Applied Metrics PS2: Matching and Weighting

Emily Case

March 8, 2022

Question 1. Regressing 1978 real earnings on the (original) treated variable yields an estimate that those who received the experiment treatment earn $818.70 more than those who weren't treated, controlling for other demographic information. The experimental effect might be different for different kinds of people, so we want to include covariates to improve the estimated experimental impact.

Question 2. Drop the experimental treatment group.

Question 3. I define `treated2` as 0 when an observation is in the CPS group, and 1 when an observation is in the experiment's control group. Running probits with `treated2` as the dependent variable on both the coarse and rich set of independent variables yields the results in Table 2.

Some combinations of covariates perfectly determine if someone is in the experimental group or not. None of the observations in the experimental sample are completely determined. There are 727 observations that are *not* in the experimental sample (i.e. they are in the CPS sample) that are completely determined using the coarse variables, and 1359 using the rich set of variables. This is important because the observations that are completely determined, we know there is no probability that they would be part of the *other* group, which would be important for propensity score matching. They are not helpful to us when we match.

Table 1: Question 1

	1978 real earnings
treated	818.70
	(487.83)
age	-145.92
	(200.76)
age2	2.80
	(3.25)
educ	206.81
	(165.45)
black	-1461.26
	(734.32)
hisp	100.48
	(958.57)
married	133.91
	(660.02)
nodegree	-405.91
	(752.08)
re74	0.09
	(0.11)
re75	0.08
	(0.12)
_cons	5648.81
	(3757.50)
N	722

Standard errors in parentheses

Table 2: Probit models

	Coarse	Rich
treated2		
age	0.2532	0.3225
	(0.03)	(0.03)
age2	-0.0045	-0.0055
	(0.00)	(0.00)
educ	0.0169	0.0178
	(0.02)	(0.02)
black	1.9899	1.9504
	(0.08)	(0.08)
hisp	0.9733	0.9775
	(0.10)	(0.11)
married	-1.1011	-0.9091
	(0.08)	(0.09)
nodegree	1.1327	1.0712
	(0.10)	(0.10)
re74		-0.0000
		(0.00)
re75		-0.0001
		(0.00)
_cons	-6.3580	-7.1081
	(0.48)	(0.51)
N	16417	16417

Standard errors in parentheses

Question 4. The descriptive statistics of the estimated propensity scores by group suggest that we do not anticipate having an issue with the common support for the coarse scores. We might need to impose the common support condition for the rich scores, since they have different maxima across groups.

Table 3: Looking at the common supports

	Minimum	Mean	Maximum
coarse scores			
CPS group	0	.01641	.6872
control group	.00008	.3873	.6872
rich scores			
CPS group	0	.01545	.7886000000000001
control group	0	.42483	.8024

The CPS comparison group is pretty different from the control group. Despite covering more or less the same support, notice that the mean propensity scores for the experimental control group and the CPS group are very different. The CPS group propensity scores are much more highly concentrated around 0.

Question 5. The histograms in Figures 1 (on page 5) and 2 (on page 5) suggest that there is not a big common support issue, but that there are underlying differences in the control group sample and the CPS sample, similarly to question 4.

Visually, it looks like the histogram for the rich scores, Figure 2, has slightly more similarities between the experimental control group and the CPS group.

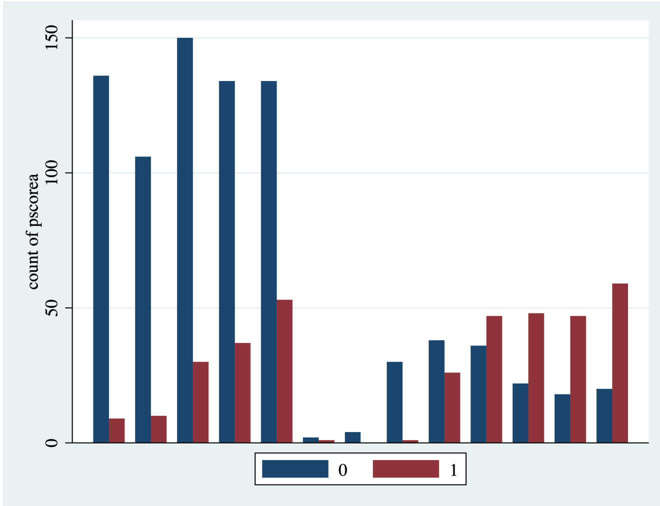

Figure 1: Counting observations of control group and CPS group in propensity score bins, using the coarse propensity scores. I omit observations in the first bin, in order to make the other bins visible.

