POL-GA 1251

Quantitative Political Analysis II

Spring 2022

Instructor:	Professor Cyrus Samii
	NYU Politics Department
	19 West 4th Street, Office 424
	cdsamii@gmail.com
Class time:	Mondays & Wednesdays, 10:00am-11:50am
Location:	19 W 4th St, Room 217.
Office hours:	See online sign up sheet
Course website:	`https://cyrussamii.com/?page_id=3413`
Teaching Assistant:	Giacomo Lemoli
	gl1759@nyu.edu
Recitation & TA Office Hours:	TBD

Overview

This course provides a current perspective on identifying and estimating causal effects in social science research. We focus on non-parametric identification methods and then non-parametric and semi-parametric estimation and frequentist inference methods. We will emphasize research design 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. With rare exception, 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 textbooks:

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

2. Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press. (Referred to as CIS.)

3. 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.)

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).

Some excellent textbooks that provide the proper statistical background for this course are the following:

- Aronow, Peter M., and Benjamin Miller. 2019. *Foundations of Agnostic Statistics*. Cambridge: Cambridge University Press. (Most direct foundation for this class.)

- Freedman, David A. 2009. *Statistical Models: Theory and Practice*. Cambridge: Cambridge University Press. (Lots of great exercises.)

- Hansen, Bruce E. 2017. *Econometrics*. Typescript, University of Wisconsin. (Free PDF on Hansen's website.)

The following provide additional insights and perspectives on causal identification and effect estimation, and I will sometimes reference them:

- Cunningham, Scott. 2021. *Causal Inference: The Mixtape*. (available from Cunningham's website)

- Hernan Miguel A., and James M. Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. (available from Hernan's website)

- Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference, Second Edition*. Cambridge: Cambridge University Press. (available from Pearl's website)

- Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. 2016. *Causal Inference in Statistics: A Primer*. West Sussex: Wiley. (corrected version available from Pearl's website)

Software

You will have the choice to work in R or Stata. It is useful to obtain fluency in both. R is great for programming estimators and algorithms "from scratch," programming simulations, and making graphics. Stata's pre-programmed estimation routines are often more reliable, and some assignments could be done using them; however Stata is very clumsy (for me) for programming, simulations, or graphics.

I encourage using RMarkdown or Stata Markdown for your assignments. This is a great investment that will pay off in the long run in terms of productivity as well as reproducibility.

- RMarkdown runs most easily through RStudio. Details here: https://rmarkdown.rstudio.com/

- Stata Markdown is a package that runs within Stata. Details here: https://data.princeton.edu/stata/markdown/

Requirements and policies

Homework

You will receive homework every week or two. You will have to submit your completed assignment within a 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 *PDF copy of your own homework* by the stated deadline to both the professor and TA.
- 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 or graphs resembling journal presentation styles. You should use a table formatting function (e.g., `outreg2` 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 or confidence intervals along with coefficients, clarify what are the dependent variables in each table or figure, and explain in footnotes to your tables or figures 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 will take place mid-way through the semester (exact date 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

A take-home 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 work during the exam period, 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.

Sessions and Topics

Required reading sometimes corresponds directly to material covered in the sessions and sometimes builds up background needed for future sessions.

1 Causal Identification

Potential outcomes, causal graphs, and definitions of causal effects.

Required reading: MHE Ch. 1; CIS Ch. 1-2; CCI Ch. 1-3; Greenland and Pearl (2017)

Further reading: Angrist and Pischke (2010); Heckman and Vytlacil (2007); Holland (1986); Freedman (1991); Pearl (2009, Ch. 3); Rosenbaum (1999); Rubin (1974); Rubin (1978); Rubin (1986).

2 Identification, Estimation, and Inference in Randomized Experiments

Estimands and estimators, bias, consistency, and efficiency. Finite and infinite populations, implications of randomization and sampling, exact distributions, and asymptotic distributions.

Required reading: CIS Ch. 6.

Further reading: CIS, rest of Part II; Athey and Imbens (2016); Blair et al. (2019); Freedman (2008); Lin (2013); Samii and Aronow (2012).

3 Agnostic Regression

4 Regression and Causal Effects

Frisch-Waugh-Lovell; omitted variable bias formula; effect heterogeneity and nonlinearity; leverage; multiple regression weights; testing restrictions.

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

Further reading: Angrist and Krueger (1999); Aronow and Miller (2019, Ch. 4); DiNardo and Lee (2011); Imbens and Wooldridge (2009).

5 Conditioning to identify causal effects

6 Conditioning via matching and weighting

Identification under conditional independence; "bad control," collider bias, and post-treatment bias; alternative matching and weighting algorithms; estimation and inference after matching.

Required reading: CCI Ch. 4, 5,7,8; CIS Ch. 12, 17-18.

