POLS 509: The Linear Model

Professor Justin Esarey
Spring 2012

January 23, 2012

Contact Information

103 Tarbutton Hall
Office Phone: 404-727-6583
e-mail: jesarey@emory.edu

Office Hours: M 1:00 -3:00 PM or by appointment

The best way to reach me is by e-mail; I have access to my e-mail via phone and computer and should respond within 24 hours of receiving your message.

Class Time and Location

Monday 9:00a - 12:00p
Tarbutton Hall 120a

Course Description

This course covers basic techniques in quantitative political analysis. It introduces students to widely used procedures for regression analysis, and provides intuitive, applied, and formal foundations for regression and more advanced methods treated later in the course sequence. Unlike POLS 508, this course will use calculus and matrix algebra rather intensively. The course covers model assumptions and techniques for addressing violations of those assumptions (e.g., heteroskedasticity, autocorrelation, multicollinearity), as well as issues of model specification, functional forms, measurement error, and endogeneity. This course builds on the prior efforts of Chris Zorn (a former faculty member now at Penn State), Eric Reinhardt, and Kyle Beardsley.

Course Materials

Required Textbooks

Damodar Gujarati and Dawn Porter, *Basic Econometrics, 5th edition.* McGraw-Hill Irwin.

Russell Davidson and James G. MacKinnon. 2004. *Econometric Theory and Methods.* Oxford.

Andrew Gelman and Jennifer Hill. 2007. *Data Analysis using Regression and Multilevel/Hierarchical Models.* Cambridge.

Computing Reference Material

Peter Dalgaard. *Introductory Statistics with R, 2nd edition.* 2008. Springer-Verlag.

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S, 4th edition.* 2002. Springer-Verlag.

Software

This course will teach material using a combination of R and Stata 11/12. R is free and available from `http://cran.r-project.org/`. Stata 12 is available in the graduate computing lab, and you may also purchase a copy through Emory's GradPlan for a reduced rate (I recommend this).

All homeworks, papers and presentation slides must be typed in LaTeX, but free software exists to make this easy to accomplish. I suggest using LyX (`http://www.lyx.org/`) in combination with MiKTeX on Windows (`http://miktex.org/`), MacTeX on Macintosh (`http://www.tug.org/mactex/`) or TeXLive on Linux (`http://www.tug.org/texlive/`). For Windows machines, this link leads to an installer that has everything you need to get started: `http://tinyurl.com/ykmzycs`.

Other

All students must have a valid Emory e-mail address and login (and access to the Blackboard website) to participate in this course.

Grading/Evaluation

Grading Scale

100%-93%: A	76.9%-73%: C
92.9%-90%: A-	72.9%-70%: C-
89.9%-87%: B+	69.9%-67%: D+
86.9%-83%: B	66.9%-60%: D
82.9%-80%: B-	>59.5%: F
79.9%-77%: C+	

Grade Components

- 20% Homeworks
- 30% Empirical Term Paper
  - 5% Presentation
  - 25% Final Draft
- 25% Midterm Exam
- 25% Final Exam

Assignments and Responsibilities

Readings

All readings are required unless marked optional, and must be read before the start of class. You don't have to fully understand everything in a reading assignment the first time you read it, but you should be familiar with the material. If you don't understand something, feel free to e-mail me with a question, or come to class with a knowledgeable question about what we've read for the day.

A note on the proper way to use the readings assigned for this class. There are two primary texts assigned for the class. Gujarati and Porter is an applied statistics book that should be comparatively easy to read and digest; I recommend that you read this material first. Davidson and MacKinnon is a high-level text that will probably not be easy to read, but is more rigorous about proving the propositions that are often advanced without proof in Gujarati and Porter. Davidson and MacKinnon also relies more strongly on matrix notation. I recommend that you not worry about assimilating 100% of the material in Davidson and MacKinnon on the first read, and to consider re-reading the appropriate sections after the class's lecture.

Research Paper

A major portion of the class's coursework will consist of writing an empirical research paper, a full paper centered around an empirical test of a novel theory and/or an improved test of an existing theory. The paper must employ statistical techniques introduced in this class to test theoretical predictions (e.g., comparative statics).

The topic of the paper is open with respect to substantive field, but the topic must be original and relevant research of topical interest to some substantive community OR (in a rare case) technical research of interest to pure statistics.

Each student in the class will schedule an appointment to meet with me during my office hours to propose a topic. This meeting must occur **before February 27, 2011**. Topics must be approved by me BEFORE work can begin on the project.

The final draft of the paper must be submitted electronically **on May 4, 2010**.

Students will present a 30-40 minute job talk-style presentation of their research during the final two class meetings; half the class will be randomly selected to present on the first day, the other half on the second day.

All papers and presentation slides must be typed in LaTeX. I suggest using LyX (`http://www.lyx.org/`) in combination with MiKTeX on Windows (`http://miktex.org/`), MacTeX on Macintosh (`http://www.tug.org/mactex/`) or TeXLive on Linux (`http://www.tug.org/texlive/`). This link leads to an installer that has everything you need to get started: `http://goo.gl/CeDzC`.

