PubPol 713 Assignment 2

Nathan Mather

February 21, 2019

1 Question 1

I generated the variables for a b and c. This is shown in my code in the appendix

2 Question 2

2.1 A

230,653 individuals are pre-selected and 244,512 are not. Of those scoring below the 475 cutoff, 60.1% are pre-selected, 39.95% are not. Of those scoring above the cutoff, 43.92% are not pre-selected and 56.08% are pre-Selected. The results are summarized in the tables below.

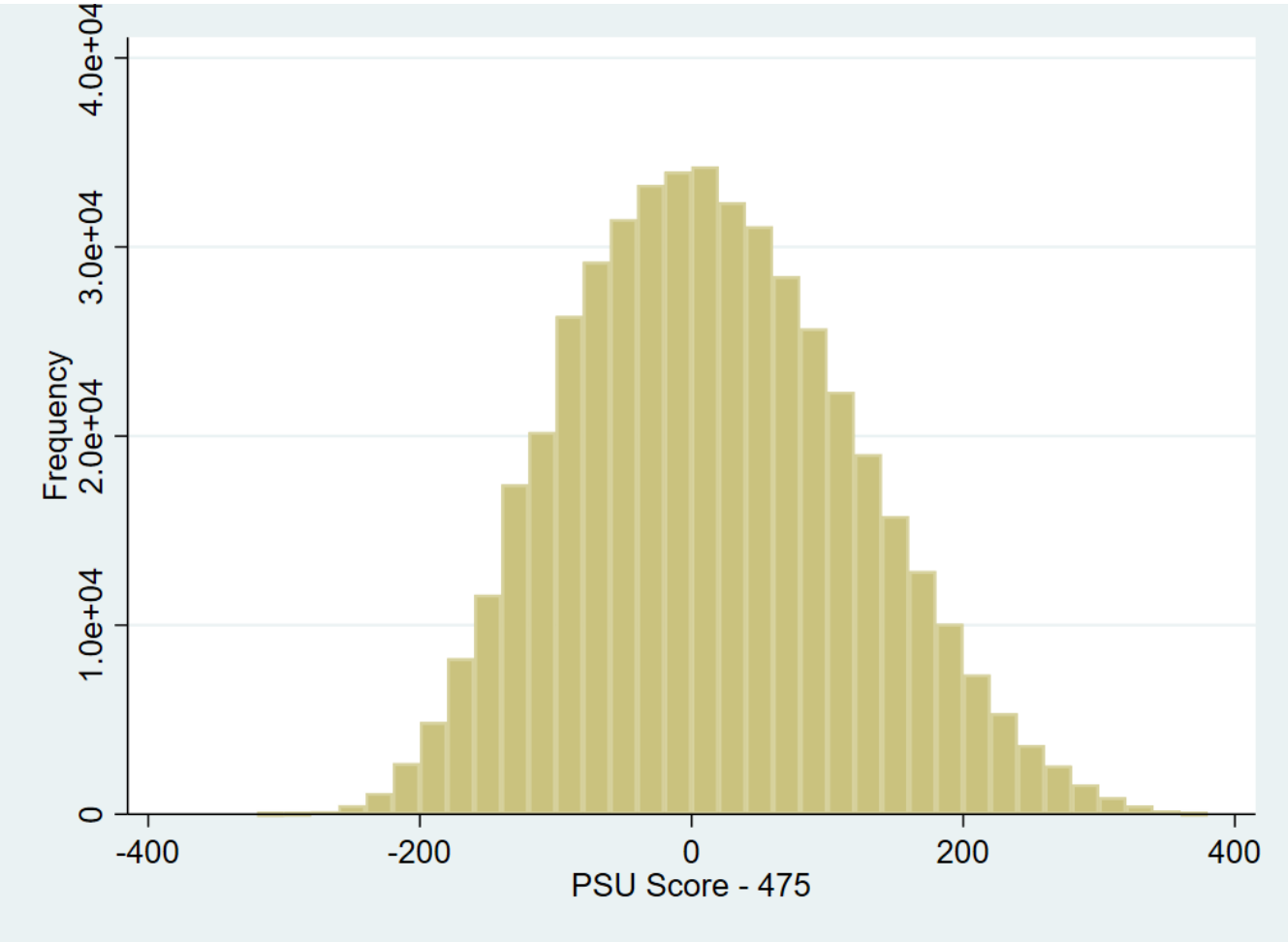
pre_sel	No.	%
Not Pre-Selected	244,512.0	51.5
Pre-Selected	230,653.0	48.5
Total	475,165.0	100.0

