PubPol 713 Assignment 3

Nathan Mather

March 29, 2019

1 Question 1

The results are displaying in the table below. The treatment effect is statistically significant.

	esum18i
Treatment indicator	562.4**
	(183.6)
Constant	9274.3***
	(139.7)
Observations	10812

Standard errors in parentheses

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

2 Question 2

First, I test the hypothesis of equal means for the treated and control group separately for each variable.

	sex	race	age	totch18	child_miss	bfeduca	ed_miss	bfyrearn	earn_miss	est10
Treatment indicator	-0.0114	0.00853	-0.226	0.000434	0.000703	0.125**	-0.00767***	-39.51	0.00185	-0.0282
	(0.00965)	(0.0159)	(0.206)	(0.0273)	(0.00556)	(0.0433)	(0.00231)	(72.07)	(0.00741)	(0.0955)
Observations	10812	10812	10812	10812	10812	10812	10812	10812	10812	10812

Standard errors in parentheses

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

Most of the variables are not statistically different with the exception of years of schooling at baseline and missing education. Even just two significant results are suggestive of a significant ex ante difference in treatment and control groups. This could indicate a problem with randomization if it was not carried out properly in the field. Even if randomization was successful, however, we may have by chance ended up with uneven characteristics. In this case, it will still be a good idea to account for these known ex post differences.

to check this in another way I run the following regression

$$treat_i = \mathbf{B}\mathbf{X_i}$$

where $\mathbf{X_i}$ is a vector of the observed characteristics from above. I then run an F test on the joint significance of all these variables to determine joint orthogonality between the treatment and control. The F test is 1.76 which translates into Prob $> F = 0.0697$. This test simply reinforces the findings from the individual t-tests.

3 Question 3

The results from running the model described are below.

	(1)
	earnings 18 months after
Treatment indicator	599.7***
	(169.0)
sex	4192.6***
	(174.0)
race	-627.5***
	(103.8)
age	-19.76*
	(8.175)
totch18	234.5***
	(63.52)
child_miss	747.1*
	(307.9)
bfeduca	737.3***
	(48.32)
ed_miss	8000.5***
	(902.5)
bfyrearn	0.634***
	(0.0244)
earn_miss	1257.8***
	(230.2)
site_num	-85.01***
	(17.28)
Constant	-694.5
	(631.8)
Observations	10812

Standard errors in parentheses

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

These are worth including because they are observable characteristics that are likely to impact outcomes.

Even in what is ex anti totally random sampling, without stratification we may end up with uneven ex post observables across treatment and control. Including these in the model ensures that these uneven observables are accounted for. As we saw in question 2, there are significant differences of observables. Furthermore, even small differences in observables can significantly impact the estimate of the treatment effect if those observables have a strong impact on the outcome.

The treatment effect is a bit higher in this model than in the treatment effect from question 1. This is a bit surprising since the treated group was more likely to have high Previous education (found in question 2), and high previous education is associated with higher earnings after treatment (as seen in the bfeduca coefficient above). If left unaccounted for, as in question 1, some of the impact of previous education would have been attributed to the treatment effect in question 1. While a bit unexpected this result is not shocking as other observables differ as well and could easily net out to a higher treatment effect from uneven previous education. For example the relationship with missing education is working in the opposite direction.

4 Question 4

Let U be the error in the regression equation, T be the treatment variable, and X our matrix of covariates. The requirement for consistency of all the estimates under OLS generally is $\mathrm{E}(U|X,T) = 0$. That is the error term in the regression needs to be zero conditional on our covariates. We are also assuming the structure of this data generating process is linear in our parameters. Since we are only interested in interpreting the coefficient on the treatment effect, lets call it T, we can actually relax that assumption a bit. We need

$$\mathrm{E}[TU|X] = 0$$

This ensures that treatment is uncorrelated with the error conditional on the X's. This relaxes the assumption because the error no longer has to be zero conditional on X it just has to be uncorrelated with T. So we could, for example, have an omitted variable that impacts Y as long as it is uncorrelated with treatment T.

