POL-GA 1251

Quantitative Political Analysis II

Spring 2018

Instructor:	Professor Cyrus Samii NYU Politics Department 19 West 4th Street, Office 424 cdsamii@gmail.com
Class time/location:	Tuesdays & Thursdays, 10:00am-12:00pm 19 West 4th Street, Room 217
Office hours:	Mon 3:00pm-5:00pm (use online sign up sheet)
Course website:	`http://cyrussamii.com/?page_id=2580`
Teaching Assistant:	Antonella Bandiera aab639@nyu.edu

Overview

This course provides a current perspective on estimating causal effects in social science research. The approach is rooted in non-parametric and semi-parametric frequentist methods. We will emphasize research design, causal identification, and robust estimation and inference.

Prerequisites and Restrictions

The course has two prerequisites. First, students should have working knowledge of probability theory, matrix algebra, and calculus at the level of POL-GA 1250, "Quant I." Second, students should have some background in writing scripts to implement statistical analyses in either R or Stata.

There is also a restriction with respect to taking the course for credit. The course provides foundational methodological training to Politics PhD students in their first or second year as part of their required sequence of courses. Only Politics PhD students will be allowed to take the course for a grade. (We do not have adequate teaching assistant and other resources to service students from other departments taking this for a grade, unfortunately.) People from other programs may audit or attend informally if space permits.

Texts

The course will draw a lot from the following two textbooks:

1. Angrist, Joshua, and Steffan Jorg Pischke. 2009. *Mostly Harmless Econometrics*. Princeton: Princeton University Press. (Referred to as MHE.)

2. Morgan, Stephen L., and Christopher Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research, Second Edition*. Cambridge, UK: Cambridge University Press. (Referred to as CCI.)

The course will mostly follow MHE, using CCI to provide more intuitive background and illustrations. I will also supplement the textbooks with notes, sections from other textbooks, and journal articles. I have listed "further reading" for each topic, and my lectures will sometimes draw on these. Readings will be available in a public Dropbox (see course website).

The books are each detailed, up to date, and they complement each other well. The books are relatively inexpensive, so students are recommended to acquire both. MHE can be mathematically difficult at times, but you are strongly encouraged to dive in, replicate proofs, and work hard to understand it. CCI provides intuition to keep you sane and grounded.

While these two textbooks are technically sound, they do gloss over details on estimation mechanics. An excellent textbook that provides the proper statistical background for this course is the following:

- Hansen, Bruce E. 2017. *Econometrics*. Typescript, University of Wisconsin.

It is freely available on Hansen's website, and there is also a PDF in our readings dropbox.

For those of you that feel like you need to refresh/improve your mathematical chops, the following is a nice workbook style development of rudiments for regression and statistical inference, with a critical eye:

- Freedman, David A. 2009. *Statistical Models: Theory and Practice*. Cambridge: Cambridge University Press.

My dear favorite graduate-level intro textbook, and the one that I used in ye olden days, is the following:

- Goldberger, Arthur S. 1991. *A Course in Econometrics*. Cambridge, MA: Harvard University Press.

Goldberger develops the theory of "agnostic" linear regression step-by-step from first principles. It's a classic, but out of print (you can get cheap used versions on Amazon though). An updated take is given in the following:

- Aronow, Peter M., and Benjamin Miller. 2015. *Theory of Agnostic Statistics*. Typescript, Yale University.

You can get it here: http://aronow.research.yale.edu/aronowmiller.pdf

The following textbooks provide some deeper theoretical insights on statistics for causal inference and I will sometimes reference them:

- Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press.

- Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference, Second Edition*. Cambridge: Cambridge University Press.

Software

You will have the choice to work in R or Stata, however you wish. Each program has its advantages, and so it is useful to obtain fluency in both. R is great for programming estimators and algorithms "from scratch," programming simulations, and making graphics. Some assignments will ask you to do that. However many of the packaged estimation routines are flawed and it does not handle very large datasets well. Stata's pre-programmed estimation routines are more reliable, and some assignments could be done using them; however Stata is very clumsy (for me) for programming, simulations, or graphics. Use whatever you want to get the tasks done in a manner that is sound and reproducible.

Requirements and policies

Homework

You will receive homework about every week. You will have to submit your completed assignment by the following week; exact deadlines will be made clear on the assignment. You can work with others, but to receive credit, your homework must comply with the following guidelines:

- You must turn in a *hard copy of your own homework* by the stated deadline.
- The assignment that you turn in must *clearly reflect your own thinking*. Sets of verbatim copies of homework will have credit reduced by half.
- Homework assignments may be hand written or typed, but they must be clearly *legible*.
- Estimates obtained from data analysis programs (e.g., Stata or R) must be *formatted properly* into tables resembling journal presentation styles. You should use a table formatting function (e.g., `outreg` or `esttab` in Stata, or `apsrtable` or `stargazer` in R). Use a reasonable (2 or at most 3) number of digits after decimal points, report standard errors along with coefficients, clarify what are the dependent variables in each table, and explain in footnotes to your tables what kinds of estimators or adjustments have been used. *Print outs of "raw" screen output or commented logs will not receive any credit*. However, you may include such output as an appendix so that the grader can troubleshoot.
- Mathematical derivations should include *all key steps with explanations* for important techniques.