PSU Score Above 475	**pre_sel**		
	Not Pre-Selected %	**Pre-Selected** %	**Total** %
Below 475	60.1	39.9	100.0
Above 475	43.9	56.1	100.0
Total	51.5	48.5	100.0

2.2 B

The normalized score (with a PSU score of 475 equal to 0) has a minimum of -314.5, a max of 375, and a mean of 14.6. The distribution does not seem to show any bunching. A histogram is shown below.

Normalized Test Score period 1



2.3 C

The results are summarized in the tables below

Group	Enrolled in college in t=1		
	No	**Yes**	**Total**
	%	%	%
Not Pre-Sel	74.4	25.6	100.0
Pre-Sel Below	88.9	11.1	100.0
Pre-Sel Above	36.3	63.7	100.0
Total	65.7	34.3	100.0

Group	Ever enrolled flag		
	No	**Yes**	**Total**
	%	%	%
Not Pre-Sel	62.2	37.8	100.0
Pre-Sel Below	80.1	19.9	100.0
Pre-Sel Above	25.5	74.5	100.0
Total	54.5	45.5	100.0

2.4 D

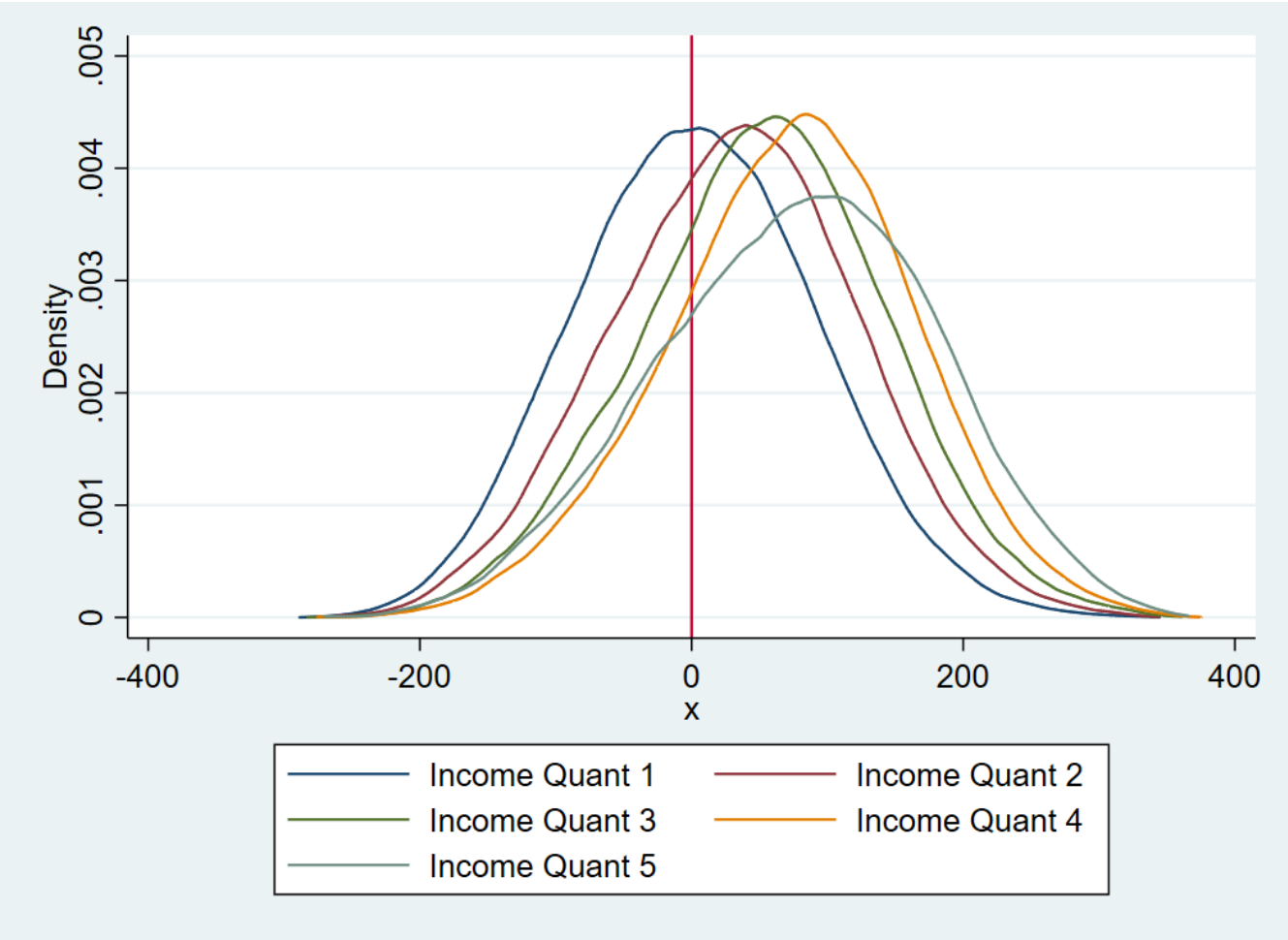
The results are summarized in the tables below

