An Introduction to Matching Methods for Causal Inference and their Implementation in Stata

Barbara Sianesi

IFS

Stata Users' Group Meeting

Berlin, June 25, 2010

(PS)MATCHING IS EXTREMELY POPULAR…

→ 240,000 entries by googling: propensity score matching

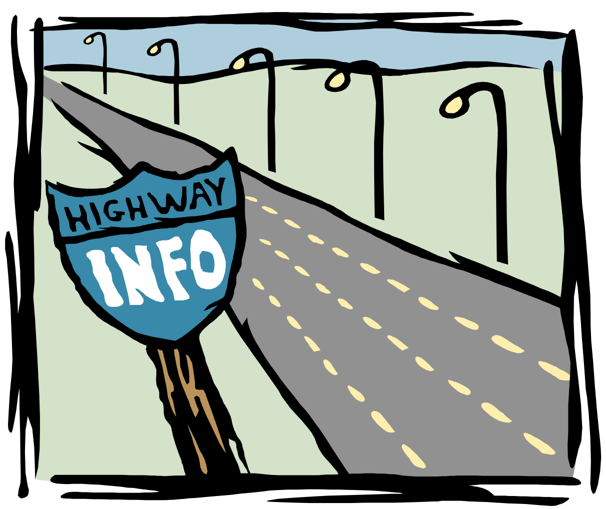
→ >8,300 downloads of –psmatch2–

among the top 1‰ research items by number of citations, discounted by citation age of the RePEc/IDEA database

→ >1,300 support emails

- Europe, US, Canada, Central + South America, former SU, Australia, Asia, Africa and the Middle East
- epidemiology, sociology, economics, statistics, criminology, agricultural economics, health economics, transport economics, public health, nutrition, paediatrics, biostatistics, finance, urban planning, geography and geosciences

Roadmap



1. The counterfactual concept of causality
2. What is matching?
3. Should we use it?
4. How do we use it?
    a. Matching estimators
    b. Practical Stata example using `psmatch2`

The Evaluation Problem

Y_{0i} , Y_{1i} → Outcome of *i* under treatment 0 and under treatment 1

$D_i \in \{0, 1\}$ → Treatment indicator

$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$ → Observed outcome of *i*

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i}) D_i$$

Causal effect on *Y* of treatment 1 relative to treatment 0 for *i*

X_i → Set of observed characteristics of *i*

Which parameter?

- ATT $= E(Y_1 - Y_0 \mid D=1) = E(Y_1 \mid D=1) - \mathbf{E(Y_0 \mid D=1)}$
- ATNT $= E(Y_1 - Y_0 \mid D=0) = \mathbf{E(Y_1 \mid D=0)} - E(Y_0 \mid D=0)$
- ATE $= E(Y_1 - Y_0) = \text{ATT} \cdot P(D=1) + \text{ATNT} \cdot P(D=0)$

Need to invoke (untestable) assumptions to identify **average unobserved counterfactuals**.

MATCHING METHODS

1. Identifying assumption: **Selection on Observables**
   (all the relevant differences between treated and non-treated are captured in X):

   ATT: $Y_0 \perp D \mid X \quad \rightarrow \quad E(Y_0 \mid X, D{=}1) = E(Y_0 \mid X, D{=}0)$
   ATNT: $Y_1 \perp D \mid X \quad \rightarrow \quad E(Y_1 \mid X, D{=}1) = E(Y_1 \mid X, D{=}0)$
   ATE: $Y_0, Y_1 \perp D \mid X$

2. To give it empirical content: **Common Support**
   (we observe participants and non-participants with the same characteristics):

   ATT: $P(D{=}1 \mid X) < 1$
   ATNT: $0 < P(D{=}1 \mid X)$
   ATE: $0 < P(D{=}1 \mid X) < 1$

$\Rightarrow$ can use the (observed) mean outcome of the non-treated to estimate the mean (counterfactual) outcome the treated would have had they not been treated.

Matching vs OLS

Matching makes the *same* identifying assumption as OLS but avoids any additional ones:

<u>COMMON SUPPORT</u> → effectively compares only comparable individuals

<u>NON-PARAMETRIC</u> → avoids potential misspecification of $E(Y_0 \mid X)$

→ allows for arbitrary X-heterogeneity in impacts $E(Y_1 - Y_0 \mid X)$

But: if OLS is correctly specified, it is more <u>efficient.</u>

Bias decomposition

$\boldsymbol{B_1}$ difference in the supports of X
Eliminated by performing matching only over Sup_{10}
NB: might recover a different causal impact: $\text{ATT}(\text{Sup}_{10}) \neq \text{ATT}(\text{Sup}_1)$ (external validity)

