Propensity Score Matching Methods

February 19, 2010

Melissa S. Kovacs, Ph.D.

ASU Center for Violence Prevention and Community Safety

ARIZONA STATE UNIVERSITY

What is Matching?

- Matching is a non-parametric or semi-parametric analogue to regression

- Used to evaluate binary treatment variables

- Uses regression methods to construct counterfactuals under an assumption of selection on observables

- One way of evaluating a treatment (other ways: instrumental variables methods, difference-in-differences methods, case matching)

In other words…

- We have one data set that we have divided into two groups: the treated and the untreated
- We can describe these two groups via numerous background characteristics
- We are going to match individuals between the two groups based on their background characteristics, and
- Test for the effect of the treatment

Why Matching?

- Our ideal is a randomized experiment
- Not often practical in social science settings
- Often want to evaluate a binary treatment
- Matching statistically mimics randomization by heavily relying on background characteristics

Matching Requires…

- A binary treatment variable
- A measurable outcome variable, either binary or continuous
- Rich data on background characteristics (more on this to come…)
- A program or outcome you wish to evaluate

A little math ….

Let Y_1 denote the outcome in the treated state

Let Y_0 denote the outcome in the untreated state

Let D be the treatment variable (0,1) denoting participation in a program

The observed outcome is then:

$Y = DY_1 + (1-D)Y_0$

Parameters of Interest (answering different policy questions….)

The impact of treatment on the treated:

$$TT = E(Y_1 - Y_0 \mid D = 1) =$$

$$E(Y_1 \mid D = 1) - E(Y_0 \mid D = 1)$$

The average treatment effect:

$$ATT = E(Y_1 - Y_0) = E(Y_1) - E(Y_0)$$

The Evaluation Problem

Clearly, we can only observe one outcome (either Y_0 or Y_1) for each individual or case in our data.

In other words, we can construct $E(Y_1 \mid D = 1)$ from data on the treated, but we cannot possibly observe $E(Y_0 \mid D = 1)$.

The result is selection bias….

In general, people do not select into the treatment randomly, or in a way that is unrelated to the untreated outcome.

We can measure the selection bias as:

$E(Y_0 \mid D = 1) - E(Y_0 \mid D = 0)$

How to Fix the Evaluation Problem? Matching!

While random assignment in an experiment would solve the evaluation problem (by directly constructing the unobserved counterfactual $E(Y_0 \mid D = 1)$ and excluding the $D = 1$ people who would otherwise have been treated from the treatment), MATCHING assumes that selection is unrelated to the untreated outcome conditional on some set of observed variables X.

Assumptions in Matching

- Conditional Independence Assumption (CIA)
- Stable Unit Treatment Value Assumption (SUTVA)
- Common support is met

Conditional Independence Assumption (CIA)

Treatment status is assumed to be random, conditional on some set of observed *X* variables

In a sense, all *X* are exogenous. *X*s affected by the treatment will invalidate the CIA.

We satisfy the CIA if our set of *X* includes all variables that affect <u>both</u> (not either) participation and outcomes

Stable Unit Treatment Value Assumption (SUTVA)

Treatment of a unit *i* affects only the outcome of unit *i*.

Assumes that the potential outcome for any particular unit following treatment is stable. OR, that it would take the same value for all other treatment allocations such that the unit receives the treatment. OR, assumes no general equilibrium effects.

SUTVA is violated when

- there are versions of each treatment varying in effectiveness
- interference between units exists (i.e. participation for one is dependent on another) (Rubin 1990)

Matching and Regression

Traditional regression analyses also assume that selection into treatment is based on observable factors and can be eliminated by conditioning on those observables.

So why do matching? Why not just regression?

Matching and Regression cont'd.

Matching does a few things that regression does not:

1. Matching does not assume a linear functional form as regression does
2. Matching highlights the support problem (more on this to come)
3. Matching weights observations differently than regression does

When to use Matching?

Depends on your data.

Matching requires rich data on observable characteristics. Matching requires data on all variables that determine BOTH outcomes and participation.

Taking a step back…

- We have one data set that we have divided into two groups: the treated and the untreated
- We can describe these two groups through numerous background characteristics
- We are going to match individuals between the two groups based on their background characteristics

Exact Matching is Ideal

Ideally, we would use the simplest method of matching – exact matching, or cell matching – to compare individuals with exactly the same values of observed variables (background characteristics) *X*.

But, this can quickly be troublesome….

The Curse of Dimensionality

If there are several discrete X, each with several values, the number of cells becomes large, and may include cells with no untreated observations to correspond to a treated observation.

I.e. with five variables that each have three values, you have $3^5 = 243$ cells.

