Political Science 6000: Regression

Professor Paolino
Wooten 137
565-2315
paolino@unt.edu

Office hours
Mon 11:00-12:00
Tues 12:00-2:00

Description

This course is a research seminar that covers regression for political science students. The goal of the course is to ensure that graduate students in political science are able to interpret, implement, and employ basic regression analysis of social science data.

This semester course is primarily devoted to regression analysis. The emphasis is on developing a more rigorous understanding of ordinary least squares (OLS), violations of regression assumptions, hypothesis testing, and the general linear model. Most of the course will be spent developing basic approaches to regression analysis when the assumptions of OLS fail. Some time will be spent on estimators other than least squares when assumption violations require these methods.

The emphasis of this course is on the acquisition and understanding of analytical techniques. It is very much an applied course, but for applications to be informed, there is a fundamental level of theoretical knowledge required. Many mistakes will be made by those who only know how to interpret findings and use "canned" computer packages. Another way of saying this is that "applied" does not mean mindlessly plugging things into Stata with no understanding of whether it should be in or not. The course emphasizes that students know exactly what is going on when they do statistical analysis.

Objectives

Students completing this course should have acquired the following skills:

1. The ability to interpret results from regression models presented in professional papers. This includes:
   - The ability to interpret coefficients.
   - The ability to make statistical inferences.
2. The ability to evaluate results from regression models presented in professional papers based upon an understanding of the Gauss-Markov assumptions. This includes:
   - The ability to explain based upon the theory and data presented in a paper whether or not the results may imply a violation of the Gauss-Markov assumptions.
   - The ability to assess the extent to which any such violations may mitigate or invalidate the paper's conclusions.
3. The ability to execute their own research projects that call for the use of OLS or other regression techniques. This includes:
   - The ability to carry out analysis that is defensible to the type of criticism outlined in point 2.
   - The ability to recognize problems in the data or the resulting analysis that does not make sense to someone with an understanding of OLS.
4. The ability to understand, both intuitively and mathematically, the OLS model, its assumptions, the effects of violations of its assumptions, and remedies for these violations. This includes:
   - An ability to derive OLS estimators.
   - An ability to manipulate formulas to demonstrate particular results.

Grading and Requirements

Grading

There will be approximately six to eight assignments, two exams, and an original research paper. The allocation of the grade is:

- Assignments: 25%
- Exam I: 25%
- Exam II: 25%
- Paper: 25%

Late assignments and papers will be penalized. Late assignments and papers will be penalized 10% per day. Papers or assignments submitted via e-mail will be considered received by the date and time stamp on the e-mail received in my e-mail inbox and followed, as soon as possible, with a hard copy. The paper will be due on the last scheduled class date (May 5th).

Assignments

Assignments and exams will cover applied and theoretical problems.

Assignments should be typewritten as much as possible. I realize that may require setting mathematical text or typographical symbols. This can be done in standard word processing software. Feel free to only use Roman letters. If you are so interested, I can give a short tutorial on how to use LaTeX for this task.

Also, any statistical output or data analysis you do should be fully interpreted and presented as though it were being sent for publication to a journal. *This means that regression output from your statistical package of choice that is copied into a word processor document is unacceptable. You should take the time to typeset the results into a meaningful table or present a well documented and coherent graphical summary of any results.* An important part of your instruction is learning how to present your work in a professional manner. Fortunately, this should not be difficult. Once you have created one table, it is usually a trivial matter to revise that table to meet new needs. If you have any questions about what to include in your data output and assignments, consult empirical work in standard journals (APSR, ISQ, AJPS, JOP, etc.) or ask.

Papers

The paper will require students to analyze and interpret regression. The paper you write should be based on your own research and interests – there is no requirement to use a particular dataset or technique. However, the paper should be original (it should not include an analysis used in previous courses, conference papers, or other writing). The following are suggested models for a paper:

- A conference paper (examples can be seen on my webpage).
- A replication article (see the replication standard for Political Analysis or at `http://gking.harvard.edu/projects/repl.shtml`)

Before beginning your paper, you are **REQUIRED** to come and talk with me about your paper. At or prior to this meeting, please provide a ONE PAGE research design and lays out the following for your paper:

- Main research question you are addressing and a sketch of the theory motivating your thesis.
- Hypothesis (or Hypotheses) you wish to evaluate in your paper.
- Data and variables to be used in the analysis.
- Tests and methods that will be used to evaluate the hypotheses.

- Tentative listing of the techniques and models you may use.

The data you use for the analysis is something that you should already have or have easy access to. Data collection is not a topic we are covering in this course and time spent building large complex datasets will detract from your ability to complete the paper adequately.

The meeting to discuss the paper should be scheduled before the 9th week of class.

Finally, even though we will cover binary dependent variables before the end of the semester, your paper should use statistical techniques dependent upon least squares methods.

