Implementing Propensity Score Matching Causal Analysis with Stata

Vivien W. Chen
Krissy Zeiser

Population Research Institute
Penn State University

2008/2/27

Estimate Causal Effect on Observational Data

- Causal Effect

  Let Y_i^1 = outcome after treatment, and Y_i^0 = outcome without treatment,

  Causal effect of unit *i*:

  $$Y_i = Y_i^1 - Y_i^0$$

  Estimated (or average) causal effect:

  $$\mathrm{E}(Y) = \mathrm{E}(Y^1 - Y^0) = \mathrm{E}(Y^1) - (Y^0)$$

- The Problem:

  It is impossible to *observe* individual treatment effect since we do not know the outcomes for untreated observations when it is under treatment, and for treated when it is not under treatment.

Counterfactual Inference

Group	Y^1	Y^0
Treatment (D=1)	observable	(counterfactual)
Control (D=0)	(counterfactual)	observable

Various effects of interest in the population of interest

- Average treatment effect

$$\mathrm{E}(Y) = \mathrm{E}(Y^1 - Y^0) = \mathrm{E}(Y^1) - (Y^0)$$

- Average treatment effect for the untreated

$$\mathrm{E}(Y^1 - Y^0 \mid D = 0) = \mathrm{E}(Y^1 \mid D = 0) - (Y^0 \mid D = 0)$$

- **Average treatment effect for the treated (ATT)**

$$\mathrm{E}\left(Y^1 - Y^0 \mid D = 1\right) = \mathrm{E}\left(Y^1 \mid D = 1\right) - \mathrm{E}\left(Y^0 \mid D = 1\right)$$

Challenges

- Selection bias: potential bias from treatment assignment/selection conditional on observed variables, due to the effects of unobserved variables, controlled with selection into treatment.

- Finite data: sample size reduces our ability to estimate causal effects by conditioning on observed variables.

Why Propensity Score?

- Adjust for (but not totally solve the problem of) selection bias.
- Minimizing the limitation from matching on many observed variables on finite data.
- Estimate counterfactual effects.

Propensity Score Matching

- Propensity score is the probability of taking treatment given a vector of observed variables.

$$p(x) = Pr[D=1|X=x]$$

- If we take individuals with the same propensity score, and divide them into two groups – those who were and weren't treated- the groups will be approximately balanced on the variables predicting the propensity score.

How it works

$\hat{p}$	Treatment	Control
0.9	$Y^1_{0.9}$	$Y^0_{0.9}$
0.7		
0.5		
0.3		
0.1		

Balancing property

- Among those with the same predicted probability of treatment ($\hat{p}$), those who get treated and not treated differ only on their error term in the propensity score equation. But this error term is approximately independent of the X's. The $\hat{p}$ treatment assignment *D* is independent of *Y*, given the strata created by X's. This is why balancing should occur.

$$Y \perp\!\!\!\perp D \mid X$$

Common support

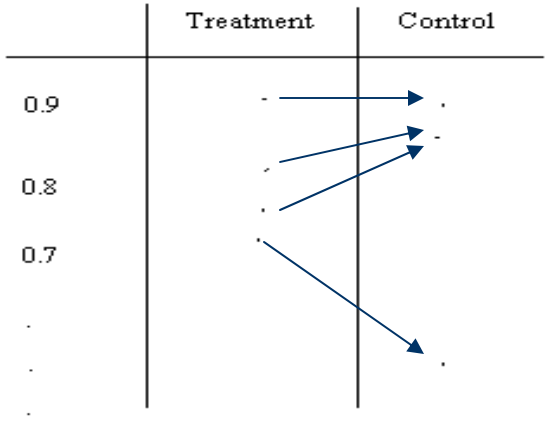
- Common support: the overlap condition for persons with the same x value in X are allowed to have a positive probability of being in treated and control groups.
- We only make inferences where we have sufficient data. Unlike ordinary regression, we don’t extrapolate outside the range of the observed data points.

