

API-208
PROGRAM EVALUATION: ESTIMATING PROGRAM EFFECTIVENESS
WITH EMPIRICAL ANALYSIS

Alexis J. Diamond
Spring 2013

Course Outline:

Program evaluation comprises a set of statistical tools designed to assess the causal effects of interventions, such as job training programs, on outcomes of interest, such as earnings. This is a methodological course, developing skills in quantitative program evaluation. We will study a variety of evaluation designs, from random assignment to quasi-experimental methods. Students will be offered the opportunity to analyze data from actual evaluations. We will analyze the strengths and weaknesses of alternative evaluation methods and practice devising evaluation designs suitable for estimating the impact of real programs.

Goals:

Evaluating the effectiveness of public programs is important, since it can help us decide which programs are working and which are not working, and why. The goal of the course is to prepare students to design, conduct, and critique empirical evaluations of public programs. We will study how to use statistical techniques to evaluate the effects of public programs, focusing experimental and quasi-experimental (observational) methods.

Course Information:

Class Meetings: Wednesdays, 11:40-2:30 at Land Auditorium, Belfer Building.

Review Sessions: TBD

STATA Review: CGIS K018, Tuesday, February 5, 10 am–12:30 pm

Faculty: Alexis Diamond, email: alexis_diamond@hks.harvard.edu
Office: Belfer, Room L-2b
Office hours: Wednesdays, 8:30–10:30 AM or by appointment

Faculty Assistant: Jennifer Hoegen, email: jennifer_hoegen@hks.harvard.edu

TF: Daniel Yew Mao Lim, email: daniellimym@gmail.com
Office Hours: Mondays, 11 AM–noon, Belfer L-2b, from Feb 4 onward

CA: Raul Galicia, email: raul_galicia@hks13.harvard.edu
Office Hours: TBD

Final Exam: Wednesday, May 8, 2–5 PM (subject to change)

Prerequisites:

This course targets mathematically inclined PhD, MPA/ID, MPP, and MPA students. Knowledge of statistical inference and regression analysis at the level of API-202 or API-210 is required: regression analysis, including instrumental variable (2SLS), qualitative dependent variable models (Probit and Logit) and panel data methods (fixed and random effects). We will make use of some elementary calculus and matrix algebra. You are free to use any statistical package of your choice. There will be a STATA tutorial provided.

Readings:

The course material is self-contained and there is no required textbook for the course. Handouts covering most of the material will be distributed in class or through the website. Required readings will be available on the course website.

Assignments and Grading:

Students will be required to write several (3 or 4) short papers over the course of the semester, evaluating program impacts with data provided. There will also be an in-class final exam, to be administered during the reading period. Assignments will count as 60 percent and the final exam/paper as 40 percent for the final grade. Students are encouraged to work in small groups. Each student must write up her or his answers separately. Please append your log files to your assignments when appropriate. Data for the assignments will be available in the course web page in STATA format. STATA (a statistical software package) may be purchased at the CMO, but is also available in the computer lab.

Code of Conduct:

All course activities, including class meetings, homework assignments, and exams, are subject to the HKS academic code and code of conduct. Please be on time and make sure that your cell phone is turned off during class.

Contents of the Course:

Required readings are preceded by a star (), and will be available on the course website. Other readings are included for your reference but they are not required material.*

1 INTRODUCTION

- 1.1 EVALUATION RESEARCH FOR PUBLIC POLICY: PURPOSE. SCOPE. EXAMPLES
- 1.2 THE FUNDAMENTAL IDENTIFICATION PROBLEM: CAUSALITY. COUNTERFACTUAL RESPONSES. HETEROGENEITY. SELECTION
- 1.3 STATISTICAL PREREQUISITES: PROBABILITY. RANDOM VARIABLES. INDEPENDENCE. MEASURES OF LOCATION. MEASURES OF DISPERSION. CONDITIONAL MEAN FUNCTION. INFERENCE

READINGS (overviews and foreshadowing important issues):

- * The New York Times, September 19, 2012, “For Weight Loss, Less Exercise May Be More,” by Gretchen Reynolds.
- * Rodrik, D., “We Shall Experiment, But How Shall We Learn?” (2008).

Angrist, J.D. and A.B. Krueger (2000), “Empirical Strategies in Labor Economics,” in A. Ashenfelter and D. Card eds. *Handbook of Labor Economics*, vol. 3. New York: Elsevier Science. Sections 1 and 2.
- Imbens, G.W. and J.M. Wooldridge (2009) “Recent Developments in the Econometrics of Program Evaluation,” *Journal of Economic Literature*, vol. 47(1), 5-86. (*excellent overview of topics covered in the course.*)
- Kapstein, E. and Kim, R. “The Social and Economic Impact of Standard Chartered Ghana” (*in reference to footnote 1 of Imbens and Woolridge.*)

2 RANDOMIZED EXPERIMENTS

- 2.1 THE ADVANTAGES OF RANDOMIZED STUDIES
- 2.2 THREATS TO INTERNAL AND EXTERNAL VALIDITY
- 2.3 METHODOLOGY. TESTING IN LARGE SAMPLES: ASYMPTOTIC DISTRIBUTION. TESTING IN SMALL SAMPLES: FISHER’S EXACT TEST. PRE-ESTIMATION DIAGNOSTICS
- 2.4 EXAMPLES

READINGS:

- * Bloom, H.S., L. L. Orr, S.H. Bell, G. Cave, F. Doolittle, W. Lin and J.M. Bos (1997), “The Benefits and Costs of JTPA Title II-A Programs,” *Journal of Human Resources*, vol. 32, 549-576.
- * The New York Times, April 26, 2012, “Is Our Adults Learning?” by David Brooks.
- * Gill M.Z., Hall A.B., (2012), “Causal Inference with Text: Judicial Identity and Appellate Opinions”, *working paper*.