You must use the APSA citation style described in the APSA Manual of Style.

Exams

There will be two exams in this class, a midterm and a final; the final will be cumulative but will focus on material learned since the previous exam. You must complete each exam within twenty-four hours of receipt, and must submit a typed LaTeX answer sheet (see research paper section above for details on software). The exams are open book and open note, but you may not consult anyone for advice on the exam. The rough timing of the exams is indicated on the course outline, and specific times will be scheduled in consultation with the class.

Attendance and Class Participation

Attendance is mandatory in this class, and as graduate students I expect that attendance will not be a problem for you. Failure to attend will influence your final grade. Active participation in lectures and discussion is expected.

Homework

Homework problem sets will be distributed during class. I encourage collaborative work on problem sets: the goal of a homework problem set is to help you learn the material and

enable you to perform well on the (non-collaborative!) research paper. With that said, simply copying another student's homework answers is not permitted and will be treated as academic dishonesty.

All homeworks must be typed in LaTeX (see research paper section above for details on software).

Course Policies

Late Work

Assignments are due at the date and time I specify for the assignment. Late homeworks and papers will be marked off at 5 percentage points for the first 24 hours late, and an additional 10 percentage points for every subsequent 24 hours late. For exams, the first hour late incurs a 5 percentage point penalty and each additional hour incurs a 10 percentage point penalty.

Late work penalties may be waived in the event of death in the immediate family (parent, spouse, sibling, or child) within 2 weeks before the due date, or in the event of an unforeseeable medical emergency affecting yourself, your spouse, or your child. Penalty waivers are at the discretion of the instructor. I may require supporting documentation.

Academic Misconduct

Cases of plagiarism on the research paper and other forms of academic misconduct will be handled according to the Emory University Honor Code, available on-line at `http://www.college.emory.edu/current/standards/honor_code.html`.

Please pay special attention to the definition of plagiarism on the Emory Honor Code web site at the link above. You may also find the Emory Writing Center's site on "Avoiding Plagiarism" helpful; this site is found at `http://www.writingcenter.emory.edu/plagiarism.html`.

If you ever have any questions about whether or how material should be cited, PLEASE contact me with your question and I can assist you. I cannot guarantee a timely response unless you contact me *at least 24 hours in advance of the time the assignment is due.*

Course Outline and Assigned Readings

- January 16: MLK Day - No Class
- January 23: Tools of Probability Theory, Matrix Algebra, and Calculus
  - Topics: basic mathematical and computational tools needed for the remainder of the course

  - Readings
    * Gujarati and Porter, Appendices A and B
    * Davidson and MacKinnon: Chapter 1 up to p. 27

- January 30: Ordinary Least Squares Regression with Matrix Algebra
  - Topics: review of OLS; statement of OLS in terms of matrix algebra; derivation of OLS estimator using multiple methods (standard, Method of Moments, residual minimization)
  - Readings
    * Gujarati and Porter: Appendix C, Sections C.1-C.3
    * Davidson and MacKinnon: Chapter 1, p. 27-end
    * *Optional*: Gujarati: Chapter 2; Chapter 3, Section 3.1; Chapter 7, Sections 7.1-7.4

- February 6: The Geometry of Linear Regression
  - Topics: vector interpretation of OLS; rank requirements for regressors; orthogonality of predictions and residuals; Frisch-Waugh-Lovell Theorem and applications; fit statistics; outliers
  - Readings
    * Gujarati and Porter: Chapter 3, Section 3.5; Chapter 7, Sections 7.5, 7..6, and 7.8; Appendix C, Section C.4-C.5
    * Davidson and MacKinnon: Chapter 2

- February 13: Assumptions and Properties of OLS
  - Topics: proof of unbiasedness + consistency; necessary assumptions to sustain unbiasedness + consistency; the VCV matrix and OLS efficiency (Gauss-Markov Theorem)
  - Readings
    * Gujarati and Porter: Chapter 3, Sections 3.2-3.4; Chapter 4, Sections 4.1-4.3; re-read Chapter 7, Section 7.1
    * Davidson and MacKinnon: Chapter 3

- February 20: Simple Hypothesis Testing and Fit Diagnostics
  - Topics: t-testing and F-testing; small-sample distributions of fit statistics; fit diagnostics for OLS
  - Readings
    * Gujarati and Porter: Chapter 8; Appendix C, Sections C.6 -C.8
    * Davidson and MacKinnon: Chapter 4

    * Chapters 9 and 10 from Kutner, Nachtsheim, Neter, and Li. *Applied Linear Statistical Models, 5th edition.* McGraw-Hill Irwin. (Distributed electronically.)

