Government 391L
Spring, 2016
Unique number: 38975

Robert C. Luskin
University of Texas at Austin
M 6:30-9:30, MEZ 1.202

Office: BAT 3.148
Phone: 650-380-0430, 512-879-9974
Office Hours: M 2:30-3:30, T 1:30-2:30, W 2:30-3:30, & by appointment

TA: Robert B. Shaffer
Office: BAT 1.118
Office Hours: TBA

Statistical Analysis in Political Science II:
Regression Models

This course is about "regression models," roughly and broadly defined as statistical models to explain some single dependent variable. The domain thus includes nonlinear as well as linear models and models for qualitative as well as quantitative dependent variables. I shall try to satisfy two clienteles: aspiring methodologists and aspiring practitioners. The course outline, below, is ambitious, and we may well not get all the way through it, but I do hope to get at least through Topic XIV.

There will be computer-based exercises to provide concrete examples, but the lectures and readings will focus on general questions of modeling, estimation, inference, and interpretation: What sorts of models imply and reflect what sorts of relationships between independent and dependent variables? What assumptions must we make, and what do they mean? How likely are the assumptions to be violated, and with what consequences? How can we tell when violations occur? What alternative assumptions might we make? What quantities should we be focusing on estimating? What estimators provide statistically desirable estimates? Where several different estimators might serve, what are their advantages and disadvantages? What do the estimates tell us, and how certainly?

The lectures and readings will treat these questions practically but abstractly, referring more to *x*'s and *y*'s than to substantive variables. There will be much mathematical notation and mathematically phrased argument and some proof and derivation. The goal is to convey a good, relatively deep understanding of the how's and why's of constructing, estimating, and interpreting the estimates of these models.

To get the taste of actual modeling and analysis, we shall examine some published examples and work through a series of mostly computer-based exercises asking you to write and analyze your own models, rooted in your own substantive interests. I shall help you find datasets, if you don't have any you are already working on or interested in. There will also be a modest term paper, to be centered on a regression model of your devising.

There will be extra sessions, yet to be scheduled, of approximately two hours per week, to review the mathematical and statistical background, go over questions and assignments, and discuss concrete applications.

The course outline follows. We shall make our way through it at a play-it-by-ear pace, depending how much attention each topic seems to need. We shall use the UT Canvas website, on which I shall post some course materials, and through which I shall send emails as necessary. Students with disabilities may request appropriate academic accommodations from the Division of Diversity and Community Engagement, Services for Students with Disabilities, 471-6259, http://www.utexas.edu/diversity/ddce/ssd/.

Prerequisites

The course requires a decent knowledge of descriptive and inferential statistics—as covered in the Government Department's Statistics I or equivalent—and a reasonable facility with ordinary algebra (but the more math you know, the better). I encourage you to consult me if you are unsure whether this is an appropriate course for you to take.

Texts

The texts, selected for simplicity without simplism, are:

Damodar N. Gujarati and Dawn C. Porter. 2008. *Basic Econometrics* (5th ed.). New York: McGraw-Hill.

Jan Kmenta. 1997. *Elements of Econometrics* (2nd ed.). Ann Arbor, MI: University of Michigan Press.

Jeffrey M. Wooldridge. 2012. *Introductory Econometrics: A Modern Approach* (5th ed.). Mason, OH: Cenage Learning.

You should read at least two of these three, though reading all three, if you can manage it, might be still better.

If you are having trouble, you might also have a look at the simpler:

A.H. Studenmund. 2010. *Using Econometrics: A Practical Guide* (6th ed.). Reading, PA: Addison-Wesley.

Topics and Reading Assignments

I Mathematical and Statistical Review

Kmenta, chs. 1-6.
Gujarati & Porter, app. A.
Wooldridge, apps. A-D.

II The Regression Game: Models and Meanings

Gujarati & Porter, chs. 1-2, secs. 7.1 (except the bit about assumptions)-7.3, and peek ahead to secs. 6.4-6.9, 7.9-7.10.

Kmenta, secs. 7.1, 7.2.
Wooldridge, ch. 1, secs. 2.1, 3.1
You may also wish to peek ahead to the readings under VII

III Classical Assumptions

Gujarati & Porter, secs. 3.2, 4.2, and 7.1 (the bit about assumptions).
Kmenta, secs. 7.1, 7.2.

IV Estimating the Coefficients: Ordinary Least Squares, Maximum Likelihood

Gujarati & Porter, secs. 3.1, 3.3-3.4, 4.1, 4.3-4.4, 7.4-7.7, pp. 217-19, apps. 3A, 4A.
Kmenta, sec. 7.3.
Wooldridge, secs. 2.2-2.6, 3.2-3.6, ch. 5.

