Political Science 590
Matching for Adjustment and Causal Inference

Jake Bowers
jwbowers@illinois.edu
http://jakebowers.org

Fall 2011

General Information

Where/When Class meets the following Wednesdays 1:30–3:30 Central Time: Sept 28, Oct 5, Oct 12, Oct 19, Nov 2, Nov 9 and Nov 16.

Office Hours Office hours are 3–4:30 on Tuesdays by appointment. I am very happy to meet with you. If you know in advance that you want to talk during office hours, please email me to reserve a 20 minute slot. I have found that making appointments for office hours leaves fewer students sitting in the hall waiting to talk with me. Please make an appointment if you want to come to office hours or if you would like to meet at times other than the office hours. I am happy to video chat, too. I do not have a camera on my desktop at the office, but I can arrange to have a camera if we schedule a video chat in advance.

This class is an introduction to statistical adjustment using matching and propensity scores as well as to how the potential outcomes framework (attributed to Neyman and since developed by Rubin and his collaborators) applies to linear models. We will also spend some time on statistical inference after such adjustment. And we will grapple with some of the questions that are current research topics in this area such as when and how one can claim to have adjusted "enough" and how to engage with concerns about unobserved confounds.

Goals and Expectations

This course aims to help you think about statistical adjustment using stratification and matching as compared to statistical adjustment using the linear model directly (adjustment by residualization).

The course ought to give you opportunities to practice producing matched designs for your data and to ask questions that puzzle you as you do this work.

The point of the course is to position you to do the future learning that is at the core of your work as an academic analyzing data.

I also hope that this course will help you continue to develop the acumen as a reader, writer, programmer and social scientist essential for your future daily life as a social science researcher.

This course does not delve deeply into the theories of causal inference, statistical inference, or algorithmns at the heart of these new attempts at statistical adjustment. Rather, through practice using tools, I hope that your curiosity is awakened and you begin to read more broadly and understand more deeply on your own.

Expectations First and foremost, I assume you are eager to learn. Eagerness, curiosity and excitement will impel your energetic engagement with the class throughout the term. If you are bored, not curious, or unhappy about the class you should come and talk with me immediately. Graduate school is not the place to waste your time on courses that are not important to you.

Second, I assume you are ready to work. Learning requires work. As much as possible I will link practice directly to application: I am not interested in evaluating your as compared to your peers. Making work about learning rather than ranking, however, will make our work that much more difficult and time consuming. You will make errors. These errors are opportunities for you to learn — some of your learning will be about how to help yourself use the computer to get the work done and some will be about statistics.

Third, I assume some previous engagement with high school mathematics, probability and statistical computing in the R statistical computing environment. If you have not used R, you are welcome to take the class, but I encourage you to get a little experience with R before the first class session.

Rules There aren't many rules for the course, but they're all important. First, read the assigned readings before you come to class. Second, ask questions when you don't understand things; chances are you're not alone. Fourth, don't miss class.

All papers written in this class will assume familiarity with the principles of good writing in Becker (1986).

All final written work will be turned in as pdf files. I will not accept Microsoft, Apple, OpenOffice, or any other proprietary format. Work turned in using those formats will not be looked at and subsequent pdf files will be considered late work.

Late Work I do not like evaluation for the sake of evaluation. Evaluation should provide opportunities for learning. Thus, if you'd prefer to spend more time using the paper assignment in this class to learn more, I am happy for you to take that time.

Incompletes Incompletes are fine in theory but terrible at the University of Illinois in practice. I urge you to avoid an incomplete in this class. If you must take an incomplete, you must give me *at least* 2 months from the time of turning in an incomplete before you can expect a grade from me. This means that if your fellowship, immigration status, or job depends on erasing an incomplete in this class, you should not leave this incomplete until the last minute.