Further reading: CIS (rest of Parts III and IV); Abadie and Imbens (2006); Abadie and Imbens (2008); Abadie and Imbens (2011); Arbour and Dimmery (2019); Bound et al. (2000, pp. 1-39); Busso et al. (2014); D'Amour et al. (2018); Dehejia and Wahba (2002); De Luna et al. (2011); Elwert (2013); Frangakis and Rubin (2002); Hainmueller (2011); Heckman (1979); Heckman et al. (1998); Hirano and Imbens (2004); Ho et al. (2007); Hyslop and Imbens (2001); Iacus et al. (2011); Imai et al. (2008); Imai and van Dyk (2004); Imbens (2000); King and Nielsen (2016); King and Zeng (2006); Lalonde (1986); Lu et al. (2001); Pearl (2009, Ch. 3, 6); Pearl et al. (2016, Ch 2-3); Pei et al. (2017); Rosenbaum and Rubin (1983); Rosenbaum (1984); Sekhon (2009); Todd (2008).

7 Robust statistical inference I

8 Robust statistical inference II

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); Barrios et al. (2012); 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); Horowitz (2001); Imbens and Kolesar (2016); Janssen (1997); Moulton (1986); Pustejovsky and Tipton (2018); Romano (1990); Young (2015a); Young (2015b).

9 Instrumental variables I

10 Instrumental variables II

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

Required reading: MHE Ch. 4; CCI Ch. 9; CIS Ch 23-24.

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 (2018).

11 Front Door Criterion

Required reading: Bellemare et al. (2017); Pearl et al. (2016, Ch. 3); Winship and Harding (2008).

12 Midterm

13 Repeated observations I

14 Repeated observations II

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

Required reading: MHE Ch. 5; CCI Ch. 11; Cunningham (2018, DID chapter).

Further reading: Abadie and Gardeazabal (2003); Abadie (2005); Abadie et al. (2010); Athey et al. (2018); Athey and Imbens (2006); Borusyak and Jaravel (2017); Bound and Solon (1999); Doudchenko and Imbens (2017); Ferman and Pinto (2019); Goodman-Bacon (2019); Miratrix (2020); Mora and Reggio (2017); Strezhnev (2018); Xu (2017).

15 Regression discontinuity I

16 Regression discontinuity II

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. (2015); Cattaneo et al. (2019); Froelich (2007); Green et al. (2009); Imbens and Kalyanaraman (2009); Imbens and Lemieux (2008); Lee and Lemieux (2010); McCrary (2008); Papay et al. (2011); Urquiola and Verhoogen (2009).

17 Mediators, moderators, and causal explanation I

18 Mediators, moderators, and causal explanation II

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

Required reading: CCI Ch. 10; Angrist et al. (2013); Imai et al. (2011).

Further reading: Acharya et al. (2016); Bullock et al. (2010); Glynn (2011); Heckman et al. (1997); Jo et al. (2011); Ludwig et al. (2011); VanderWeele (2008); VanderWeele (2015).

19 Distributional effects

Quantile treatment effect; minimum absolute deviations; rank invariance.

Required reading: MHE Ch. 7; Chernozhukov et al. (2013).

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

20 Multiple endpoints

Index and mean effects; multiple comparisons adjustments.

Required reading: Anderson (2008).

Further reading: Casey et al. (2011); Caughy et al. (2015); Clingingsmith et al. (2009); Farcomeni (2008); Gibson et al. (2011); Kling and Liebman (2004);O'Brien (1984); Romano and Wolf (2007); Shaffer (1995).

21 Missing data and attrition

Bounds; inverse probability weighting; imputation.

Required reading: CCI Ch. 12; Gerber and Green (2012, Ch. 7).

Further reading: Aronow et al. (2015); Horton and Kleinman (2007); King et al. (2001); Lee (2009); Manski (1995, Ch. 2); Jones (1996); Puma et al. (2009); Vansteelandt et al. (2010).

22 Interference and spillover effects

Required reading: Aronow et al. (2020).

Further reading: Aronow and Samii (2017); Hudgens and Halloran (2008).

23 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).

24 Machine learning and causal inference

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

Required reading: Athey and Imbens (2017).

Further reading: Athey and Imbens (2015); Belloni et al. (2014); Chernozhukov et al. (2017); Imai and Ratkovic (2012); Imai and Strauss (2011); Kleinberg et al. (2015); Samii et al. (2015). Van der Laan and Rose (2011); Wager and Athey (2015).

25 Generalization and external validity

Unconfounded location; external validity; generalizability; target validity.

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. (2019); Gechter (2015); Greenland (1994); Hernan and VanderWeele (2011); Hotz et al. (2005); Rubin (1992); Westreich et al. (2019).

26 Structure and identification

Policy effects versus structural parameters; equilibrium effects; interpretation, interpolation, and extrapolation.

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

Further reading: Acemoglu et al. (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).

27 Take-home final exam

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 (2013). Explaining charter school effectiveness. *American Economic Journal: Applied Economics 5*(4), 1–27.

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.

