Introduction to propensity score matching

Stephen R. Porter

Department of Leadership, Policy and Adult and Higher Education
North Carolina State University
srporter@ncsu.edu

November 11, 2013

What it PSM?

Method to estimate a causal effect of a treatment when randomization is not possible

Best way to think about it is data preprocessing rather than statistical analysis

1. Units are dropped until the remaining units in the treated and control groups look similar
2. Then a statistical analysis is performed on the pruned data

Other ways (such as weighting), that keep all units, but the creation of weights still occurs prior to statistical analysis

Why has it become so popular?

Growing focus on causal effects in many fields, rather than description or association

Other approaches to causal inference are more burdensome

- Experiments - randomization of treatment
- RD - special assignment variable with cutoff
- DID/ITS - repeated observations over time
- IV - special assignment variable (instrument)

PSM is only technique that can (potentially) handle non-random assignment of treatment on virtually any cross-sectional dataset

Crucial assumption underlying PSM

[PSM] should only be applied if the underlying identifying assumption can be credibly invoked based on the informational richness of the data and a detailed understanding of the institutional set-up by which selection into treatment takes place ... Caliendo & Kopeinig (2008), p. 32

Identifying assumption: you observe all variables that drive both treatment and outcome

If you have to have all those covariates ...

Why not use some version of the general linear model?