Income quintile for year 1	Enrolled in college in t=1		
	No	**Yes**	**Total**
	%	%	%
1	64.7	35.3	100.0
2	54.3	45.7	100.0
3	47.7	52.3	100.0
4	43.0	57.0	100.0
5	51.2	48.8	100.0
Total	55.8	44.2	100.0

Income quintile for year 1	Ever enrolled flag		
	No	**Yes**	**Total**
	%	%	%
1	56.4	43.6	100.0
2	44.4	55.6	100.0
3	35.9	64.1	100.0
4	29.7	70.3	100.0
5	31.3	68.7	100.0
Total	44.7	55.3	100.0

3 Question 3

Below is a table showing the distribution of PSU scores by income quantile. There is also a verticle line at zero indicating the point to look for discontinuities. The distributions all appear to change smoothly across zero. This supports the needed assumptions and RD design used since income does not appear to change discontinuously at the cutoff.



4 Question 4

Below I have the results from table three of the paper replicated. I used the same regression equation and bandwidth of 44. The main coefficient of interest here is the "PSU score Above 475". In column 1 this coefficient shows that, for pre-selected students, being above the cutoff implies an increase of 17.5 percentage points in the probability of enrolling in college immediately after the test. One potential threat to interpreting this relationship as causal is that passing the 475 threshold may provide some benefit other than loan eligibility. An example discussed in the paper is higher probability of acceptance to schools if students score above 475. Column two tests this possibility with a placebo test on non-selected students. Non selected students are never eligible for the loans and so any observed effect of passing 475 would have to be from something other than loan eligibility. Here be find no significant effect for being above the 475 mark. This is what we would expect since these students are not eligible for loans anyway and it supports the interpretation of the first result as causal.

Table 3 Replication

	(1)	(2)
PSU Score Above 475	0.175***	0.00273
	(0.00611)	(0.00556)
PSU Score - 475	0.00160***	0.00163***
	(0.000147)	(0.000133)
PSU Score if Above 475	0.00222***	0.000866***
	(0.000238)	(0.000223)
Constant	0.182***	0.159***
	(0.00387)	(0.00367)
Bandwidth	44	44

Standard errors in parentheses

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

In addition to the straight replication, I also ran these regressions using an updated bandwidth selection technique and local linear regression (Calonico, Cattaneo, & Titiunik, 2014). The results are similar and can be found in the table below. "RD_Estimate" is comparable to "PSU Score Above 475" in the table above.

Table 3 With Alternative Methods

	(1b)	(2b)
RD_Estimate	0.176***	0.00281
	(0.00668)	(0.00569)
Bandwidth	44.61	50.29

Standard errors in parentheses

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

5 Question 5

The results of the 2SLS estimates are consistent with what we found in question 4. Column (1) shows that the probability of ever going to college increase 16 percentage points when a student is ever eligible for the loan program holding other variables constant. Column 2 shows that even in a first stage regression, among non-preselected high income students, ever passing the 475 mark does not significantly predict college enrollment holding other variables constant.

