

Statistical Methods for Cross-National Research

Adam Przeworski

December 21, 2004

This is an advanced course devoted to conceptual, statistical, and practical problems entailed in evaluating effects of institutions, policies, or events on any kind of outcomes. The course will proceed from big philosophical to minute practical issues. After we become familiar with some examples in the recent literature, we will examine different concepts of causality. Then we will proceed to different estimators of causal effects.

Students will be given a small master dataset, *SELECT05*, and access to a large collection of cross-national data covering the period 1946-2002. Most of the statistical work will be done in *STATA*, but we will also use *LIMDEP* and perhaps *R*.

The prerequisite (no exceptions) for the course is Quant II or equivalent. Students must have basic familiarity with instrumental variables estimators, non-linear regression models, and time-series-cross-section models, as well as be at ease with computer operations. The number of students will be limited to the number of available computers.

There will be weekly assignments during a part of the course. A short final paper will be the main basis for grading. The pace of the course will depend on student progress.

If possible, read in the order indicated.

Note: The course will meet twice a week during the first seven weeks of the semester, so that students will have the remainder of the period to write papers.

1. Introduction

Generic problems in cross-national research. Effects of institutions, policies, or events. Effects of causes and causes of events. Causality and counterfactuals. Identifying causal effects. Experimental studies. Observational studies. Historical studies. Biases. Estimators. Endogeneity and manipulation.

2. Substantive Example: Political Institutions and Economic Development

Random selection. Non-random selection. Endogeneity in the causal sense. Endogeneity in the statistical sense. Endogeneity is selection. Endogeneity and manipulability. Identifying causal effects. Instrumental variables. Finding instruments.

Read:

Przeworski, Adam. 2004. "The Last Instance? Are Institutions the Primary Cause of Economic Development?" *European Journal of Sociology* 15: 165-188.

Banerjee, Abhijit, and Lakshmi Iyer. 2002. "History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." Working Paper. Department of Economics, MIT.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *The Quarterly Journal of Economics* 117: 1231-1294.

Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. "Political Institutions and Human Capital in Economic Development." Ms. Department of Economics, Harvard University.

3. Introduction to Data and Software.

Select05 file. Large data set. Stata, Limdep, R. Papers.

Read: Codebooks (on class page).

Assignment: Using each software open Select05 data file. Compare means of G, YG, and POPG for REG=1 and REG=0. Regress YG on REG and on REG and LEVEL. What did you learn?

4. Causality and Counterfactuals

Can causality be assessed without counterfactuals? The logical problem with counterfactuals. Metalinguistic approach. Semantic approach. Causality and manipulability. Disciplining counterfactuals. The Rubin model. The Dawid critique. Logical validity vs statistical identification.

Read:

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81: 945-960.

Dawid, A.P. 2000. "Causal Inference without Counterfactuals." *Journal of the American Statistical Association* 95: 407-424.

Pearl, Judea. 2000. "Epilogue" to *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press. Pages 331-358.

Berk, Richard A. 2004. *Regression Analysis: A Constructive Critique*. Thousand Oaks: Sage. Section 10.5.

Assignment: Choose a problem to study. Write a few of pages motivating and defining this problem.

5. Evaluating Causal Effects: Preview of the Issues

What is a causal effect? The hypothesis of no effect. Average treatment effect. Models of treatment effects. Response schedules. What is an intervention? Studying causality statistically. Can one evaluate effects without identifying causes? Experiments, quasi-experiments, and history. Random assignment vs. structural modeling.

Read:

Rosenbaum, Paul R. 2002. *Observational Studies*. New York: Springer-Verlag. 2nd edition. Sections 2.1, 2.2, 2.3, 2.4.1, 2.4.2, 2.4.5, 2.5.1, 2.5.2, 2.5.3.

Berk, Richard A. 2004. *Regression Analysis: A Constructive Critique*. Thousand Oaks: Sage. Chapter 5.

Heckman, James J. 2004. "The Scientific Model of Causality." Working Paper. Department of Economics, University of Chicago.

6. Biases

Why biases? Which biases? Baseline bias. Self-selection bias. Post-treatment effect. Distance. SUTVA.