V Forecasting, Fit, and Inference

Gujarati & Porter, secs. 3.5, 7.5, 5.2-5.10, 8.1-8.8
Kmenta, sec. 7.4.
Wooldridge, ch. 4, sec. 6.4.

VI Matters of Interpretation

Gujarati & Porter, sec. 6.1-6.3, 7.2-7.3, 7.6, app 6A.2.
Wooldridge, secs. 6.1, 6.3.
Gary King. 1986. How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science. *American Journal of Political Science*, 35: 1032-46.
Robert C. Luskin. 1991. Abusus Non Tollit Usum: Standardized Coefficients, Correlations, and R^2s. *American Journal of Political Science*, 35: 1032-46.

VII Trouble in Regression City: Heteroskedasticity and Autoregression (Introducing Generalized Least Squares)

Gujarati & Porter, chs. 11-12
Kmenta, ch. 8, sec. 9.2.
Wooldridge, secs. 8.1-8.4, 10.1, 10.3, 11.2, 12.1-12.3

VIII More Trouble: Collinearity and Related Ills

Gujarati & Porter, ch. 10.
Kmenta, sec. 10.3.

IX Still More Trouble: Correlations of Regressor and Disturbance

Gujarati & Porter, secs. 7.7, 13.2-13.5.
Kmenta, sec. 10.4.
Wooldridge, sec. 9.4

X Diagnostics, Specification Tests, and Specification Searching

Gujarati & Porter, secs. 5.12, 13.1, 13.6-13.10.
Wooldridge, secs. 9.1-9.2.

XI Linear Regression Encore (in Matrix Notation)

Gujarati & Porter, app B.
Kmenta, secs. 10.1, 10.2, 12.1.
Wooldridge, app. E.

XII Nonlinear Models

Gujarati & Porter, secs. 6.3-6.9, 7.10-7.11.
Kmenta, sec. 11.3.
Wooldridge, secs. 6.4, 9.3.

XIII Models with Discrete Regressors (including ANOVA and ANCOVA)

Gujarati & Porter, ch. 9.
Kmenta, pp. 460-73.
Wooldridge, secs. 7.1-7.4.

XIV Models for Discrete or Limited Dependent Variables (Introducing Logit, Probit, Tobit, and Poisson Regression)

Gujarati & Porter, ch. 15.
Kmenta, sec. 11.5.
Wooldridge, secs. 7.5-7.7, 8.5, 17.1-17.4.

XV Some Models for Time-Series: Lagged Variables and Distributed Lags

Gujarati & Porter, ch. 17.
Kmenta, sec. 11.4.
Wooldridge, sec. 10.2

XVI Pooled Time-Series of Cross-Sections and Seemingly Unrelated Regressions

Gujarati & Porter, ch. 16.
Kmenta, secs. 12.2-12.3.
Wooldridge, ch. 13.

Assignments and Grading

There will be two exams, a series of exercises, and a modest term paper (effectively, a final, freer-form exercise). The exams will be in-class and closed-book. The exercises will be a

mix of pen-and-paper and computer-based, the former to help cement the math, the latter, calling for you to write and analyze models of your own, to provide a taste of actual modeling. The term paper should propose, justify, detail, estimate, and interpret the results from a regression model of your devising, using data of your choosing. It may build on but must go well beyond the exercises. The grading will be mostly on the basis of the evidence of your mastery of the course material, rather than on the paper's substantive merits.

I offer two options for the aggregating the grades on the exercises, exam, and term paper into a course grade (to include a plus and minus, as merited). Under option A —for students aiming to become methodologists or sophisticated practitioners—the exams will count for 60% (30% apiece), the exercises collectively for 25% (each receiving equal weight), and the term paper for 15%. Under Option B—for students aiming to become competent rather than sophisticated practitioners, much less methodologists—the exams will count for 25% (10% apiece), the exercises collectively for 25% (each receiving equal weight), and the term paper for 50%. The maximum course grade under Option B will be A-.

Option B is intended to give students who have trouble with the more abstract and mathematical material but nonetheless learn something about how to write, estimate, and interpret these models a better chance of earning a higher grade. The A- ceiling is necessary because a good performance at that level still misses some important depth of understanding. Students may choose between Options A and B as late as one day after the second exam—though I strongly recommend making a provisional choice sooner, in the interest of making the most efficient allocation of effort.