

ECON 717A: Problem Set 2

Alex von Hafften

March 6, 2022

1 Write-Up

Problem 0

I drop observations with `sample` equal to 3. This step dropped 2,490 observations.

Problem 1

I regressed earnings in 1978 on treatment with and without the covariate of age, age squared, education, indicators for black, Hispanic, married, and no degree, and earnings in 1974 and 1975. The treatment effect is \$886.30 without covariates and \$818.70 with covariates with both statistically significant at the 10 percent level. It is important to include covariate even in experimental data because we get a more precise estimate for the treatment effect.

VARIABLES	(1) re78	(2) re78
treated	886.3* (488.1)	818.7* (487.8)
age		-145.9 (200.8)
age_2		2.799 (3.246)
educ		206.8 (165.5)
black		-1,461** (734.3)
hisp		100.5 (958.6)
married		133.9 (660.0)
nodegree		-405.9 (752.1)
re74		0.0871 (0.106)
re75		0.0840 (0.119)
Constant	5,090*** (277.4)	5,649 (3,757)
Observations	722	722
R-squared	0.005	0.045

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Problem 2

I drop observations with `sample` equal to 1 and `treated` equal to 1. This step dropped 297 observations.

Problem 3

I define `in_control` equal to one if `sample` equals one and zero otherwise. The probit estimation for the propensity scores are the coarse scores and the rich scores are below.

VARIABLES	(1) in_control	(2) in_control
age	0.253*** (0.0293)	0.322*** (0.0316)
age_2	-0.00453*** (0.000493)	-0.00548*** (0.000530)
educ	0.0169 (0.0181)	0.0178 (0.0183)
black	1.990*** (0.0778)	1.950*** (0.0796)
hisp	0.973*** (0.103)	0.978*** (0.106)
married	-1.101*** (0.0826)	-0.909*** (0.0869)
nodegree	1.133*** (0.100)	1.071*** (0.104)
re74		-1.07e-06 (8.60e-06)
re75		-5.76e-05*** (9.56e-06)
Constant	-6.358*** (0.483)	-7.108*** (0.509)
Observations	16,417	16,417
Comparison group obs. completely determined	727	1359
Control group obs. completely determined	0	0

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The completely determined observations are 727 and 1359 comparison group observations that have propensity scores with almost zero.

Problem 4

The table below shows descriptive statistics of `pscorea` and `pscoreb`.

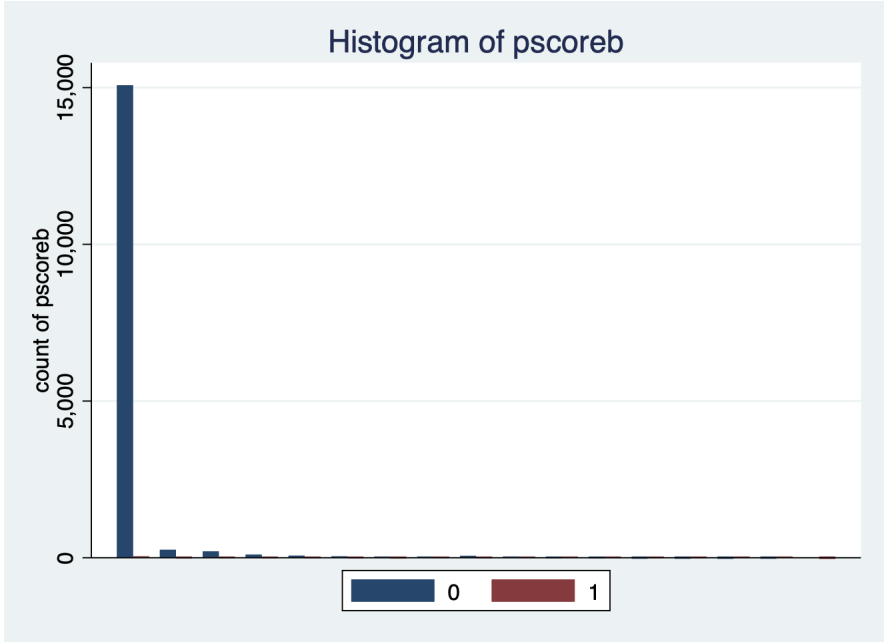
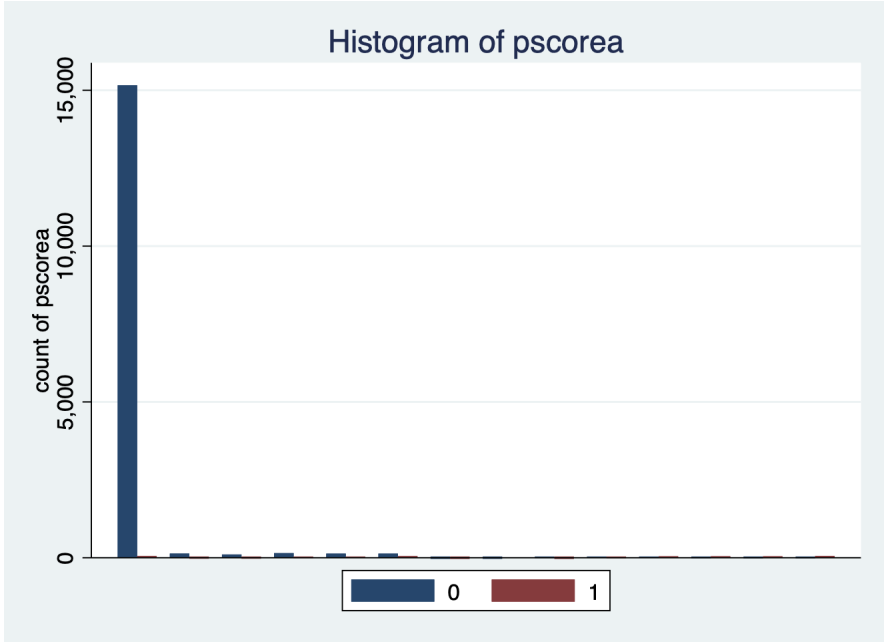
	Mean	SD	Min	Median	Max	N
0	0.02	0.07	0.00	0.00	0.69	15,992
1	0.39	0.23	0.00	0.47	0.69	425
Total	0.03	0.09	0.00	0.00	0.69	16,417