Matching Algorithms

- Decide which matching algorithm to use and match sample
  - **Nearest Neighbor (NN)**
  - **Caliper & Radius**
  - **Stratification & interval**
  - **Kernal & Local Linear**

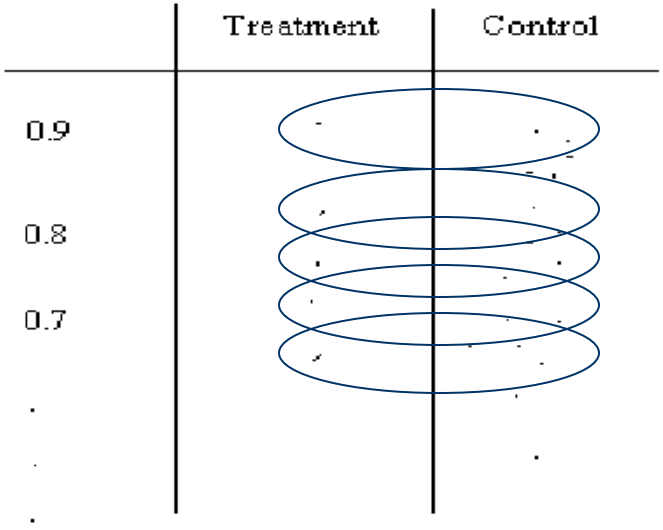
Nearest Neighbor (NN)

- A case in control group is matched to a treated case based on the closest propensity score.



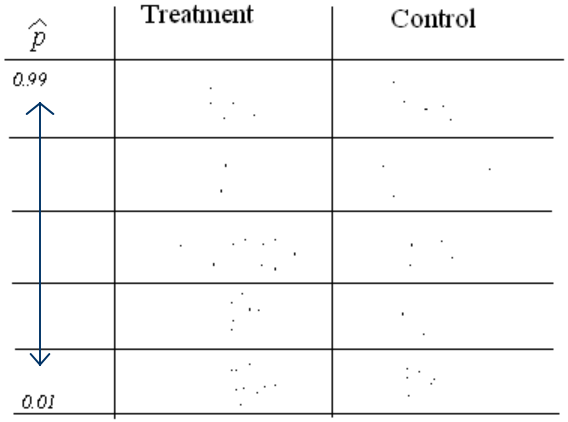
Caliper & Radius Matching

- Use a tolerance level on the maximum propensity score distance (caliper) to avoid the risk of bad matches; match with the NN within the caliper.
- The radius matching is to use not only use the closest NN within each caliper, but all the individuals in control group within the caliper.



Stratification & Interval Matching

- Use a set of interval (or strata) to divide the common support of propensity score, then match treatment and control cases within each interval/strata. Some researchers suggest 5 strata are enough to remove 95% of bias associated with covariates (Cochrane & Chambers 1965).
- The average treatment effect is then the mean of the interval-specific treatment effect, weighted by the number of cases in the treatment interval/strata.



Kernel Matching

- Use weighted averages of all cases in the control group to estimate counterfactual outcomes. The weight is calculated by the propensity score distance between a treatment case and all control cases.
- The closest control cases are given the greatest weight.

(For more detail, see Heckman, Ichimura, & Todd 1998; Smith & Todd 2005)

To Achieve A Matching Equation

- **pscore** lwgrand nsib male ethmin ethhuk ethmain ethoth edexp liked_s liked_t man lwgp7 lw2par paed wkmom twkid sc sick conflict accept strict senio prefboy talk paexp w2paexp schpar comsoc fin1- misfinc age1112 age13 age1415 misage famsup undemo, **pscore(mypscore) comsup numblo (5) level (0.001) blockid (myblock) logit**

You May Get the Unmatched Result:

Results

```
**************************************************************************
Step 2: Test of balancing property of the propensity score
Use option detail if you want more detailed output
**************************************************************************

Variable edexp is not balanced in block 6

Variable paexp is not balanced in block 6

Variable w2paexp is not balanced in block 6

The balancing property is not satisfied

Try a different specification of the propensity score
```

Inferior of block of pscore	live w/ grandparent 0	1	Total
0	2,183	191	2,374
.1	1,951	261	2,212
.15	423	91	514
.2	257	95	352
.4	107	120	227
.6	256	582	838
.8	30	136	166
Total	5,207	1,476	6,683

```
Note: the common support option has been selected
```