To fix …

Matching estimators that do not require exact matching pose a solution to the Curse of Dimensionality

Inexact matching procedures reduce the dimension by defining a distance metric on *X* and matching on the distance rather than the *X*.

Propensity Scores

If you can match on X, you can also match on the propensity score (Rosenbaum and Rubin 1983).

The propensity score is notated as:

$$P(X) = \Pr(D = 1 \mid X)$$

Alternatively stated, two groups with the same probability of participation will be in the treated and untreated groups in equal proportions.

Frequently estimated using a parametric model such as logit or probit.

Interval Matching

Interval matching divides the range of propensity scores into a fixed number of intervals (of any length). The estimate takes the difference between the mean outcomes of the treated and untreated units in each interval.

Nearest Neighbor Matching

Very common form of matching – each treated individual is matched to the nearest untreated individual.

Matching with replacement occurs when a given treated observation forms the counterfactual for more than one untreated observation.

Matching without replacement, however, can result in poor results if there are few treated observations or few untreated observations in your sample.

Nearest Neighbor Matching cont’d.

Matching without replacement keeps variance low at the cost of potential bias, while matching with replacement keeps bias low at the cost of larger variance.

To use replacement or not depends on the ratio of treated to the untreated in your data (Frohlich 2004 and Hansen 2004).

This is inefficient – we don’t use all the data we have collected.

Kernel Matching

Treated individuals are matched to a weighted sum of individuals who have similar propensity scores.

Kernel matching can be more efficient – some methods use all the data from the untreated group (i.e. Gaussian kernel) while others use only a certain probability range or bandwidth (i.e. Epanechnikov kernel)

Other Methods

- Local linear matching
- Caliper matching
- Mahalanobis distance matching
- Inverse probability weighting

Choosing a Matching Method

All matching estimators are consistent – as the sample gets arbitrarily large, the units being compared get arbitrarily close to one another with respect to their background characteristics.

When comparison observations are few, nearest neighbor without replacement is a bad idea.

When comparison observations are many (and evenly distributed), nearest neighbor with replacement will make good use of your rich data.

Choosing cont'd.

When comparison observations are many and are asymptotically distributed, kernel matching helpfully uses the additional data where it exists, but will not use bad matches.

If observations have *P(X)* near zero or one, local linear matching makes sense.

Matching Assumes Common Support

Logically, if you are going to estimate the counterfactual for an individual by matching someone to that person, then your match must have similar characteristics to the person in the counterfactual state.

If not, you have not met the common support condition.

The common support problem has implications for how data is collected, if you have a chance to influence data collection (this would be ideal).

Fixing the Common Support Problem

Most common method: discard all observations with extreme estimates (below the two minima and above the two maxima)

- This is simple, but, potentially fine matches near the boundary may be lost
- Interior regions where the common support condition fails are not excluded

Alternatively, caliper matching and kernel matching drop extreme matches as a matter of form.

Variable Selection

We choose which *X* variables to include with the CIA in mind.

There is no deterministic variable selection procedure to tell you which variables to choose.

Let theory and institutional knowledge be your guides.

Recall that you want to include all variables that affect BOTH participation and outcomes.

Balancing Tests

Balancing tests determine whether observations with the same propensity score have the same distribution of observable covariates independent of treatment status.

Different Balancing Tests

- Blocking based on the propensity score (Dehejia and Wahba 1999 and 2002)
- Standardized differences (Rosenbaum and Rubin 1985)
- Hoteling T-squared test
- Regression-based balancing test (Smith, unpublished evidence)

Problems with Standard Errors

Frequently, matching leads to added variance through:

1. Estimating the propensity scores
2. Matching with replacement

Evidence exists that these factors do add significant variance in matching models (Heckman, Ichimura and Todd 1998)

Bootstrapping – an Imperfect Fix

Bootstrapped standard errors can account for matching's added variance (Efron 1993), but is not helpful for nearest neighbor matching due to a lack of smoothness (Abadie and Imbens 2006)

Bootstrapping is valid for kernel matching, local linear, etc. that are based on smooth non-parametric regressions

Summing Up…

- With matching, we are more concerned with the quality of the match than the outcome parameter
- Matching has useful, practical applications
- Matching avoids a linearity assumption inherent in regressions
- Matching highlights the common support problem that regression does not
- You must observe variables that affect both participation and outcome
- Like other methods, it is useful to consider the sensitivity of matching estimators and procedure before data collection

Melissa S. Kovacs, Ph.D.

Email: mkovacs1@asu.edu

Special thanks to Jeff Smith, University of Michigan, for providing some material for this workshop, and to the Center for Violence Prevention and Community Safety for making this possible.

ASU®
Center for Violence Prevention
and Community Safety

ARIZONA STATE UNIVERSITY