The Statistics of Causal Inference in the Social Sciences
Political Science C236A
Statistics C239A

Professor Jasjeet Singh Sekhon

Lecture: Mondays 12-3
791 Barrows

Professor Jasjeet S. Sekhon
sekhon@berkeley.edu
HTTP://sekhon.berkeley.edu
Office: Barrows Hall 750C
Office Hours: To be announced

Sören R. Künzel, TA
srk@berkeley.edu
Section: Fridays 10-12, Barrows 791
Office Hours: To be announced

Description

Approaches to causal inference using the potential outcomes framework. Covers randomized experiments with and without noncompliance, observational studies with and without ignorable treatment assignment, instrumental variables, regression discontinuity, sensitivity analysis, and permutation inference. Applications are drawn from a variety of fields including political science, economics, sociology, public health, and medicine.

Prerequisites

A multivariate regression course with linear algebra, upper division probability and statistics, and familiarity with high-level programming languages. For example, PS231a *and* PS231b or Statistics 133, 134, and 135 are minimum requirements. Experience with applications is helpful. The programming language used in the course is **R**. If you need to review material, please consult David Freedman's excellent *Statistical Models: Theory and Practice*.

Evaluation

Final grades will be based on a series of homework assignments (30% of final grade), a midterm (30%), a final exam (30%), and class and section participation (10%). The final exam requirement can be waived if students continue to the second part of the course in the spring and write a research paper. The final will be more mathematically demanding than the midterm.

Weekly readings and homework assignments are the norm. It is highly recommended that students form study groups in order to complete the homework assignments. Although it is rec-

ommended that people work together in order to complete the assignments, students must hand in their own individual answers. Photocopies and other reproductions of someone else's answers are not acceptable. Students should hand in the answers to the problem sets, and all computer code written to find those answers.

During exams, students are not allowed to communicate or cooperate with anyone in any way about exam. Any questions should be asked directly to me and the GSI. To repeat: for exams, one is not allowed to use study groups, online help forms, the writing center, or any other form of help aside from those that are explicitly allowed on the exam instructions. If in doubt, ask.

Incompletes: incompletes will not be granted unless there are exceptional circumstances.

Date for the midterm: November 13–20.

We will be using Piazza to help maintain a healthy discussion and to provide help.

Course Software and Books

The programming language for this course is the **R**. It is available for download from: `http://www.r-project.org/`. **R** is open source software (released under the GNU public license).

The books listed below are available at various online bookstores.

Required book:

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

Additional books that you may want. These books are not required, but most purchase them because we assume that you have access to them when needed.

- Freedman, David A. 2010. *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*. David Collier, Jasjeet S. Sekhon, and Philip B. Stark, Editors. Cambridge University Press. ISBN-10 0521123909. Note: this is a collection of articles that you can also find elsewhere.

The following books may be of additional help:

- Angrist, Joshua D. and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

- Morgan, Stephen L. and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press. ISBN-10: 0521671930.

- Rosenbaum, Paul R. 2002. *Observational Studies*. Springer-Verlag. 2nd edition. ISBN 0387989676.

- Rosenbaum, Paul R. 2009. *Design of Observational Studies*. Springer-Verlag. ISBN-10 1441912126.

  This book is less technical than Rosenbaum's *Observational Studies*. The book is also available online via Cal's contract with Springer-Verlag: `http://www.springerlink.com/content/978-1-4419-1212-1/#section=627451&page=1`

Course outline

1. Causality

   The potential outcomes framework for causal inference.

   - Imbens and Rubin: Chapter 1 and 2.
   - Holland (1986) "Statistics and Causal Inference"
   - Little and Rubin (2000) "Causal Effects in Clinical and Epidemiological Studies via Potential Outcomes"

   Extra reading:

   - Sekhon (2004a): "Quality Meets Quantity: Case Studies, Conditional Probability and Counterfactuals"
   - Winship and Morgan (1999) "The Estimation of Causal Effects from Observational Data"

2. Statistical Modeling: Foundations and Limitations

   - Imbens and Rubin Chapters 3–4.
   - Freedman Chapter 1 (Freedman 1995): "Some Issues in the Foundations of Statistics: Probability and Model Validation."
   - Freedman Chapter 2 (Freedman 2003): "Statistical Assumptions as Empirical Commitments."
   - Freedman Chapter 3 (Freedman 1991): "Statistical Models and Shoe Leather."