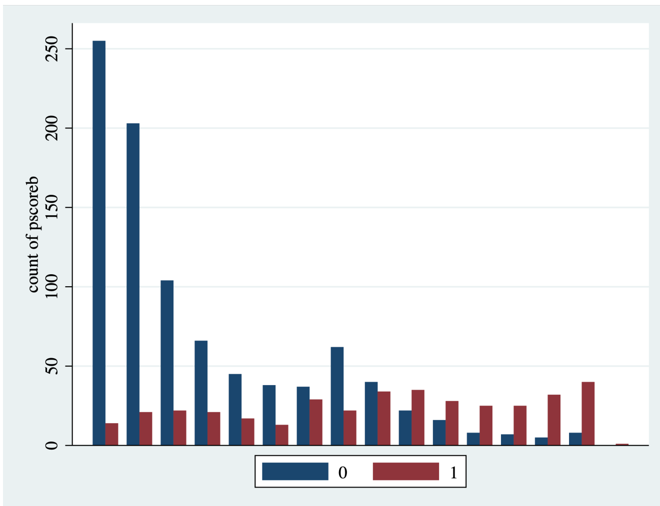

Figure 2: Counting observations of control group and CPS group in propensity score bins, using the rich propensity scores. I omit observations in the first bin, in order to make the other bins visible.

Question 6. Impose common support condition, no replacement.
No observations were dropped using the coarse scores, and 7 were dropped using the rich scores. The observations off the support (for rich scores) have propensity scores that are above 0.7886, which is the minimum of 0.7886 and 0.8024 (the two maxima propensity scores).

The experimental impact from Table 1 was \$818.70. The results for coarse and rich scores are below, in Table 4. Using coarse scores, the experimental group is estimated to have \$4,439 less real earnings in 1978 compared to the CPS group. Using rich scores, they are estimated to have \$2,340 less than the CPS group. We want the differences to be close to zero... both these scores are clearly very far away from zero.

The rich scores do better by about \$2,000, but again, still pretty far away from zero.

Table 4: Using `psmatch2` (questions 6 and 7)

	No replacement		**Replacement**	
	Difference	S.E.	Difference	S.E.
coarse scores				
Unmatched	-9756.6	470.2	-9756.6	470.2
ATT	-4439.1	486.5	-3677	934.5
rich scores				
Unmatched	-9756.6	470.2	-9756.6	470.2
ATT	-2340.8	449.4	-1516	707.6

Question 7. Repeat question 6 but with replacement.
When we allow for nearest neighbors to be reused for matching, the estimated differences for coarse and rich scores (in Column 2 of Table 4) decrease compared to when we did not allow for replacement. They are still pretty far off from zero, but we are closer to zero as a result of more flexibility allowing better matches.

This makes sense when you think about the high concentration of CPS observations with propensity scores close to zero. Without replacement, when we want to match the control group scores with *high* propensity scores, we run out of CPS observations with high propensity scores, and have to make worse matches. When we allow replacement, it gets a little bit better.

Question 8. The standardized differences for raw data and for the rich scores

Table 5: Standardized differences (question 8)

	Raw data	Single NN with replacement	Reduced bias by
`re74`	126.32	-9.1	92.83%
`re75`	141.35	-7.34	94.8%

using single nearest neighbor matching with replacement are in Table 5 on page 7. When we match using the rich scores, the bias is reduced significantly (see the third column).

Question 9 and 10. Results for the Gaussian kernel and local linear matching are in Table 6 on page 8. Compared to question 6 and 7, using a Gaussian kernel generally worsens the difference estimates, while using local linear matching improves them.

In Gaussian kernels, the bandwidth of 0.02 is the best in the sense that it is close to zero, but we also do notice the trade off of a higher standard error as a result. As bandwidth increases, the difference estimates get larger in magnitude. Here we want a smaller bandwidth.

Using local linear matching, in absolute value the bandwidth of 2.0 is closest to zero, and performs better than any of the difference estimates so far.

Table 6: Using `psmatch2` (questions 9 and 10)

	Q9: Kernel		**Q10: LLR**	
	Difference	S.E.	Difference	S.E.
Unmatched	-9756.61	470.16	-	-
ATT ban = 0.02	-2349.09	658.53	-1980.88	707.62
ATT ban = 0.2	-7043.43	337.06	-2421.42	707.62
ATT ban = 2.0	-9772.13	289.25	993.88	707.62

Question 11 and 12. Results for 11 and 12 are in Table 7 on page 9. Running a linear regression of real earnings in 1978 using all of the observations in the control group and the CPS comparison group in column 1 shows an estimated bias of -\$1,853.39, which is higher than our closest efforts from matching.

If instead we use only the untreated (CPS) observations, I get a very similar estimated bias of -\$.

Question 13. Using the formulas for the ATT using inverse probability weighting on page 19 of the notes, I get an average treatment effect on the treated of -\$1420.38 for unscaled weights and -\$1996.47 using scaled weights . Normalizing the weights to sum to one in the sample is the "preferred" version.

Table 7: Using regressions instead of matching (questions 11 and 12)

	Pooled	CPS only
treated2	-1853.39	
	(383.07)	
age	-235.53	-252.05
	(40.54)	(41.43)
age2	1.86	2.04
	(0.55)	(0.56)
educ	163.31	166.60
	(28.42)	(28.71)
black	-832.38	-773.88
	(208.66)	(215.05)
hisp	-114.15	-168.23
	(215.29)	(219.47)
married	199.44	244.14
	(148.05)	(150.49)
nodegree	296.58	330.68
	(176.62)	(179.40)
re74	0.29	0.30
	(0.01)	(0.01)
re75	0.47	0.47
	(0.01)	(0.01)
_cons	7757.18	7908.36
	(726.89)	(740.47)
N	16417	15992

Standard errors in parentheses