Homework will be graded for points as indicated on each assignment and count toward 50% of your grade.

Mid-term exam

An in-class mid-term exam is provisionally scheduled during class on March 9 (to be confirmed). The mid-term serves the purpose of evaluating individual progress, which in turn helps me to understand where to place emphasis for the remainder of the semester. If you are unable to make it to the exam, you must provide notice *at least a week prior* so that we can arrange an alternative time. The mid-term will count toward 15% of your grade.

Final exam

An in-class final exam will be scheduled during the final examination period. The final also serves the purpose of evaluating individual progress, which in turn allows me to provide individualized recommendations on where students should apply effort to strengthen their methodological foundations. If you are unable to make it to the exam, you must provide notice *at least a week prior* so that we can arrange an alternative time. The final will count toward 25% of your grade.

Attendance and participation

Attendance and participation in class discussions is *required* and counts toward 10% of your grade.

Special needs

Students with special needs should come to office hours or schedule an appointment with the instructor to discuss possible accommodation.

Topics

Topics listed below will be covered in around 1-2 class sessions each. Required reading sometimes corresponds directly to material covered in the sessions and sometimes builds up background needed for future sessions. Most of the required reading comes from MHE and CCI, although the topics covered toward the end of the semester will draw on other texts that will be made available.

1 Identification, estimation, and inference

Identification concepts: randomization and the experimental ideal, manipulability, potential outcome model of causal effects, average treatment effects, dose-response functions. *Estimation concepts*: estimands and estimators, bias, consistency, and efficiency. *Inference concepts*: finite and infinite populations, data generating process, implications of randomization and sampling, stochasticity, choosing an "error term," exact distributions, asymptotic distributions and central limit theorem.

Required reading: MHE Ch. 1-2; CCI Ch. 1-3.

Further reading: Angrist and Krueger (1999); Angrist and Pischke (2010); DiNardo and Lee (2011); Freedman (1991); Heckman and Vytlacil (2007); Holland (1986); Imbens and Wooldridge (2009); Imbens and Rubin (2011, Ch. 1-3); Manski (1995, Ch. 1); Pearl (2009, Ch. 1); Rothman (1976); Rosenbaum (1999); Rosenbaum (2002, Ch. 1-3); Rubin (1974); Rubin (1978); Rubin (1986).

2 Regression mechanics

Core decompositions and their relation to regression; Frisch-Waugh-Lovell; conditional independence/exogeneity; omitted variable bias formula; effect heterogeneity and nonlinearity; regression and matching; regression and simple contrasts; influence and leverage; testing restrictions on coefficients.

Required reading: MHE Ch. 3; CCI Ch. 4,6; Aronow and Samii (2016).

Further reading: Angrist and Krueger (1999); Freedman (2008a); Freedman (2008b); Imbens and Rubin (2011, Ch. 7); Lin (2013); Schochet (2010).

3 Notions of bias

Confounding; post-treatment bias; sample selection bias; aggregation/effect heterogeneity biases; misspecification/extrapolation bias and model dependence; measurement error biases.

Required reading: CCI Ch. 8; Bound et al. (2000, pp. 1-39); Heckman (1979); Lalonde (1986); Rosenbaum (1984); Samii (2016).

Further reading: Frangakis and Rubin (2002); Heckman et al. (1998); Hyslop and Imbens (2001); Imai et al. (2008); King and Zeng (2006); Pearl (2009, Ch. 3, 6); Pei et al. (2017).

4 Control and balance via matching and weighting

Identification under conditional exogeneity; alternative matching and weighting algorithms; estimation and inference after matching.

Required reading: CCI Ch. 5,7.

Further reading: Abadie and Imbens (2006); Abadie and Imbens (2008); Abadie and Imbens (2011); Busso et al. (2014); D'Amour et al. (2017); Dehejia and Wahba (2002); De Luna et al. (2011); Ghosh (2018); Hainmueller (2011); Hirano and Imbens (2004); Ho et al. (2007); Iacus et al. (2011); Imai and van Dyk (2004); Imbens (2000); King and Nielsen (2016); Lu et al. (2001); Rosenbaum and Rubin (1983); Sekhon (2009); Todd (2008).

5 Robust statistical inference

Clustering, autocorrelation, and spatial dependence; Moulton's problem; heteroskedasticity and cluster robust standard errors; bootstrapping; estimating the exact randomization variance; permutation tests.