   For extra readings see the rest of the Freedman volume, especially:

   - Freedman Chapter 20 (Freedman 2008a): "On Types of Scientific Inquiry: The Role of Qualitative Reasoning."
   - Freedman Chapter 14 (Freedman and Humphreys 1996): "The Grand Leap" (of graphical models).
   - Freedman Chapter 15 (Freedman 2004): "On Specifying Graphical Models for Causation, and the Identification Problem."

3. Randomized Experiments and Controlling Bias in Observational Studies

   Properties of experiments, basic implementations, and illustrations of observational studies based on approximate experimental design.

   - Imbens and Rubin Chapter 6.
   - Neyman (1923/1990): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." *Statistical Science* 5, 465–472.
   - Rubin (1990) "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies," *Statistical Science* 5, 472-480.

   Extra readings:

- Rubin (2006) Chapters 1 and 2:
  "William G. Cochran's Contributions to the Design, Analysis and Evaluation of Observational Studies"
  Cochran and Rubin (1973): "Controlling Bias in Observational Studies: A Review"
- Rosenbaum (2002b) Chapter 2
- Przeworski (In Press) "Is the Science of Comparative Politics Possible?"
- Cox (1958): *Planning of Experiments*. Chapters 1 and 2.
- Cochran (1965): "The Planning of Observational Studies of Human Populations"
- Cochran (1983): Chapters 1 and 7

4. Randomization Inference

   Fisherian and permutation Inference, and the Lady Tasting Tea

   - Imbens and Rubin Chapter 5
   - Fisher (1935, ch 1–2): *Design of Experiments*. `http://tinyurl.com/c9tj2hy`
   - Rosenbaum (2002b, ch 2): *Observational Studies*
   - Rosenbaum (2002a): "Covariance adjustment in randomized experiments and observational studies." *Statistical Science* 17 286–327 (with discussion).

   Extra reading:

   - Freedman Chapter 8: "What is the Chance of an Earthquake?"
   - Bowers and Panagopoulos (2011): "Fisher's Randomization Mode of Statistical Inference, Then and Now."
   - Attributable effects: Rosenbaum (2002b, 188–194).
   - Pitman (1937b): "Significance Tests Which May be Applied to Samples From any Populations"
   - Pitman (1937a): "Significance Tests Which May be Applied to Samples from any Populations. II. The Correlation Coefficient Test"
   - Pitman (1938): "Significance Tests which can be Applied to Samples from any Populations. III. The Analysis of Variance Test"

5. Univariate Matching Methods for Controlling Bias in Observational Studies

   Experimental and observational studies where assignment to treatment is done on observables. Stratification and matching.

   - Rubin (2006) Chapters 3 to 5:
     "Matching to Remove Bias in Observational Studies" Rubin (1973a)
     "The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies" Rubin (1973b)
     "Assignment to a Treatment Group on the Basis of a Covariate"
   - Rosenbaum (2002b) Chapter 3.1–3.3

6. The Propensity Score

   Logistic regression and the fundamentals of propensity score matching

- Handout on general linear models
- Rubin (2006) Chapters 10, 11 and 14 all with Paul R. Rosenbaum:
  "The Central Role of the Propensity Score in Observational Studies" Rosenbaum and Rubin (1983)
  "Assessing Sensitivity to an Unobserved binary Covariate in an Observational Study with Binary Outcome"
  "The Bias Due to Incomplete Matching"
- Sekhon (2004b): The Varying Role of Voter Information Across Democratic Societies
- Morgan and Harding (2006): "Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice"

Also see Rosenbaum and Rubin (1984); Rubin and Thomas (2000).

7. Regression Discontinuity Design

   - Thistlethwaite and Campbell (1960): "Regression-Discontinuity Analysis: An alternative to the ex post facto experiment"
   - Lee (2008): "Randomized Experiments from Non-random Selection in U.S. House Elections"
   - Caughey and Sekhon (2011): "Elections and the Regression-Discontinuity Design: Lessons from Close U.S. House Races, 1942–2008"
   - Hahn, Todd, and van der Klaauw (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design"

   Extra reading:

   - Dunning (2008): "Improving Causal Inference: Strengths and Limitations of Natural Experiments." *Political Science Quarterly* 61(2):282–293 2008.

