Applied Econometrics: Mostly Harmless Big Data

MIT 14.387
Spring 2017

J. Angrist (angrist@mit.edu)
Victor Chernozhukov (vchern@mit.edu)
John Firth (jfirth@mit.edu)

This course covers empirical strategies for applied micro research. Our agenda includes regression and matching, instrumental variables, differences-in-differences, regression discontinuity designs, standard errors, and the analysis of high-dimensional data sets, a.k.a. big data.

Course requirements and grading: Students are expected to do the readings. In addition, there are three graded problem sets, which must be submitted on time to be graded for credit (the course is graded pass/fail). The atmosphere is informal, but we ask you to put all electronic devices away when class is in session. We encourage questions and class discussion – we'll be asking you questions too!

We use the following text:

J.D. Angrist and J.S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2009 (MHE).

Examples and review material come in part from:

J.D. Angrist and J.S. Pischke, *Mastering 'Metrics: The Path from Cause to Effect*, Princeton University Press, 2015 (MM).

Articles are posted on Stellar at 14.387 Class home.

ANGRIST READINGS

I. REGRESSION RECAP

MM, Chapters 1-2
MHE, Chapters 1-2 and 3.1-3.2

Bad control and measurement error

MM, 6.1-6.2; Appendix to Chapter 6

Z. Griliches, "Estimating the Returns to Schooling – Some Econometric Problems," *Econometrica* 45 (January 1977).

Limited dependent variables and marginal effects

MHE, Section 3.4.2

II. MATCHING AND TRAINING

Matching

MHE, Section 3.3.1

J. Angrist, "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica* 66(2), 1998, 249-288.

A. Abadie and G. Imbens, "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica* 74(1), 2006, 235-267.

G. Imbens, "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *The Review of Economics and Statistics,* 86(1), 2004, 4-29.

The propensity score

MHE, Sections 3.3.2-3.3.3

J. Angrist and A. Krueger, "Empirical Strategies in Labor Economics," Chapter 23 in *The Handbook of Labor Economics, Volume III*, Elsevier, 1999; Section 2.2-2.3.

O. Ashenfelter, "Estimating the Effect of Training Programs on Earnings," *The Review of Economics and Statistics* 60, 1978, 47-57.

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," *The American Economic Review* 76, September 1986, 604-62.

J. Heckman and J. Hotz, "Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social programs: The Case of Manpower Training," *Journal of the American Statistical Association* 84(408), 1989, 862-874.

P. Rosenbaum and R. Rubin, "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score," *Journal of the American Statistical Association* 79(387), 1984, 516-524.

R. Dehejia and S. Wahba, "Causal Effects in Nonexperimental Studies: Re-evaluating the Evaluation of Training Programs," *Journal of the American Statistical Association* 94(448), 1999, 1053-1062.

J. Smith and P. Todd, "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, 125(1), 2005, 305-353.

P. Kline, "Oaxaca-Blinder as a Reweighting Estimator," *The American Economic Review* 101, May 2011, 532-37.

J. Hahn, "On the Role of the Propensity Score in Efficient Estimation of Average Treatment Effects," *Econometrica* 66(2), 1998, 315-331.

J. Angrist and J. Hahn, "When to Control for Covariates? Panel Asymptotics for Estimates of Treatment Effects," *The Review of Economics and Statistics* 86, February 2004, 58-72.

K. Hirano, G. Imbens, and G. Ridder, "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica* 71(4), 2003, 1161-1189.

T. Cook, W.R. Shadish, and V.C. Wong, "Three Conditions Under Which Experiments and Observational Studies produce Comparable Causal Estimate: New Findings from Within-Study Comparisons," *J. Policy Analysis and Management* 27 (Fall 2008).

D. Card, J. Kluve, and A. Weber, "Active Labour Market Policy Evaluations: A Meta-analysis," *The Economic Journal* 120 (November 2010), F452-F477.

A. Abadie and G. Imbens, "Matching on the Estimated Propensity Score." *Econometrica* 84(2), 2016, 781-807.

A. Abdulkadiroglu, J. Angrist, Y. Narita, and P. Pathak, "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation," NBER Working Paper No. 21705, November 2015.

