Political Science 590
Potential Outcomes Inference (aka "the matching class")

Jake Bowers
jwbowers@uiuc.edu
http://www.umich.edu/~jwbowers/ps590.html

Fall 2007

General Information

Class meets the following Wednesdays 1:30–3:30 CST: Sept 26, Oct 3, Oct 10, Oct 24, Oct 31, Nov 7 and Nov 14.

Jake's office hours are 4-6 on Wednesdays. Alert him if you'd like to talk "face-to-face" via iChat (preferred video chatting method because it is higher quality) or Skype.

This class is an introduction to matching and propensity scores as well as to how the potential outcomes framework (invented by Neyman, developed by Rubin and his collaborators) applies to linear models such as regression.

Goals and Expectations

I assume some previous engagement with probability, linear models, statistical computing, and most importantly research design.

By the end of the course, I expect that students will have had some practice creating matched samples from datasets of interest to them and will be able to reason about the circumstances in which covariance adjustment (i.e. "controlling for" using regression) is, or is not, reasonable.

I'll calculate your grade for the course this way: TBA

Paper Each of you will write a paper which you will turn in three weeks after the end of class. This paper is an opportunity for you to use the ideas from this class to help you make some causal inference in your own substantive work. You may also write a methodological paper. This is a relatively new set of methods and there is ample room for interesting simulation studies let alone theoretical work.

We will have several assignments oriented around your paper to (1) give you practice with the techniques under discussion and (2) push your paper along so that the quality of papers turned in at the end is high.

Exercises We may have a few, short, computer exercises. The idea of these exercises will be to make more concrete and immediately useful the ideas and concepts and algorithms we learn about in class.

Involvement Quality seminar participation does not mean "talking a lot." It includes turning in assignments on time; thinking and caring about the material and expressing your thoughts respectfully and succinctly in class.

Books

Required: Morgan, S. L. and Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research (Analytical Methods for Social Research)*. Cambridge University Press See http://www.wjh.harvard.edu/~cwinship/cfa.html for some links and background reading)

Rubin, D. B. (2006). *Matched sampling for causal effects*. Cambridge University Press, Cambridge

Recommended: Becker, H. S. (1986). *Writing for Social Scientists: How to Start and Finish Your Thesis, Book, or Article*. University of Chicago Press

Becker, H. S. (1998). *Tricks of the trade : how to think about your research while you're doing it*. University of Chicago Press, Chicago, Ill

Berk, R. (2004). *Regression Analysis: A Constructive Critique*. Sage

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press (particularly chapters 9,10 and 23 see `http://www.stat.columbia.edu/~gelman/arm/`).

Rosenbaum, P. (2002a). *Observational Studies*. Springer-Verlag, second edition (see `http://www-stat.wharton.upenn.edu/~rosenbap/index.html` for lots of papers and presentations).

Computing

In this class, I will be using the R statistical language. You are free to use other languages although I suspect you will find it easier to learn R unless you are already a code ninja in some other language and are ready to re-implement some rather complicated matching techniques. You may want to get a book to help you learn R. I used Venables, W. and Ripley, B. (2002). *Modern Applied Statistics with S-PLUS*. Springer, New York, 4th edition, but there are many more available out there now.

Schedule

Note: This schedule is preliminary and subject to change. If you miss a class make sure you contact me or one of your colleagues to find out about changes in the lesson plans or assignments.

1 Conceptual Preliminaries (Sept 26)

What is a potential outcome? How does this reconceptualization clarify or obfuscate the process of making an argument for causal inference? Or for theory testing?

