Implementing Propensity Score Matching Estimators with STATA

Barbara Sianesi

University College London
and
Institute for Fiscal Studies

E-mail: barbara_s@ifs.org.uk

Prepared for

UK Stata Users Group, VII Meeting
London, May 2001

BACKGROUND: THE EVALUATION PROBLEM

POTENTIAL-OUTCOME APPROACH

Evaluating the **causal effect** of some treatment on some outcome Y experienced by units in the population of interest.

Y_{1i} $\rightarrow$ the outcome of unit i if i were exposed to the treatment

Y_{0i} $\rightarrow$ the outcome of unit i if i were not exposed to the treatment

$D_i \in \{0, 1\}$ $\rightarrow$ indicator of the treatment actually received by unit i

$Y_i = Y_{0i} + D_i\,(Y_{1i} - Y_{0i})$ $\rightarrow$ the actually observed outcome of unit i

X $\rightarrow$ the set of pre-treatment characteristics

CAUSAL EFFECT FOR UNIT *i*

$$Y_{1i} - Y_{0i}$$

THE 'FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE'

impossible to *observe* the individual treatment effect

$\rightarrow$ impossible to make causal inference without making generally untestable assumptions

Under some assumptions:

estimate the *average* treatment effect at the population, or at a sub-population, level:

- average treatment effect
- average treatment effect on the untreated
- **AVERAGE TREATMENT EFFECT ON THE TREATED**:

$$E(Y_1 - Y_0|D=1) = E(Y_1|D=1) - E(Y_0|D=1)$$

Need to construct the counterfactual $E(Y_0 \mid D=1)$ – the outcome participants would have experienced, on average, had they not participated.

$E(Y_0 \mid D=0)$?

In non-experimental studies:

need to adjust for confounding variables

MATCHING METHOD

1. assume that all relevant differences between the two groups are captured by their observables X:

$$\boxed{Y_0 \perp D \mid X} \qquad \text{(A1)}$$

2. select from the non-treated pool a control group in which the distribution of observed variables is as similar as possible to the distribution in the treated group

For this need:

$$\boxed{0 < Prob\{D{=}1 \mid X{=}x\} < 1 \quad \text{for } x \in \tilde{X}} \qquad \text{(A2)}$$

$\Rightarrow$ matching has to be performed over the common support region

PROPENSITY SCORE MATCHING

$$p(x) \equiv Pr\{D{=}1|X{=}x\}$$

A1) & A2) $\Rightarrow$

$$Y_0 \perp D \mid p(X) \quad \text{for } X \text{ in } \tilde{X}$$

OVERVIEW: TYPES OF MATCHING ESTIMATORS

- pair to each treated individual *i* some *group* of 'comparable' non-treated individuals and then
- associate to the outcome of the treated individual i, y_i, the (*weighted*) outcomes of his 'neighbours' j in the comparison group:

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

where:

- $C^0(p_i)$ is the set of neighbours of treated i in the control group
- $w_{ij} \in [0, 1]$ with $\sum_{j \in C^0(p_i)} w_{ij} = 1$

  is the weight on control j in forming a comparison with treated i

<u>Two broad groups of matching estimators</u>

→ individual neighbourhood

→ weights

Associate to the outcome y_i of treated unit i a 'matched' outcome given by

1. the outcome of the most observably similar control unit

   ⇒ TRADITIONAL MATCHING ESTIMATORS:

   one-to-one matching

$$C^0(p_i) = \left\{ j : |p_i - p_j| \;=\; \min_{k \in \{D=0\}} \{|p_i - p_k|\} \right\}$$

$$w_{ik} = 1(k=j)$$

2. a weighted average of the outcomes of more (possibly all) non-treated units where the weight given to non-treated unit j is in proportion to the closeness of the observables of i and j

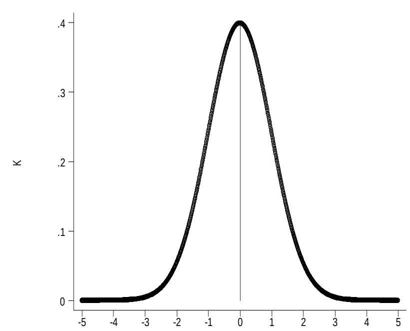
   ⇒ SMOOTHED WEIGHTED MATCHING ESTIMATORS:

   kernel-based matching

$C^0(p_i) = \{D=0\}$ (for gaussian kernel)

$$w_{ij} \propto K\left(\frac{p_i - p_j}{h}\right)$$

$K(.)$
- non-negative
- symmetric
- unimodal



Preparing the dataset

Keep only one observation per individual

Estimate the propensity score on the *X*'s
e.g. *via* `probit` or `logit`
and retrieve either the predicted probability or the index

Necessary variables:

- ✓ the 1/0 dummy variable identifying the treated/controls
- ✓ the predicted propensity score
- ✓ the variable identifying the outcome to be evaluated
- ✓ [optionally: the individual identifier variable]

ONE-TO-ONE MATCHING WITH REPLACEMENT (WITHIN CALIPER)