Required reading: MHE Ch. 8.

Further reading: Aronow et al. (2015); Athey et al. (2017); Barrios et al. (2012); Bell and McCaffrey (2002); Bertrand et al. (2004); Cameron et al. (2008); Cameron et al. (2009); Chung and Romano (2013); Conley (1999); Efron and Tibshirani (1993); Freedman (2009, Ch. 8); Hansen (2007); Horowitz (2001); Imbens and Kolesar (2012); Janssen (1997); Moulton (1986); Pustejovsky and Tipton (2017); Romano (1990); Samii and Aronow (2012); Young (2015a); Young (2015b).

6 Instrumental variables

Exclusion restriction; valid first stage; principal strata; local average treatment effect (LATE); weak instrument; sensitivity analysis.

Required reading: MHE Ch. 4; CCI Ch. 9.

Further reading: Abadie (2003); Angrist et al. (2000); Angrist et al. (1996); Baum et al. (2003); Bazzi and Clemens (2013); Bound et al. (1995); Conley et al. (2010); Deaton (2010); Heckman and Urzua (2009); Imbens (2010); Imbens and Rosenbaum (2005); Kolesar et al. (2011); Sovey and Green (2011); Staiger and Stock (1997); Stock et al. (2002); Young (2017).

7 Repeated observations

Adjusting for unobserved heterogeneity via fixed effects and difference-in-differences; synthetic control; event studies.

Required reading: MHE Ch. 5; CCI Ch. 11; Abadie and Gardeazabal (2003); Borusyak and Jaravel (2017).

Further reading: Abadie (2005); Abadie et al. (2010); Athey et al. (2017); Athey and Imbens (2006); Beck and Katz (2001); Bound and Solon (1999); DellaVigna and La Ferrara (2010); Dube et al. (2011); Fisman (2001); Green et al. (2001); Imai and Kim (2012); Imbens and Rubin (2011, Ch. 31, 33); Mora and Reggio (2013).

8 Regression discontinuity (RD)

Forcing variables; sharp and fuzzy RD; conditional average treatment effect (CATE); local linearity, bandwidth, and non-parametric regression; kernel weighting; multiway discontinuities; checks for sorting around cut-points; endogenous forcing variables; measurement error in forcing variables.

Required reading: MHE Ch. 6; CCI Ch. 11.

Further reading: Card et al. (2012); Cattaneo et al. (2017); Caughy and Sekhon (2011); Froelich (2007); Green et al. (2009); Imbens and Kalyanaraman (2009); Imbens and Lemieux (2008); Lee and Card (2008); Lee and Lemieux (2010); McCrary (2008); Papay et al. (2011); Urquiola and Verhoogen (2009).

9 Moderators, mediators, and causal explanation

Moderators and effect heterogeneity; mediators and mechanisms; sequential ignorability.

Required reading: Acharya et al. (2016); CCI Ch. 10; Angrist et al. (2011); Bullock et al. (2010); Imai et al. (2011).

Further reading: Glynn (2011); Heckman et al. (1997); Jo et al. (2011); Ludwig et al. (2011); VanderWeele (2008); VanderWeele (2015).

10 Distributional effects and quantile regression

Quantile treatment effect; minimum absolute deviations; rank invariance.

Required reading: MHE Ch. 7.

Further reading: Bitler et al. (2006); Chernozhukov and Hansen (2005); Heckman et al. (1997); Koenker and Hallock (2000).

11 Multiple endpoints

Index and mean effects; multiple comparisons adjustments.

Required reading: Anderson (2008); Caughy et al. (2015); Romano and Wolf (2007).

Further reading: Casey et al. (2011); Clingingsmith et al. (2009); Farcomeni (2008); Gibson et al. (2011); Kling and Liebman (2004);O'Brien (1984); Shaffer (1995).

12 Missing data and attrition

Bounds; inverse probability weighting; imputation.

Required reading: CCI Ch. 12; Gerber and Green (2012, Ch. 7); Horton and Kleinman (2007); King et al. (2001); Lee (2009); Manski (1995, Ch. 2); Vansteelandt et al. (2010).

Further reading: Aronow et al. (2015); Jones (1996); Puma et al. (2009); Samii (2011).

13 Limited dependent variable effects

Structural versus causal estimands.

Required reading: Angrist (2001); Beck (2015).

Further reading: Davidson and MacKinnon (2004, Ch. 10-11) Fox (2002); Freedman (2006); Greene (2004); Hubbard et al. (2010); Imbens and Rubin (2011, Ch. 8); Liang and Zeger (1986); Van der Laan and Rose (2011, Ch. 7, 11, 16-17); Wooldridge (2002, Ch. 15, 19-20).

14 Machine learning and causal inference

Data-driven estimation; ensemble methods; machine learning for implementing CIA, characterizing effect heterogeneity, and discovering instruments.

