Political Science 6321: Regression

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

Office hours
Tues 12:30-2:00
Wed 12:00-1:30

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 classical linear regression model. Most of the course will be spent developing basic approaches to regression analysis when the assumptions of OLS fail.

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. Students who only how to interpret findings and use "canned" computer packages will make many mistakes when conducting their own analysis. Another way of saying this is that "applied" does not mean mindlessly plugging things into a software package 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 and present 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: 30%
- Exam I: 20%
- Exam II: 25%
- Paper: 25%

The research paper will be due on May 3rd. Late assignments and papers will be penalized 10% per day.

Assignments

Assignments and exams will cover applied and theoretical problems.

Assignments should be typewritten and submitted through Canvas. 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 arrange a short tutorial on how to use LaTeX for this task. I also require that you submit a file with the computer code that you use for assignments. The code file should be written so that anyone could reproduce perfectly the results from your assignment by working from the originial data and changing *only* file or directory names.

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 and will result in a 10-pt deduction from your grade for that week's assignment. 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. Stata has programs that will put your output into reasonably well-formatted tables or figures.* An important part of your instruction is learning how to present your work in a professional manner. More importantly, an important part of scientific research is the ability to communicate your results to others in a clear manner. If you cannot make it easy for others to understand what you have done, then you have not completely met the demands of scientific research. Again, this should not be difficult. Once you have written code for a table or figure, you should be able to modify that code to produce an informative table or figure.

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; though it is permissible, with the permission of both instructors, to use a research design paper from another class as the basis of the paper for this class). 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. Prior to this meeting, please provide a ONE OR TWO PAGE research design that 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.

Please submit this research design summary to me by March 11. The meeting to discuss the paper should be scheduled before the 9th week of class. Note, because we will not cover limited dependent variables (e.g. logit and probit) this semester, your paper should use statistical techniques dependent upon least squares methods. You should, however, consider designs that may require methods that we will not cover until the second half of the semester. Our meeting will allow me to advise you on the feasibility of such methods for your project.

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.

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, please advise me as far in advance as possible. I reserve the right to drop you from class after two unexcused absences.

Disabilities Accommodation

The University of North Texas makes reasonable academic accommodation for students with disabilities. Students seeking reasonable accommodation must first register with the Office of Disability Access (ODA) to verify their eligibility. If a disability is verified, the ODA will provide you with a reasonable accommodation letter to be delivered to faculty to begin a private discussion regarding your specific needs in a course. You may request reasonable accommodations at any time, however, ODA notices of reasonable accommodation should be provided as early as possible in the semester to avoid any delay in implementation. Note that students must obtain a new letter of reasonable accommodation for every semester and must meet with each faculty member prior to implementation in each class. Students are strongly encouraged to deliver letters of reasonable accommodation during faculty office hours or by appointment. Faculty members have the authority to ask students to discuss such letters during their designated office hours to protect the privacy of the student. For additional information, refer to the Office of Disability Access website at https://studentaffairs.unt.edu/office-disability-access. You may also contact ODA by phone at (940) 565-4323.

Academic Integrity

According to UNT Policy 06.003, Student Academic Integrity, academic dishonesty occurs when students engage in behaviors including, but not limited to cheating, fabrication, facilitating academic dishonesty, forgery, plagiarism, and sabotage. A finding of academic dishonesty may result in a range of academic penalties or sanctions ranging from admonition to expulsion from the University. Students should review the policy, which may be located at https://policy.unt.edu/policy/06-003. Faculty are required to submit reports of violations of academic dishonesty even in instances that do not result in sanctions. There will be no deviation from this policy.

Emergency Notification & Procedures

UNT uses a system called Eagle Alert to quickly notify you with critical information in the event of an emergency (i.e., severe weather, campus closing, and health and public safety emergencies like chemical spills, fires, or violence). In the event of a university closure, please refer to Canvas for contingency plans for covering course materials.

Books and other materials

Required:

Wooldridge, Jeffrey M. *Introductory Econometrics: A Modern Approach.* With Stata code and data for examples available at http://fmwww.bc.edu/gstat/examples/wooldridge/wooldridge.html and http://fmwww.bc.edu/ec-p/data/wooldridge/datasets.list.html, respectively.