8. Multivariate Matching

   Mahalanobis distance, and Equal Percent Bias Reduction

   - Rubin (2006) Chapters 8 and 9:
     "Bias Reduction Using Mahalanobis-Metric Matching" Rubin (1980)
     "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies" Rubin (1979)

9. Genetic Matching

   Automatic balance optimization, evaluating balance and the LaLonde controversy

   - Diamond and Sekhon (2013): "Genetic Matching for Estimating Causal Effects"
   - Sekhon and Grieve (2012): "A Matching Method for Improving Covariate Balance in Cost-Effectiveness Analyses."
   - LaLonde (1986) [JSTOR]
   - Dehejia and Wahba (1999) [JSTOR]
   - Smith and Todd (2001)

10. Natural Experiments

- Sekhon and Titiunik (2012): "When Natural Experiments Are Neither Natural Nor Experiments"

11. Random Forests

- Breiman (2001): "Random forests"

12. Conditional Average Treatment Effect (CATE)

- Athey and Imbens (2015): "Machine Learning for Estimating Heterogeneous Causal Effects"
- Athey and Imbens (2016): "Recursive partitioning for heterogeneous causal effects"
- Hill (2011): "Bayesian nonparametric modeling for causal inference"
- Green and Kern (2012): "Modeling heterogeneous treatment effects in survey experiments with Bayesian additive regression trees"
- Wager and Athey (2015): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests"
- Künzel, Sekhon, Bickel, and Yu (2017): "Meta-learners for Estimating Heterogeneous Treatment Effects using Machine Learning"

13. Instrumental Variables (IV)

- Angrist and Krueger (2001): "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments"
- Angrist, Imbens, and Rubin (1996) "Identification of Causal Effects Using Instrumental Variables"
- Heckman (1997) "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations"

Application, and use of randomization inference to correct an issue:

- Imbens and Rosenbaum (2005): "Robust, Accurate Confidence Intervals with a Weak Instrument: Quarter of Birth and Education," *Journal of the Royal Statistical Society*, Series A, vol 168(1), 109–126.
- Angrist and Krueger (1991): "Does compulsory school attendance affect earnings?" *Quarterly Journal of Economics* 1991; 106: 979–1019.
- Bound, Jaeger, and Baker (1995): "Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Regressors is Weak," *JASA* 90, June 1995, 443–450.

14. (Regression) Adjustment to Experimental Data

- Lin et al. (2013): "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique." `http://tinyurl.com/9378kmk`
- Miratrix, Sekhon, and Yu (2013b):Adjusting Treatment Effect Estimates by Post-Stratification in Randomized Experiments

Extra readings:

- Freedman (2008c): "On regression adjustments to experimental data."
- Freedman (2008b): "Randomization does not justify logistic regression"
- Freedman (2008d): "On regression adjustments in experiments with several treatments"

15. Blocking in Experiments

    - Lock Morgan and Rubin (2012): "Rerandomization to improve covariate balance in experiments"
    - Imbens (2011): "Experimental Design for Unit and Cluster Randomized Trials"
    - Higgins and Sekhon (2013): "Improving Experiments by Optimal Blocking: Minimizing the Maximum Within-block Distance"
    - Miratrix, Sekhon, and Yu (2013a): "Adjusting Treatment Effect Estimates by Post-Stratification in Randomized Experiments"

16. External Validity: Better LATE Than Nothing

    - Deaton (2009): "Instruments of Development: Randomization in the tropics, and the search for the elusive keys to economic development"
    - Imbens (2009): "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)"
    - Hartman, Grieve, Ramsahai, and Sekhon (2015): "From SATE to PATT: Combining Experimental with Observational Studies to Estimate Population Treatment Effects"

17. Sensitivity Analysis for Hidden Bias and other Helpful Suggestions

    - Rosenbaum (2002b): chapters on sensitivity analysis
    - Rosenbaum (2005): why more data (even big data) isn't necessarily better. Also links Mill to Fisher.

18. Application: Fixing Broken Experiments and a Controversy

    - Gerber, Alan S. and Donald P. Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." American Political Science Review 94(3): 653 663.
    - Imai, Kosuke. "Do Get-Out-The-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments." American Political Science Review
    - Green and Gerber Reply
    - Bowers, Jake and Ben Hansen. 2005. "Attributing Effects to A Cluster Randomized Get-Out-The-Vote Campaign."

