

EMPIRICAL STRATEGIES IN APPLIED ECONOMETRICS

Institute for the Study of Labor
April 2002

Summer School in Labor Economics
J. Angrist

This short-course covers topics in econometrics and discusses empirical modeling strategies that I have found especially useful for applied research. The main theoretical ideas are illustrated with examples.

OUTLINE AND READINGS

I. AGNOSTIC REGRESSION

G. Chamberlain, "Panel Data," Chapter 22 in *The Handbook of Econometrics*, Volume II, Amsterdam: North-Holland, 1983.

II. CAUSAL REGRESSION AND REGRESSION VS. MATCHING

J. Angrist and A. Krueger, "Empirical Strategies in Labor Economics," Chapter 23 in O. Ashenfelter and D. Card, eds., *The Handbook of Labor Economics*, Volume III, North Holland, 1999.

J. Angrist, "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, March 1998.

III. ESTIMATING THE EFFECT OF TRAINING PROGRAMS

R. Lalonde, "The Promise of Public Sector-Sponsored Training Programs," *The Journal of Economic Perspectives* 9 (Spring 1995), 149-168.

O. Ashenfelter and D. Card, "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs on Earnings," *The Review of Economics and Statistics* 67 (1985): 648-66.

R. LaLonde, "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review* 76 (September 1986): 604-620.

J. Heckman and J. Hotz, "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social programs: The Case of Manpower Training," *JASA* 84 (1989): 862-8.

R. Dehejia and S. Wahba, "Causal Effects in Nonexperimental Studies: Re-evaluating the Evaluation of Training Programs," *JASA* 94 (Sept. 1999).

IV. INSTRUMENTAL VARIABLES

A. Models with constant effects; Wald, grouping, and two-sample IV

J. Angrist and A. Krueger, "Instrumental Variables and the Search for Identification," *Journal of Economic Perspectives*, Fall 2001.

W. Newey, "Generalized Method of Moments Specification Testing," *Journal of Econometrics* 29 (1985), 229-56.

J. Angrist, "Grouped Data Estimation and Testing in Simple Labor Supply Models," *Journal of Econometrics*, February/March 1991.

J. Angrist and A. Krueger, "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," *JASA* 87 (June 1992), 328-36.

- J. Angrist and A. Krueger, "Split-Sample Instrumental Variables Estimates of the Returns to Schooling," *JBES*, April 1995.
- J. Angrist, "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, June 1990.

B. IV with heterogeneous potential outcomes

- G. Imbens and J. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, March 1994.
- J. Angrist, G. Imbens, and D. Rubin, "Identification of Causal effects Using Instrumental Variables," with comments and rejoinder, *JASA*, 1996.
- J. Angrist and G. Imbens, "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *JASA*, June 1995.
- J. Angrist and A. Krueger, "Does Compulsory Schooling Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics* 106, November 1991, 979-1014.

V. MISCELLANEOUS TOPICS

A. Limited dependent variables and quantile treatment effects

- A. Abadie, "Semi-Parametric Estimation of Instrumental Variable Models for Causal Effects," NBER Technical Working Paper No. 261, September 2000.
- J. Angrist, "Estimation of Limited-Dependent Variable Models with Binary Endogenous Regressors: Simple Strategies for Empirical Practice," *The Journal of Business and Economic Statistics*, January 2001.
- A. Abadie, J. Angrist, and G. Imbens, "Instrumental Variables Estimation of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, November, 2001.
- J. Angrist and W. Evans, "Children and their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review*, June 1998, 450-477.

B. Clustering and the Moulton problem

- B. Moulton, "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics* 32 (1986), pp. 385-97.
- K. Liang, and Scott L. Zeger, "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika* 73 (1986), 13-22.
- C.M. Schmidt, Rob Baltussen, and Rainer Sauerborn, "The Evaluation of Community-Based Interventions: Group-Randomization, Limits and Alternatives," IZA DP No. 206, October 2000.
- Z. Feng, P. Diehr, A. Peterson, and D. McLerran, "Selected Statistical issues in Group Randomized Trials," *Annual Review of Public Health* 22 (2001), 167-87.