This is convincingly satisfied by the randomization of treatment and control. Error is zero conditional on treatment without any variables since as the sample tends to infinity covariates X should tend to be equally distributed to treatment and control. However, the idea here is that controlling for them anyway will improve the finite sample estimates.

5 Question 5

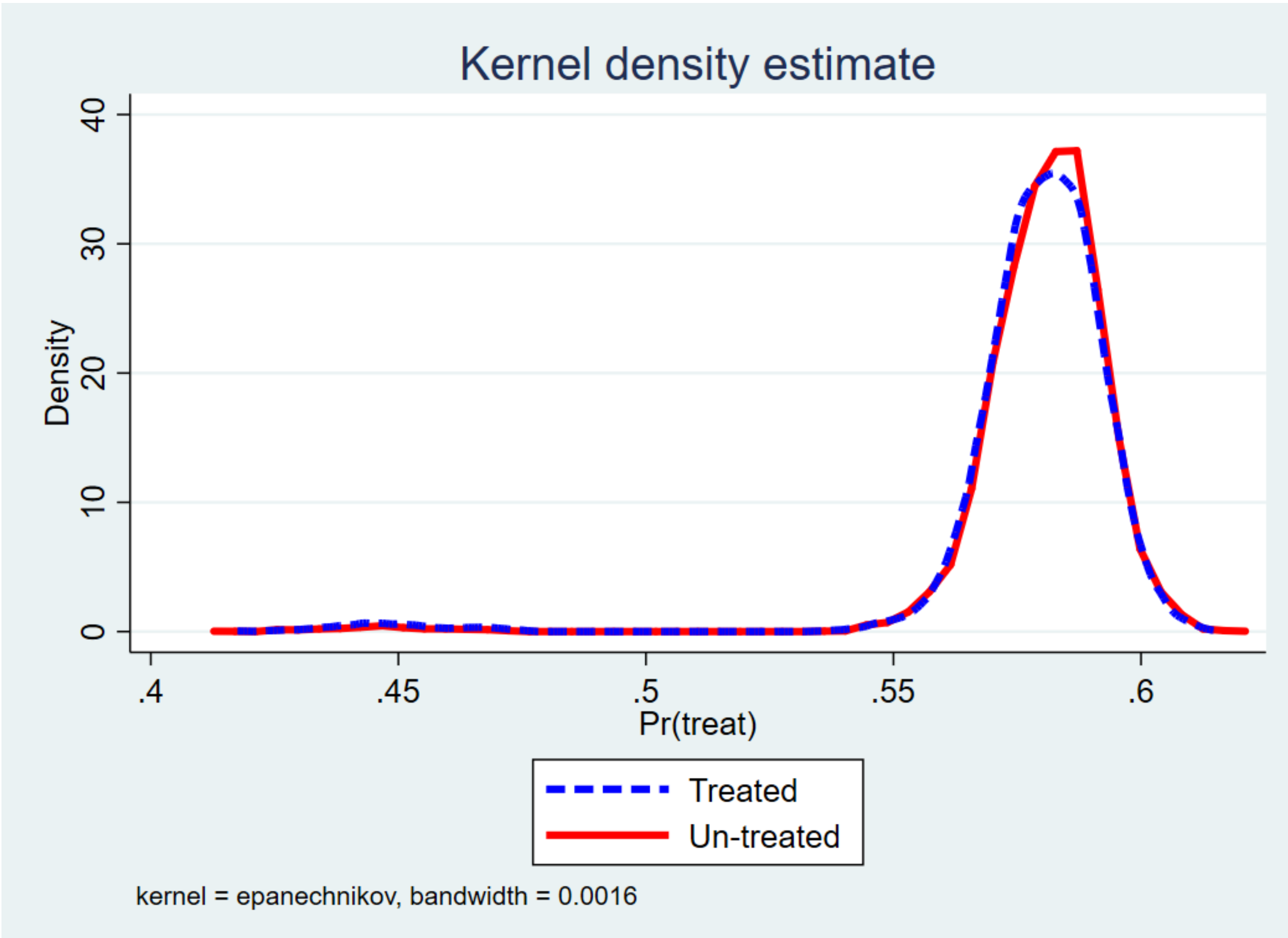
The results of my propensity score regression are below