Endogenous stratification

B. Hansen, "The Prognostic Analogue of the Propensity Score," *Biometrika* 95(2), 2008, 481-488.

A. Abadie, M. Chingos, and M. West, "Endogenous Stratification in Randomized Experiments," NBER Working Paper No. 19742, December 2013 (revised April 2014).

III. INSTRUMENTAL VARIABLES

Part 1

2SLS with constant effects; the Wald estimator, grouped data

MM, Chapter 3
MHE, Section 4.1
J. Angrist and A. Krueger, "Instrumental Variables and the Search for Identification," *Journal of Economic Perspectives* 15(4), 2001, 69-85.
J. Angrist, "Grouped Data Estimation and Testing in Simple Labor Supply Models," *Journal of Econometrics* 47(2), 1991, 243-266.
J. Angrist, "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *The American Economic Review*, June 1990, 313-336.

Two-Sample IV and related estimators

MHE, Section 4.3

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," *Journal of the American Statistical Association* 87(418), June 1992, 328-336.
J. Angrist and A. Krueger, "Split-Sample Instrumental Variables Estimates of the Returns to Schooling," *Journal of Business & Economic Statistics* 13(2), April 1995, 225-235.
A. Inoue and G. Solon, "Two-Sample Instrumental Variables Estimators," *The Review of Economics and Statistics* 92, 2010, 557-561.
Pacini, David, and Frank Windmeijer, "Robust inference for the Two-Sample 2SLS estimator." *Economics Letters* 146, 2016, 50-54.

2SLS details

2SLS mistakes: *MHE*, Section 4.6.1
The bias of 2SLS: *MHE*, Section 4.6.4

J. Angrist, G. Imbens, and A. Krueger, "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics* 14(1), 1999, 57-67.
Flores-Lagunes, Alfonso, "Finite-Sample Evidence on IV Estimators with Weak Instruments," *Journal of Applied Econometrics* 22, 2007, 677-694.
M. Kolesar, "Estimation in an Instrumental Variables Model with Treatment Effect Heterogeneity," Princeton Department of Economics, mimeo, November 2013.

Part 2

IV with heterogeneous potential outcomes

MHE, Section 4.4

G. Imbens and J. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica* 62(2), March 1994, 467-475.
J. Angrist, G. Imbens, and D. Rubin, "Identification of Causal Effects Using Instrumental Variables," with comments and rejoinder, *JASA* 91(434), 1996, 444-455.

A. Abadie, "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variables Models," *Journal of the American Statistical Association* 97, March 2002, 284-292.

A. Abadie, "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics* 113, 2003, 231-263.

A. Abadie, J. Angrist, and G. Imbens. "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica* 70(1), 2002, 91-117.

J. Angrist, "Instrumental Variables in Experimental Criminological Research: What, Why, and How," *Journal of Experimental Criminological Research* 2, 2005, 1-22.

J. Angrist, S. Cohodes, S. Dynarski, P. Pathak and C. Walters, "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice," *Journal of Labor Economics* 34(2), April 2016.

Models with variable, continuous, and multiple treatments

MHE, Section 4.5.3

J. Angrist and G. Imbens, "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity," *Journal of the American Statistical Association* 90(430), 1995, 431-442.

J. Angrist and A. Krueger, "Does Compulsory Schooling Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics* 106, November 1991, 979-1014.

D. Card, "The Causal Effect of Education on Earnings," *The Handbook of Labor Economics, Volume IIIA*, Elsevier Science Publishers, 1999.

J. Angrist, G. Imbens, K. Graddy, "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies* 67(3), July 2000, 499-528.

D.E. Powers and S.S. Swinton, "Effects of Self-Study for Coachable Test Item Types, " *Journal of Educational Psychology* 76, 1984, 266-78.

L. Behaghel, B. Crepon, and M. Gurgand, "Robustness of the Encouragement Design in a Two-Treatment Randomized Control Trial," IZA DP No. 7447, June 2013.

A. Abdulkadiroğlu, J. Angrist, P. Hull, and P. Pathak, "Charters without lotteries: Testing takeovers in New Orleans and Boston," *The American Economic Review* 106(7), 2016, 1878-1920.

J. Angrist, P. Hull, P. Pathak, and C. Walters, "Interpreting Tests of School VAM Validity," *AER Papers and Proceedings*, May 2016.

J. Angrist, P. Hull, P. Pathak, and C. Walters, "Leveraging Lotteries for School Value-Added: Testing and Estimation," forthcoming in the *QJE*.