$\boldsymbol{B_2}$ difference of the distribution of X over Sup_{10}
Eliminated since matching reweighs D=0 data to equate the distribution of X in the D=1 sample

$\boldsymbol{B_3}$ difference in unobservables
Matching just as biased as OLS (internal validity)

$\Rightarrow$ Matching focuses on comparability in terms of observables,
i.e. on constructing a suitable comparison group by carefully matching treated and non-treated on X / reweighting the non-treated to realign their X

BUT we don't need matching to make OLS less parametric…

FULLY INTERACTED OLS

`film` or `margins, dydx(treated) over(treated)`

$$Y = m_0(X_1, X_2) + \delta D + \delta_1(X_1 D) + \delta_2(X_2 D) + \delta_{12}(X_1 X_2 D) + e$$

$$\beta_{ATT} = \delta + \delta_1 \bar{X}_{1|D=1} + \delta_2 \bar{X}_{2|D=1} + \delta_{12}(\overline{X_1 X_2})_{|D=1}$$

$$\beta_{ATNT} = \delta + \delta_1 \bar{X}_{1|D=0} + \delta_2 \bar{X}_{2|D=0} + \delta_{12}(\overline{X_1 X_2})_{|D=0}$$

$$\beta_{ATE} = \delta + \delta_1 \bar{X}_1 + \delta_2 \bar{X}_2 + \delta_{12}(\overline{X_1 X_2})$$

Can F-test for presence of heterogeneous effects.

STILL, matching (≠ OLS) highlights comparability of groups

Check matching quality

- Propensity score
  - more ‘structural’ model
  - more flexible specification
  - probit/logit
  - probability/index/odds ratio

- Matching
  - metric: X, $\hat{p}(X)$ or $\{X, \hat{p}(X)\}$
  - type of matching
  - smoothing parameters
  - common support

- Assessment of matching quality

CAN we get the two groups balanced?

STRENGTHS AND WEAKNESSES

☺ **Advantages** ☺

- controls for selection on observables and on observably heterogeneous impacts
- non-(or semi-) parametric:
  no specific form for outcome equation, decision process or either unobservable term
- Sup_{10}: compare only comparable people and help in determining which results most reliable
- flexible and easy

☹ **Disadvantages** ☹

- selection on observables: matching as good as its *X*'s
- Sup_{10}: if impact differs across treated, restricting to Sup_{10} may change parameter being estimated
  $\rightarrow$ unable to identify ATT
- data hungry

OPERATIONALISING MATCHING METHODS

Curse of dimensionality

- impose linearity in the parameters (regression analysis)
- choose a distance metric
  - ❖ Euclidean, Mahalanobis, etc.
  - ❖ **Propensity Score** $e(x) \equiv P(D=1 | X=x)$

$$\boxed{X \perp D \mid e(X)}$$

$$\boxed{\begin{array}{c} (Y_1\,,\,Y_0) \perp D \mid X \quad \text{and} \quad 0 < e(X) < 1 \\ \Rightarrow \quad (Y_1\,,\,Y_0) \perp D \mid e(X) \end{array}}$$

Overview of Matching Estimators

1. pair to each treated *i* some group of 'comparable' non-treated individuals
2. associate to the outcome y_i of treated *i*, a matched outcome $\hat{y}_i$ given by the (weighted) outcomes of his 'neighbours' in the comparison group:

$$\hat{y}_i = \sum_{j \in C^0(p_i)} w_{ij} y_j$$

- $C^0(p_i)$ = set of neighbours of treated *i* in the D=0 group
- w_{ij} = weight on non-treated *j* in forming a comparison with treated *i*, where $\sum_{j \in C^0(p_i)} w_{ij} = 1$

General form of the matching estimator for ATT (within S_{10}):

$$\hat{ATT} = \frac{1}{\#(D=1 \cap S_{10})} \sum_{i \in \{D_i=1 \cap S_{10}\}} \{y_i - \hat{y}_i\} = E(Y \mid \text{treated on } S_{10}) - E(Y \mid \text{matched non-treated})$$

TRADITIONAL MATCHING ESTIMATORS

One-to-one matching

- with or without replacement
- nearest neighbour or within caliper

SIMPLE SMOOTHED MATCHING ESTIMATORS

- *K*-nearest neighbours
  - with or without replacement
  - nearest neighbour or within caliper
- radius matching

WEIGHTED SMOOTHED MATCHING ESTIMATORS

- kernel-based matching
- local linear regression-based matching
- ❑ bandwidth choice
- ❑ kernel choice