19. Synthetic Cohorts
    When good matches cannot be found: create a new unit

    - Abadie, Diamond, and Hainmueller (2010): Synthetic control methods for comparative case studies: Estimating the effect of California]s tobacco control program
    - Abadie and Gardeazabal (2003): "The Economic Costs of Conflict: a Case-Control Study for the Basque Country"

20. Full and Optimal Matching

- Rosenbaum (1991, 1989)
- Hansen (2004)

21. Matching Examples Using Observational Data
Political Science

- Gordon and Huber: "The Effect of Electoral Competitiveness on Incumbent Behavior"
- Gilligan and Sergenti: "Evaluating UN Peacekeeping with Matching to Improve Causal Inference"
- Lenz and Ladd: "Exploiting a Rare Shift in Communication Flows: Media Effects in the 1997 British Election"
- Simmons and Hopkins (2005): "The Constraining Power of International Treaties: Theory and Methods"

Economics

- Galiani, Gertler, and Schargrodsky (2005): "Water for Life: The Impact of the Privatization of Water Services on Child Mortality"
- Imbens, Rubin, and Sacerdote (2001): "Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence from a Survey of Lottery Players"
- Angrist (1998): "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants."

Other

- Christakis and Iwashyna (2003): "The Health Impact of Health Care on Families: A matched cohort study of hospice use by decedents and mortality outcomes in surviving, widowed spouses"
- Rubin (2001): "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation"

22. Application: Voting Irregularities

- Wand, Shotts, Sekhon, Walter R. Mebane, Herron, and Brady (2001): The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida
- Mebane and Sekhon (2004): Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data
- Herron and Wand (2007): Assessing Partisan Bias in Voting Technology: The Case of the 2004 New Hampshire Recount
- Sekhon (2004c): The 2004 Florida Optical Voting Machine Controversy: A Causal Analysis Using Matching

23. Pre-Test Problems

- Diaconis (1985): "Theories of Data Analysis: From Magical Thinking Through Classical Statistics"
- Freedman (1983): "A Note on Screening Regression Equations"

References

A. Abadie and J. Gardeazabal. The economic costs of conflict: a case-control study for the basque country. *American Economic Review*, 92(1), 2003.

A. Abadie, A. Diamond, and J. Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(490), 2010.

J. Angrist and A. Krueger. Does compulsory school attendance affect earnings? *Quarterly Journal of Economics*, 106:979–1019, 1991.

J. D. Angrist. Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica*, 66(2):249–288, March 1998.

J. D. Angrist and A. B. Krueger. Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, 15(4):69–85, 2001.

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

S. Athey and G. Imbens. Machine Learning for Estimating Heterogeneous Causal Effects. *arXiv preprint arXiv:1504.01132*, 2015.

S. Athey and G. W. Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27):7353–60, 2016. ISSN 1091-6490. doi: 10.1073/pnas.1510489113. URL `http://www.ncbi.nlm.nih.gov/pubmed/27382149{%}5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4941430`.

J. Bound, D. Jaeger, and R. Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous regressors is weak. *Journal of the American Statistical Association*, 90:443–450, 1995.

J. Bowers and C. Panagopoulos. Fisher's randomization mode of statistical inference, then and now. Working Paper, 2011.

L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

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

N. A. Christakis and T. I. Iwashyna. The health impact of health care on families: A matched cohort study of hospice use by decedents and mortality outcomes in surviving, widowed spouses. *Social Science & Medicine*, 57(3):465–475, 2003.

W. G. Cochran. The planning of observational studies of human populations (with discussion). *Journal of the Royal Statistical Society, Series A*, 128:234–255, 1965.

W. G. Cochran. *Planning and analysis of observational studies*. John Wiley and Sons, New York, 1983. Edited posthumously by L. E. Moses and F. Mosteller.

W. G. Cochran and D. B. Rubin. Controlling bias in observational studies: A review. *Sankhya, Series A*, 35(4):417–446, 1973.

D. R. Cox. *Planning of Experiments*. Wiley, New York, 1958.

A. Deaton. Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development. NBER Working Paper 14690, 2009.

R. Dehejia and S. Wahba. Causal effects in non-experimental studies: Re-evaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448):1053–1062, 1999.

P. Diaconis. Theories of data analysis: From magical thinking through classical statistics. In D. Hoaglin, F. Mosteller, and J. Tukey, editors, *Exploring Data Tables, Trends and Shapes*, pages 1–36. Wiley, New York, 1985.

A. Diamond and J. S. Sekhon. Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, 95(3):932–945, 2013.

T. Dunning. Improving causal inference: Strengths and limitations of natural experiments. *Political Science Quarterly*, 61(2):282–293, 2008.