C. The propensity score paradox

- J. Hahn, "On the Role of the Propensity Score in Efficient Estimation of Average Treatment Effects," *Econometrica* 66, March 1998.
- J. Angrist and J. Hahn, "When to Control for Covariates? Panel-Asymptotic Results for Estimates of Treatment Effects," NBER Technical Working Paper, May 1999.

EMPIRICAL STRATEGIES IN APPLIED ECONOMETRICS

Institute for the Study of Labor
April 2002

Summer School in Labor Economics
J. Angrist

Agenda

I. Agnostic regression – a place to start

Regression as Best Linear Predictor (BLP) for the Conditional Expectation Function (CEF)
Review of large sample theory for OLS estimates

II. Causal regression (our main occupation); regression vs. matching

Linking a regression model with a causal model
The experimentalist view of non-experimental research
Matching to estimate the effect of treatment on the treated
Theoretical comparison of regression and matching
The Angrist (1998) study of the effects of voluntary military service

III. Estimating the effect of training programs – the mother of econometric evaluation problems

Why training programs are hard to evaluate
The Ashenfelter and Card (1985) training evaluation
The credibility of non-experimental training evaluations; Lalonde (1986)
Use of the propensity score in evaluation research
The Dehejia and Wahba (1999) propensity-score study

IV. Instrumental variables (nature's stream of experiments)

A. Constant-effects models

IV and omitted variables bias: long regression without the controls
The Wald estimator and grouped data
Two-sample IV and related methods
The Angrist (1990) study of the effects of Vietnam-era military service

B. IV with heterogeneous potential outcomes

Local average treatment effects; internal vs. external validity
The *compliers* concept; identification of effects on the treated
Models with variable treatment intensity; examples
The Angrist and Krueger (1991) schooling study

V. Miscellaneous topics

- A. Limited dependent variables and quantile treatment effects
- B. Clustering and the Moulton problem
- C. The propensity score paradox

EMPIRICAL STRATEGIES IN APPLIED ECONOMETRICS

Institute for the Study of Labor
April 2002

Summer School in Labor Economics
J. Angrist

Review Questions

1. Discuss the relationship between regression and matching, as described below:
 - a. Suppose all covariates are discrete and you are trying to estimate a treatment effect. Prove that if the regression model for covariates is saturated, then matching and regression estimates will estimate the same parameter (i.e., have the same *plim*) in either of the following two cases: (i) treatment effects are independent of covariates; (ii) treatment assignment is independent of covariates.
 - b. Propose a weighted matching estimator that estimates the same thing as regression.
 - c. Why might you prefer regression estimates over matching estimates, even if you are primarily interested in the effect of treatment on the treated?
 - d. (extra credit) Calculate matching and regression estimates in the empirical application of your choice. Discuss the difference between the two estimates with the aid of a figure like the one used in Angrist (1998) for this purpose.
2. Discuss the link between causal effects and structural parameters in a Bivariate Probit model of the relationship between divorce and female labor force participation. The purpose of the model is to determine whether female employment strengthens a marriage or encourages divorce more by making it easier for women to live independently.

Organize your discussion as outlined below:

- a. Explain in words why the causal effect of employment on divorce is difficult to determine. Is the problem here primarily one of identification or estimation? Could you design an experiment to answer the question of interest?
- b. Write the potential outcomes and potential treatment assignments in your causal model in terms of latent indices with unobserved random errors in a structural model.
- c. What should the population be for this study? What does it mean for employment to be “endogenous” in the structural model? How about in the causal model?
- d. Show how to use the Probit structural parameters and distributional assumptions to calculate LATE, the population average treatment effect (ATE), and the effect of treatment on the treated (ETT). Which of these parameters are identified without distributional assumptions?
- e. Discuss the relationship between the three average causal effects, LATE, ATE, and ETT. Can you say which is likely to be largest and which is likely to be smallest?
- f. (extra credit) Compare OLS with Probit and IV with Bivariate Probit in the application of your choice (as in Angrist, 2001).