Arbour, D. and D. Dimmery (2019). Permutation weighting. *arXiv 1901.01230v1 [stat.ME]*.

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., D. Eckles, C. Samii, and S. Zonszein (2020). Spillover effects in experimental data. In J. Druckman and D. P. Green (Eds.), *Handbook of Experimental Political Science*, Cambridge. Cambridge University Press.

Aronow, P. M. and B. T. Miller (2019). *Foundations of Agnostic Statistics*. Cambridge: Cambridge University Press.

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. and C. Samii (2017). Estimating average causal effects under general interference, with application to a social network experiment. *Annals of Applied Statistics 11*(4), 1912–1947.

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. W. Imbens, and K. Khosravi (2018). Matrix completion methods for causal panel data models. *arXiv arXiv:1710.10251v2 [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.

Athey, S. and G. W. Imbens (2016). The econometrics of randomized experiments. Tyepscript, Stanford University. Available at https://www.povertyactionlab.org/handbook-field-experiments.

Athey, S. and G. W. Imbens (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives 31*(2), 3–32.

Banerjee, A. V., R. Chattopadhyay, 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. Manuscript, Massachussetts Institute of Technology, Indian Institute of Management, Yale University, and Rajasthan Police.

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.

Bellemare, M. F., J. R. Bloem, and N. Wexler (2017). The paper of how: Estimating treatment effects using the front-door criterion. Typescript, University of Minnesota.

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.

Blair, G., J. Cooper, A. Coppock, and M. Humphreys (2019). Declaring and diagnosing research designs. Available at: https://declaredesign.org/declare.pdf.

Borusyak, K. and X. Jaravel (2017). Revising event study designs, with an application to the estimation of the marginal propensity to consume. Manuscript, 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. Lee, Z. Pei, and A. Wever (2015). Inference on causal effects in a generalized regression kink design. *Econometrica 83*(6), 2453–2483.

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 (2019). *A Practical Introduction to Regression Discontinuity Designs: Foundations and Extensions*. 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.

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

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, and W. Newey (2017). Double/debiased/neyman machine learning of treatment effects. Typescript, Massachussetts Institute of Technology.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica 81*(6), 2205–2268.

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*.

Cunningham, S. (2018). *Causal Inference: The Mixtape*.

D'Amour, A., P. Ding, A. Feller, L. Lei, and J. Sekhon (2018). Overlap in observational studies with high-dimensional covariates. *arXiv 1711.02582v3 [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 (2019). From local to global: External validity in a natural fertility natural experiment. *Journal of Business and Economic Statistics* (forthcoming).

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.

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

Doudchenko, N. and G. W. Imbens (2017). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. *arXiv 1610.07748v2 [stat.AP]*.

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

Elwert, F. (2013). Graphical causal models. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Research*, pp. 245–273. Dordrecht: Springer Netherlands.

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.

Ferman, B. and C. Pinto (2019). Inference in differences-in-differences with few treated groups and heteroskedasticity. *Review of Economics and Statistics* (forthcoming).

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

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.

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.

Goodman-Bacon, A. (2019). Difference-in-differences with variation in treatment timing. Typescript, Vanderbilt University.

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.

Greenland, S. and J. Pearl (2017). Causal diagrams. *Wiley StatsRef: Statistics Reference Online*, 1–10.

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.

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. *Handbook of Econometrics 5*(52), 3159–3228.

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.

Hudgens, M. G. and M. E. Halloran (2008). Toward causal inference with interference. *Journal of the American Statistical Association 103*(482), 832–842.

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., 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 (2016). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics 98*(4), 701–712.

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 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.

Miratrix, L. (2020). Using simulation to analyze interrupted time series designs. *arXiv 2002.05746v1*([stat.ME]), 2205–2268.

Mora, R. and I. Reggio (2017). Alternative diff-in-diffs estimators with several pretreatment periods. *Econometric Reviews DOI: 10.1080/07474938.2017.1348683*.

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.

Pearl, J., M. Glymour, and N. P. Jewell (2016). *Causal Inference in Statistics: A Primer*. West Sussex: Wiley.

Pei, Z., J.-S. Pischke, and H. Schwandt (2017). Poorly measured confounders are more useful on the left than on the right. *NBER Working Paper 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 (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business and Economic Statistics 36*(4), 672–683.

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. 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.

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. (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.

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.

Strezhnev, A. (2018). Semiparametric weighting estimators for multi-period difference-in-differences designs. Typescript, New York University.

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.

Westreich, D., J. K. Edwards, C. R. Lesko, S. R. Cole, and E. A. Stuart (2019). Target validity and the hierarchy of study designs. *American Journal of Epidemiology* (forthcoming).

Winship, C. and D. J. Harding (2008). A mechanism-based approach to the identification of age–period–cohort models. *Sociological Methods & Research 36*(3), 362–401.

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.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis 25*, 57–76.

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. (2018). Consistency without inference: Instrumental variables in practical application. Manuscript, London School of Economics.