- Nearest-neighbour matching

Treated unit *i* is matched to that non-treated unit *j* such that:

$$|p_i - p_j| = \min_{k \in \{D=0\}} \{|p_i - p_k|\}$$

- Caliper matching

For a pre-specified $\delta>0$, treated unit *i* is matched to that non-treated unit *j* such that:

$$\delta > |p_i - p_j| = \min_{k \in \{D=0\}} \{|p_i - p_k|\}$$

If none of the non-treated units is within δ from treated unit *i*, *i* is left unmatched.

```
. psmatch treated, on(score) cal(.01)
    [id(serial)] [outcome(wage)]
```

Creates:

1) `_times` → number of times used

use `_times` as frequency weights to identify the matched treated and the (possibly repeatedly) matched controls

2) `_matchdif` → pairwise difference in score

`. sum _matchdif, det` for matching quality

If `id(idvar)` specified

3) `_matchedid` → the `idvar` of the matched control

If `outcome(outcomevar)` specified:

→ directly calculates and displays:

```
Mean wage of matched treated  = 640.39
Mean wage of matched controls = 582.785
Effect  = 57.605
Std err = 74.251377
Note: takes account of possibly repeated use
      of control observations but NOT of
      estimation of propensity score.
T-statistics for H0: effect=0 is .77581053
```

KERNEL-BASED MATCHING

Idea

associate to the outcome y_i of treated unit i
a matched outcome given by a kernel-weighted average of the outcome of all non-treated units,
where the weight given to non-treated unit j is in proportion to the closeness between i and j:

$$\hat{y}_i = \frac{\sum_{j \in \{D=0\}} K\left(\frac{p_i - p_j}{h}\right) y_j}{\sum_{j \in \{D=0\}} K\left(\frac{p_i - p_j}{h}\right)}$$

Control j's outcome y_j is weighted by

$$w_{ij} = \frac{K\left(\frac{p_i - p_j}{h}\right)}{\sum_{j \in \{D=0\}} K\left(\frac{p_i - p_j}{h}\right)}$$

Option `smooth(outcomevar)` creates:

`_moutcomevar` $\rightarrow$ the matched smoothed `outcomevar` $\hat{y}_i$

Bandwidth *h* selection

a central issue in non-parametric analysis
→ trade-off bias-variability

Kernel *K* choice

- Gaussian $K(u) \propto \exp(-u^2/2)$
  uses all the non-treated units

```
. psmatch treated, on(score) cal(0.06)
   smooth(wage)
```

```
Mean wage of matched treated  =   642.70352
Mean wage of matched controls =   677.1453
Effect  =-34.441787
```

- Epanechnikov $K(u) \propto (1-u^2)$ if $|u|<1$ (zero otherwise)
  uses a moving window within the D=0 group, i.e. only those non-treated units within a fixed caliper of h from p_i: $|p_i - p_j| < h$

```
. psmatch treated, on(score) cal(0.06)
   smooth(wage)epan
```

Common support

if not ruled out by the option `nocommon`, common support is imposed on the treated units:

treated units whose *p* is larger than the largest *p* in the non-treated pool are left unmatched.

```
. psmatch treated, on(score) cal(0.06)
    smooth(wage) [epan] nocommon
```

SMOOTHING THE TREATED TOO

For kernel-based matching:

for each $i \in \{D=1\}$,
smooth non-parametrically $E(Y|D=\mathbf{1}, P(X)=p_i) \equiv \hat{y}_i^s$

(to be used instead of the observed y_i)

```
. psmatch treated, on(score) cal(0.06)
    smooth(wage) [epan] [nocommon] both
```

In addition to

`_moutcomevar` $\rightarrow$ the matched smoothed `outcomevar` $\hat{y}_i$

option `both` creates:
`_soutcomevar` $\rightarrow$ the treated smoothed `outcomevar` $\hat{y}_i^s$

E.g.

```
. psmatch treated, on(score) cal(0.06)
   smooth(wage) both
```

```
Mean wage of matched treated    =   642.9774
Mean wage of matched controls   =   677.1453
Effect  =  -34.167822
```

MAHALANOBIS METRIC MATCHING

Replace $p_i – p_j$ above with $d(i,j) = (\boldsymbol{P}_i - \boldsymbol{P}_j)' \mathbf{S}^{-1} (\boldsymbol{P}_i - \boldsymbol{P}_j)$

where

- $\boldsymbol{P}_i$ is the (2×1) vector of scores of unit i
- $\boldsymbol{P}_j$ is the (2×1) vector of scores of unit j
- **S** is the pooled within-sample (2×2) covariance matrix of $\boldsymbol{P}$ based on the sub-samples of the treated and complete non-treated pool.

Useful in particular for multiple treatment framework

```
. psmatch treated, on(score1 score2) cal(.06)
   [smooth(wage)] [epan] [both] [nocommon]
```

ESSENTIAL REFERENCES

- **Propensity score matching**

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.

- **Caliper matching**

Cochran, W. and Rubin, D.B. (1973), “Controlling Bias in Observational Studies”, *Sankyha*, 35, 417-446.

- **Kernel-based matching**

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 distance matching**

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