Then You Have to Adjust Your Covariates to Get A Matched Equation

Results

```
**********************************************************************
Step 2: Test of balancing property of the propensity score
Use option detail if you want more detailed output
**********************************************************************

The balancing property is satisfied

This table shows the inferior bound, the number of treated
and the number of controls for each block
```

Inferior of block of pscore	live w/ grandparent 0	1	Total
0	4,717	849	5,566
.1	2,999	391	3,390
.15	586	122	708
.2	321	119	440
.4	186	252	438
.6	400	816	1,216
.8	30	126	156
Total	9,239	2,675	11,914

```
**********************************************************************
End of the algorithm to estimate the pscore
**********************************************************************
```

To Create a Matched Sample & Average Treatment Effect for the Treated (ATT) in Stata

- **Nearest Neighbor: “attnd” or “attnw”**

  attnd all3p *lwgrand* nsib male ethmin ethhuk ethmain ethoth liked_s liked_t man lwgp7 paed wkmom twkid sc sick conflict accept strict senio prefboy talk schpar comsoc fin1- misfinc age1112 age13 age1415 misage famsup undemo**, comsup bootreps(1000) dots logit**

ATT Result from NN Matching:

Results

```
. h pscore

. h attnw,

. h attnd

. h attnw

. use paa3

. attnd all3p lwgrand  nsib male ethmin ethhuk ethmain ethoth liked_s liked_t man lwgp7 pae
> c sick conflict accept strict senio prefboy talk schpar comsoc fin1- misfinc age1112 age1
> e famsup undemo, comsup bootreps(1000) dots logit


 The program is searching the nearest neighbor of each treated unit.
 This operation may take a while.


ATT estimation with Nearest Neighbor Matching method
(random draw version)
Analytical standard errors

------------------------------------------------------------------------------
n. treat.    n. contr.         ATT     Std. Err.           t
------------------------------------------------------------------------------
     2102         1269       0.210         0.038       5.522
------------------------------------------------------------------------------
Note: the numbers of treated and controls refer to actual
nearest neighbour matches
```

To Create a Matched Sample & Average Treatment Effect for the Treated (ATT) in Stata (continued)

- Caliper/radius matching

  attr all3p *lwgrand* nsib male ethmin ethhuk ethmain ethoth liked_s liked_t man lwgp7 paed wkmom twkid sc sick conflict accept strict senio prefboy talk schpar comsoc fin1- misfinc age1112 age13 age1415 misage famsup undemo**, comsup bootreps(1000) dots logit radius(0.005)**

- Stratification/interval matching

  atts all3p *lwgrand* nsib male ethmin ethhuk ethmain ethoth liked_s liked_t man lwgp7 paed wkmom twkid sc sick conflict accept strict senio prefboy talk schpar comsoc fin1- misfinc age1112 age13 age1415 misage famsup undemo**, pscore (**psout**) blockid(**myblock**) comsup bootreps(1000) dots**

To Create a Matched Sample & Average Treatment Effect for the Treated (ATT) in Stata (continued)

- Kernel matching

  attk all3p *lwgrand* nsib male ethmin ethhuk ethmain ethoth liked_s liked_t man lwgp7 paed wkmom twkid sc sick conflict accept strict senio prefboy talk schpar comsoc fin1-misfinc age1112 age13 age1415 misage famsup undemo**, comsup bootreps(1000) dots logit**

Sensitivity Analysis

- Test robustness & unmeasured bias

  psmatch2 *lwgrand* nsib male ethmin ethhuk ethmain ethoth liked_s liked_t man lwgp7 paed wkmom twkid sc sick conflict accept strict senio prefboy talk schpar comsoc fin1- misfinc age1112 age13 age1415 misage famsup undemo**,** **outcome (**all3p**) noreplace logit**

- gen **difscore** = all3p - _all3p if _treat==1& _support==1

  * difscore is the difference in treatment effect between treated and untreated

- **rbounds** difscore, gamma (1 (0.1)2)

  * Rosenbaum bounds takes the difference in the response variable between treatment and control cases as difscore.

