.
- Gujarati, 4 A.1 and 7 A.4

Week Eight: Catch-up, Review, and Midterm March 11th, 13th (**Mid-term on the 13th**)

Spring Break **March 18th, 20th**

Part II: Violations of Gauss-Markov

Week Nine: Non-spherical Errors March 25th, 27th

Heteroskedasticity, GLS and FGLS, White robust standard errors.

- Greene, 5th ed.: 10.1–10.3, 11.1–11.2, 11.4–11.6; 6th ed.: 8.1–8.9.
- Gujarati: 11.1 to 11.6 and 12.1 to 12.11.
- Kennedy: Chapter 8.

Week Ten: Simultaneous Equation Models April 1st, 3rd

- Greene, 5th ed.: 15.1–15.5; 6th ed.: 13.1–13.5.
- Gujarati: 18.1 to 18.5.
- Kennedy: Chapter 10.

Week Eleven: Autocorrelation and Time Series Data April 8th, 10th

- Greene, 5th ed.: 12.1–12.3, 12.7–12.9; 6th ed.: 19.1–19.3, 19.7–19.9.
- Gujarati: 21.1 to 21.10 and 17.8 to 17.10.
- Kennedy: Chapter 17.

Part III: Special Topics

Week Twelve: Time-Series Cross-Section Data April 15th, 17th

Panel data: fixed effects, random effects, Hausman specification tests. Problems with panel heteroskedasticity. Panel corrected standard errors.

- Greene, 5th ed.: 13.1–13.4, 13.7, 13.9; 6th ed: 9.1–9.4, 9.6.

- Gujarati: 16.1–16.6.

Week Thirteen: Dichotomous Dependent Variable Models April 22nd, 24th

Logit and probit.

- Greene, 5th ed.: 21.1–21.4, 21.6–21.7; 6th ed.: 23.1–23.4, 23.8–23.12.
- Gujarati: 15.1 to 15.4.
- Kennedy: Chapter 15.

Week Fourteen: Discrete and Limited Dependent Variable Models April 29th, May 1st

Censored models, selection models, count models, duration models, resources for methodologists.

- Greene, 5th ed.: 21.9, 22.1–22.6; 6th ed.: 25.2, 24.1–24.3, 24.5, 25.6.
- Kennedy: Chapter 16.

Final, take home exam May 6th–16th