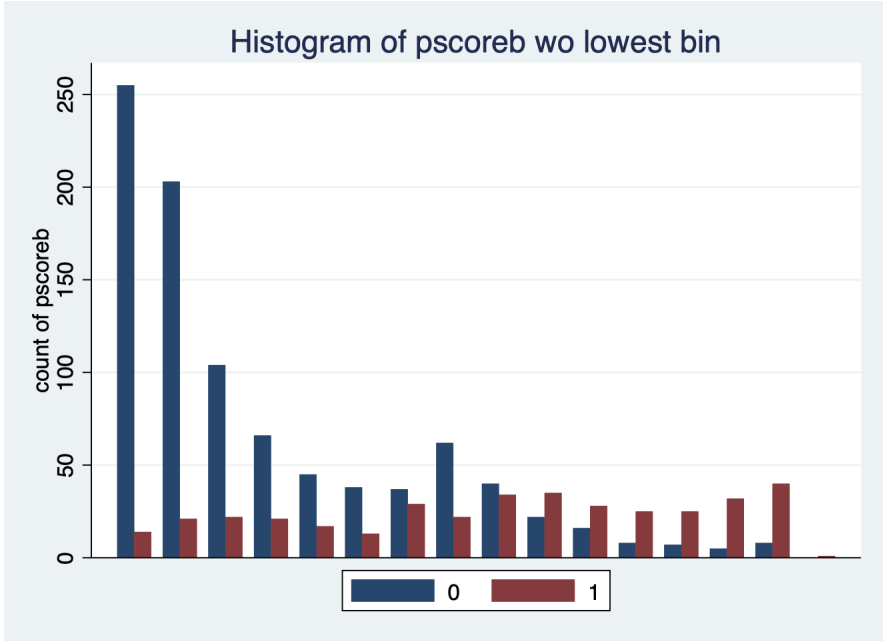
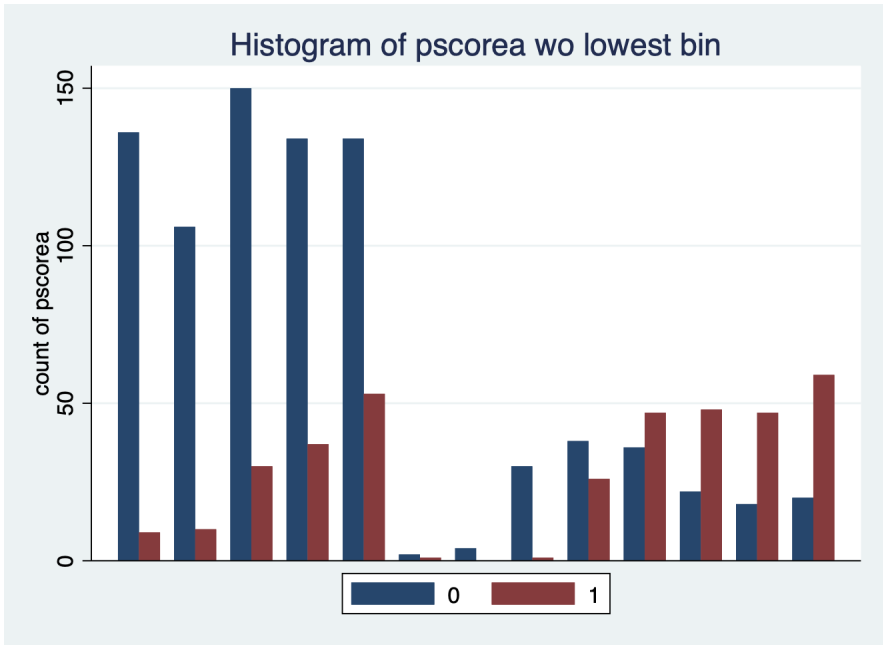
	Mean	SD	Min	Median	Max	N
0	0.02	0.06	0.00	0.00	0.79	15,992
1	0.42	0.25	0.00	0.46	0.80	425
Total	0.03	0.10	0.00	0.00	0.80	16,417