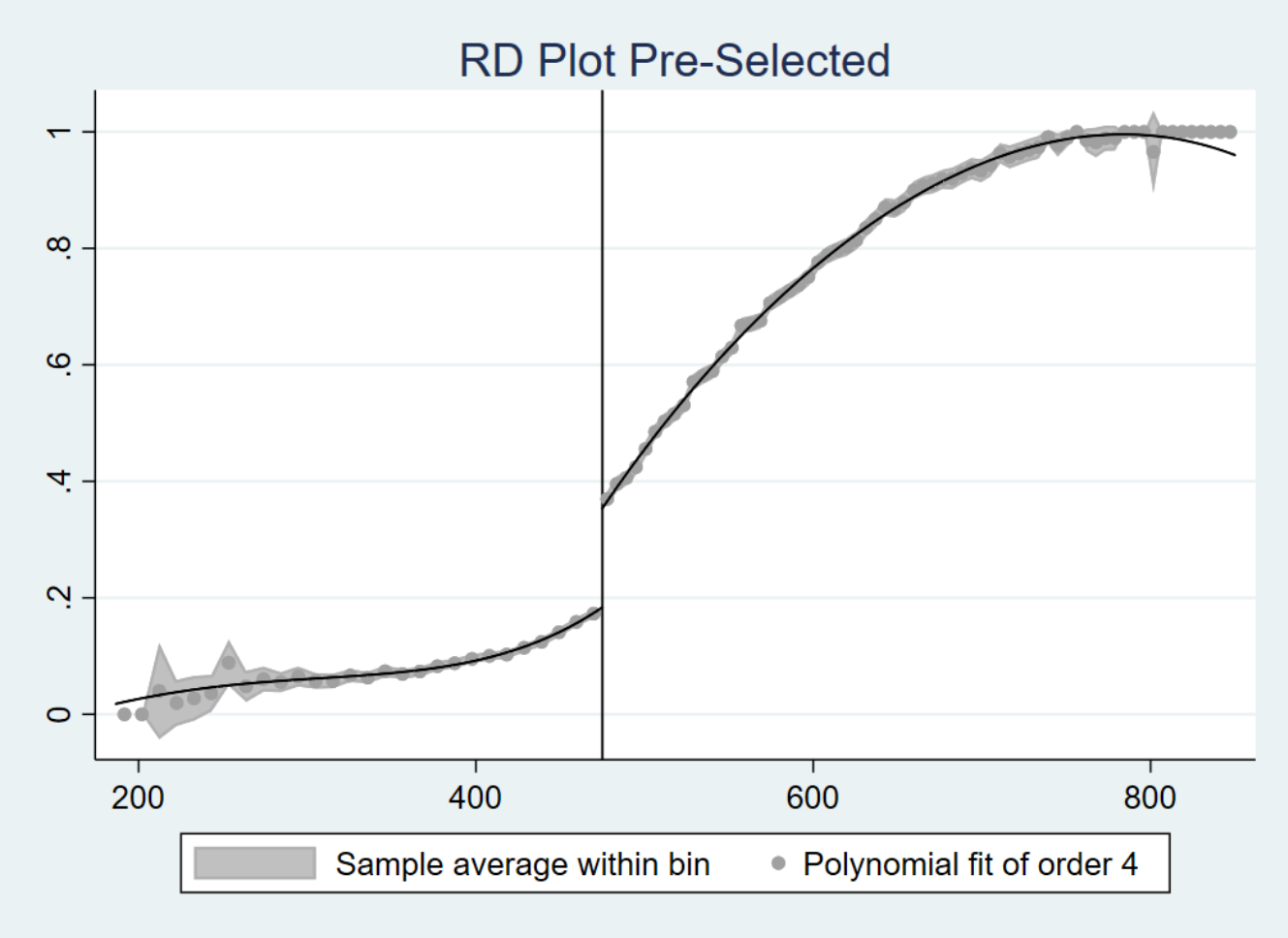
Table 4 Replication

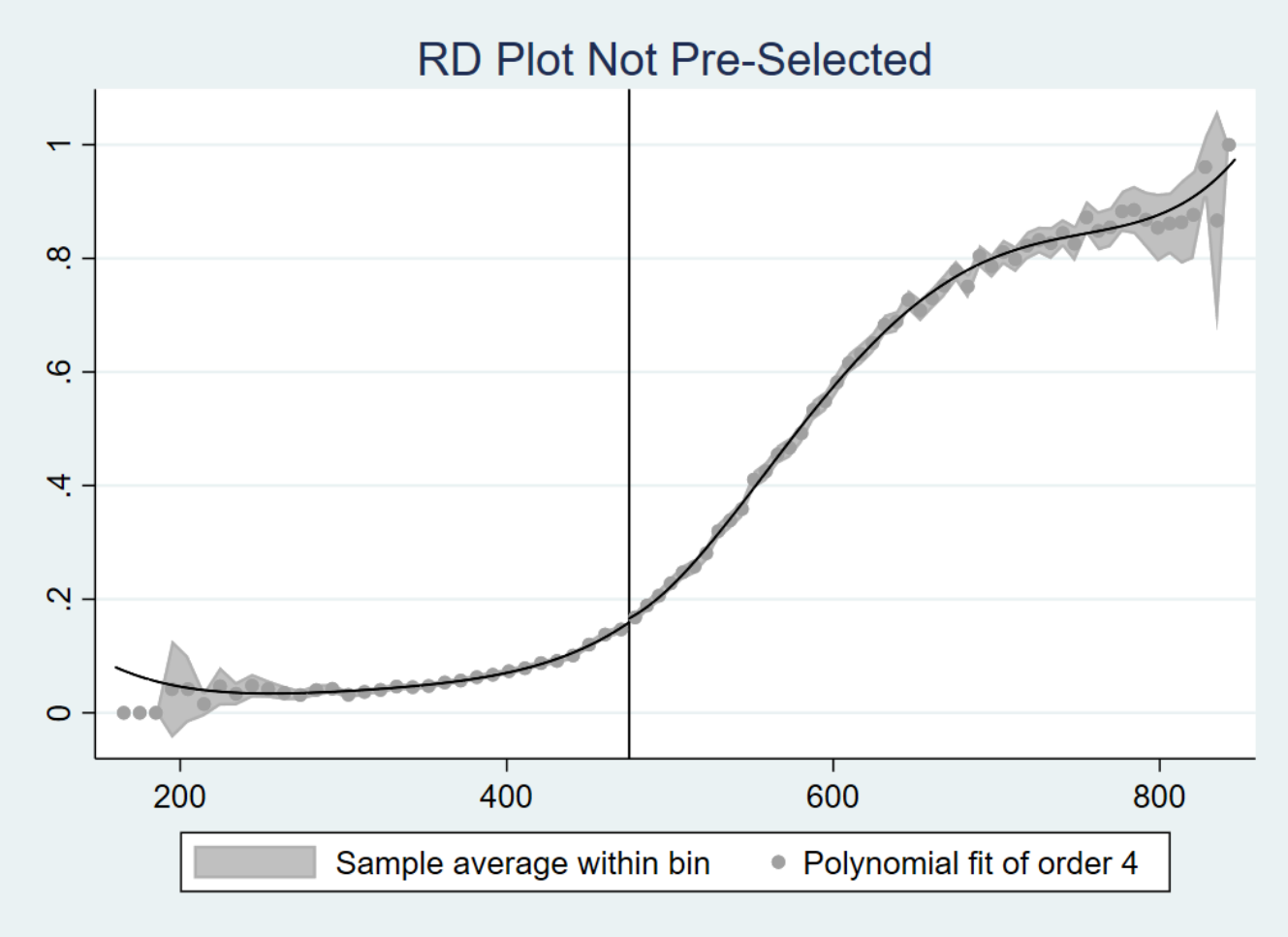
	(1 FS)	(1)	(2)
PSU Score Above 475	0.866***		0.00870
	(0.00332)		(0.00692)
PSU Score - 475	0.00176***	0.00270***	0.00283***
	(0.000122)	(0.000187)	(0.000172)
PSU Score if Above 475	-0.00176***	0.00155***	0.000620*
	(0.000122)	(0.000258)	(0.000273)
Ever eligible flag		0.154***	
		(0.00773)	
Constant	0.134***	0.316***	0.300***
	(0.00332)	(0.00540)	(0.00469)
Bandwidth			