Required reading: Belloni et al. (2014); Imai and Strauss (2011); Samii et al. (2015).

Further reading: Athey and Imbens (2015); Green and Kern (2012); Imai and Ratkovic (2012); Kleinberg et al. (2015); Pearl (2009, Ch. 2); Van der Laan and Rose (2011); Wager and Athey (2015).

15 Generalization and external validity

Unconfounded location; external validity; generalizability; prediction error.

Required reading: Imbens (2010).

Further reading: Aronow and Carnegie (2013); Angrist and Fernandez-Val (2010); Angrist and Rokkanen (2014); Bisbee et al. (2015); Dehejia et al. (2017); Gechter (2015); Greenland (1994); Hernan and Vander-Weele (2011); Hotz et al. (2005); Meager (2016); Rubin (1992).

16 Structure and identification

Policy effects versus structural parameters; spill-over and equilibrium effects.

Required reading: Acemoglu (2010); Angrist and Pischke (2010); Heckman (2010).

Further reading: Acemoglu et al. (2015); Aronow and Samii (2015); Banerjee et al. (2017); Brollo and Nannicini (2012); Chassang et al. (2012); Chetty (2009); Rosenzweig and Wolpin (2000); Todd and Wolpin (2006); Wolpin (2013).

References

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics 113*, 231–263.

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies 72*(1), 1–19.

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 105*(490), 493–505.

Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the Basque country. *The American Economic Review 93*(1), 113–132.

Abadie, A. and G. W. Imbens (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica 74*(1), 235–267.

Abadie, A. and G. W. Imbens (2008). On the failure of the bootstrap for matching estimators. *Econometrica 76*(6), 1537–1557.

Abadie, A. and G. W. Imbens (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business and Economic Statistics 29*(1), 1–11.

Acemoglu, D. (2010). Theory, general equilibrium, and the political economy of development. *Journal of Economic Perspectives 24*(3), 17–32.

Acemoglu, D., C. Garcia-Jimeno, and J. A. Robinson (2015). State capacity and economic development: A network approach. *The American Economic Review 105*(8), 2364–2409.

Acharya, A., M. Blackwell, and M. Sen (2016). Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review* (In Press).

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervetion: A reevaluation of the Abecedarian, Perry Preschool and Early Training projects. *Journal of the American Statistical Association 103*(484), 1481–1495.

Angrist, J. D. (2001). Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. *Journal of Business and Economic Statistics 19*(1), 2–16.

Angrist, J. D. and I. Fernandez-Val (2010). ExtrapoLATE-ing: External validity and overidentification in the LATE framework. *NBER Working Paper 16566*.

Angrist, J. D., K. Graddy, and G. W. Imbens (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Review of Economic Studies 67*, 499–527.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*(434), 444–455.

Angrist, J. D. and A. B. Krueger (1999). Empirical strategies in labor economics. In O. C. Ahsenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3. Amsterdam: North Holland.

Angrist, J. D., P. A. Pathak, and C. R. Walters (2011). Explaining charter school effectiveness. *NBER Working Paper Series 17332*.

Angrist, J. D. and J.-S. Pischke (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Angrist, J. D. and M. Rokkanen (2014). Wanna get away? regression discontinuity estimation of exam school effects away from the cutoff. Unpublished Manuscript, MIT and Columbia University.

Aronow, P. M. and A. Carnegie (2013). Beyond LATE: Estimation of the average treatment effect with an instrumental variable. *Political Analysis 21*, 492–506.

Aronow, P. M., A. Coppock, A. S. Gerber, D. P. Green, and H. L. Kern (2015). Combining double sampling and bounds to address non-ignorable missing outcomes in randomized experiments. Manuscript, Yale University, Columbia University, and Florida State University.

Aronow, P. M. and C. Samii (2015). Estimating average causal effects under general interference. Manuscript, Yale University and New York University.

Aronow, P. M. and C. Samii (2016). Does regression produce representative estimates of causal effects? *American Journal of Political Science 60*(1), 250–267.

Aronow, P. M., C. Samii, and V. A. Assenova (2015). Cluster robust variance estimation for dyadic data. *Political Analysis 23*(4), 564–577.

Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2017). Matrix completion methods for causal panel data models. *arXiv 1710.10251 [math.ST]*.

Athey, S. and G. W. Imbens (2006). Identification and inference in nonlinear difference-in-difference models. *Econometrica 74*(2), 431–497.

Athey, S. and G. W. Imbens (2015). Recursive partitioning for heterogeneous causal effects. Unpublished Manuscript, Stanford University.

Banerjee, A. V., R. Chattopadhay, E. Duflo, D. Keniston, and N. Singh (2017). The efficient deployment of police resources: Theory and new evidence from a randomized drunk driving crackdown in india. Typescript, Massachussetts Institute of Technology.