Participation We will be doing hands-on work nearly every class meeting. I will lecture very little and instead will pose problems of statistical theory, research design, and data, which will require us to confront and apply the reading that prepared us for the day's work. I anticipate that you'll work in small groups at your sites, asking me and/or the group questions via chat or via microphone as you proceed. I will break away to draw on the board (or perhaps a twiddla whiteboard) or demonstrate on my own computer now and then if everyone is running into the same problem.

Quality seminar participation does not mean "talking a lot." It includes thinking and caring about the material and expressing your thoughts respectfully and succinctly in class. It means asking questions that show that you have done the reading and thought about the reading. It also may mean organizing yourselves to have extra meetings during the week to go over the commented classwork and/or readings and/or your papers.

Paper Each of you will write a paper which you will turn in three weeks after the end of class. The goal of the paper is to allow you to practice writing a technical report comparing a linear model (covariance adjustment) to matching (post-stratification adjustment) as methods of adjustment. You can think of it as an appendix to a paper that justifies your choice of analytic strategy.

The paper for this course will take a specific form that I roughly outline here.

1. Find a linear model in which a comparative claim is assessed: for example, the claim may be that two groups differ on values of their outcome "controlling for" covariates. The ideal linear model would be one that you've used in other work — a (g)lm that is a centerpiece of a dissertation chapter, a conference paper, a seminar paper.

2. Ensure that you can reproduce the linear model output table.

3. Re-specify and execute the analysis in the matching paradigm — pretending as if you had never run that linear model before. What are the potential outcomes? Would you choose the same covariates? What would statistical inference mean? What justifies the statistical inference strategy you choose? Do you need to execute a sensitivity analysis? Will you use a linear model after matching for further covariance adjustment?

4. Explain why the two analyses differ (or not). How do the two analyses reflect differently on your substantive and theoretical concerns that motivated the covariance adjustment in the first place?

If you would rather write a different paper, you'll need to get explicit consent from me. I encourage you to contact me to discuss the shape of your paper as the class goes along.

Grading I'll calculate your grade for the course this way: 100% the paper.

Because moments of evaluation are also moments of learning in this class (and not moments of competition or ranking), I do not curve. If you all perform at 100%, then I will give you all As.

I have abolished the participation part of the grade for this class because it is clear to me that there is no easy way to evaluate quality participation via our current video conferencing systems. However, if you have not worked through the classwork in class (and in the outside meetings that I encourage you to have at each site), it will be difficult for you to write the paper for the class. Thus,

Books

Required: Rosenbaum, P. R. (2010). *Design of Observational Studies*. Springer (pdf free to download from campus ip addresses or via campus library springerlink subscriptions: `http://www.springerlink.com/content/978-1-4419-1212-1/contents/`)

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/`).

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)

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

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

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.

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.

Important: Come to class with your laptops (if you have them) with R installed.

Sept 28 – Adjustment by Simple Stratification

Read: Rosenbaum Chap 1, 3, 7

Due in Class: Come to class prepared to tell us about a relationship between an outcome and an explanatory variable that you think needs adjustment and a proposal for a variable (or set of variables) that are the putative confounders.

Do: Engage with a short introduction to the course. Iron out technical problems. Try out the ideas in the reading about stratification for adjustment. Compare to covariance adjustment.

Recommended: (Gelman and Hill, 2007, Chap 9 and 10) especially on (a) adjusting for post-treatment variables and (b) discussion of issues of interpolation, extrapolation, and covariance adjustment depending on functional form.

Oct 5 – Adjustment by Multivariate Matching

Read: Rosenbaum Chap 8, 9; Hansen (2004)

Do: Today involves lots of engagement with the details of the craft. Make propensity scores, distance matrices, calipers, deal with missing data on covariates. Make matched sets. Assess balance.

Recommended: Rosenbaum Chap 13

Hansen (2011) for an example walk-through of a matched analysis.

Ho et al. (2007) [esp. their discussion of model sensitivity, for example their Fig 2]

Three different ideas about balance testing: (1) Imai et al. (2008); (2) Sekhon (2007a)[1]; (3) Hansen and Bowers (2008) Hansen (2008) or for a less mathematical version of the same argument (Bowers, 2011, §3)[2].