  * gamma: log odds of differential assignment due to unobserved heterogeneity.

Optional Stata syntax to diagnose the matched outcomes:

- "**pstest**" calculates several measures of the balancing of the variables in varlist before and after matching.
- "**psgraph**" graphs the propensity score histogram by treatment status.
- **bootstrap r(att) : psmatch2** treatment varlist**, out**(outvar)

Predicting Interscholastic Sports Participation

Treatment = interscholastic1

Observed covariates for matching:

non-athletic participation
intramural participation
Race (white ref.)
- Asian
- black
- Hispanic
- multiracial
- American Indian

SES
Family Structure (two-parent ref.)
- cohabiting parents
- single parent

father's education
mother unemployed
father unemployed
English is a second language
student employed
participates in non-school sport
GPA-10th grade
ever held back
Academic Track
- college prep
- vocational

reading test score

School-level variables
% minority
% free lunch
% deviance problems
school size
minimum GPA requirement
Urbanity (suburban ref.)
- urban
- rural

% lep
% college prep
% sports available
Educational expectations * 10th

Calculating Propensity Scores

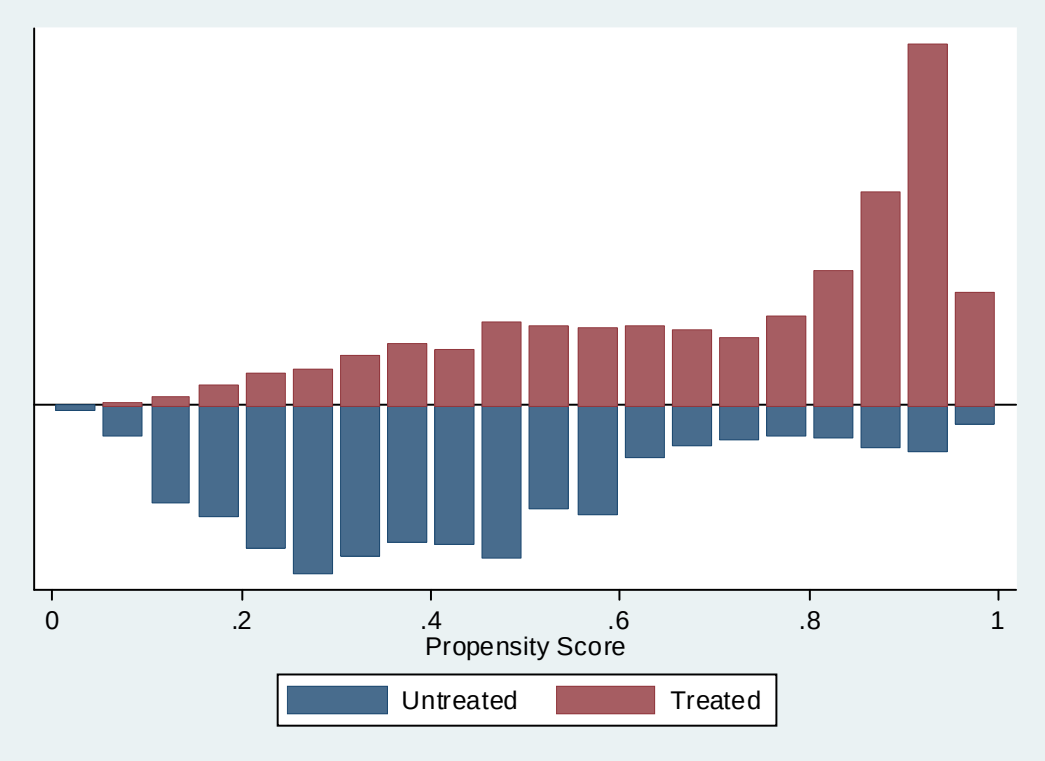
pscore interscholastic1 gpa10 nonathleticc ///
intramuralcc asian black hispanic multiracial ///
amindian ses cohabitparent singleparent ///
otherparent mothered fathered munemployed ///
funemployed secondlang heldback collegeprep ///
vocational readingtest mathtest edex10 employed ///
nonss pminority pfreelunch pdanger schoolsize ///
mingpa urban rural plep pcollegeprep mpavailact ///
edex_gpa, pscore(propensity) blockid(strat) detail ///
logit comsup