We almost always assume a linear functional form; suppose you estimate

$$Y = B_0 + B_1 X + u$$

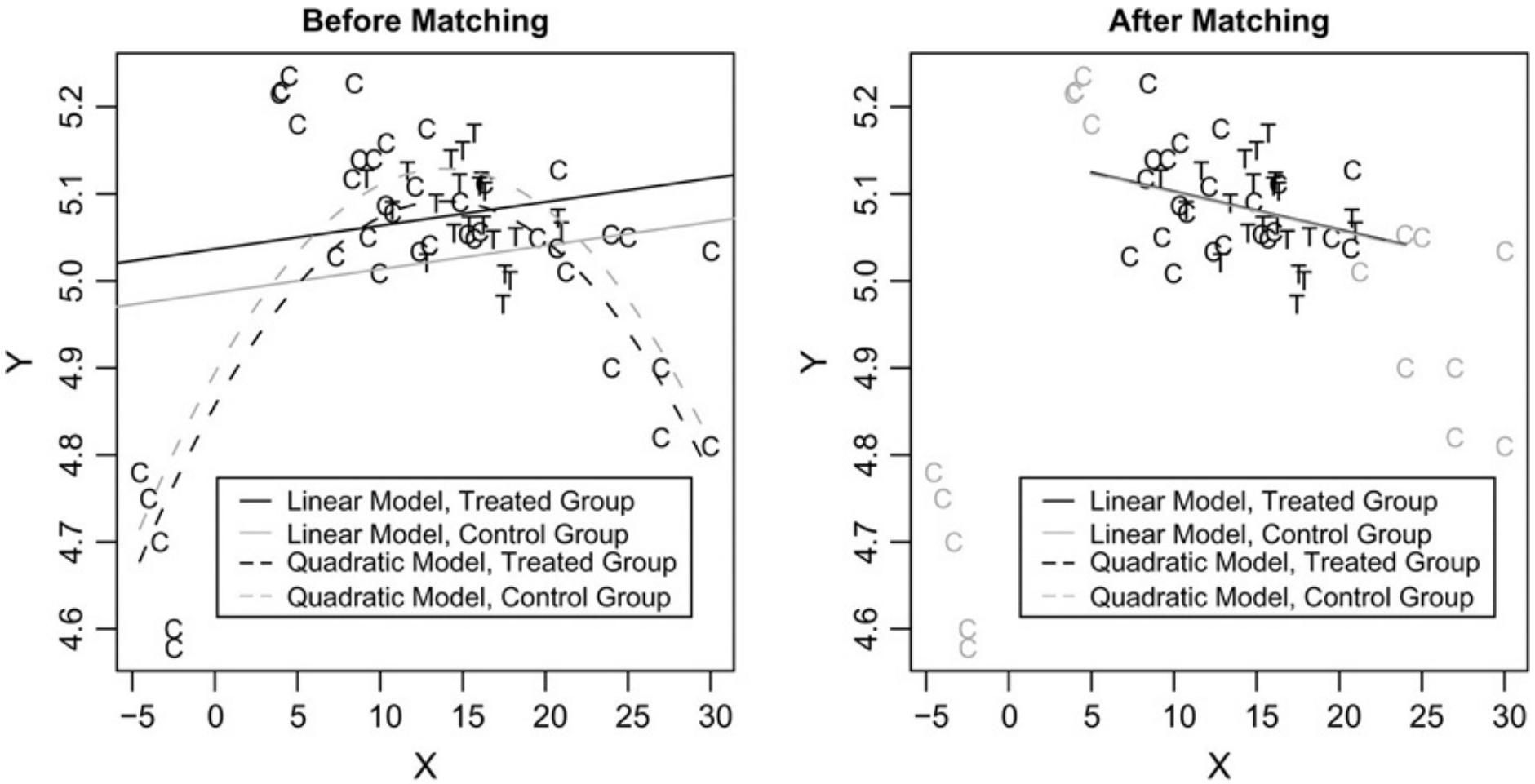
but the true model is

$$Y = B_0 + B_1 X + B_2 X^2 + u$$

You are actually estimating

$$Y = B_0 + B_1 X + \overbrace{X^2 + v}^{u}$$

Extrapolation and common support



Source: Ho et al. (2007) *Political Analysis*, 15:199-236

Model dependence

- That one OLS/HLM model you see in a journal article is the final version of dozens or hundreds of models run by the author
- Very easy to alter model specification to get the "right" result
- When done properly, PSM requires that you create the matched dataset before you begin analysis
  - Matching model not driven by "right" results
  - Research shows that analytic model at second stage not affected by model specification when using matched data

Rubin causal model of potential outcomes

Treatment: developmental (remedial) math at college entry

Each person has two potential outcomes for Y_i (dependent variable)

- $Y_i(0)$ college math performance when assigned to control (do not participate in dev ed)
- $Y_i(1)$ college math performance when assigned to treatment (participate in dev ed)

$$UTE_i \equiv Y_i(1) - Y_i(0)$$

This is the effect of the jobs program for each individual, or the Unit Treatment Effect

Fundamental problem of causal inference

Very important concept: we do not observe both $Y(0)$ and $Y(1)$ for the same unit i

- One of these is the actual outcome that occurs
- The other we do not observe: the **counterfactual** outcome

We do observe $Y_i(1)$ for those who take treatment

- Can use to calculate the mean outcome of Y under treatment for those who take the treatment, or $Y_i(1) \mid T_i = 1$

But to estimate treatment effect, we need:

- Mean outcome of Y under control condition for those who take the treatment, or $Y_i(0) \mid T_i = 1$

The problem as missing data and the solution

To estimate causal effect, we need $Y_i(0) \mid T_i = 1$

So causal inference is about finding a credible estimate of $Y_i(0) \mid T_i = 1$

In experiments we use $Y_i(0) \mid T_i = 0$. Why?

In PSM, we find control units that are "clones" of treated units

- Use $Y_i(0) \mid T_i = 0$ for this subgroup as an estimate of $Y_i(0) \mid T_i = 1$
- Assume their outcome is good counterfactual for treated units

Main assumption for PSM

$$Y(1), Y(0) \perp T \mid X$$

Someone's potential outcomes are independent of their treatment status, once the set of covariates X are taken into account

- Called uncounfoundedness, conditional independence, exogeneity
- Not saying that Y is independent of T; expect T to affect Y
- Units cannot end up in treatment or control based on what their potential outcomes might be
- E.g., choose dev ed because it will improve performance (low ability student) and avoiding it because it won't improve performance (high ability student)