Barrios, T., R. Diamond, G. W. Imbens, and M. Kolesar (2012). Clustering, spatial correlations, and randomization inference. *Journal of the American Statistical Association 107*(498), 578–591.

Baum, C. F., M. E. Schaffer, and S. Stillman (2003). Instrumental variables and GMM: Estimation and testing. *Stata Journal 3*(1), 1–31.

Bazzi, S. and M. Clemens (2013). Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics 5*(2), 152–186.

Beck, N. (2015). Estimating grouped data models with a binary dependent variable and fixed effects: What are the issues? Unpublished Manuscript, New York University.

Beck, N. and J. N. Katz (2001). Throwing out the baby with the bath water: A comment on Green, Kim, and Yoon. *International Organization 55*(2), 487–495.

Bell, R. M. and D. F. McCaffrey (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology 28*(2), 169–181.

Belloni, A., V. Chernozhukov, and C. B. Hansen (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives 28*(2), 29–50.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics 119*(1), 249–275.

Bisbee, J., R. H. Dehejia, C. Pop-Eleches, and C. Samii (2015). Local instruments, global extrapolation: External validity of the labor supply-fertility local average treatment effect. *NBER Working Paper 21663*.

Bitler, M. P., J. B. Gelbach, and H. W. Hoynes (2006). What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review 96*(4), 988–1012.

Borusyak, K. and X. Jaravel (2017). Revisiting event study designs. Typescript, Harvard University and Stanford University.

Bound, J., C. Brown, and N. Mathiowetz (2000). Measurement error in survey data. Technical Report Report No. 00-450, Population Studies Center, University of Michigan.

Bound, J., D. A. Jaeger, and R. M. Baker (1995). Problems with instrumental variables estimation when correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association 90*(430), 443–450.

Bound, J. and G. Solon (1999). Double trouble: on the value of twins based estimation of the returns to schooling. *Economics of Education Review 18*, 169–182.

Brollo, F. and T. Nannicini (2012). Tying your enemy's hands in close races: The politics of federal transfers in Brazil. *American Political Science Review 106*(4), 742–761.

Bullock, J. G., D. P. Green, and S. E. Ha (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology 98*(4), 550–558.

Busso, M., J. DiNardo, and J. McCrary (2014). New evidence on the finite sample properties of propensity score reweighting and matching estimators. *The Review of Economics and Statistics 96*(5), 885–897.

Cameron, A., J. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics 90*(3), 414–427.

Cameron, A., J. Gelbach, and D. L. Miller (2009). Robust inference with multi-way clustering. Manuscript, University of California, Berkeley.

Card, D., D. S. Lee, Z. Pei, and A. Webber (2012). Nonlinear policy rules and the identification of causal effects in a generalized regression kink design. *NBER Working Paper 18564*.

Casey, K., R. Glennerster, and E. Miguel (2011). Reshaping institutions: Evidence on aid implacts using a pre-analysis plan. *NBER Working Paper Series 17012*.

Cattaneo, M. D., N. Idrobo, and R. Titiunik (2017). *A Practical Introduction to Regression Discontinuity Designs*. Cambridge Elements: Quantitative and Computational Methods for Social Science. Cambridge: Cambridge University Press.

Caughy, D., A. Dafoe, and J. Seawright (2015). Global tests of complex hypotheses: A nonparametric framework for testing elaborate theories. Unpublished Manuscript, MIT, Yale University, and Northwestern University.

Caughy, D. and J. S. Sekhon (2011). Elections and the regression discontinuity design: Lessons from close US House races, 1942-2008. *Political Analysis 19*(4), 385–408.

Chassang, S., G. Padro-I-Miquel, and E. Snowberg (2012). Selective trials: A principle-agent approach to randomized controlled experiments. *American Economic Review 102*(4), 1279–1309.

Chernozhukov, V. and C. B. Hansen (2005). An IV model of quantile treatment effects. *Econometrica 73*(1), 245–261.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reducted form methods. *Annual Review of Economics 1*, 451–487.

Chung, E. and J. P. Romano (2013). Exact and asymptotically robust permutation tests. *The Annals of Statistics 41*(2), 484–507.

Clingingsmith, D., A. I. Khwaja, and M. Kremer (2009). Estimating the impact of the hajj: Religion and tolerance in Islam's global gathering. *Quarterly Journal of Economics 124*(3), 1133–1170.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics 92*, 1–45.

Conley, T. G., C. B. Hansen, and P. E. Rossi (2010). Plausibly exogenous. *The Review of Economics and Statistics forthcoming*.

D'Amour, A., P. Ding, A. Feller, L. Lei, and J. Sekhon (2017). Overlap in observational studies with high-dimensional covariates. *arXiv :1711.02582v2 [math.ST]*.

Davidson, R. and J. G. MacKinnon (2004). *Econometric Theory and Methods*. Oxford: Oxford University Press.

