27) Propensity Score Matching (PSM)

Vitor Kamada

April 2018

Reference

Caliendo, M., Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. Journal of Economic Surveys 22(1): 31–72.

Cameron and Trivedi (2005): Ch 25.1 to 25.4, and 25.8

Conditional Independence Assumption (CIA)

$$y_0, y_1 \perp D$$

$$y = \alpha D + u$$

$$y_0, y_1 \perp D|x$$

$$y = x'\beta + \alpha D + u$$

$$F(y_j|x, D = 1) = F(y_j|x, D = 0) = F(y_j|x)$$

$$F(u_j|x, D = 1) = F(u_j|x, D = 0) = F(u_j|x)$$

Potential Outcome Approach (Rubin, 1974)

$$\tau_i = y_1 - y_0$$

$$\tau_{ATE} = E(\tau) = E[y_1 - y_0]$$

$$\tau_{ATT} = E(\tau|D = 1)$$

$$= E[y_1|D = 1] - E[y_0|D = 1]$$

$$E[y_1|D = 1] - E[y_0|D = 0]$$

$$= \tau_{ATT} + E[y_0|D = 1] - E[y_0|D = 0]$$

Unconfoundedness and Common Support

$$y_0, y_1 \perp D|x$$

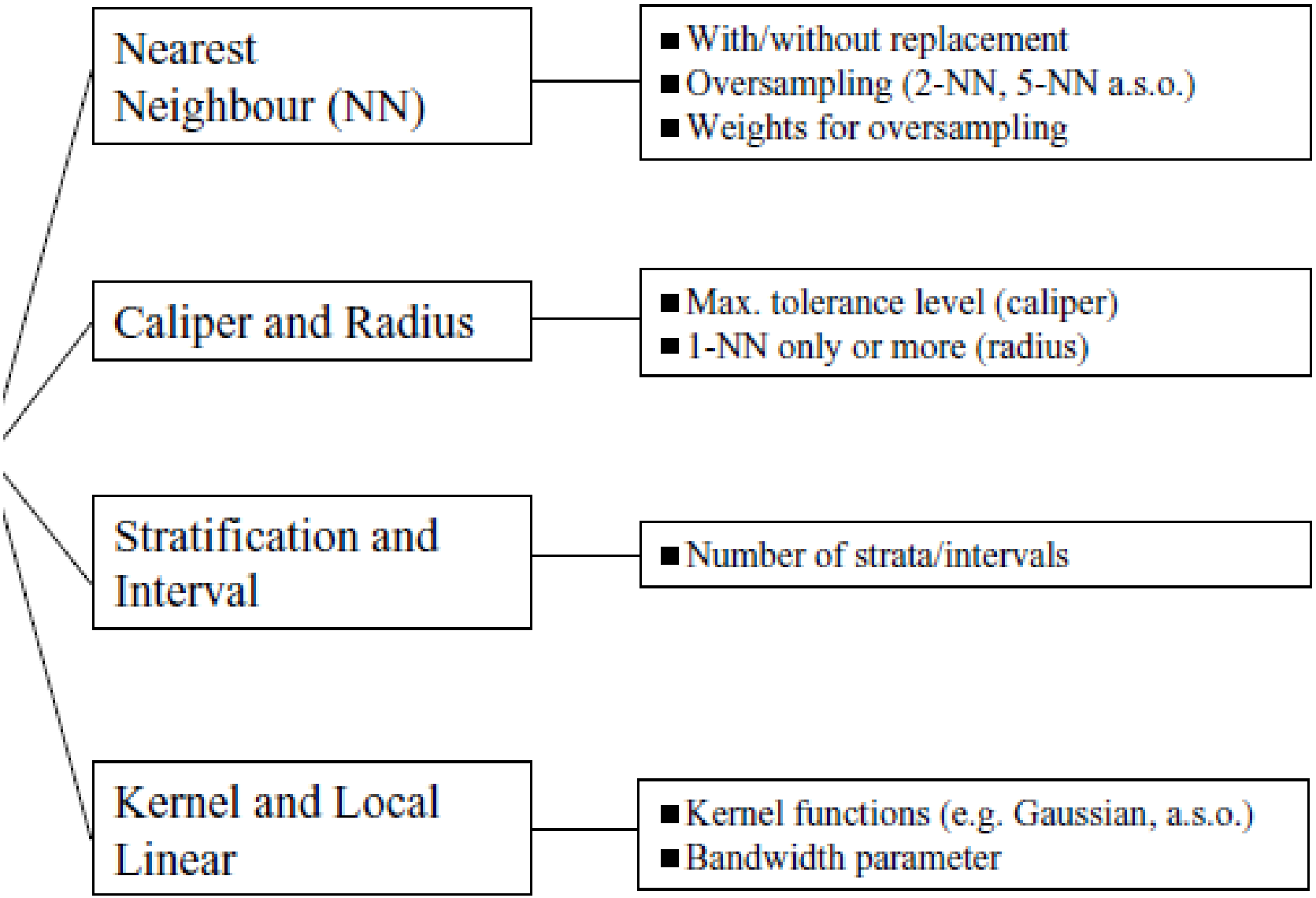
$$0 < Pr(D = 1|x) < 1$$

$$y_0, y_1 \perp D|p(x)$$

$$\tau_{ATT}^{PSM} =$$

$$E\{E[y_1|D = 1, p(x)] - E[y_0|D = 0, p(x)]\}$$

Matching Algorithms



Kernel Matching

$$ATET = \frac{1}{N_T} \sum_{i \in \{D=1\}} [y_{1,i} - \sum_j w(i,j) y_{0,j}]$$

$$w(i,j) = \frac{K(x_j - x_i)}{\sum_{j=1}^{N_{c,i}} K(x_j - x_i)}$$

Nearest-Neighbor and Radius Matching

$$A_j(x) = \{j | x_j \in c(x_i)\}$$

$$A_i(x) = \{j | min_j \parallel x_i - x_j \parallel\}$$

$$A_i(p(x)) = \{p_j | min_j \parallel p_i - p_j \parallel\}$$

$$A_i(p(x)) = \{p_j | \parallel p_i - p_j \parallel < r\}$$

Stratification or Interval Matching

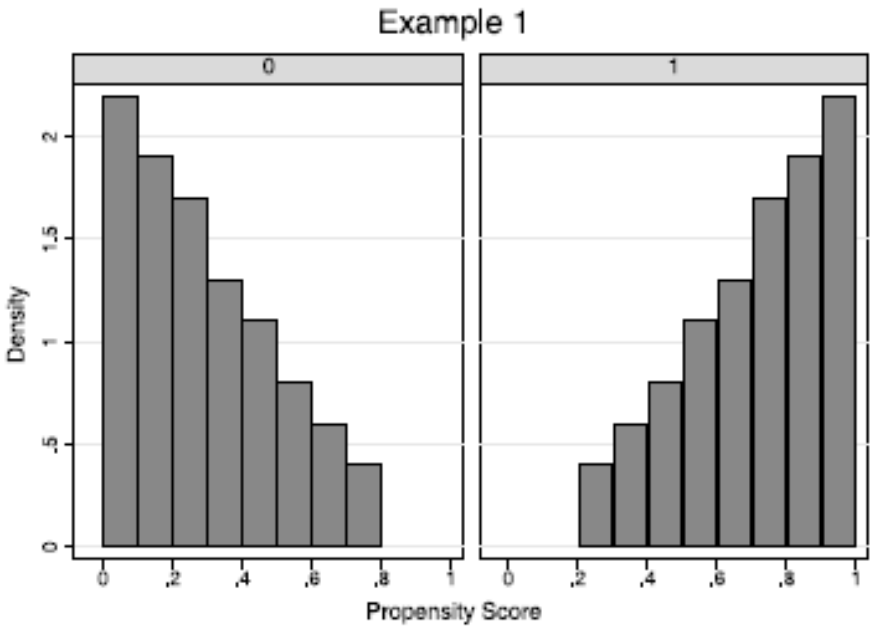
$$ATET_b^S = \frac{1}{N_b^T} \sum_{i \in I(b)} Y_{1i} - \frac{1}{N_b^C} \sum_{j \in I(b)} Y_{0i}$$