Attendance

It should go without saying that in a class of this size your attendance is easily noted and therefore required. If you are unable to make a class or will be late, advise the instructor as far in advance as possible. I reserve the right to drop you from class after two absences.

Books and other materials

Three books have been ordered. They are:

Required:

Wooldridge, Jeffrey M. 2009. *Introductory Econometrics: A Modern Approach*, 4th ed.

Optional:

Chiang, Alpha. 1984. *Fundamental Methods of Mathematical Economics*. New York: McGraw-Hill.
Kennedy, Peter. 2008. A Guide to Econometrics. 6th edition. Cambridge: MIT Press.

Students find the Kennedy book to be extremely helpful, so I encourage students to buy it. There will be occasions in which Kennedy will be extraordinarily helpful; use it as you see fit.

Alpha Chiang is a good source for mathematical topics that may not be well covered in some of the other texts. There are good (albeit expensive) references for optimization, differential equations, and other topics. Buy it only if you intend to continue using complex statistical techniques or pursue formal modeling (e.g., game theory).

In general you should use the books as reference materials. If I have not indicated a reading from a book, it does not mean that a topic is not covered in it. Thus, use the texts as reference materials for each week and take the indicated readings as a guide. It is my expectation that you will find additional texts and resources to supplement the assigned course materials. Some sources will be indicated in course lectures and notes as I see fit.

Students will also be expected to utilize other advanced sources for data and statistical reference. These include JSTOR journal databases, Statlib, POLMETH, and other resources to which you will be directed. I intend to focus primarily on material from the main text, but I may also assign readings from other sources as supplements to the main text.

Computing and Course Materials

A major component of applied statistics is using computers and data to implement models and test hypotheses. Toward that end, students should be expected to utilize a variety of statistical packages for their work. While no one statistical program will be used for this course, familiarity with several is important because each has its own niche. I will primarily use Stata, and datasets from the Wooldridge book are available for Stata, but other good statistics packages include:

Package	Website	Cost
Stata	http://www.stata.com	Available at UNT
R	http://cran.us.r-project.org	Free
S-plus	http://www.insightful.com	Available at UNT
SAS	http://www.sas.com	Available at UNT
Rats	http://www.estima.com	Available at UNT
Ox	http://www.oxmetrics.com	Free console version
GAUSS	http://www.aptech.com	Available at UNT

Any others are acceptable (and the problem set data are available in ascii form), and you should *not* express reservation about learning additional programs–it is an impediment to your research and teaching.

For home use (i.e., those who hate to purchase licenses), R, the GNU clone of S-plus, is becoming quite popular among political scientists, and many others (see http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html?_r=1\&em). It is free, has great graphics and, is well documented. For econometric analysis and time series, Ox or Rats are notable options.

As this is not an introductory course, I will not spend time leading tutorials on statistical software in class. I am happy to have them arranged outside of class. I will also make programs an code available on an "as needed" basis in several statistical packages to demonstrate techniques. Note that somethings are easier to do in different packages and then convert the data to another format for analysis (I often construct variables and datasets in Stata and then move them to other software for analysis.) Your best resource for learning and implementing new methods is your peers.

Problem sets and other course material will be distributed from the class's blackboard site. This will include notes, problem sets, example code, answers to problem sets, etc.

Course Outline

Readings listed for each week are required. My lectures and discussion will parallel these readings. We have a full schedule of topics to cover, so it is essential that you are prepared for class–and this includes being prepared to derive results for the class. I will regularly post or hand out lecture notes on the material we are covering. Problem sets and supplemental readings will be assigned during class meetings.

Week 1: Introduction, Math for Social Scientists (Jan. 20)

- *IE*, Ch. 1 & Appendices A-C.

Week 2: Review of the Simple Linear Model (Jan. 27)

- *IE*, Ch. 1 & 2

Week 3: Multiple Regression (Feb. 3)

- *IE*, Ch. 3

Weeks 4: OLS Inference and Asymptotics (Feb. 10)

- *IE*, Ch. 4-5.

Week 5: OLS with Qualititative Information (Feb. 17)

- *IE*, Ch. 7

Week 6: Specification issues (Feb. 24)

- *IE*, Ch. 9

Week 7: Heteroscedasticity (Mar. 3)

- *IE*, Ch. 8

Week 8: Mid-term Exam (Mar. 10)

Weeks 9: Time series and serial correlation (Mar. 24)

- *IE*, Ch. 10.

Week 10: More Time Series (Mar. 31)

- *IE*, Ch. 11-12.

Weeks 11-12: Panel Data (Apr. 7-14)

- *IE*, Ch. 13-14.

Week 13: Systems of equations, identification, non-fixed regressors (Apr. 21)

- *IE*, Ch. 15-16.

Week 14: Limited Dependent Variables (Apr. 28)

- *IE*, Ch. 17.

Week 15: Regression with Matrix Algebra (May 5)

- *IE*, Appendices D & E.

Final Exam: Take home during finals week