De Luna, X., I. Waernbaum, and T. S. Richardson (2011). Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika 98*(4), 861–875.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature 48*, 424–455.

Dehejia, R. H., C. Pop-Eleches, and C. Samii (2017). From local to global: External validity in a natural fertility natural experiment. *NBER Working Paper 21459*.

Dehejia, R. H. and S. Wahba (2002). Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics 84*(1), 151–161.

DellaVigna, S. and E. La Ferrara (2010). Detecting illegal arms trade. *American Economic Journal: Economic Policy 2*, 26–57.

DiNardo, J. and D. S. Lee (2011). Program evaluation and research designs. *Handbook of Labor Economics 4a*(463-536).

Dube, A., E. Kaplan, and S. Naidu (2011). Coups, corporations, and classified information. *The Quarterly Journal of Economics 126*(3), 1375–1409.

Efron, B. and R. J. Tibshirani (1993). *An Introduction to the Bootstrap*. Boca Raton: Chapman and Hall/CRC.

Farcomeni, A. (2008). A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Statistical Methods in Medical Research 17*, 347–388.

Fisman, R. (2001). Estimating the value of political connections. *American Economic Review 91*(4), 1095–1102.

Fox, J. (2002). Cox proportional-hazards regression for survival data. Online appendix to *An R and S-PLUS Companion to Applied Regression*.

Frangakis, C. E. and D. B. Rubin (2002). Principal stratification in causal inference. *Biometrics 58*, 21–29.

Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological Methodology 21*, 291–313.

Freedman, D. A. (2006). On the so-called "Huber sandwich estimator" and "robust standard errors". *The American Statistician 60*(4), 299–302.

Freedman, D. A. (2008a). On regression adjustments in experiments with several treatments. *The Annals of Applied Statistics 2*(1), 176–196.

Freedman, D. A. (2008b). On regression adjustments to experimental data. *Advances in Applied Mathematics 40*(1), 180–193.

Freedman, D. A. (2009). *Statistical Models: Theory and Practice*. Cambridge: Cambridge University Press.

Froelich, M. (2007). Regression discontinuity design with covariates. *IZA Discussion Paper Series 3024*.

Gechter, M. (2015). Generalizing the results from social experiments: Evidence from mexico and india. Unpublished Manuscript, Pennsylvania State University.

Gerber, A. S. and D. P. Green (2012). *Field Experiments: Design and Analysis*. New York, NY: Norton.

Ghosh, D. (2018). A note on strict functional overlap in causal inference problems with high-dimensional covariates. Typescript, University of Denver.

Gibson, J., D. McKenzie, and S. Stillman (2011). The impacts of international migration on remaining household members: Omnibus results from a migration lottery program. *The Review of Economics and Statistics 93*(4), 1297–1318.

Glynn, A. (2011). The product and difference fallacies for indirect effects. *American Journal of Political Science 56*(1), 257–269.

Green, D. P. and H. L. Kern (2012). Modeling heterogenous treatment effects in survey experiments with Bayesian additive regression trees. *Public Opinion Quarterly 76*(3), 491–511.

Green, D. P., S. Y. Kim, and D. H. Yoon (2001). Dirty pool. *International Organization 55*(2), 441–468.

Green, D. P., T. Y. Leong, H. L. Kern, A. S. Gerber, and C. W. Larimer (2009). Testing the accuracy of regression discontinuity analysis using experimental benchmarks. *Political Analysis 17*(4), 400–417.

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics Journal 7*, 98–119.

Greenland, S. (1994). Invited commentary: A critical look at some popular meta-analytic methods. *American Journal of Epidemiology 140*(3), 290–296.

Hainmueller, J. (2011). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis 17*(4), 400–417.

Hansen, C. B. (2007). Asymptotic properties of a robust variance matrix estimator for panel data when t is large. *Journal of Econometrics 141*, 597–620.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature 48*(2), 356–398.

Heckman, J. J., H. Ichimura, J. Smith, and P. Todd (1998). Characterizing selection bias using experimental data. *Econometrica 66*(5), 1017–1098.

Heckman, J. J., J. Smith, and N. Clements (1997). Makings the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *Review of Economic Studies 64*(4), 487–535.

Heckman, J. J. and S. Urzua (2009). Comparing IV with structural models: What simple IV can and cannot identify. *NBER Working Paper Series 14706*.

Heckman, J. J. and E. J. Vytlacil (2007). Econometric evaluation of social programs, parts I and II. *Handbook of Econometrics 6B*, 4779–5143.

Hernan, M. A. and T. J. VanderWeele (2011). Compound treatments and transportability of causal inference. *Epidemiology 22*(3), 368–377.

Hirano, K. and G. W. Imbens (2004). The propensity score with continuous treatments. Manuscript, University of Miami and University of California, Berkeley.