- Morgan and Winship, Chap 1,2,8
- Little, R. and Rubin, D. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annual Review of Public Health*, 21:121–145
- Brady, H. and Seawright, J. (2004). Framing social inquiry: From models of causation to statistically based causal inference. Working Paper
- Rosenbaum, P. R. (1999). Choice as an alternative to control in observational studies (with discussion). *Statistical Science*, 14(3):259–304 (skipping the discussion section)

Recommended

- Holland, P. W. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association*, 81:945–970
- Freedman, D. A. (1991). Statistical models and shoe leather (Disc: P315-358). *Sociological Methodology*, 21:291–313
- Rubin, Chap 1 & 2 (i.e.
  - Rubin, D. B. (1984). William G. Cochran's contributions to the design, analysis, and evaluation of observational studies. In *W. G. Cochran's Impact on Statistics*, pages 37–69. Wiley (New York)
  - Cochran, W. G. and Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyā, Series A, Indian Journal of Statistics*, 35:417–446)

Assignments **DUE IN THIS FIRST CLASS:** Come to this first class with no more than two pages on which you've answered the following questions: What research question will motivate your class paper? Why should anyone care about this question (i.e. what are the implications of answering your research question one way or another on some theory)? What is the causally important variable (i.e. the "treatment")? How do "doses" of the treatment get to the units of analytical interest? How are the potential outcomes defined in your case? Thinking about Rosenbaum (1999), name at least two "control" groups for your "treatment" group. How might the biases which occur when you compare your treatment group(s) to your control group(s) differ from one another?

DUE NEXT CLASS: Come to the next class with no more than two pages detailing your research design (i.e. the dataset you plan to use and the particular observations you plan to choose to observe within it (e.g. the experiment you plan to run or the discontinuity in the world you plan to exploit)) and your measurement strategy following Figure 1 of Adcock, R. and Collier, D. (2001). Measurement validity: A shared standard for qualitative and measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review*, 95(3):529–546.

2 Random Assignment as the "Reasoned Basis" for Causal and Probabilistic Inference and *The Lady Tasting Tea* (Oct 3)

Back to basics. Statements about causality and probability can made most confidently when we know a lot about the random process by which either outcomes OR treatments occur. For social scientists, this tends to mean treatment assignment versus the "data generating process" because we tend to have weak theories about outcomes and, in the lab anyway, we have actual physical control over treatment. Matching and propensity scores are built on the metaphor of the randomized experiment: in fact, one could say that matching is an attempt to make manifest a latent random assignment. Thus, before we learn about matching, we must first learn about random assignment.

- Fisher, R. (1935). *The design of experiments. 1935.* Oliver and Boyd, Edinburgh, Chapter 2
- Rosenbaum, P. (2002a). *Observational Studies.* Springer-Verlag, second edition, Chapter 2
- Splawa-Neyman, J., Dabrowska, D., and Speed, T. (1990). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465–472
- Rubin, D. B. (1990). [on the application of probability theory to agricultural experiments. essay on principles. section 9.] comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science*, 5(4):472–480 (on super-population based inference)
- Rubin, D. (1978). Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, 6(1):34–58 (on Bayesian based inference)
- Rubin, Chap 24 (i.e. Rubin, D. B. (1991). Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, 47:1213–1234) (comparing different modes of inference)

Recommended Some more development of Bayesian approaches given random assignment.

- Rubin, D. (2004). Teaching Statistical Inference for Causal Effects in Experiments and Observational Studies. *Journal of Educational and Behavioral Statistics*, 29(3):343
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis.* Chapman and Hall/CRC, 2nd ed. edition, Chapter 7.1–7.7

Assignment **DUE IN THIS CLASS** Turn in mini-research designs.

3 Causal Inference using Linear Regression (Oct 10)

Linear regression is the dominant metaphor for how we assess our theories with respect to observation. Under what circumstances can we think of regression coefficients as telling us something about comparisons among potential outcomes (with or without control of treatment assignment)?