Summary

Think of causal effect as difference between factual and counterfactual outcome

Because we only observe the factual outcome, have to find credible measure of what would have happened to T's under the control condition

Randomization of treatment is one way to achieve this

PSM is another: we use propensity scores to create a group of C's that are observably similar to T's

But whether this provides a credible measure of the counterfactual depends on selection process and available data

Determining the matched comparison group

We could try exact matching, e.g., to find a match for me find someone who is/has

- Age 48, white, male, Ph.D., married, two kids, owns home, same income, same responses to set of attitudinal survey questions, etc.
- Very difficult to find matches; may only be able to match a handful of treated units
- "Curse of dimensionality"

Turns out that the predicted probablity of treatment from a model using covariates to predict treatment captures all of the information in those covariates

- So if we match T's and C's with same $\hat{p}$ and compare distributions of covariates for the two groups, they will look (almost) the same

1. Understand the selection process

By far the most important step - credibility of analysis rests on whether you can make a case for unconfoundedness

Can think of three sets of variables

1. Affect outcome only
2. Affect treatment only
3. Affect both - this is the set that matters

Theory, previous research, and particular context should guide variable choice

- If crucial covariates are unavailable, have to make a case using proxy variables
- Simply matching on a set of covariates without a discussion of the selection process is not very convincing

2. Estimate the selection model

Propensity score is simply the predicted probability from a logistic regression model

What is the goal of the model?

- It is *not* to maximize pseudo R^2 or % correctly predicted
- Goal: achieve balance on covariates between treated and controls
- May have to consider nonlinear and interaction terms to achieve this