	(1)
	earnings 18 months after
Treatment indicator	
sex	-0.0270
	(0.0253)
race	0.00660
	(0.0152)
age	-0.00119
	(0.00119)
totch18	0.000646
	(0.00924)
child_miss	0.0308
	(0.0449)
bfeduca	0.00927
	(0.00704)
ed_miss	-0.238
	(0.131)
bfyrearn	-0.00000104
	(0.00000355)
earn_miss	0.0140
	(0.0336)
site_num	-0.000415
	(0.00252)
Constant	0.134
	(0.0912)
Observations	10812

Standard errors in parentheses

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

6 Question 6



The densities do not appear to differ dramatically. This makes sense as it is based on a randomized trial. However there is some discrepancy around the peak which is consistent with the mild differences we observed above. In terms of the range of values in the density, the support of the treatment and control group don't exactly match but are extremely close (I would never expect them to match exactly). They are close enough that no observations will be excluded for lack of common support in the propensity score matching estimate.

7 Question 7

I did both a propensity score matching model with the "Nearest Neighbor matching" and, as a comparison, I ran an exact nearest neighbor matching model treating bfyrearn age bfeduca totch18 as continuous and adjusting for bias introduced by multiple continuous variables.

	Propensity score matching	Nearest Neighbor match
ATET		
r1vs0.Treatment indicator	469.8*	414.9
	(231.8)	(221.0)
Observations	10812	9764

Standard errors in parentheses

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

The results of the propensity score matching are smaller than what we found in question 1. This is different than the OLS results in question 3. This result controls for ex post differences in observables between the treatment and the control. This differs from the OLS estimate in that it does not assume linearity in the parameters for the estimate of the ATET. That being said the propensity score assumes a functional form for the impact of observables on the probability of treatment and so it is still parametric.

8 Question 8

The simulated confounders gave an ATT estimate of 470.923 with a standard error of 241.364.

a) we are assuming that selection is a function of observables and this simulated confounder only. In other words that this satisfies the conditional independence assumption once we include this binary confounder. Let C be the confounder, then $\text{E}[y_{ji}|X,T,C] = \text{E}[Y_{ij}|X,C] for j = 0,1$

b) The simulated confounder gave an ATT estimate of 692.847 with a standard error of 283.202. This is nt too far off from the original estimate which suggests that if there is a confounder of roughly equal importance to sex, which I used for the simulated probabilities, our analysis is still in the right ball park. I think Ideally I would run this simulation for a grid of probabilities and determine at what point our conclusion from the estimates would drastically differ. Then, I can compare the distribution of this hypothetical confounder to any omitted variables the may be biasing the result.

c) The strength is that it helps us think about the magnitude of how violations to our untestable assumptions would impact our estimates. The limitation is that it can't tell us anything about the probability of any of these confounding variables actually existing. For that, we need to use the variables we have as a reference and theory for what we think impacts our outcome.

