ECON 672 Winter 2022 Problem Set #2

Jon Holder

February 3, 2022

Question 1

The log file shows that the data files is read into Stata. Table 1 shows the summary statistics in the data set.

	Mean	SD	Min	Max	Num Obs
Sex	.4568225	.4981443	0	1	20601
Age	29.30702	10.52068	16	82	20601
Race	1.646037	.8114741	1	4	20601
Earnings 18 months later	9513.183	9546.872	0	93951	14192
Earnings at initial survey	2810.803	4050.78	0	64436	17043
Years of educations at initial survey	11.42821	1.776717	7	18	20254
Childer under 18yr at initial survey	1.17241	1.438531	0	20	18659

Table 1: *Summary Statistics*

Question 2

See log file.

Question 3

See log file.

Question 4

Table 2 model 1 shows estimates of the impact of the experimental assignment to the treatment group without covariates. OLS estimates that the treatment group resulted in $562.40 increase in average earnings. The estimate is statistically significant at the 1% level.

Question 5

Table 2 model 2 shows estimates of the impact of the experimental assignment to the treatment group with covariates. OLS estimates that the treatment group resulted in $577.80 increase in average earnings. The estimate is statistically significant at the 5% level. This estimate clustered at the site level. The inclusion of the covariates lowered the estimated impact, however the result remains of a similar magnitude and is remains statistically different from zero.

	(1) Earnings 18 month later	(2) With covariates
In treatment group	562.4**	577.8*
	(184.2)	(249.8)
adj. R^2	0.001	0.174
F	9.322	.
N	10812	10812

Standard errors in parentheses

Estimate with covariates are clustered by site location

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: *OLS Regression Estimates*

Question 6

The estimates from the nearest 10 neighbor propensity score produce similar estimates to the linear regression estimates. There is an increase in estimated effect about $32 higher in the propensity score matching model. The propensity score finds matches based on the probability of treatment from the probit model which used the covariates. This is distribution is non-linear. The OLS model linearly estimates a value based on the covariates switching on/off. This would produce some differences between the estimates.

Question 7

	(1) Basic Earnings	(2) With Covariates
treatment	601.9***	584.4***
	(176.7)	(170.5)
adj. R^2	0.001	0.083
F	11.60	34.78
N	10812	10812

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: *Difference-in-Difference Estimates*

Question 8

Table 3 shows a difference in means estimate between the JTPA control group and the JTPA treatment. The DID models estimate that enrollment in the JTPA treatment resulted in an approximately $600 increase in average earnings when compared to control group. Adding covariates to the basic DID model reduced the impact of the program towards the basic OLS estimate with covariates. Both DID estimates are significant at the 1% level. The results of the DID are in the neighborhood of the propensity score matching results. The propensity score ATT estimate is \$594.16 with the basic DID estimate of \$601.90 and DID estimate with covariates of \$584.40.

Question 9

See log file.

Question 10

See log file.

Question 11

See Table 4.

	(1)
	Window
At or below cutoff	1126.4
	(962.6)
adj. R^2	0.001
F	1.369
N	329

Standard errors in parentheses
Cutoff at \$2,650
Window is -/+ \$500
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: *Regression Discontinuity via Difference in Means Estimates*

Question 12

The estimate shown in Table 4 estimate the impact of receiving treatment on earnings 18 months after the treatment for participants with prior earnings in the -/+ \$500 window around \$2,650. The model estimates that receiving the treatment causes a \$1,126.40 increase in earnings although the estimate is not statistically different from zero.

Question 13

The result of Table 4 is not surprising because the treatment in this setting is hypothetical. The estimate indicates that this treatment had no impact.

Question 14

See Table 5.

Question 15

	(1) All Data	(2) Window
At or below cutoff	-3217.6***	21449.5
	(597.8)	(20043.3)
Earnings 1 year prior	0.596***	11.37*
	(0.0808)	(4.647)
β Interaction	1.518***	-5.884
	(0.236)	(7.711)
adj. R^2	0.127	0.013
F	154.4	3.725
N	3757	329

Standard errors in parentheses

Cutoff at $2,650

Window is -/+ $500

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: *Regression Discontinuity Estimates*

Table 5 shows regression discontinuity estimates that allow both the y-intercept and slope on both sides of the discontinuity to vary. The estimates using the full data set show a statistically and economically significant result that the treatment reduces future earnings. However, the economic effect becomes substantially positive and statistically indistinguishable from zero when reducing the window to a sample that is more localized to the treatment.

The results of both estimates are not surprising. The full dataset shows the importance of making sure the comparison groups are comparable. The windowed estimate shows the impact of the psuedotreatment was not different from zero.

I prefer the windowed estimate. It seems more reasonable that participants in that narrow pretreatment income band are very comparable. I expect that the differences between the full treatment group and full control would prevent a valid comparison for this type of treatment.