Important to only include covariates that are pre-treatment

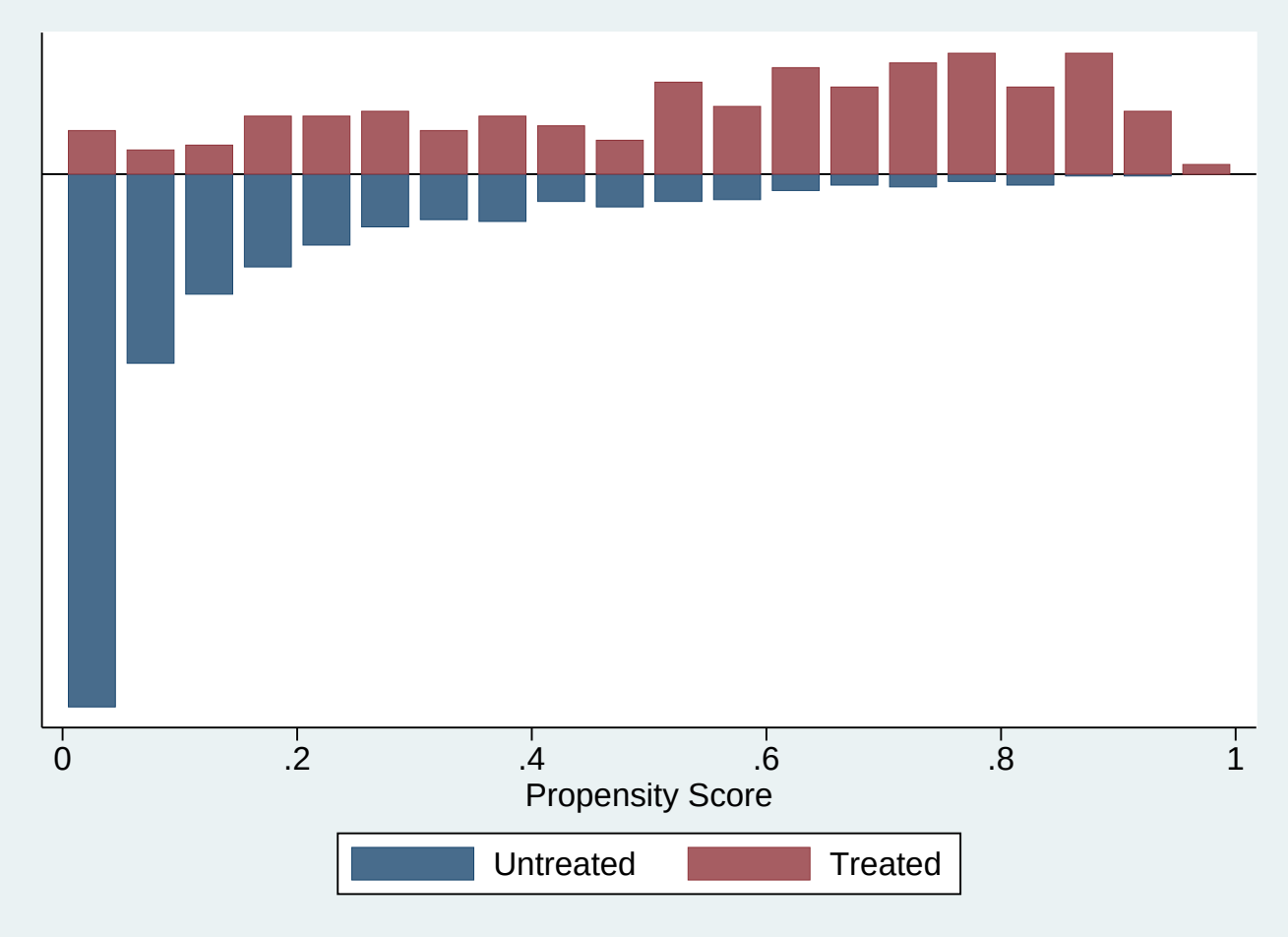
3. Assess common support

Fancy way of saying that you have treated and control units with similar probabilities of treatment

Best way to assess is graphically; should be included in any report

- Often treated will have on average higher probabilities than controls
- It is easy to lose many cases at this stage

3. Assess common support



Source: Data from Guo and Fraser (2010), Example 5.9.2

4. Use algorithm to match controls to treated

There are a wide variety of ways to match

- Key: face a tradeoff between bias and efficiency
- Limit matches to very good matches
  - End up with small sample size and little bias
- Allow matches to differ
  - Larger sample size, but probably some bias, because matches are not as good

- Define a caliper of a specific width, such as .02
- Search for nearest neighbor only among C's within the the caliper
  - Out of these C's, choose the nearest