Ho, D. E., K. Imai, G. King, and E. A. Stuart (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis 15*(3), 199–236.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association 81*(396), 945–960.

Horowitz, J. L. (2001). The bootstrap. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, Chapter 52, pp. 3159–3228. Amsterdam: Elsevier.

Horton, N. J. and K. P. Kleinman (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *American Statistician 61*(1), 79–90.

Hotz, V., G. W. Imbens, and J. H. Mortimer (2005). Predicting the efficacy of future training programs using past experiments at other locations. *Journal of Econometrics 125*, 241–270.

Hubbard, A. E., J. Ahern, N. L. Fleischer, M. Van der Laan, S. A. Lippman, N. Jewell, T. Bruckner, and W. A. Satariano (2010). To GEE or not to GEE. *Epidemiology 21*(4), 467–474.

Hyslop, D. R. and G. W. Imbens (2001). Bias from classical and other forms of measurement error. *Journal of Business and Economic Statistics 19*(4), 475–481.

Iacus, S. M., G. King, and G. Porro (2011). Causal inference without balance checking: coarsened exact matching. *Political Analysis 19*(4), 1–24.

Imai, K., L. Keele, D. Tingley, and T. Yamamoto (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review 105*(4), 765–789.

Imai, K. and I. Kim (2012). Understanding and improving linear fixed effects regression models for causal inference. Manuscript, Princeton University.

Imai, K., G. King, and E. A. Stuart (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society, Series A 171*(2), 481–502.

Imai, K. and M. Ratkovic (2012). Estimating treatment effect heterogeneity in randomized program evaluation. *Annals of Applied Statistics* (In press.).

Imai, K. and A. Strauss (2011). Estimation of heterogenous treatment effects from randomized experiments, with application to the optimal planning of the get-out-the-vote campaign. *Political Analysis 19*, 1–19.

Imai, K. and D. A. van Dyk (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association 99*(467), 854–866.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika 87*, 706–710.

Imbens, G. W. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature 48*, 399–423.

Imbens, G. W. and K. Kalyanaraman (2009). Optimal bandwidth choice for the regression discontinuity estimator. *NBER Working Paper Series 14726*.

Imbens, G. W. and M. Kolesar (2012). Robust standard errors in small samples: Some practical advice. *NBER Working Paper Series 18478*.

Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615–635.

Imbens, G. W. and P. R. Rosenbaum (2005). Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. *Journal of the Royal Statistical Society, Series A 168*(1), 109–126.

Imbens, G. W. and D. B. Rubin (2011). Causal inference in statistics and social sciences. Manuscript, Harvard University.

Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

Janssen, A. (1997). Studentized permutation tests for non-i.i.d. hypotheses and the generalized Behrens-Fisher problem. *Statistics and Probability Letters 36*, 9–21.

Jo, B., E. A. Stuart, D. P. MacKinnon, and A. D. Vinokur (2011). The use of propensity scores in mediation analysis. *Multivariate Behavioral Research 46*, 425–452.

Jones, M. P. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association 91*(433), 222–230.

King, G., J. Honaker, A. Joseph, and K. Scheve (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review 95*(1), 49–69.

King, G. and R. Nielsen (2016). Why propensity scores should not be used for matching. Unpublished Manuscript, Harvard University.

King, G. and L. Zeng (2006). The dangers of extreme counterfactuals. *Political Analysis 14*(2), 131–159.

Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer (2015). Prediction policy problems. *American Economic Review: Papers and Proceedings 105*(5), 491–495.

Kling, J. R. and J. B. Liebman (2004). Experimental analysis of neighborhood effects on youth. Manuscript, Princeton University and Harvard University.

Koenker, R. and K. F. Hallock (2000). Quantile regression: An introduction. Manuscript, University of Illinois at Urbana-Champaign.

Kolesar, M., R. Chetty, J. Friedman, E. Glaeser, and G. W. Imbens (2011). Identification and inference with many invalid instruments. Manuscript, Harvard University.

Lalonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review 76*(4), 604–620.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies 76*, 1071–1102.

Lee, D. S. and D. Card (2008). Regression discontinuity inference with specification error. *Journal of Econometrics 142*(2), 655–674.

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*, 281–355.

Liang, K. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Annals of Applied Statistics 7*(1), 295–318.

Lu, B., E. Zanutto, R. Hornik, and P. R. Rosenbaum (2001). Matching with doses in an observational study of a media campaign against drug abuse. *Journal of the American Statistical Association 96*(456), 1245–1253.

Ludwig, J., J. R. Kling, and S. Mullainathan (2011). Mechanism experiments and policy evaluations. *Journal of Economic Perspectives 25*(3), 17–38.

Manski, C. F. (1995). *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard University Press.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

Meager, R. (2016). Aggregating distributional treatment effects: A bayesian hierarchical analysis of the microcredit literature. Typescript, London School of Economics.