External Validity

J. Angrist, "Treatment Effect Heterogeneity in Theory and Practice," *The Economic Journal* 114, March 2004, C52-C83.

P. Oreopoulos, "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter," *The American Economic Review* 96, March 2006, 152-175; see also Oreopoulos' Corrigendum posted on the AER web site.

J. Angrist, V. Lavy, and Analia Schlosser, "Multiple Experiments for the Causal Link Between the Quantity and Quality of Children," *Journal of Labor Economics* 28(4), October 2010, 773-824.

J. Angrist and I. Fernandez-Val, "ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework," in *Advances in Econometrics Theory and Applications*, Tenth World Congress, Volume III, 2013.

Spec Tests Come LATEly

MHE, Section 4.2.2

G. Imbens and D. Rubin, "Estimating Outcome Distributions for Compliers in Instrumental Variable Models, *Review of Economic Studies* 64, 1997, 555-574.
T. Kitagawa, "A Test for Instrument Validity," *Econometrica* 83(5), 2015, 2043-2063.
M. Huber, "Testing the Validity of the Sibling Sex Ratio Instrument," *Labour* 29(1), 2015, 1-14.
Clement de Chaisemartin, "Tolerating Defiance: LATE Without Monotonicity," forthcoming in Quantitative Economics, 2017.
Angrist, Lavy and Schlosser on QQ, JOLE 2010 (see "External Validity" above)
J. Angrist, P. Hull, P. Pathak, and C. Walters, "Interpreting Tests of School VAM Validity," *The American Economic Review: Papers and Proceedings* (May 2016).

IV. REGRESSION-DISCONTINUITY DESIGNS

Basics

MM, Chapter 4
MHE, Chapter 6

T. Cook, "Waiting for Life to Arrive: A History of the Regression-Discontinuity Design in Psychology, Statistics, and Economics," *Journal of Econometrics* 142(2), 2008, 636-654.
G. Imbens and T. Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics* 142(2), 2008, 615-635.
D. Lee, "Randomized Experiments from Non-Random Selection in U.S. House Elections," *Journal of Econometrics* 142(2), 2008, 675-697.
J. Angrist and V. Lavy, "Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement," *The Quarterly Journal of Economics* 114(2), May 1999, 533-575.
A. Abdulkadiroglu, J. Angrist, and P. Pathak, "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," *Econometrica* 82(1), January 2014, 137-196.

Regression Kinks

D. Card, D. Lee, Z. Pei, and A Weber, "Inference on Causal Effects in a Generalized Regression Kink Design," *Econometrica* 83 (November 2015).
P. Ganong and S. Jäger, "A Permutation Test for the Regression Kink Design," Harvard University Department of Economics, mimeo, March 2015.

Extrapolation

J. Angrist and M. Rokkanen, "Wanna Get Away? RD Identification Away from the Cutoff," *Journal of the American Statistical Association*, December 2015.
Y. Dong and A. Lewbel. "Identifying the effect of changing the policy threshold in regression discontinuity models," *Review of Economics and Statistics* 97(5), 2015, 1081-1092.
M. Rokkanen, "Exam Schools, Ability, and the Effects of Affirmative Action: Latent Factor Extrapolation in the Regression Discontinuity Design" Columbia University Department of Economics, mimeo, March 2015.
R. Maynard, K. Couch, C. Wing, and T. Cook, "Strengthening the Regression Discontinuity Design Using Additional Design Elements: A Within-Study Comparison," *JPAM* 32(4), 2013, 853-877.

Nonpara-ʻmetrics

J. Hahn, P. Todd, and W. van der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica* 69(1), 2001, 201-209.
J. Ludwig and D. Miller, "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *The Quarterly Journal of Economics* 122, 2007, 159-208.
B. Frandsen, M. Frölich, and B. Melly, "Quantile Treatment Effects in the Regression Discontinuity Design," *Journal of Econometrics* 168(2), 2012, 382-395.
G. Imbens and K. Kalyanaraman, "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies* 79, 2012, 933–959.
S. Calonico, M. Cattaneo, and R. Titiunik, "Robust Nonparametric Confidence Intervals for Regression Discontinuity Designs," *Econometrica* 82(6), 2014, 2295-2326.
S. Calonico, M. Cattaneo and R. Titiunik, "Optimal Data-Driven Regression Discontinuity Plots," *Journal of the American Statistical Association* 110(512), December 2015, 1753-1769.