9 Appendix

9.1 Stata Code

```
*** * Do file for assignment 3 of pp 713

clear all
set more off, perm

* input directory
global dir "C:\Users\Nmath_000\Documents\MI_school\Second Year\PP 713\ps3"

* output directory
global outdir "C:\Users\Nmath_000\Documents\Code\courses\PP 713\ps3_tex\"

* load in data
use "$dir\njs_data_pp713.dta"



**************
* Question 1 *
**************

* clear stored models
eststo clear

* regresss earnins on treatment to get ttest
eststo: reg esum18i treat

* save table
esttab using "$outdir\ps3_table_1.tex",  nonumbers replace label se

* lear stored models
eststo clear


**************
* Question 2 *
**************
*lear stored models
eststo clear

* create varlist of variabls for q2
local q2vars sex race age totch18 child_miss bfeduca ed_miss bfyrearn earn_miss site_num
foreach y of varlist `q2vars' {

eststo: reg `y' treat

}

di `qu_q2vars'

esttab using "$outdir\ps3_table_2.tex", ///
mtitles( "sex" "race" "age" "totch18" "child_miss" "bfeduca" "ed_miss" "bfyrearn" "earn_miss") ///
 nonumbers replace label se ///
keep(treat)

eststo clear

reg treat `q2vars'
display e(F)


**************
* Question 3 *
**************

* clear stored models
eststo clear

* regresss earnins on treatment to get ttest
eststo: reg esum18i treat `q2vars'
```

```

* save table
esttab using "$outdir\ps3_table_3.tex", mtitles("earnings 18 months after") replace label se

* lear stored models
eststo clear


**************
* Question 5 *
**************


* clear stored models
eststo clear

* regresss earnins on treatment to get ttest
eststo: probit treat `q2vars'

* save table
esttab using "$outdir\ps3_table_5.tex",  mtitles("earnings 18 months after") replace label se

* lear stored models
eststo clear

**************
* Question 6 *
**************

* get predicted values
predict pr_score




kdensity  pr_score if treat == 1, lc(red) lw(thick) plot(kdensity pr_score if treat == 0, lc(blue) lp(dash) lw(thick)) legend(order(2 "Treated" 1 "Un-treated"))


graph export "$outdir\6_kdens.png" , replace

* check for common support
summ pr_score if treat == 1
summ pr_score if treat == 0




**************
* question 7 *
**************

* clear stored models
eststo clear

 * do it with propensity score matcing and nearest neighbor sample
eststo:  teffects psmatch (esum18i) (treat sex race age totch18 child_miss bfeduca ed_miss bfyrearn site_num earn_miss, probit), atet

  teffects psmatch (esum18i) (treat sex race age totch18 child_miss bfeduca ed_miss bfyrearn site_num, probit), atet

* do it with actual nearest neighbor matching
capture noisily teffects nnmatch (esum18i sex race age totch18 child_miss bfeduca ed_miss bfyrearn earn_miss site_num) ///
 (treat),  biasadj(bfyrearn age bfeduca) ematch(totch18 sex race child_miss  ed_miss earn_miss site_num) atet osample(nomatch_1)

 capture noisily teffects nnmatch (esum18i sex race age totch18 child_miss bfeduca ed_miss bfyrearn earn_miss site_num) ///
```

```
 (treat) if nomatch_1 == 0,  biasadj(bfyrearn age bfeduca totch18) ematch(sex race
child_miss  ed_miss earn_miss site_num) atet osample(nomatch_2)

 capture noisily teffects nnmatch (esum18i sex race age totch18 child_miss bfeduca ed_miss
bfyrearn earn_miss site_num) ///
 (treat) if nomatch_2 == 0 &  nomatch_1 == 0,  biasadj(bfyrearn age bfeduca totch18) ematch(
sex race child_miss  ed_miss earn_miss site_num) atet osample(nomatch_3)

  eststo: teffects nnmatch (esum18i sex race age totch18 child_miss bfeduca ed_miss bfyrearn
 earn_miss site_num) ///
 (treat) if nomatch_2 == 0 &  nomatch_1 == 0 & nomatch_3 == 0,  biasadj(bfyrearn age bfeduca
 totch18) ematch(sex race child_miss  ed_miss earn_miss site_num) atet



  esttab using "$outdir\ps3_q7_table.tex", ///
 nonumbers replace label se   mtitles("Propensity score matching" "Nearest Neighbor match")


 * clear stored models
eststo clear





**************
* question 8 *
**************

ssc install sensatt
sensatt esum18i treat sex race age totch18 child_miss bfeduca ed_miss bfyrearn site_num
earn_miss, p(sex) reps(100) boot
```