Duflo, E., R. Glennerster and M. Kremer (2008), “Using Randomization in Development Economics Research: A Toolkit,” in T.P. Schultz and J.A. Strauss eds. *Handbook of Development Economics*, vol. 4. New York: Elsevier Science.

LaLonde, R. (1986), “Evaluating the Econometric Evaluation of Training Programs with Experimental Data,” *American Economic Review*, vol. 76, 604-620.

Rosenbaum, P.R. (1995), *Observational Studies*. New York: Springer-Verlag. Chapter 2 (*on reserve at HKS Library*).

3 INTRODUCTION TO OBSERVATIONAL STUDIES

3.1 COMPARISON OF EXPERIMENTAL AND OBSERVATIONAL STUDIES

3.2 APPROXIMATING EXPERIMENTS WITH OBSERVATIONAL DATA

3.3 STUDY DESIGN

3.4 EXAMPLES

READINGS:

- * Freedman, D.A. (edited by Collier, D. and Sekhon, J.), *Statistical Models and Causal Inference*. New York: Cambridge University Press. Editors’ Introduction: “Inference and Shoe Leather.”
- * Rosenbaum, P.R., “Reasons for Effects”, *Chance*, (2005), vol 18, 5–10.
- * Rubin, D.B. (2001), “Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation”, *Health Services & Outcomes Research Methodology*, vol 2, 169–188.

Rosenbaum, P.R. (2009), *Design of Observational Studies*. New York: Springer-Verlag. Chapter 1, sections 1-6 (*highly recommended and on reserve at HKS Library*).

4 MATCHING AND REGRESSION

4.1 IDENTIFICATION: SIMPSON’S PARADOX. SELECTION ON OBSERVABLES

4.2 MATCHING ESTIMATORS: MATCHING ON COVARIATES. PROPENSITY SCORE METHODS GENETIC MATCHING

4.3 REGRESSION

4.4 EXAMPLES

READINGS:

- * Dehejia, R.H. and S. Wahba (1999), “Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs,” *Journal of the American Statistical Association*, vol. 94, 1053-1062.

- * Diamond, A. and Sekhon, J., “Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies,” *Review of Economics and Statistics*, forthcoming.)
- Abadie, A. and G.W. Imbens (2006), “Large Sample Properties of Matching Estimators for Average Treatment Effects,” *Econometrica*, vol. 74, 235-267.
- Abadie, A. and G.W. Imbens (2008), “On the Failure of the Bootstrap for Matching Estimators,” *Econometrica*, vol. 76, 1537-1557.
- Cochran, W.G., (1968), “The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies,” *Biometrics*, vol. 24, 295-313.
- Hainmueller, J. (2012) “Entropy Balancing: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies.” *Political Analysis*.
- Härdle, W and O. Linton (1994), “Applied Nonparametric Methods,” in R. F. Engle and D. L. McFadden eds. *Handbook of Econometrics*, vol. 4. New York: Elsevier Science.
- Heckman, J.J., H. Ichimura and P.E. Todd (1997), “Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme,” *Review of Economic Studies*, vol. 64, 605-654.
- Imbens, G.W. (2004), “Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review,” *Review of Economics and Statistics*, vol. 86(1), 4-29.
- Rosenbaum, P.R. (2002), *Observational Studies*. New York: Springer-Verlag. Chapter 3.
- Rosenbaum, P.R., and D. B. Rubin (1983), “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, vol. 70, 41-55.
- Rubin, D.B. (2006), *Matched Sampling for Causal Effects*. Cambridge: Cambridge University Press.
- White, H. (1980), “Using Least Squares to Approximate Unknown Regression Functions,” *International Economic Review*, vol. 21, 149-170.

5 ROBUSTNESS, SENSITIVITY, FALSIFICATION

5.1 ASSESSING THE CONFOUNDING EFFECTS OF UNOBSERVED FACTORS

5.2 A FORMAL SENSITIVITY TEST

READINGS:

- * Rosenbaum, P.R. (2005), “Sensitivity Analysis in Observational Studies,” *Encyclopedia of Statistics in Behavioral Science*, vol. 4, 1809-1814.

Imbens, G.W. (2003), “Sensitivity to Exogeneity Assumptions in Program Evaluation,” *American Economic Review (Papers & Proceedings)*, vol. 93(2), 126-132.

Rosenbaum, P.R. (2002), *Observational Studies*. New York: Springer-Verlag. Chapter 4.