Optional:

- Kennedy, Peter. 2008. *A Guide to Econometrics*. Cambridge: MIT Press.
- Chiang, Alpha. 1984. *Fundamental Methods of Mathematical Economics*. New York: McGraw-Hill.
- Moore, Will H. and David A. Siegel. 2013. *A Mathematics Course for Political & Social Research*. Princeton University Press.
- Heiss, Florian. *Using R for Introductory Econometrics.* http://www.urfie.net/read/index.html#page/1. This book provides R code for reproducing most of the examples from Wooldridge.
- Monogan, James E. 2015. *Political Analysis Using R* (access through UNT library web-site).

Students have found the Kennedy book to be extremely helpful, so I encourage students to buy it and use as you see fit.

Chiang and Moore and Siegal are good sources for mathematical topics that may not be well covered in some of the other texts. These 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). If your interest is more limited, then the appendix of Wooldridge should provide enough math background for this course.

Finally, the Heiss and Monogan books provide useful treatments of using R for basic regression. While we'll use Stata for analysis in this course, R is a powerful tool for econometric analysis and, if you continue onto more sophisticated methods, often provides the ability to conduct analysis that is unavailable through Stata.

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, POLMETH, and other resources to which you will be directed as appropriate. 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.

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 lecture notes and sample code on the material we are covering. Problem sets and supplemental readings will be assigned during class meetings.

Week 1: Introduction and Math for Social Scientists (Jan. 17)

This lecture provides an overview of some basic mathematical and statistical properties that will be useful for econometrics.

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

Week 2: Review of the Simple Linear Model (Jan. 24)
This lecture looks at the bivariate regression model as a way of highlighting important properties of least squares estimation and the Gauss-Markov theorem.

- *IE*, Ch. 1 & 2

Week 3: Multiple Regression (Jan. 31)
This lecture extends last week's examination of the bivariate regression model to models with multiple explanatory variables.

- *IE*, Ch. 3
- Bivariate regression assignment due

Week 4: OLS Inference and Asymptotics (Feb. 7) This lecture examines the small and large-sample properties of Ordinary Least Squares.

- *IE*, Ch. 4-5 & Appendix C.

Week 5: Functional Form and OLS with Nominal Information (Feb. 14)
This lecture discusses how to use nominal variables as explanatory variables to model both main and interaction effects.

- *IE*, Ch. 6-7
- Multiple regression assignment due.

Week 6: Heteroscedasticity (Feb. 21)
This lecture discusses the violation of the constant variance assumption and introduces various diagnostic and corrective approaches.

- *IE*, Ch. 8
- Nominal variable regression assignment due.

Week 7: Specification and Data issues (Feb. 28)
This lecture introduces various ways to evaluate model specification and examines consequences of measurement error and missing data.

- *IE*, Ch. 9
- Heteroskedasticity assignment due.

Week 8: Mid-term Exam (Mar. 6)

$\Longrightarrow$ *No class on Mar. 13 during Spring Break*

Week 9: Time series and serial correlation (Mar. 20)
This lecture discusses the consequences of violating the independent errors assumption and introduces diagnostic and corrective approaches.

- *IE*, Ch. 10.

Week 10: More Time Series (Mar. 27)
This lecture extends discussion of considerations when modeling time series data.

- *IE*, Ch. 11-12.

Weeks 11-12: Panel Data (Apr. 3-10)

These lectures examine analysis of data that has both cross-sectional and time series properties. Primary attention will be given to fixed and random effects models.

- *IE*, Ch. 13-14.
- (Apr.3) Time series assignment due.

Week 13: Systems of equations, identification, non-fixed regressors (Apr. 17)

This lecture explores the problem of endogenous regressors, with a primary focus upon instrumental variables approaches to accounting for endogeneity.

- *IE*, Ch. 15-16.
- Panel Data assignment due.

Week 14: Regression with Matrix Algebra (Apr. 24)

This lecture introduces concepts from matrix algebra and examines the multiple regression model using matrix algebra notation.

- *IE*, Appendices D & E.

Week 15: Paper presentations (May 1)

- Students should prepare 20-minute presentations of their research papers.

Final Exam: Take home during finals week