- Berk, R. (2004). *Regression Analysis: A Constructive Critique*. Sage, Chap 5
- Morgan and Winship, Chap 3, 5, 6, 7
- Gelman and Hill, Chap 9,10, 23
- Rosenbaum, P. R. (2002c). Covariance adjustment in randomized experiments and observational studies. *Statistical Science*, 17(3):286–327

Assignment **DUE NEXT CLASS** Pretend that the values for your causal variable of interest were randomly assigned by you. Use randomization inference either implementing the covariance adjustment strategy of Rosenbaum (2002c) or some version the sum statistic from (Rosenbaum, 2002a, Chapter 2) on either (1) your variables of interest from the dataset that you proposed to use in your research design sketch OR (2) on data that you simulate/fake but which share relevant characteristics with the actual/real data you plan to gather. Compare your results to the "standard" linear model as applied to your question. Explain why the two sets of results differ (or not). What are the different (and similar) assumptions you had to make in order to make your results reflect meaningfully on a causal inference? (It is ok if these assumptions are untenable in your particular situation right now.) Show one graph (or table) that will help you diagnose whether a linear model additive model (including potential confounds as covariates additively) is a good idea or not using the frameworks for evaluating the usefulness of regression from Gelman and Hill, Berk and Morgan and Winship.[Shoot for 6 pages][1]

4 Matching and Propensity Scores (Oct 24)

How can we make a strong case for ignorability when we don't even know *how* the many possible confounds are doing their dastardly confounding? (Notice we still assume here that we can list and observe the potential confounds.)

- Rosenbaum, P. (2002a). *Observational Studies*. Springer-Verlag, second edition, Chap 3.1–3.3
- Morgan and Winship, Chap 4, 5.3–5.5
- Morgan, S. and Harding, D. (2006). Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice. *Sociological Methods & Research*, 35(1):3
- Caliendo, M. and Kopeinig, S. (forthcoming). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*
- Rubin, Chap 10,17, 26
  - Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55
  - Rubin, D. B. and Thomas, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics*, 52:249–264
  - Rubin, D. (1997). Estimating Causal Effects from Large Data Sets Using Propensity Scores. *Annals of Internal Medicine*, 127(8):757–763

[1] I am willing to allow people to substitute "predictive Bayesian inference" for "randomization inference" here. You'll want to do a stripped down version of what occurs in Imbens and Rubin (1997) (see the week on IV estimation). Imai has several papers implementing this kind of inference: see in particular this application to a vote turnout experiment `http://imai.princeton.edu/research/manifesto.html`

Recommended

- Rubin, Chap 12, 13,
  - Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387):516–524
  - Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38
- Rosenbaum, P. R. (2001b). Effects attributable to treatment: Inference in experiments and observational studies with a discrete pivot. *Biometrika*, 88(1):219–231
- Abadie, A. and Imbens, G. (2006). Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, 74(1):235–267
- Abadie, A. and Imbens, G. (2004). On the Failure of the Bootstrap for Matching Estimators. *NBER, Unpublished Manuscript*
- Imbens, G. (2004). Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review. *Review of Economics and Statistics*, 86(1):4–29
- Rosenbaum, P. R. (2002b). Attributing effects to treatment in matched observational Studies. *Journal of the American Statistical Association*, 97:183–192
- Joffe, M. and Rosenbaum, P. (1999). Invited commentary: propensity scores. *Am J Epidemiol*, 150(4):327–333
- Rosenbaum, P. (2001a). Observational studies: Overview. *International Encyclopedia of the Social and Behavioral Sciences*
- ?Sekhon, Jas. ?forthcoming? [as yet untitled review of matching]. *Annual Review.*

Assignment **DUE IN THIS CLASS:** Data analysis “as if” randomized, exercises.

DUE NEXT CLASS: Write down the formula that defines the causal effect that you care to estimate using some version of potential outcomes notation. Explain what you’ve written in plain language — for an audience who is statistics savvy but not deeply into the jargon of the potential outcomes framework. Generate two different proposals for a propensity score model for your assignment process. Justify them substantively (i.e. given what you know about your assignment process, how is your propensity model sensible) and also statistically (i.e. is this model a good fit to your assignment process?) Estimate both proposals. Do pair matching using the propensity score. Try it both with and without replacement. Provide balance metrics for before and after each matching. Choose one of the four matchings as best in your particular case and justify your choice. Estimate a treatment effect and associated confidence interval using this matching. Interpret your effect in a substantively meaningful way. Now, estimate effects using the other matchings. Do they differ? If they differ, can you see the effects of observed confounders? [shoot for 8 pages]

5 Optimal, Full and Genetic Matching (Oct 31)

Making better use of your data: avoiding throwing away observations, establishing higher standards for balance (and arguing about these things).

- Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, 99(467):609–618
- Diamond, A. and Sekhon, J. S. (2005). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. Unpublished manuscript

- Imai, K. and van Dyk, D. A. (2004). Causal inference with generalized treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, 99(467):854–866
- Sekhon, J. and Titiunik, R. (2007). Exploiting tom delay: A new method for estimating incumbency advantage and the effect of candidate ethnicity on turnout. Unpublished manuscript
- Hansen, B. B. and Bowers, J. (2007b). Covariate balance in simple, stratified and clustered comparative studies. Unpublished manuscript

Recommended

- Rosenbaum, P. (2002a). *Observational Studies.* Springer-Verlag, second edition, Chap 10. See this for more details about optimal matching.
- Hansen, B. B. and Klopfer, S. O. (2005). Optimal full matching and related designs via network flows. Technical Report 416, Statistics Department, University of Michigan. See this for more details about full, optimal matching.
- Mebane, W. and Sekhon, J. S. (2007). Genetic optimization using derivatives: The rgenoud package for r. Unpublished manuscript. See this for more details about the genetic algorithm behind genetic matching
- Sekhon, J. S. (2007). Multivariate and propensity score matching software with automated balance optimization: The matching package for r. Unpublished manuscript. See this for more details and examples of how to use the genetic matching software
- Ho, D., Imai, K., King, G., and Stuart, E. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15:199–236. See this for a perspective on using matching to enhance the stability of covariance adjustment without using matched sets per se.
- Imai, K., King, G., and Stuart, E. (2008). Misunderstandings among experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society, Series A*, 171(2):1–22. See this, especially pages 17 to 22, for a criticism of balance testing and cites to similar criticisms from the clinical trials literature.
- Sekhon, J. S. (2006). Alternative balance metrics for bias reduction in matching methods for causal inference. Unpublished manuscript. See this, for an assessment of different balance testing metrics as applied in genetic matching.

Assignment **DUE IN THIS CLASS** Your propensity, pair, matching.

DUE IN THE NEXT CLASS: Repeat the previous assignment, but now, pull out all the stops to simultaneously diminish bias and increase efficiency. Use both genetic matching and full/optimal matching. Compare the effects, confidence intervals, and balance tests you get from using each algorithm.

6 Instrumental Variables and Encouragement Designs (Nov 7)

What happens when random assignment does not or cannot directly control the "dose" of treatment received? Does the potential outcomes framework have anything to say about this?

- Morgan and Winship, Chap 6 and 7
- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables (Disc: p456-472). *Journal of the American Statistical Association*, 91:444–455baker bio
- Rosenbaum, P. R. (1996). Identification of causal effects using instrumental variables: Comment. *Journal of the American Statistical Association*, 91(434):465–468

- Greevy, R., Silber, J. H., Cnaan, A., and Rosenbaum, P. R. (2004). Randomization inference with imperfect compliance in the ace-inhibitor after anthracycline randomized trial. *Journal of the American Statistical Association*, 99(465):7–15
- Imbens, G. W. and Rosenbaum, P. R. (2005). Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1):109+
- Barnard, J., Frangakis, C., Hill, J., and Rubin, D. (2003). Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City. *Journal of the American Statistical Association*, 98(462):299–324

Applications The Vote 98 Controversy:

- Gerber, A. S. and Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review*, 94:653–663
- Imai, K. (2005). Do get-out-the-vote calls reduce turnout? the importance of statistical methods for field experiments. *American Political Science Review*, 99(2)
- Gerber, A. S. and Green, D. P. (2005). Correction go Gerber and Green (2000), replication of disputed findings, and reply to imai (2005). *American Political Science Review*, 99(2):301–313
- Hansen, B. B. and Bowers, J. (2007a). Attributing effects to a cluster randomized get-out-the-vote campaign. Unpublished manuscript

