SYLLABUS
Political Research and Analysis

PS-2030
Spring 2017
4430 Wesley W Posvar Hall – W 9:30-11:55

INSTRUCTOR : Dr. Michaël Aklin
DEPARTMENT : Department of Political Science, University of Pittsburgh
OFFICE : Posvar Hall, Room #4814
EMAIL : aklin@pitt.edu
OFFICE HOURS : Monday 10-12pm, or by appointment

Course Objectives

This course follows PS2010 and PS2020. It is an introduction to modern quantitative tools in political science. What I want is for the studenst to:

- Understand the basic tools of quantitative studies
- Know how to apply these tools
- See where the discipline is heading

The class focuses on the linear model and how to improve the (causal) inferences we make when using it. The last classes are devoted to other models, such as logistic regressions.

Requirements

Your final grade is based on the following components.

Four homework (15% each). They can be done in R or Stata. I'll generally be using Stata, but you should feel free to use R. You should bring your laptop (and install). I strongly encourage you to submit your homework in LaTeX. Your homework should be clean. This means: no copy-paste from Stata/R, tables should be numbered and labeled (e.g. *Table 2: The effect of democracy on GDP.*), variables should have proper labels (e.g. *GDP per Capita* instead of *gdpcap2005*), pretty figures, and so forth. To generate tables in LaTeX(or Word, for that matter), check out packages such as *estout* (for Stata) and *stargazer* (for R).

Final paper + an appendix (20%+20%=40%). Due at the end of the semester (though I'll ask for updates throughout the semester; see calendar below). You need to submit a paper (20%) that employs some of the tools we covered in class (e.g. linear regression). The paper should be structured and presented like an article. This means that it should have an introduction, a very brief literature review that situates the paper in current scholarship, a

theory, testable hypotheses, a research design (+ data), results, and a conclusion. Grading will be based on the quality of the paper overall (structure, clarity, etc.), the quality of its research design, and the discussion of the results. This doesn't mean that the design has to be perfect. However, I do want to see that you have thought through the possible threats to the validity of your findings.

In addition, you need to submit a technical appendix (20%). This appendix will contain all the data work that didn't make it in the main paper. This includes summary statistics as well as additional tests that are typically conducted to assess and show the validity of your empirical analysis. More on this during the semester.

Late submission penalty: 5% of your final grade per day.

The tentative schedule (subject to change):

- Wed, Jan 25: distribution of homework 1
- Wed, Feb 1: submission of homework 1; distribution of homework 2
- Web, Feb 8: submission of homework 2
- Wed, Feb 15: paper proposal (hard copy due in class). Should contain: a research question, a research design, and the data you plan to use (dependent and independent variable(s)).
- Wed, March 15: distribution of homework 3
- Wed, March 29: submission of homework 3
- Wed, April 5: distribution of homework 4
- Wed, April 12: submission of homework 4
- Wed, April 26: submission of the final paper and the appendix (hard copy and by email) by the beginning of the class.

Textbooks & Readings

There are a number of useful textbooks out there. You should definitely acquire at least one of them. The one I'll follow most closely is Wooldridge's 5th edition:

- (hereafter **W**): Jeffrey M. Wooldridge. *Introductory Econometrics*. South-Western Cengage Learning, 5th edition, 2012

You should be able to find recent editions at an affordable price (I see second hand copies on amazon for $30-50). This being said: you could also acquire a different book if you want (see list below). However, it will be your job to find the corresponding chapters.

If you don't want to use **W**, then you should look at some of these books:

- Good alternative (very clear but more concise than **W**). Be careful: don't confuse it with his other textbooks. David A Freedman. *Statistical Models: Theory and Practice*. Cambridge University Press, New York, 2009
- Good introductory book that also covers some probability theory: Sean Gailmard. *Statistical Modeling and Inference for Social Science*. Cambridge University Press, New York, 2014
- Advanced: Russell Davidson and James G. MacKinnon. *Econometric Theory and Methods*. Oxford University Press, Oxford, 2003
- Advanced (but easier than Davidson and MacKinnon): Fumio Hayashi. *Econometrics*. Princeton University Press, Princeton, 2000
- More advanced than **W**: Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT press, Cambridge, MA, 2002
- Useful to understand the intuitions behind basic concepts of econometrics. Peter Kennedy. *A Guide to Econometrics*. Blackwell Publishing, Malden, MA, 5th edition edition, 2003
- A good reference: William H. Greene. *Econometric Analysis*. Pearson, New York, 6th edition, 2008
- To refresh your notions of probability theory: Morris H. DeGroot and Mark J. Schervish. *Probability and Statistics*. Addison-Wesley, Boston, 2002

Besides textbooks, the following books are likely to be useful throughout your studies:

- Great introduction to causal inference (**MW**): Steven L. Morgan and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, Cambridge, 2007
- Good presentation of modern econometric techniques (**AP**): Joshua Angrist and Steffan J. Pischke. *Mostly Harmless Econometics*. Princeton University Press, Princeton, 2008

Misc.