Standard errors in parentheses

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

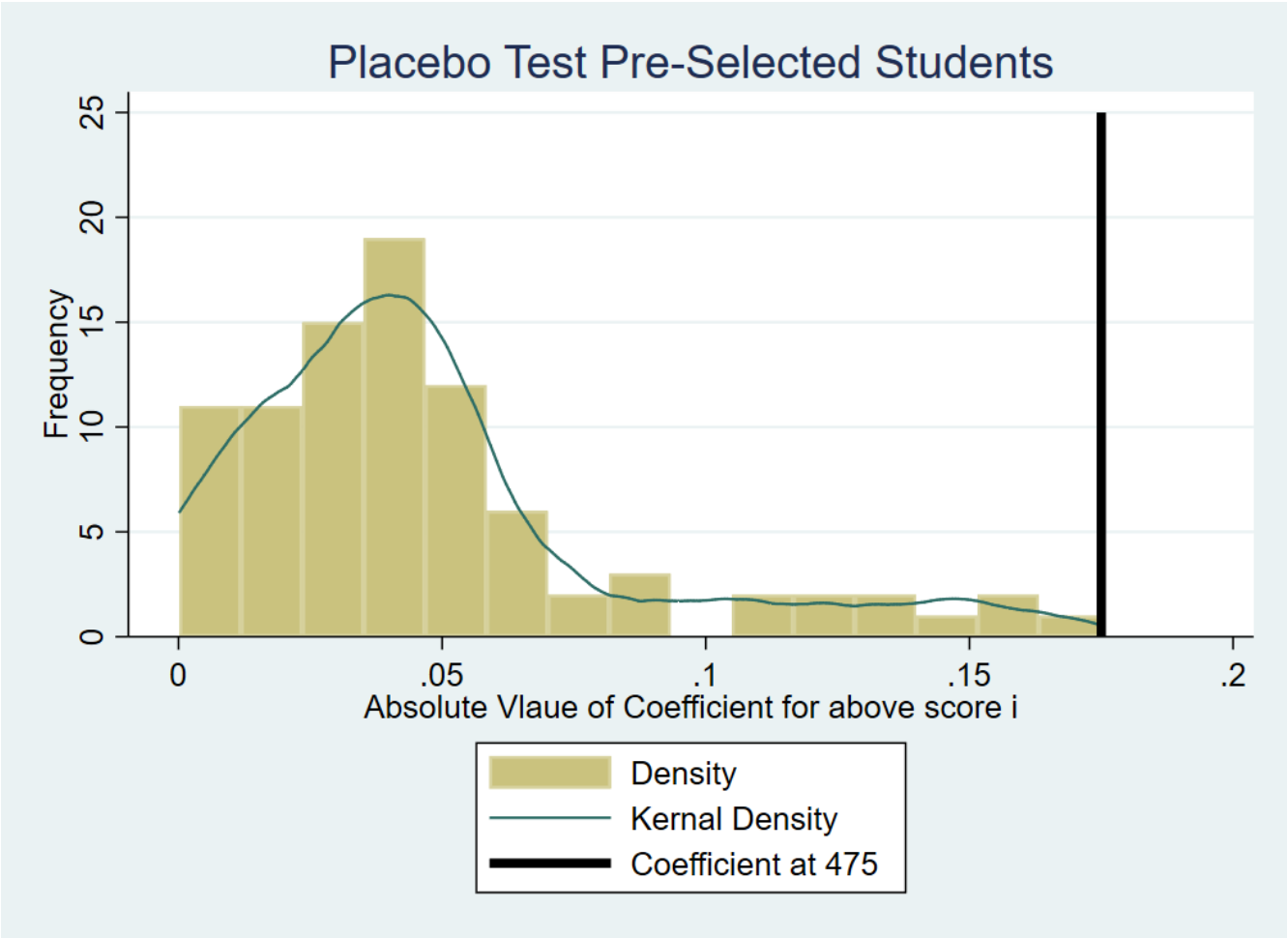
6 Question 6

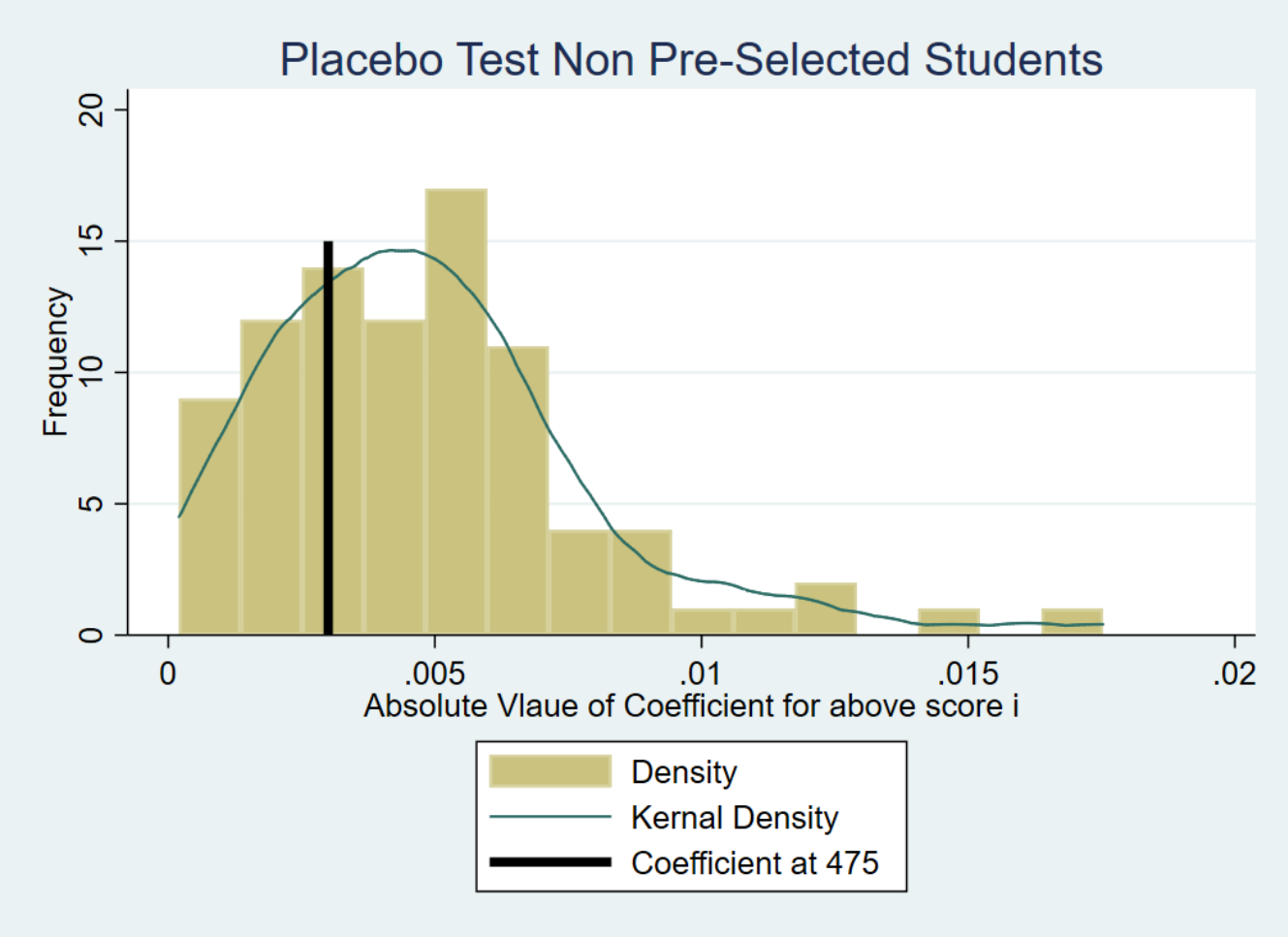

RD Plot Pre-Selected
0
.2
.4
.6
.8
1
200
400
600
800
Sample average within bin
Polynomial fit of order 4



7 Question 7

None of the placebo effects are larger than the true effect for the pre-selected students. For non-Preselected students, the coefficient at the 475 cutoff is greater in magnitude than only 22 of the 88 placebo cut points. These two points together suggest these results are legitimate.