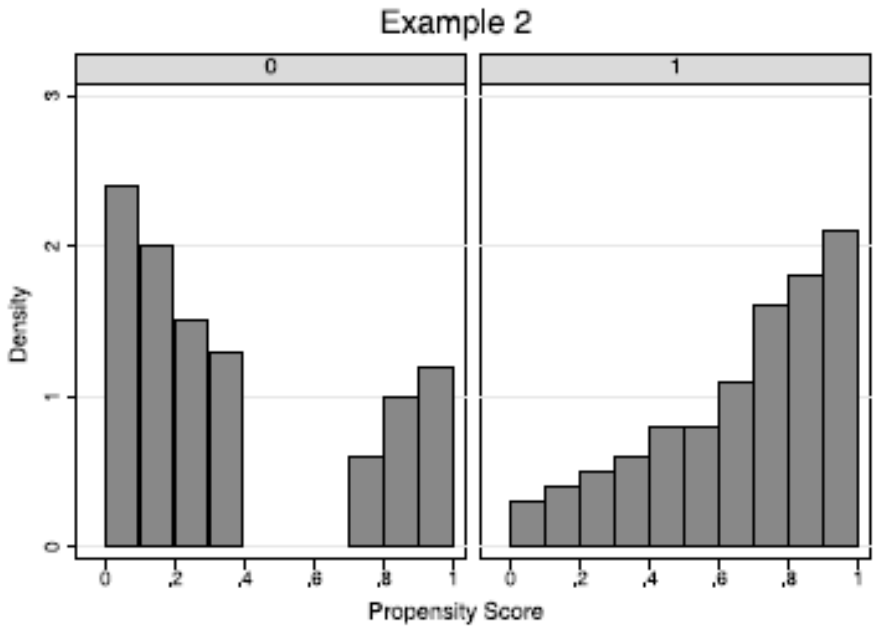
$$ATET^S = \sum_{b=1}^{B} ATET_b^S [\sum_{i \in I(b)} D_i / \Sigma D_i]$$

Trade-offs in Terms of Bias and Efficiency

Decision	Bias
Nearest neighbour matching:	
multiple neighbours/single neighbour	(+)/(−)
with caliper/without caliper	(−)/(+)
Use of control individuals:	
with replacement/without replacement	(−)/(+)
Choosing method:	
NN matching/Radius matching	(−)/(+)
KM or LLM/NN methods	(+)/(−)
Bandwidth choice with KM:	
small/large	(−)/(+)

Trimming the Common Support





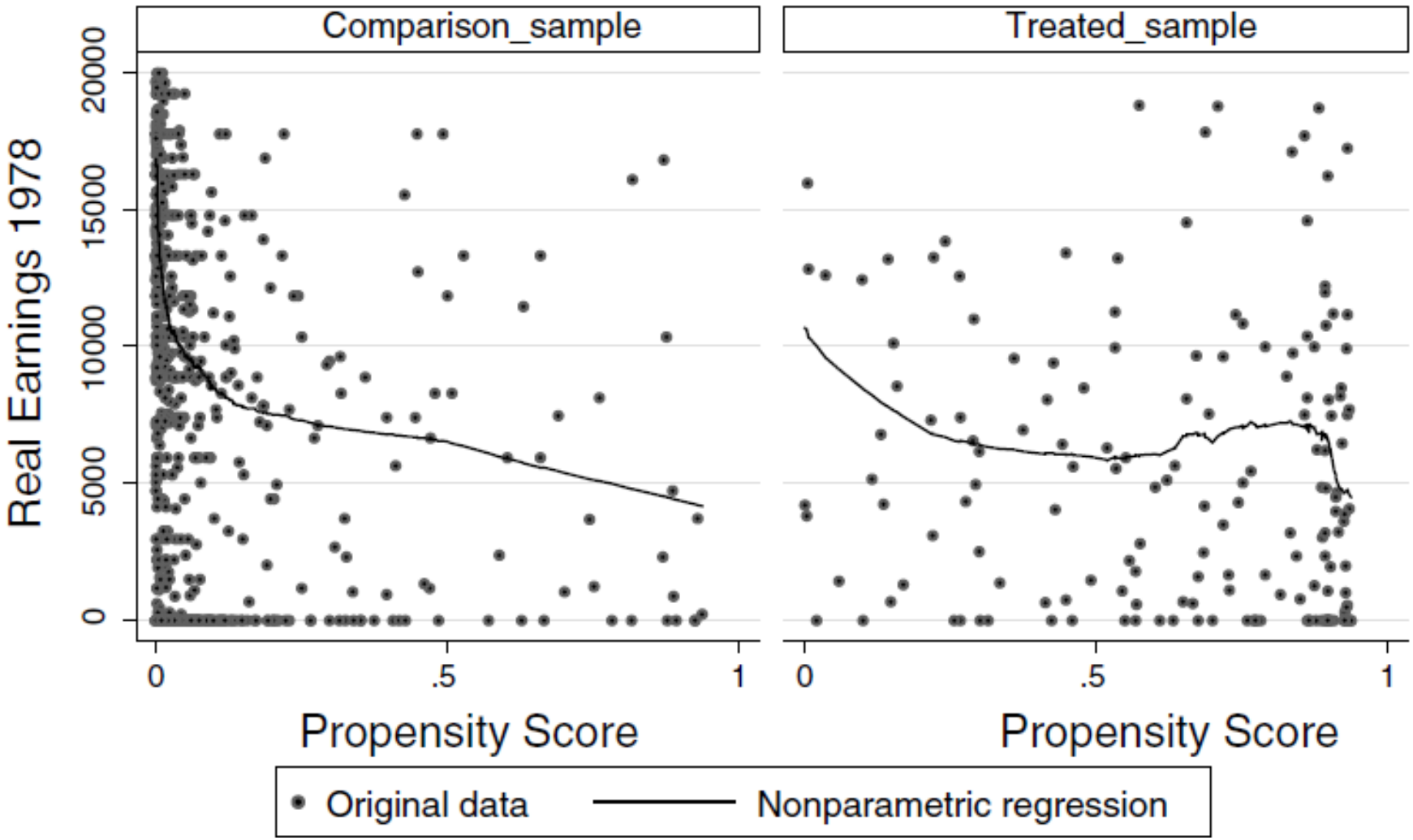
Dehejia and Wahba (1999)