Cheating/plagiarism will not be tolerated. Students suspected of violating the University of Pittsburgh Policy on Academic Integrity, from the February 1974 Senate Committee on Tenure and Academic Freedom reported to the Senate Council, will be required to participate in the outlined procedural process as initiated by the instructor. A minimum sanction of a zero score for the quiz or exam will be imposed.

View the complete policy at `www.cfo.pitt.edu/policies/policy/02/02-03-02.html`.

Class Schedule

The schedule is subject to change. Also, this is not a weekly schedule, i.e. each topic can last more (rarely less) than a week.

1. Introduction

Introduction to the class. Potential outcome framework. Causal identification. Treatment effects. Math, statistics, and Stata review. Conducting a research project.

To read:

- **MW**, chapter 1, 2
- Jonathan Nagler. Coding style and good computing practices. *Political Science and Politics*, 28:488–492, 1995

2. Regression - Basics

Linear model. Ordinary least squares. Bivariate regression.

To read:

- **W**, chapter 2.

3. Regression - Advanced

Multiple regression. Mechanics of regression.

To read:

- **W**, Appendix D, E.1 , chapter 3 up to to 3.3 (included), Appendix 3A

4. Inference with Regression

Sampling distribution. Standard errors. Hypothesis testing. Confidence intervals. P-values. Multiple comparison. F statistic. R^2. Review of the assumptions of regression.

To read:

- **W**, Appendix C.5, C.6, E.2, 3.4, and 4 (entire chapter)
- The following are short articles:
  - On power: Martin Krzywinski and Naomi Altman. Points of significance: Power and sample size. *Nature Methods*, 10(12):1139–1140, 2013

  - Multiple comparison: Martin Krzywinski and Naomi Altman. Points of significance: Comparing samples – part II. *Nature Methods*, 11(4):355–356, 2014
  - On p-values with multiple comparisons: Naomi Altman and Martin Krzywinski. Points of significance: P values and the search for significance. *Nature Methods*, 14 (1):3–4, 2017

- On confidence intervals: Rink Hoekstra, Richard D. Morey, Jeffrey N. Rouder, and Eric-Jan Wagenmakers. Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, 21(5):1157–1164, 2014
- Optional: Andrew Gelman and Eric Loken. The garden of forking paths. Working Paper, November 2013

5. Problems and Solutions: Heteroskedasticity and Multicollinearity

Heteroskedasticity. Robust standard errors. Multicollinearity. Qualitative variables.

To read:

- **W**, 8 up to 8.3 (included)

6. Problems and Solutions: Model Specification

Specifying a functional form. Outliers. Nonlinear effects. Interaction effects.

To read:

- **W**, 6.2, 7 up to 7.4 (included)
- Thomas Brambor, William R. Clark, and Matt Golder. Understanding interaction models: Improving empirical analyses. *Political Analysis*, 14(1):63, 2006

7. Problems and Solutions: Omitted Variable Bias and Endogeneity (1)

Endogeneity. Omitted variable bias. Instrumental variables. Two-staged least square.

To read:

- **W**, 15
- Allison J. Sovey and Donald P. Green. Instrumental variables estimation in political science: A readers' guide. *American Journal of Political Science*, 55(1):188–200, 2011
- Optional: **AP**, 4

Example:

- Daron Acemoglu, Simon Johnson, and James A. Robinson. The colonial origins of comparative development: An empirical investigation. *The American Economic Review*, 91 (5):1369–1401, 2001
- Edward Miguel, Shanker Satyanath, and Ernest Sergenti. Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, 112(4):725–753, 2004
- Oeindrila Dube, Omar Garcia-Ponce, and Kevin Thom. From maize to haze: Agricultural shocks and the growth of the mexican drug sector. CGD Working Paper 355, February 2014

8. Problems and Solutions: Omitted Variable Bias and Endogeneity (2)

Natural experiment. Regression discontinuity. Randomized controlled trial.

To read:

- **AP**, 1
- Thad Dunning. Improving causal inference: Strengths and limitations of natural experiments. *Political Research Quarterly*, 61(2):282–293, 2008
- Optional:
  - **AP**, 6
  - Jasjeet S. Sekhon and Rocio Titiunik. When natural experiments are neither natural nor experiments. *American Political Science Review*, 106(1):35–57, 2 2012

Example:

- Melissa Dell. The persistent effects of peru's mining mita. *Econometrica*, 78(6):1863–1903, 2010

9. Problems and Solutions: Temporal Variation

Serial correlation. Cross section and time series. Fixed effects. Differences-in-differences.

To read:

- **W**, 10
- Optional:
  - **W**, 12, 13
  - **AP**, 5

Example:

- Arindrajit Dube, Oeindrila Dube, and Omar Garcia-Ponce. Cross-border spillover: U.S. gun laws and violence in mexico. *American Political Science Review*, 107(3):397–417, 8 2013

10. Other Models (1)

Linear probability model. Logit. Probit. Maximum likelihood.

To read:

- **W**, 17 up to 17.1 (included)

11. Other Models (2)

Ordered logit. Survival models.

To read:

- Notes to be distributed in class.
- Ben Ansell. The political economy of ownership: Housing markets and the welfare state. *American Political Science Review*, 108(2):383–402, 2014

12. Extra Topics

Possibly: matching, etc.