R. A. Fisher. *Design of Experiments*. Hafner, New York, 1935.

D. A. Freedman. A note on screening regression equations. *The American Statistician*, 10:454–461, 1983.

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

D. A. Freedman. Some issues in the foundation of statistics. *Foundations of Science*, 1:19–39, 1995.

D. A. Freedman. Statistical assumptions as empirical commitments. In T. G. Blomberg and S. Cohen, editors, *Punishment and Social Control: Essays in Honor of Sheldon Messinger*, pages 235–254. Aldine de Gruyter, 2nd edition, 2003.

D. A. Freedman. On specifying graphical models for causation, and the identification problem. *Evaluation Review*, 26(4):267–93, 2004.

D. A. Freedman. On types of scientific inquiry: The role of qualitative reasoning. In J. M. Box-Steffensmeier, H. E. Brady, and D. Collier, editors, *Oxford Handbook of Political Methodology*, pages 300–318. Oxford University Press, 2008a.

D. A. Freedman. Randomization does not justify logistic regression. *Statistical Science*, 23(2): 237–249, 2008b.

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

D. A. Freedman. On regression adjustments to experimental data. *Advances in Applied Mathematics*, 40(2):180–193, 2008d.

D. A. Freedman and P. Humphreys. The grand leap. *British Journal for the Philosophy of Science*, 47:113–123, 1996.

S. Galiani, P. Gertler, and E. Schargrodsky. Water for life: The impact of the privatization of water services on child mortality. *Journal of Political Economy*, 113(1):83–120, 2005.

M. J. Gilligan and E. J. Sergenti. Evaluating un peacekeeping with matching to improve causal inference. *Quarterly Journal of Political Science*, 3(2):89–122, 2008.

S. Gordon and G. Huber. The effect of electoral competitiveness on incumbent behavior. *Quarterly Journal of Political Science*, 2(2):107–138, 2007.

D. P. Green and H. L. Kern. Modeling heterogeneous treatment effects in survey experiments with Bayesian additive regression trees. *Public Opinion Quarterly*, 76(3):nfs036, 2012. ISSN 0033362X. doi: 10.1093/poq/nfs036.

J. Hahn, P. Todd, and W. van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69:201–209, 2001.

B. B. Hansen. Full matching in an observational study of coaching for the sat. *Journal of the American Statistical Association*, 99(467):609–618, 2004.

E. Hartman, R. Grieve, R. Ramsahai, and J. S. Sekhon. From sample average treatment effect to population average treatment effect on the treated: combining experimental with observational studies to estimate population treatment effects. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(3):757–778, 2015.

J. Heckman. Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *The Journal of Human Resources*, 32(3):441–462, 1997.

M. C. Herron and J. Wand. Assessing partisan bias in voting technology: The case of the 2004 new hampshire recount. *Electoral Studies*, 26(2):247–261, 2007.

M. Higgins and J. S. Sekhon. Improving experiments by optimal blocking: Minimizing the maximum within-block distance. Working Paper, 2013.

J. L. Hill. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011. ISSN 1061-8600. doi: 10.1198/jcgs.2010.08162.

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

G. Imbens. Better late than nothing: Some comments on deaton (2009) and heckman and urzua (2009). NBER Working Paper 14896, April 2009.

G. W. Imbens. Experimental design for unit and cluster randomized trials. Working Paper, 2011.

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

G. W. Imbens, D. B. Rubin, and B. I. Sacerdote. Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players. *American Economic Review*, 91(4):778–794, Sep. 2001.

S. Künzel, J. Sekhon, P. Bickel, and B. Yu. Meta-learners for estimating heterogeneous treatment effects using machine learning. *arXiv preprint arXiv:1706.03461*, 2017.

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

D. S. Lee. Randomized experiments from non-random selection in u.s. house elections. *Journal of Econometrics*, 142(2):675–697, 2008.

G. S. Lenz and J. M. Ladd. Exploiting a rare shift in communication flows: Media effects in the 1997 british election. *American Journal of Political Science*, 53(2):394–410, 2009.

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

R. J. A. Little and D. B. Rubin. Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annual Review of Public Health*, 21:121–145, 2000.

K. Lock Morgan and D. B. Rubin. Rerandomization to improve covariate balance in experiments. *Annals of Statistics*, 40(2):1263–1282, 2012.

W. R. J. Mebane and J. S. Sekhon. Robust estimation and outlier detection for overdispersed multinomial models of count data. *American Journal of Political Science*, 48(2):391–410, 2004.

L. W. Miratrix, J. S. Sekhon, and B. Yu. Adjusting treatment effect estimates by post-stratification in randomized experiments. *Journal of the Royal Statistical Society, Series B*, 75(2):369–396, 2013a.