5. Assess match quality

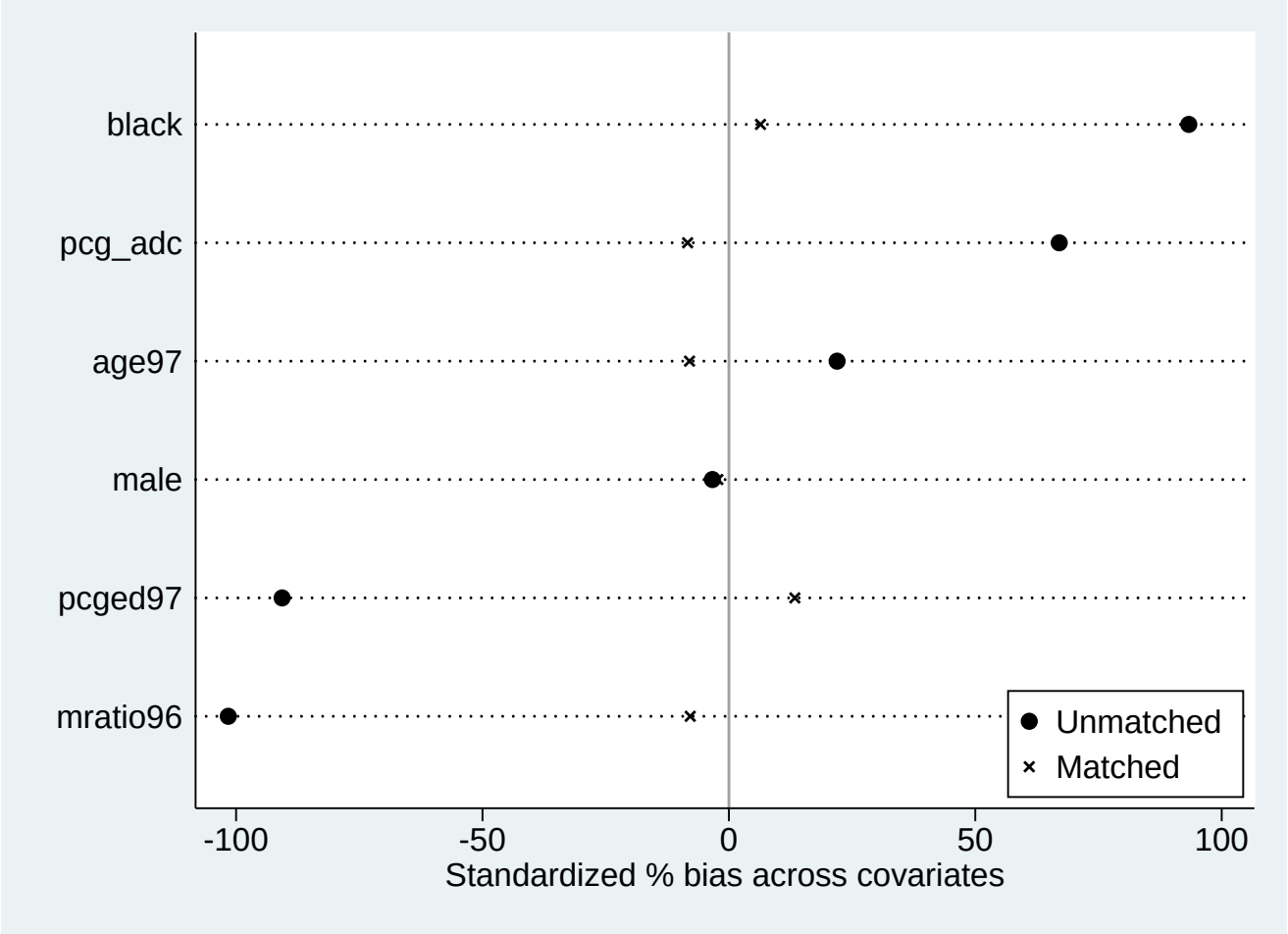
Goal of matching process is to yield two groups that look alike on a given set of covariates

Several ways to see if this is the case

- Simple t-test
- Standardized bias - difference between sample means as a percentage of square root of variances

$$100 * \frac{\bar{X}_T - \bar{X}_C}{\sqrt{(s_T^2 + s_C^2)/2}}$$

5. Assess match quality



Source: Data from Guo and Fraser (2010), Example 5.9.2

6. Estimate the treatment effect

Choice between difference of means or multivariate model using treatment dummy variable

We are replicating a random experiment; with experimental data we usually just conduct a t-test or ANOVA

- Commands in Stata such as *psmatch2* calculate the treatment effect as a difference in means

Some methodologists advocate a multivariate model

- Necessary if imbalance between covariates; idea is that OLS will control for any remaining differences

Suggested readings

Methods

- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31-72.
- Guo, S., & Fraser, M.W. (2010). *Propensity Score Analysis: Statistical Methods and Applications*. Thousand Oaks, CA: Sage Publications
- Ho, D.E. et al. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*(3), 199-236.
- Stuart, E.A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1-21.

Applications

- Gasper, J. et al. (2011). Switching schools: Revisiting the relationship between school mobility and high school dropout. *American Educational Research Journal, 49*(3), 487-519.
- Retelsdorf et al. (2012). Reading development in a tracked school system: A longitudinal study over 3 years using propensity score matching. *British Journal of Educational Psychology, 82*, 647-671.