- February 27: The Central Limit Theorem and Hypothesis Testing using Simulation **Last Day for Exam One Material**

  – Topics: power analysis; fit statistics in large samples; hypothesis testing with simulation; dummy variables; polynomial terms; DV and IV transformations

  – Readings

    * Gujarati and Porter: Chapter 6; Chapter 7, Section 7.9-7.10; Chapter 9
    * Murphy et al.: *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests, Third Edition.* Chapters 1 and 2.
    * Brambor, Thomas, William Clark and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14: 63-82
    * Gelman and Hill: Chapter 7

- March 5: Problems and Solutions - Omitted Variables, Multicollinearity, and Heteroskedasticity

  – Topics: effect of model misspecification, correlated regressors, and heteroskedasticity on estimates; specification diagnostics and tests; White's robust standard errors; FGLS

  – Readings

    * Gujarati and Porter: Chapter 10; Chapter 11; Chapter 13, Sections 13.1-13.3
    * Davidson and MacKinnon: Chapter 5, Section 5.5; re-read Chapter 3, Section 3.7
    * Reread Chapters 9 and 10 from Kutner, Nachtsheim, Neter, and Li. *Applied Linear Statistical Models, 5th edition.* McGraw-Hill Irwin. (Distributed electronically.)
    * Long, J. Scott and Laurie H. Ervin. *The American Statistician* 54(3): 217-224. "Using Heteroskedasticity Consistent Standard Errors in the Linear Regression Model."
    * Clarke, Kevin A. "The Phantom Menace: Omitted Variable Bias in Econometric Research." *Conflict Management and Peace Science* 22(4): 341-352.
    * Achen, Christopher H. "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong." *Conflict Management and Peace Science* 22(4): 327-339.
    * Oneal, John R. and Bruce Russett. "Rule of Three, Let It Be? When More Really Is Better." *Conflict Management and Peace Science* 22(4): 293-310.
    * Davidson and MacKinnon: Chapter 7, Sections 7.1-7.5

- March 12: Spring Break (no class)
- March 19: Problems and Solutions - Measurement Error and Endogeneity
  - Topics: effect of measurement error on estimates; instrumental variables and two-stage least squares analysis
  - Readings
    * Gujarati and Porter: Chapter 13, Section 13.5; Chapter 20
    * Davidson and MacKinnon, Chapter 8
- March 26: Problems and Solutions - Spatial and Temporal Correlation
  - Topics: fixed and random effects models; panel-corrected standard errors; dynamic models; FGLS
  - Readings
    * Gujarati and Porter: Chapter 16; Chapter 17, Sections 17.1-17.3, 17.8
    * Beck, Nathaniel and Jonathan N. Katz. 1995. "What to do (and not to do) with time-series cross-section data." *American Political Science Review* 89(3): 634-647.
    * Wilson, Sven, and Daniel M. Butler. 2007. "A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications." *Political Analysis* 15: 101-123.
    * Davidson and MacKinnon: Chapters 7 and 13
    * *Optional:* Judson Ruth A. and Ann L. Owen. 1999. "Estimating dynamic panel data models: A guide for macroeconomists." *Economic Letters* 65:9–15.
- April 2: Problems and Solutions - Spatial and Temporal Correlation Part II
  - Topics: hierarchical approaches to spatial and temporal correlation.
  - Readings
    * Gelman and Hill: Chapters 11-12, Chapter 13, Section 13.1
- April 9: Problems and Solutions - Non-linear Models
  - Topics: maximum likelihood; introduction to logit/probit
  - Readings
    * Gujarati and Porter: Chapters 14 and 15
    * Davidson and MacKinnon: Chapter 10, Sections 10.1-10.5; Chapter 11, Sections 11.1-11.3
- April 16: Problems and Solutions - Neural Network Models and Missing Data
  - Readings

    - Chapter 13, Section 6 from Kutner, Nachtsheim, Neter, and Li.
    - Chapter 8, Section 10 from Venables and Ripley, *Modern Applied Statistics with S.*
    - Gelman and Hill: Chapter 25
    - Royston, Patrick. 2004. "Multiple imputation of missing values." *Stata Journal* 4(3): 227-241. URL: `http://www.stata-journal.com/sjpdf.html?articlenum=st0067`
    - "Stata Library: Multiple Imputation using ICE." URL: `https://www.ats.ucla.edu/stat/Stata/library/ice.htm`

- April 23: Student Conference Day One
- April 30: Student Conference Day Two
- TBD: **Final Exam**

Students with Disabilities

Emory University complies with the regulations of the Americans with Disabilities Act of 1990 and offers accomodations to students with disabilities. All students with special requests or need for accomodations should make this request to Prof. Esarey as soon as possible.[1] Documentation from the Emory Office of Disability Services is required; see url-http://www.ods.emory.edu/students.htm for more details.

Syllabus Change Policy

The policies of this syllabus may be changed by Prof. Esarey with advance notice.

[1]This statement is quoted from the Office of Faculty Resources for Disabilities website at `http://www.portals.emory.edu/sylideas.html`.