MAHALANOBIS-METRIC MATCHIING

combine the *W*'s into a distance measure and then match on the resulting scalar:

$$d(i,j) = (\boldsymbol{W}_{\mathrm{i}} - \boldsymbol{W}_{\mathrm{j}})\, V^{-1}\, (\boldsymbol{W}_{\mathrm{i}} - \boldsymbol{W}_{\mathrm{j}})'$$

Implementing the Common Support requirement

- caliper
- at the boundaries
- trimming

Checking matching quality

Check (and possibly improve on) balancing of observables

$$\boxed{D \perp X \mid \hat{p}(X)}$$

- for each variable
- overall measures

Inference

- naïve variance
- bootstrapping
- Abadie-Imbens standard errors

Leuven and Sianesi (2003) `psmatch2` suite

```
psmatch2 depvar [indepvars] [if exp] [in range] [,
    outcome(varlist)
    pscore(varname) logit odds index
    neighbor(integer) ties
    noreplacement descending
    caliper(real)
    radius
    kernel
    llr
    kerneltype(type) bwidth(real)
    spline nknots(integer)
    mahalanobis(varlist) add pcaliper(real)
    common trim(real)
    ate
    ai]

psgraph
pstest
```

THE IMPACT OF THE NSW DEMONSTRATION

Very famous data in the evaluation literature,

combining treatment and controls from a <u>randomised</u> evaluation of the NSW Demonstration with non-experimental individuals drawn from various sources.

LaLonde (1986), Dehejia, R.H. and Wahba, S. (1999), Smith, J. and Todd, P. (2005) with response, rejoinder, final thoughts.

Also used by Ichino and Becker (2002) and Abadie, Drukker, Leber Herr and Imbens (2001) to illustrate their respective Stata matching programs.

Here we use the NSW male treated with male comparisons drawn from the PSID.

To keep in mind: experimental impact estimate on real earnings is +$886*

WRAPPING UP…

SELECTION ON UNOBSERVABLES

- Set of conditioning X matters
  $\Rightarrow$ better data help a lot!

SELECTION ON OBSERVABLES

- avoid use of functional forms in constructing counterfactual
  $\Rightarrow$ (matching $\approx$ fully interacted OLS) $\succ$ simple OLS
  Matching *versus* simple OLS:
  no mis-specification bias; ATT *versus* ATNT

- compare comparable people
  $\Rightarrow$ matching $\succ$ fully interacted OLS
  Matching *versus* fully interacted OLS:
  highlights actual comparability of groups, hence reliability (& relevance) of estimates

SELECTED REFERENCES

A comprehensive review

Imbens, G. (2004), ‘Semiparametric estimation of average treatment effects under exogeneity: a review’, *Review of Economics and Statistics*, 86, 4-29.

The propensity score

Rosenbaum, P.R. and Rubin, D.B. (1983), “The Central Role of the Propensity Score in Observational Studies for Causal Effects”, *Biometrika*, 70, 41-55.

Rosenbaum, P.R. and Rubin, D.B. (1984), “Reducing Bias in Observational Studies Using Sub-Classification on the Propensity Score”, *Journal of the American Statistical Association*, 79, 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.

Dehejia, R.H. and Wahba, S. (1999), “Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programmes”, *Journal of American Statistical Association*, 94, 1053-1062.

Heckman, J.J., Ichimura, H. and Todd, P.E. (1997), “Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme”, *Review of Economic Studies*, 64, 605-654.

Heckman, J.J., Ichimura, H. and Todd, P.E. (1998), “Matching as an Econometric Evaluation Estimator”, *Review of Economic Studies*, 65, 261-294.

Mahalanobis-metric matching

Rubin, D.B. (1979), “Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies”, *Journal of the American Statistical Association*, 74, 318-328.

Rubin, D.B. (1980), “Bias Reduction Using Mahalanobis-Metric Matching”, *Biometrics*, 36, 293-298.

Multiple treatments

Imbens, G.W. (2000), “The Role of Propensity Score in Estimating Dose-Response Functions”, *Biometrika*, 87, 706-710.

Lechner, M. (2001), Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption, in: Lechner, M., Pfeiffer, F. (eds), *Econometric Evaluation of Labour Market Policies*, Heidelberg: Physica/Springer, 43-58.

Inference/Efficiency issues

Hahn, J. (1998), “On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects,” *Econometrica*, 66, 315-331.

Hirano, K., G. Imbens, and G. Ridder (2003), “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score,” *Econometrica*, 71, 1161-1189.

Abadie, A. and Imbens, G. (2006), “Large Sample Properties of Matching Estimators for Average Treatment Effects”, *Econometrica*, 74, 235-267.