Distribution of Propensity Scores by Stratum and Treatment Status (Males)

	Control			Treatment		
Stratum	N	Mean	Std. Dev.	N	Mean	St. Dev
1	83	.1486	.029	18	.1708	.023
2	133	.2551	.029	41	.2516	.028
3	131	.3492	.031	77	.3562	.030
4	138	.4491	.031	102	.4559	.027
5	98	.5510	.028	116	.5503	.030
6	72	.6855	.058	223	.7010	.060
7	70	.8900	045	603	.9029	.044

Histograms of Propensity Scores

psgraph, treated (interscholastic1) pscore(propensity)



Finding the ATT: Stratification

atts gpa12 interscholastic1 nonathleticc intramuralcc asian black hispanic multiracial amindian ses ///
cohabitparent singleparent otherparent mothered ///
fathered munemployed funemployed secondlang ///
gpa10 heldback collegeprep vocational ///
readingtest mathtest edex10 employed nonss ///
pminority pfreelunch pdanger schoolsize mingpa ///
urban rural plep pcollegeprep fpavailact edex_gpa ///
if missingv1 == 0, pscore(propensity) blockid(strat)

Finding the ATT: Nearest-Neighbor

attnd gpa12 interschlastic1 nonathleticc intramuralcc ///
asian black hispanic multiracial amindian ses ///
cohabitparent singleparent otherparent mothered ///
fathered munemployed funemployed secondlang ///
gpa10 heldback collegeprep vocational ///
readingtest mathtest edex10 employed nonss ///
pminority pfreelunch pdanger schoolsize mingpa ///
urban rural plep pcollegeprep fpavailact edex_gpa ///
if missingv1 == 0, pscore(propensity)

Finding the ATT: Kernel Matching

attk gpa12 interscholastic1 interscholastic1 gpa10 ///
nonathleticc intramuralcc asian black hispanic ///
multiracial amindian ses cohabitparent ///
singleparent otherparent mothered fathered ///
munemployed funemployed secondlang heldback ///
collegeprep vocational readingtest mathtest ///
edex10 employed nonss pminority pfreelunch ///
pdanger schoolsize mingpa urban rural plep ///
pcollegeprep mpavailact edex_gpa if ///
missingv1== 0, pscore(propensity) **boot reps(200)**

Results: Predicting 12th Grade GPA

Method	N treatment	N control	ATT	Std. error	T
Nearest-neighbor	1260	348	.040	.082	.487
Stratification	1260	854	.051	.058	.876
Kernel matching	1260	848	.049	.052	.959

Results: OLS Regression: Predicting 12th Grade GPA (sig. results only)

	Model 1	Model 2	Model 3
Interscholastic Participation	.157**	.048	.046
Non-athletic participation	.120**	.037**	.033**
Intramural participation	-.107**	-.001	.001
Family Structure (two-parent ref.)			
cohabiting parents	-.247**	-.150**	-.151**
single parent	-.229**	-.103**	-.102**
other household structure	-.334**	-.215**	-.220**
Participates in non-school sport	-.044**	-.033**	-.030*
Constant	2.851**	1.309**	1.528**
Observations	1909	1909	1909
R-squared	.17	.45	.46

Propensity Score in Other Statistical packages

- SAS
- S-Plus
- R
- SPSS
- MPlus (Complier Average Causal Effect, *cace*)

(See the handout for how to implement propensity score matching in SAS and R)

References

- Rosenbaum, P. & D. Rubin. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70:41-55.
- Heckman et al. 1998 [to check!]
- Morgan, S. & C. Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Science*. NY: Cambridge University Press.
- Harding, D. 2003. Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping Out and Teenage Pregnancy. *The American Journal of Sociology* 109(3):676-719.
- Hong, G. & S. Raudenbush. 2005. Effect of Kindergarten Retention Policy on Children's Cognitive Growth inReading and Matheatics. *Educational Evaluation and Policy Analysis*. 27(3): 205-224
- More on references, check:

http://help.pop.psu.edu/help-by-statistical-method