Variable	Definition	Treated	Control
AGE	Age in years	25.82	34.85
EDUC	Education in years	10.35	12.12
NODEGREE	1 if EDUC < 12	0.71	0.31
BLACK	1 if race is black	0.84	0.25
HISP	1 if Hispanic	0.06	0.03
MARR	1 if married	0.19	0.87
U74	1 if unemployed in 1974	0.60	0.10
U75	1 if unemployed in 1975	0.71	0.09
RE74	Real earnings in 1974 (in 1982 $)	2,096	19,429
RE75	Real earnings in 1975 (in 1982 $)	1,532	19,063
RE78	Real earnings in 1978 (in 1982 $)	6,349	21,554
D	1 if received training (treatment)	1.00	0.00
Sample size		185	2,490

Training Impact

Method	Definition	Estimate	St. Error[a]
Treatment–control comparison	$\overline{\text{RE78}}_{D=1} - \overline{\text{RE78}}_{D=0}$	−15,205	656
Control function estimator	$\widehat{\alpha}$ from OLS regression (25.76)	218	768
Before–after comparison	$\overline{\text{RE78}}_{D=1} - \overline{\text{RE75}}_{D=1}$	4,817	625
Differences-in-differences	$\widehat{\alpha}$ from OLS regression (25.77)	2,326	749
Propensity score	See Section 25.8.4	995	–

Post-treatment Earnings against Propensity Score



Distribution of Propensity Score Using Dehejia and Wahba's (1999) Specification

Minimum $\widehat{p}(\mathbf{x})$	Treated	Untreated	Total
0.000364	9	960	969
0.10	10	56	66
0.20	14	33	47
0.40	24	22	46
0.60	33	7	40
0.80	95	8	103
Total	185	1086	1271

[a] From the second row, for example, the propensity score lies between 0.10 and 0.20 for 10 treated and 56 untreated individuals.

Training Impact: Estimates of ATET

Matching Procedure	Number Treated	Number in Control	ATET	Standard Error	% of $1794
Dehejia and Wahba (2002) specification[a]					
Nearest neighbor	185	53	2385	1209[c]	133
Radius, $r = 0.001$	54	517	−7815	1118[d]	−436
Radius, $r = 0.0001$	24	92	−9333	2282[d]	−520
Radius, $r = 0.00001$	15	19	−2200	2986[d]	−123
Stratification	185	1086	1452	1041[c]	81
Kernel	185	1058	1309	975[c]	73
Dehejia and Wahba (1999) specification[b]					
Nearest neighbor	185	57	560	1098[c]	31
Radius, $r = 0.001$	57	583	−9358	997[d]	−522
Radius, $r = 0.0001$	27	76	−7847	2066[d]	−437
Radius, $r = 0.00001$	16	13	223	4551[d]	12
Stratification	185	1146	2156	814[c]	120
Kernel	185	1146	1518	890[c]	85

[a] Logit Model: Pr[treat = 1] = h(CONSTANT, AGE, AGE^2, EDU, EDU^2, MARRIED, NODEGREE, BLACK, HISPANIC, RE74, RE74^2, RE75, U74, U75, U74*HISPANIC).

[b] Logit Model: Pr[treat = 1] = h(CONSTANT, AGE, AGE^2, EDU, EDU^2 MARRIED, NODEGREE, BLACK, HISPANIC, RE74, RE74^2, RE75, RE75^2, RE74*RE75, U74*BLACK).

[c] Bootstrapped standard errors with 200 replications.

[d] Analytical standard errors.