Read:

Reread the relevant sections of Heckman 2004.

King, Gary, and Langche Zeng. 2002. "When Can History be Our Guide? The Pitfalls of Counterfactual Inference." <http://GKing.Harvard.Edu>

7. Estimators: An Overview

An overview of estimators. Biases they correct. Importance of assumptions. Robustness under different assumptions.

Read: (These texts cover similar grounds. If you read any two, it should be enough)

Duffo, Esther. 2002. "Empirical Methods." Class notes. Department of Economics, MIT.

Persson, Torsten and Guido Tabellini. 2003. *The Economic Effects of Constitutions*. Cambridge: MIT. Chapter 5.

Winship, Christopher, and Stephen L. Morgan. 1999. "The Estimation of Causal Effects from Observational Data." *Annual Review of Sociology* 25: 659-707.

Angrist, Joshua D. and Alan B. Krueger. 1999. "Empirical Strategies in Labor Economics." Chapter 23 in O. Ashenfelter and D. Card (eds.), *The Handbook of Labor Economics*, vol III. North Holland.

8. Non parametric Estimators

Estimands. Assumptions. Matching on observables. The Rubin-Rosenbaum theorem. Optimal matching. Sensitivity analysis. Pitfalls of matching. Regressions on observables. Relation to research design debates.

Read:

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press. Chapter 18.

Rosenbaum, Paul R. 2002. *Observational Studies*. New York: Springer-Verlag. 2nd edition. Section 3.2 and Chapter 10.

Imbens, Guido W. 2002. "Semiparametric Estimation of Average Treatment Effect under Exogeneity: A Review." Working Paper. Department of Economics, University of California at Berkeley.

Becker, Sascha O. and Andrea Ichino. 2002. "Estimation of average treatment effects based on propensity scores." *The Stata Journal* 7: 1-19.

Assignment: Implement file "Match.do" using your problem. Interpret the results.

9. Parametric Estimators: Instrumental Variables

Why instruments? Definition of an instrument. The two assumptions. 2SLS. When does IV solve the problem?

Read:

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press. Chapter 5.

or, more advanced,

Greene, William H. 2002. *Econometric Analysis*. Fourth Edition. Upper Saddle River, NJ: Prentice Hall. Sections 16.1-16.5.

Heckman, James J. 1996. "Instrumental Variables: A Cautionary Tale." Technical Working Paper No. 185. Cambridge, MA: National Bureau of Economic Research.

or (This is the published version, but I find the working paper clearer.),

Heckman, James J. 1997. "Instrumental Variables: A Study in Implicit Behavioral Assumptions Used in Making Program Evaluations." *The Journal of Human Resources* 32: 441-462.

Assignment: Apply IV using your problem. Interpret the results.

10. Parametric and Semi-Parametric Estimators: Selection on Unobservables

Heckman's method. Varieties and extensions. Univariate selection. Multinomial selection. Bivariate selection. Static and dynamic selection models. Distributional assumptions. Semi-parametric estimation.

Read:

Przeworski, Adam, Michael R. Alvarez, José Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Well-being in the World, 1950-1990*. New York: Cambridge University Press. Appendix 1.

Greene, William H. 2002. *Econometric Analysis*. Fourth Edition. Upper Saddle River, NJ: Prentice Hall. Section 21.4 (study the preceding sections if you do not understand this one).

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press. Sections 17.1-17.6.

Assignment: Write the same as file "Dprobit.lim" either in *STATA* or *R*. Apply Heckit using your problem. Interpret the results..

11. Fixed Effects and Difference-in-difference Estimators

Using natural experiments to identify causal effects. Difference-in-difference. Fixed effects. Rival hypotheses. Measurement errors.

Read:

Reread the relevant sections of Duflo 2002 and Angrist and Krueger 1999.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press. Sections 6.3.1 and 10.6.4.

Assignment: Implement a DiD estimator using your problem. Interpret the results.

12. Overview

What have we learned about the impact of political institutions on economic development? What have we learned about evaluating causal effects? Is the science of comparative politics possible? Research design in cross-national studies.

Assignment: Collect all you did and write the paper.