These descriptive statistics suggest that imposing the common support condition will not drop many observations. For `pscorea`, the minimum is zero for both groups and the maximum is 0.69 for both groups. This suggests that there are compared observations within the comparison group to the control group. Similarly for `pscoreb`, the minimum is zero for both groups and the maximum is around 0.8 for both groups. These descriptive statistics suggest that the CPS group is not very comparable to the experimental control group. That is, these descriptive statistics highlight the importance of matching because the control group has a much higher mean propensity score than the comparison group (i.e., 0.39 vs. 0.02 for `pscorea` and 0.42 vs. 0.02 for `pscoreb`).

Problem 5

Below are histograms (based on the provided code) for `pscorea` and `pscoreb` across the CPS comparison group (i.e., 0) and the experimental control group (i.e., 1) with and without the lowest bin (i.e., `pscorex` $\in (0.0, 0.05)$). Looking at the first two histograms, most of the comparison group observations have a very low estimated propensity score (confirming the descriptive statistics), so the comparison group is not very comparable to the control group. I added histograms without the lowest bin because many observations in the comparison group have estimated propensity scores of basically zero impairing our ability to evaluate the common support condition. Looking at the second two histograms, the common support condition seems to be well satisfied with comparison group observations at all propensity levels of the control group. These histograms highlight concerns about matching without replacement because the number of observations in the control group with high propensity scores is larger than the number of observations in the comparison group with large propensity scores (i.e. on the right of the graphs the red bars are taller than the blue bars). Matching without replacement would cause these the control group observations to be matched with comparison group observations with significantly lower propensity scores.





Problem 6

Using single nearest neighbor matching without replacement and common support condition, I estimate the non-experimental bias reported in the table below. The non-experimental bias is significant and negative indicating that even after matching observations in the control group that significantly lower earnings in 1978. The non-experimental bias is lower when matching based on the rich propensity scores. No observations are dropped when matching with coarse propensity scores. Seven observations are dropped when matching with fine propensity scores.

	Unmatched	ATT for <code>pscorea</code>	ATT for <code>pscoreb</code>
Difference	-9756.610000000001	-4439.07	-2340.78
SE	470.16	486.48	449.37

Problem 7

Using single nearest neighbor matching with replacement and common support condition, I estimate the non-experimental bias reported in the table below. The non-experimental bias is still significant and negative but it is lower than without replacement. As mentioned in problem 5, this reduction in the non-experimental bias is due the relatively fewer comparison group observations with high propensity scores, so that there's more non-experimental bias when matching without replacement.

	Unmatched	ATT for <code>pscorea</code>	ATT for <code>pscoreb</code>
Difference	-9756.610000000001	-3677.03	-1515.99
SE	470.16	934.5	707.62

Problem 8

	<code>re74</code>	<code>re75</code>
Raw Data	1.26	1.41
Matched by <code>pscoreb</code> using Single Nearest Neighbor	1.16	0.98

Problem 9

	Bandwidth = 0.02	Bandwidth = 0.2	Bandwidth = 2.0
Difference	-2349.09	-7043.43	-9772.130000000001
SE	658.53	337.06	289.25

Problem 10

	Bandwidth = 0.02	Bandwidth = 0.2	Bandwidth = 2.0
Difference	-1980.88	-2421.42	993.88
SE	707.62	707.62	707.62

Problem 11

VARIABLES	(1) re78
in_control	-1,853*** (343.2)
age	-235.5*** (40.08)
age_2	1.858*** (0.550)
educ	163.3*** (28.53)
black	-832.4*** (193.9)
hisp	-114.2 (213.9)
married	199.4 (150.8)
nodegree	296.6* (174.3)
re74	0.291*** (0.0152)
re75	0.471*** (0.0153)
Constant	7,757*** (726.7)
Observations	16,417
R-squared	0.483

Robust standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

	Mean	SD	Min	Median	Max	N
0	14,846.66	6,608.11	241.16	15,369.08	25,559.22	15,992
1	5,090.05	4,227.39	-1,476.13	3,915.22	31,078.45	425
Total	14,594.08	6,737.87	-1,476.13	15,111.14	31,078.45	16,417

Problem 12

VARIABLES	(1) re78
age	-252.0*** (40.92)
age_2	2.041*** (0.561)
educ	166.6*** (28.79)
black	-773.9*** (199.6)
hisp	-168.2 (218.9)
married	244.1 (153.1)
nodegree	330.7* (176.9)
re74	0.299*** (0.0152)
re75	0.470*** (0.0153)
Constant	7,908*** (739.7)
Observations	15,992
R-squared	0.476
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

	Mean	SD	Min	Median	Max	N
0	14,846.66	6,654.46	162.96	15,370.95	25,693.43	15,992
1	6,941.60	4,269.45	275.72	5,770.59	33,124.62	425
Total	14,642.01	6,721.77	162.96	15,110.59	33,124.62	16,417

Problem 13