Mora, R. and I. Reggio (2013). Treatment effect identification using alternative parallel assumptions. Manuscript, Universidad Carlos III de Madrid.

Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics 32*, 385–397.

O'Brien, P. C. (1984). Procedures for comparuing samples with multiple endpoints. *Biometrics 40*(4), 1079–1087.

Papay, J. P., J. B. Willett, and R. J. Murnane (2011). Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics 161*(2), 203–207.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference, Second Edition*. Cambridge: Cambridge University Press.

Pei, Z., J.-S. Pischke, and H. Schwandt (2017). Poorly measured confounders are more useful on the left than on the right. *NBER Working Papers 23232*.

Puma, M. J., R. B. Olsen, S. H. Bell, and C. Price (2009). What to do when data are missing in group randomized trials. Technical report 2009-0049, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, Washington D.C.

Pustejovsky, J. E. and E. Tipton (2017). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business and Economic Statistics*, doi:10.1080/07350015.2016.1247004.

Romano, J. P. (1990). On the behavior of randomization tests without a group invariance assumption. *Journal of the American Statistical Association 85*(411), 686–692.

Romano, J. P. and M. Wolf (2007). Control of generalized error rates in multiple testing. *The Annals of Statistics 35*(4), 1378–1408.

Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, Series A 147*(5), 656–666.

Rosenbaum, P. R. (1999). Choice as an alternative to control in observational studies. *Statistical Science 14*(3), 259–304.

Rosenbaum, P. R. (2002). *Observational Studies, 2nd Ed.* New York: Springer.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Rosenzweig, M. R. and K. I. Wolpin (2000). Natural "natural experiments" in economics. *Journal of Economic Literature 38*, 827–874.

Rothman, K. J. (1976). Causes. *American Journal of Epidemiology 104*(6), 587–592.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*(5), 688–701.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics 6*(1), 34–58.

Rubin, D. B. (1986). Comment: Which ifs have causal answers. *Journal of the American Statistical Association 81*(396), 961–962.

Rubin, D. B. (1992). Meta-analysis: Literature synthesis or effect-size surface estimation. *Journal of Educational and Behavioral Statistics 17*(4), 363–374.

Samii, C. (2011). Weighting and augmented weighting for causal inference with missing data: new directions. Manuscript, Columbia University and Yale University.

Samii, C. (2016). Causal empiricism in quantitative research. *Journal of Politics 78*(3), 941–955.

Samii, C. and P. M. Aronow (2012). On equivalencies between design-based and regression-based variance estimators for randomized experiments. *Statistics and Probability Letters 82*(2), 365–370.

Samii, C., L. Paler, and S. Daly (2015). Retrospective causal inference with machine learning ensembles: An application to ex-combatant recidivism in colombia. Unpublished Manuscript, New York University, University of Pittsburgh, and Notre Dame University.

Schochet, P. Z. (2010). Is regression adjustment supported by the Neyman model for causal inference? *Journal of Statistical Planning and Inference 140*(1), 246–259.

Sekhon, J. S. (2009). Opiates for the matches: Matching methods for causal inference. *Annual Review of Political Science 12*(1), 487–508.

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology 46*, 561–584.

Sovey, A. J. and D. P. Green (2011). Instrumental variables estimation in political science: A reader's guide. *American Journal of Political Science 55*(1), 188–200.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics 20*(4), 518–529.

Todd, P. (2008). Matching estimators. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics, 2nd Edition*, Hampshire. Palgrave MacMillan. (online access).

Todd, P. E. and K. I. Wolpin (2006). Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *The American Economic Review 96*(5), 1384–1417.

Urquiola, M. and E. Verhoogen (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review 99*(1), 179–215.

Van der Laan, M. and S. Rose (2011). *Targeted Learning: Causal Inference for Obervational and Experimental Data*. New York, NY: Springer.

VanderWeele, T. J. (2008). Simple relations between principal stratification and direct and indirect effects. *Statistics and Probability Letters 78*, 2957–2962.

VanderWeele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford: Oxford University Press.

Vansteelandt, S., J. Carpenter, and M. G. Kenward (2010). Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology 6*(1), 37–48.

Wager, S. and S. Athey (2015). Estimation and inference of heterogenous treatment effects using random forests. Unpublished Manuscript, Stanford University.

Wolpin, K. I. (2013). *The Limits of Inference Without Theory*. Cambridge, MA: MIT Press.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Young, A. (2015a). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. Unpublished Manuscript, London School of Economics.

Young, A. (2015b). Improved, nearly exact, statistical inference with robust and clustered covariance matrices using effective degrees of freedom corrections. Unpublished Manuscript, London School of Economics.

Young, A. (2017). Consistency without inference: Instrumental variables in practical application. Typescript, London School of Economics.