Rosenbaum, P.R. and D.B. Rubin (1983), “Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome,” *Journal of the Royal Statistical Society. Series B*, vol. 45(2), 212-218.

6 DIFFERENCE-IN-DIFFERENCES ESTIMATORS

6.1 IDENTIFICATION: SELECTION ON TIME-INVARIANT CHARACTERISTICS

6.2 DIFFERENCE-IN-DIFFERENCES AS A FIXED-EFFECTS ESTIMATOR

6.3 SYNTHETIC CONTROL METHODS

6.4 EXAMPLES

READINGS:

* Abadie, A., A. Diamond and J. Hainmueller (2010), “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program,” *Journal of the American Statistical Association*, vol. 105, 493-505.

* Abadie, A. and J. Gardeazabal (2003), “The Economic Costs of Conflict: A Case Study of the Basque Country,” *American Economic Review*, vol. 93(1), 113-132.

* Card, D. (1990), “The Impact of the Mariel Boatlift on the Miami Labor Market,” *Industrial and Labor Relations Review*, vol. 44, 245-257.

* Card, D. and A.B. Krueger (1994), “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania,” *American Economic Review*, vol. 84, 772-793.

Abadie, A. (2005), “Semiparametric Difference-in-Differences Estimators,” *Review of Economic Studies*, vol. 72, 1-19.

Ashenfelter, O. and D. Card (1985), “Using the Longitudinal Structure of Earnings to Estimate the Effects of Training Programs,” *Review of Economics and Statistics*, vol. 67, 648-660.

Duflo E. (2001), “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment,” *American Economic Review*, vol. 91, 795-813.

Meyer, B.D. (1995), “Natural and Quasi-Experiments in Economics,” *Journal of Business & Economic Statistics*, vol. 13, 151-161.

7 INSTRUMENTAL VARIABLES

- 7.1 IDENTIFICATION: USING EXOGENOUS VARIATION IN TREATMENT INTAKE GIVEN BY INSTRUMENTS. IMPERFECT COMPLIANCE IN RANDOMIZED STUDIES
- 7.2 METHODOLOGY: THE WALD ESTIMATOR. LOCAL AVERAGE TREATMENT EFFECTS. 2SLS
- 7.3 CAUSAL RESPONSES FOR COMPLIERS
- 7.4 EXAMPLES

READINGS:

- * Angrist, J.D., G.W. Imbens and D.B. Rubin (1996), “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association*, vol. 91, 444-472.
 - * Angrist, J.D. (1990), “Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records,” *American Economic Review*, vol. 80, 313–336.
 - * Hainmueller, J and Kern, H.L. (2009), “Opium for the Masses: How Foreign Media Can Stabilize Authoritarian Regimes”, *Political Analysis*, vol. 17, 377–399.
 - * Imbens, G. (2010), “Better LATE Than Nothing,” *Journal of Economic Literature*, vol. 48, 399–423.
- Abadie, A. (2003), “Semiparametric Instrumental Variable Estimation of Treatment Response Models,” *Journal of Econometrics*, vol. 113, 231-263.
- Angrist J.D. and A. Krueger (1991), “Does Compulsory School Attendance Affect Schooling and Earnings?,” *Quarterly Journal of Economics*, vol. 106, 979-1014.

8 FURTHER TOPICS

- 8.1 DISTRIBUTIONAL EFFECTS
- 8.2 OTHER APPROACHES: THE REGRESSION DISCONTINUITY DESIGN. NONPARAMETRIC BOUNDS. NONPARAMETRIC BOUNDS IN EXPERIMENTAL STUDIES WITH ATTRITION

READINGS:

- * Koenker, R. and Hallock, K.F. (2001), “Quantile Regression”, *Journal of Economic Perspectives*, vol. 15, no. 4, 143–156.
- * Lee, D.S, and T. Lemieux (2010), “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature*, vol. 48, 281-355.

- Abadie, A. (2002), "Bootstrap Tests for the Effects of a Treatment on the Distribution of an Outcome Variable," *Journal of the American Statistical Association*, vol. 97, 284-292.
- Abadie, A., J.D. Angrist and G.W. Imbens (2002), "Instrumental Variable Estimation of the Effects of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, vol. 70, 91-117.
- Hahn, J., P. Todd and W. van der Klaauw (2001), "Identification and Estimation of Treatment Effects with a Regression Discontinuity Design," *Econometrica*, vol. 69, 201-209.
- Hainmueller, J. and Hazlett, C. (2012), "Kernel Regularized Least Squares: Moving Beyond Linearity and Additivity Without Sacrificing Interpretability" *working paper*.
- Imbens, G.W. and T. Lemieux (2008), "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, vol. 142, 615-635.
- Manski, C. (1995), *Identification Problems in the Social Sciences*. Cambridge: Harvard University Press. Chapter 2.
- Rubin, D.B. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *Annals of Statistics*, vol. 6, 34-58.
- Trochim, W. (1990), "The Regression-Discontinuity Design," in L. Sechrest, E. Perrin and J. Bunker eds. *Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data*. Washington, D.C.: U.S. Dept. of HHS, Agency for Health Care Policy and Research.