Recommended

- Imbens, G. and Rubin, D. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*, 25(1):305–327
- Frangakis, C. and Rubin, D. (2002). Principal Stratification in Causal Inference. *Biometrics*, 58(1):21–29
- Angrist, J. and Krueger, A. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *The Journal of Economic Perspectives*, 15(4):69–85
- Bound, J., Jaeger, D., and Baker, R. (1995). Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak. *Journal of the American Statistical Association*, 90(430)
- Baker, S. (2001). Rethinking historical controls. *Biostatistics*, 2(4):383–396

Assignment **DUE IN THIS CLASS:** Your full and genetic matching exercises.

DUE IN THE NEXT CLASS: Do you have an instrumental variable in your research design? If so, explain and justify with data analysis how it meets the criteria for an instrument. If you do not have an IV variable in your research design, explain what one would look like if you could see one. Generate a fake IV (or perhaps use one of the variables existing in your dataset as an IV even thought you cannot fully justify it). Explain and justify with data analysis how your fake/temporary IV meets the criteria for an instrument (or does not). Complete an IV estimation using one of the matchings that you've already produced. Explain why we'd do this rather than produce a new matching especially for an IV estimation? Justify your choice of mode of inference here: if you follow the predictive Bayesian approach or the randomization approach, or even the 2SLS approach, justify why you made that choice and compare your results to one of the other approaches.

7 Sensitivity Analysis with Matching and Linear Models (Nov 14)

When there is some doubt about the control over the assignment of values of the causally important variable (i.e. treatment assignment) (as is commonplace in observational studies), those who make causal claims face special burdens of persuasion. Sensitivity analysis refers to the set of techniques that attempt to assess the extent to which unobserved confounds may be biasing a particular causal effect.

Canonical Example and Extensions

- Cornfield, J., Haenszel, W., Hammond, E., Lilienfeld, A., Shimkin, M., and Wynder, E. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute*
- Rosenbaum, P. R. (2005b). Sensitivity analysis in observational studies. *Encyclopedia of Statistics in Behavioral Science*, 4:1809–1814
- Rosenbaum, P. (2002a). *Observational Studies*. Springer-Verlag, second edition, Chapter 4.1 and 4.2
- Imbens, G. (2003). Sensitivity to Exogeneity Assumptions in Program Evaluation. *The American Economic Review*, 93(2):126–132

Multiple Control Groups

- Rosenbaum, P. (1987). The Role of a Second Control Group in an Observational Study. *Statistical Science*, 2(3):292–306
- Rosenbaum, P. (1988). Sensitivity analysis for matching with multiple controls. *Biometrika*, 75:577– 581
- Rosenbaum, P. (2002a). *Observational Studies*. Springer-Verlag, second edition, Chapter 8

Recommended

- Rosenbaum, P. (2002a). *Observational Studies*. Springer-Verlag, second edition, Chapter 4 (for lots of technical development of sensitivity analysis in many different cases)
- Rosenbaum, P. R. (2005a). Heterogeneity and causality: Unit heterogeneity and design sensitivity in observational studies. *The American Statistician*, 59(2):147–152 (for a good discussion of heterogeneity and sensitivity analysis).
- Clarke, K. (2006). Practical sensitivity analysis. Unpublished manuscript
- Berk (2004), Chapter 11 (for a very nice overview of problems with regression with some mention of sensitivity analysis, propensity scores, instrumental variables, etc..)
- Aakvik, A. (2001). Bounding a Matching Estimator: The Case of a Norwegian Training Program. *Oxford Bulletin of Economics and Statistics*, 63(1):115–143 (on Bounding Effects)

Assignment **DUE IN THIS CLASS:** Your IV exercise.

DUE IN 3 WEEKS: Your research paper — a paper which must include some kind of formal sensitivity analysis of your estimated causal effect.