Heaping

D. Almond, J. Doyle, A. Kowalski, and H. Williams, "Estimating the Marginal Returns to Medical Care: Evidence from At-Risk Newborns," *The Quarterly Journal of Economics* 125(2), 2010, 591-634.
A. Barreca, M. Guildi, J. Lindo, and G. Waddell, "Saving Babies? Revisiting the Effect of Very Low Birthweight Classification," *The Quarterly Journal of Economics* 126(4), November 2011, 2117-2123.
D. Almond, *et al*., "Reply to Barreca, *et al*.," Same issue.
Y. Dong, "Regression Discontinuity Applications with Rounding Errors in the Running Variable," *Journal of Applied Econometrics*, 2014.
F. Gerard, M. Rokkanen, and C. Rolfe, "Identification and Inference in Regression Discontinuity Designs with a Manipulated Running Variable," Columbia University Department of Economics, mimeo, December 2015.

V. DIFFERENCES-IN-DIFFERENCES

MM, Chapter 5
MHE, Chapter 5

A. Abadie, A. Diamond, and J. Hainmueller, "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association* 105(490), 2010.
N. Doudchenko and G. Imbens, "Balancing, Regression, Difference-in-differences and Synthetic Control Methods: A Synthesis," NBER Working Paper No. 22791, October 2016.
A. Abadie, "Semiparametric Differences-in-Differences Estimators," *The Review of Economic Studies* 72, 2012, 1-19.
S. Athey and G. Imbens, "Identification and Inference in Nonlinear Difference-in-Difference Models," *Econometrica* 74, 2006, 431-97.
C. de Chaisemartin and X. D'Haultfoeuille, "Fuzzy Differences-in-Differences," mimeo, October 2015.

VI. NON-STANDARD STANDARD ERROR ISSUES

Review of large-sample theory

MHE, Section 3.1.3

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

W.K. Newey and D. McFadden, "Large Sample Estimation an Hypothesis and Testing," Chapter 36 in *The Handbook of Econometrics*, Volume IV, Amsterdam: North-Holland, 1994.

Finite-sample issues

MHE, Chapter 8

A. Chesher and I. Jewitt, "The Bias of a Heteroskedasticity-Consistent Covariance Matrix Estimator," *Econometrica* 55, September 1987, 1217-1222.

Moulton, B., "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics* 32(3), 1986, 385-397.

M. Bertrand, E. Duflo, and S. Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119, February 2004, 249-275.

Hansen, C., "Asymptotic Properties of a Robust Variance estimator for Panel Data When T is Large", *Journal of Econometrics* 141(2), 2007, 597-620.

Hansen, C., "Generalized Least Squares Inference in Panel and Multilevel Models with Serial Correlation and Fixed Effects," *Journal of Econometrics* 140(2), 2007, 670-694.

C. Cameron, J. Gelbach, and D. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *The Review of Economics and Statistics* 90(3), August 2008, 414-427.

C. Cameron and D. Miller, "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources* 50(2), February 2015, 317-372.

G. Imbens and M. Kolesar. "Robust Standard Errors in Small Samples: Some Practical Advice," *Review of Economics and Statistics* 98(4), 2016, 701-712.

A. Abadie, G. Imbens, and F. Zheng, "Inference for Misspecified Models with Fixed Regressors," *Journal of the American Statistical Association* 109(508), December 2014, 1601-1614.

M. Cattaneo, M. Jansson, and W. Newey, "Treatment Effects with Many Covariates and Heteroskedasticity, " University of Michigan Department of Economics, Working Paper, February 2016.

A. Abadie, S. Athey, G Imbens, and J. Wooldridge, "Finite Population Causal Standard Errors," NBER Working Paper No. 20325, November 2014.

Permutation Inference

A. Young, "Improved, Nearly Exact, Statistical Inference with Robust and Clustered Covariance Matrices using Effective Degrees of Freedom Corrections," London School of Economics, mimeo, November 2015.

M. Cattaneo, B. Frandsen and R. Titiumik, "Randomization Inference in the Regression Continuity Design: An Application of Party Advantages in the U.S. Senate," *Journal of Causal Inference* 3(1), 2015, 1-24.