L. W. Miratrix, J. S. Sekhon, and B. Yu. Adjusting treatment effect estimates by post-stratification in randomized experiments. *Journal of the Royal Statistical Society, Series B*, 75(2):369–396, 2013b.

S. L. Morgan and D. J. Harding. Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological Methods & Research*, 35(1):3–60, 2006.

J. Neyman. On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 5(4):465–472, 1923/1990. Trans. Dorota M. Dabrowska and Terence P. Speed.

E. J. G. Pitman. osignificance tests which may be applied to samples from any populations. ii. the correlation coefficient test. *Supplement to the Journal of the Royal Statistical Society*, 4(2):pp. 225–232, 1937a. ISSN 14666162.

E. J. G. Pitman. Significance tests which may be applied to samples from any populations. *Supplement to the Journal of the Royal Statistical Society*, 4(1):pp. 119–130, 1937b. ISSN 14666162.

E. J. G. Pitman. Significance tests which can be applied to samples from any populations. iii. the analysis of variance test. *Biometrika*, 29(3–4):322–335, 1938.

A. Przeworski. Is the science of comparative politics possible. In C. Boix and S. C. Stokes, editors, *Oxford Handbook in Comparative Politics*. Oxford University Press, New York, In Press.

P. R. Rosenbaum. Optimal matching for observational studies. *Journal of the American Statistical Association*, 84(408):1024–1032, 1989.

P. R. Rosenbaum. A characterization of optimal designs for observational studies. *Journal of the Royal Statistical Society, Series B*, 53(3):597–610, 1991.

P. R. Rosenbaum. Covariance adjustment in randomized experiments and observatio nal studies (with discussion). *Statistical Science*, 17:286–327, 2002a.

P. R. Rosenbaum. *Observational Studies*. Springer-Verlag, New York, 2nd edition, 2002b.

P. R. Rosenbaum. Heterogeneity and causality: Unit heterogeneity and design sensitivity in observational studies. *The American Statistician*, 59:147–152, 2005.

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

P. R. Rosenbaum and D. B. Rubin. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387):516–524, 1984.

D. B. Rubin. Matching to remove bias in observational studies. *Biometrics*, 29:159–184, 1973a.

D. B. Rubin. The use of matching and regression adjustment to remove bias in observational studies. *Biometrics*, 29:185–203, 1973b.

D. B. Rubin. Using multivariate sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74:318–328, 1979.

D. B. Rubin. Bias reduction using mahalanobis-metric matching. *Biometrics*, 36(2):293–298, 1980.

D. B. Rubin. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science*, 5(4):472–480, 1990.

D. B. Rubin. Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology*, 2(1):169–188, 2001.

D. B. Rubin. *Matched Sampling for Causal Effects*. Cambridge University Press, New York, 2006.

D. B. Rubin and N. Thomas. Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association*, 95:573–585, 2000.

J. S. Sekhon. Quality meets quantity: Case studies, conditional probability and counterfactuals. *Perspectives on Politics*, 2(2):281–293, 2004a.

J. S. Sekhon. The varying role of voter information across democratic societies. Working Paper, 2004b.

J. S. Sekhon. The 2004 florida optical voting machine controversy: A causal analysis using matching. Working Paper, 2004c.

J. S. Sekhon and R. Grieve. A non-parametric matching method for bias adjustment with applications to economic evaluations. *Health Economics*, 21(6):695–714, 2012.

J. S. Sekhon and R. Titiunik. When natural experiments are neither natural nor experiments. *American Political Science Review*, 106(1):35–57, 2012.

B. A. Simmons and D. J. Hopkins. The constraining power of international treaties: Theory and methods. *American Political Science Review*, 99(4):623–631, 2005.

J. A. Smith and P. E. Todd. Reconciling conflicting evidence on the performance of propensity score matching methods. *AEA Papers and Proceedings*, 91(2):112–118, 2001.

D. L. Thistlethwaite and D. T. Campbell. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6):309–317, 1960.

S. Wager and S. Athey. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *arXiv preprint arXiv:1510.04342*, pages 1–43, 2015. URL `http://arxiv.org/abs/1510.04342`.

J. N. Wand, K. W. Shotts, J. S. Sekhon, J. Walter R. Mebane, M. C. Herron, and H. E. Brady. The butterfly did it: The aberrant vote for buchanan in palm beach county, florida. *American Political Science Review*, 95(4):793–810, 2001.

C. Winship and S. Morgan. The estimation of causal effects from observational data. *Annual Review of Sociology*, 25:659–707, 1999.