Oct 12 – Statistical Inference for Matched/Post-stratified Designs: Fisher's Randomization Inference

Statistical inference for post-stratified observational study designs is based on statistical inference for pre-stratified randomized study designs. Today we get the basics down: hypothesis tests of Fisher's sharp null hypothesis of no effects in a pair-randomized study and in a matched observational version of the same study.

Read: Berk (2004, Chap 4) on general requirements for statistical inference; Rosenbaum chap 2; Imbens and Rubin (2009, Chap 17)

Do: Explore Fisher's framework for statistical inference about causal effects in randomized studies; review what "statistical inference" means.

Recommended: Bowers and Panagopoulos (2011)[3] provides (what I hope is) a nice introduction to the Fisher approach for political scientists.

Morgan and Harding (2006) in particular their "Matching as Stratification" section.

Imbens and Rubin (2009, Chap 5–8) on varieties of statistical inference for randomized experiments.

---

[1] http://sekhon.berkeley.edu/papers/SekhonBalanceMetrics.pdf

[2] I'll try to scan this if you want to read it. Let me know. Otherwise, you can download a pre-copy-edited version from http://jakebowers.org

[3] http://jakebowers.org/PAPERS/BowPan-Fisher.pdf

Oct 19 – Statistical Inference for Matched/Post-stratified Designs: Neyman's Randomization Inference and Model Based Approaches

Neyman proposed another solution to the fundamental problem of causal inference. His idea, which relies on averages, fits very naturally within the least-squares linear modelling framework. The question of the application of statistical inference based on the linear regression model to matched designs is a current research question. See the recommended reading for some of the different perspectives.

Read: Imbens and Rubin (2009, Chap 6,7,17)

Do: Produce tests, intervals, and estimates for average treatment effects in randomized and post-stratified studies.

Recommended: Freedman (2008b,a, 2007, 2006) Suggesting that even the large sample statistical inference from using linear regression in randomized experiments is biased. Also arguing that the Huber-White standard errors are not a good idea.

Rosenbaum (2002a) and Bowers and Panagopoulos (2011) showing how covariance adjustment is compatible with Fisher's randomization inference (and thus can be unproblematic after matching).

Schochet (2009); Green (2009) Suggesting that in large samples these biases worried about by Freedman ought not to worry us. Lin (2011) provides some useful proofs supporting the idea that linear regression with "robust" standard errors provides a useful large-sample way to do statistical inference about average treatment effects.

(Imbens and Rubin, 2009, Chap 6–8) Suggesting, similarly to Green and Schochet, that regression is fine for statistical inference in experiments (and further suggesting the use of the Huber-White robust standard errors).

Abadie and Imbens (2004) suggesting that the bootstrap is not a good approach with matched designs.

Extra Reading For advanced reading on the latest in statistical theory for statistical inference for "matching estimators" (which include but are not restricted to post-stratified studies) see:

Hansen (2009) for theory using randomization-inference.

Abadie and Imbens (2009) for a large-sample, Normal theory approach.

Nov 2 – Sensitivity Analysis

Read: Rosenbaum chap 14; Hosman et al. (2010)

Do: Execute a sensitivity analysis of a statistical inference (before and after matching).

Recommended: Imbens (2003)

Nov 9 – Advances in Multivariate Matching: Beyond Binary Treatment

Read: Rosenbaum 11; Lu et al. (2011)

Do: Non-bipartite matching.

Recommended: Imai and van Dyk (2004)

Nov 16 – Advances in Multivariate Matching: Matching with Longitudinal Data

Read: Rosenbaum 12

Stuff that was painfully left out but which is important

Here are just a few extra citations to launch self-study of aspects of matching which we did not cover in our class.

The class elected to focus on matching for longitudinal problems for the last class. We thus are unable to cover other approaches to matching that have been developed by political methodogists such as Genetic Matching Diamond and Sekhon (2006); Sekhon (2007b) or Coarsened Exact Matching Iacus et al. (2009, 2011) let alone many other applied and theoretical topics in matching such as the work establishing causal interpretation of the propensity score (cited in the Rosenbaum textbook), or the alternative approaches to causal inference based on weighting by functions of the propensity score such as those arising from work by Jamie Robins Glynn and Quinn (2010), let alone alternative conceptualizations of causal relations such at those developed by Judea Pearl Pearl (2000).

DUE IN 3 WEEKS: Your final paper.

References

Abadie, A. and Imbens, G. (2004). On the Failure of the Bootstrap for Matching Estimators. *NBER, Unpublished Manuscript*.

Abadie, A. and Imbens, G. (2009). Matching on the estimated propensity score.

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.

Bowers, J. (2011). Making effects manifest in randomized experiments. In Druckman, J. N., Green, D. P., Kuklinski, J. H., and Lupia, A., editors, *Cambridge Handbook of Experimental Political Science*, chapter 32. Cambridge University Press, New York, NY.

Bowers, J. and Panagopoulos, C. (2011). Fisher's randomization mode of statistical inference, then and now. Unpublished manuscript.

Diamond, A. and Sekhon, J. (2006). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies.

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. (2007). On regression adjustments in experiments with several treatments. *Annals of Applied Statistics (To Appear)*.

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

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

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Glynn, A. and Quinn, K. (2010). An introduction to the augmented inverse propensity weighted estimator. *Political Analysis*, 18(1):36.

Green, D. P. (2009). Regression adjustments to experimental data: Do david freedman's concerns apply to political science? Unpublished Manuscript.

Hansen, B. (2008). Comment: The essential role of balance tests in propensity-matched observational studies. *Statistics in Medicine*, 27(12).

Hansen, B. (2009). Propensity score matching to recover latent experiments: diagnostics and asymptotics. Technical Report 486, Statistics Department, University of Michigan.

Hansen, B. and Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 23:219.

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

Hansen, B. B. (2011). Propensity score matching to extract latent experiments from nonexperimental data: A case study.

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.

Hosman, C. A., Hansen, B. B., and Holland, P. W. (2010). The sensitivity of linear regression coefficients's confidence limits to the omission of a confounder. *The Annals of Applied Statistics*, 4(2):849–870.

Iacus, S., King, G., and Porro, G. (2009). Causal inference without balance checking: Coarsened exact matching. *Retrieved September*, 13:2010.

Iacus, S., King, G., and Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106(493):345–361.

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.

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.

Imbens, G. (2003). Sensitivity to Exogeneity Assumptions in Program Evaluation. *The American Economic Review*, 93(2):126–132.

Imbens, G. and Rubin, D. (2009). Causal inference in statistics. Unpublished book manuscript. Forthcoming at Cambridge University Press.

Lin, W. (2011). Agnostic notes on regression adjustments to experimental data: reexamining freedman's critique. Unpublished manuscript.

Lu, B., Greevy, R., Xu, X., and Beck, C. (2011). Optimal nonbipartite matching and its statistical applications. *The American Statistician*, 65(1):21–30.

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

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.

Pearl, J. (2000). *Causality : Models, Reasoning, and Inference*. Cambridge University Press, reprinted with corrections edition.

Rosenbaum, P. R. (2002a). Covariance adjustment in randomized experiments and observational studies. *Statistical Science*, 17(3):286–327.

Rosenbaum, P. R. (2002b). *Observational Studies*. Springer-Verlag, second edition.

Rosenbaum, P. R. (2010). *Design of Observational Studies*. Springer.

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

Schochet, P. (2009). Is regression adjustment supported by the neyman model for causal inference. *Journal of Statistical Planning and Inference*.

Sekhon, J. (2007a). Alternative balance metrics for bias reduction in matching methods for causal inference. *Survey Research Center, University of California, Berkeley*.

Sekhon, J. (2007b). Multivariate and propensity score matching software with automated balance optimization: The matching package for r. *Journal of Statistical Software*, 10(2):1–51.