Placebo Test Pre-Selected Students
Frequency
0
5
10
15
20
25
0
.05
.1
.15
.2
Absolute Vlaue of Coefficient for above score i
Density
Kernal Density
Coefficient at 475



8 Appendix

8.1 Stata Code

```
1
2   *** * Do file for assignment 2 of pp 713
3
4   clear all
5   set more off, perm
6
7   * input directory
8   global dir "C:\Users\Nmath_000\Documents\MI_school\Second Year\PP 713\ps2"
9
10  * output directory
11  global outdir "C:\Users\Nmath_000\Documents\Code\courses\PP 713\ps2_tex\"
12
13  * load in data
14  use "$dir\ps2_dataset.dta"
15
16
17  ****************************
18  *1. Created needed variables *
19  ****************************
20
21  * a. indicator for scoreing above 475 in year one
22  gen pass_1 = (psut1 >= 475 & psut1 != .)
23  replace pass_1 =. if psut1 == .
24  tab(pass_1)
25
26    label define pass_1L 0 "Below 475" 1 "Above 475"
27    label values pass_1 pass_1L
28
29
30  * b. gnerate pre selected variable based on income quintile
31  gen pre_sel = (qqt1 <= 4 & qqt1 != .)
32  tab qqt1 pre_sel
33
34    label define pre_selL 0 "Not Pre-Selected" 1 "Pre-Selected"
35    label values pre_sel pre_selL
36
37
38
39  * c. running score centered at 475
40  gen r_score_1 = psut1 - 475
41
42  * make interaction variable
43  gen pass_score_1 = pass_1 * r_score_1
44
45  * make some labels
46  label variable r_score_1    "PSU Score - 475"
47  label variable pass_1       "PSU Score Above 475"
48  label variable pass_score_1 "PSU Score if Above 475"
49
50  ****************************
51  * 2. descriptive stats
52  ******************************
53
54  * check if anyone doesn't have a value for PSU in period one
55  count if psut1 == .
56  * none, no need to worry about that
57
58  ******
59  * a. *
60  ******
61
62  * pre selected individuals in period 1
63  * and proportion of psu takers that are preselected
64  tab pre_sel
65
66  * save it for latex
67  tabout pre_sel using "$outdir\tab2a.tex", ///
68  replace ///
69  style(tex) font(bold)  cells(freq col)
70
```

```
* proportiion above and below cutoff that are preselectd
tab pass 1 pre sel, r nof

* save it for latex
tabout pass_1 pre_sel using "$outdir\tab2aii.tex", ///
replace ///
style(tex) font(bold)  cells(row)


****
*B.*
****
* the forcing variable
summarize r_score_1
hist r_score_1, freq width(20) start(-320)

* save plot
graph export "$outdir\2b_hist.png" , replace


****
*C.*
****
* Rates of immediate ennrollement and ever enrollment by group/

* make labes
  label define yesno 0 "No" 1 "Yes"
  label values enrolt1 yesno
 label values everenroll1 yesno
  * make by group
gen Group = pre sel
replace Group = 2 if pre_sel == 1 & pass_1 == 1

  label define GroupL 0 "Not Pre-Sel" 1 "Pre-Sel Below" 2 "Pre-Sel Above"
  label values Group GroupL

 * check tables
tab Group enrolt1 , r nof
tab Group everenroll1 , r nof

* save them for latex
tabout Group enrolt1 using "$outdir\tab2ci.tex", replace ///
style(tex) font(bold)  cells(row)

tabout Group everenroll1 using "$outdir\tab2cii.tex", replace ///
style(tex) font(bold)  cells(row)


****
*D.*
****


 * check tables
tab qqt1 enrolt1 , r nof
tab qqt1 everenroll1 , r nof

* save them for latex
tabout qqt1 enrolt1 using "$outdir\tab2di.tex", replace ///
style(tex) font(bold)  cells(row)

tabout qqt1 everenroll1 using "$outdir\tab2dii.tex", replace ///
style(tex) font(bold)  cells(row)




*****************************
* 3. Checking Assumptions
*****************************
```

```
141
142
143   * make plot of distributions by income quantile
144   twoway                                  ///
145      (kdensity r_score_1 if qqt1 == 1 ) ///
146      (kdensity r_score_1 if qqt1 == 2 ) ///
147      (kdensity r_score_1 if qqt1 == 3 ) ///
148      (kdensity r_score_1 if qqt1 == 4 ) ///
149      (kdensity r_score_1 if qqt1 == 5 ) ///
150      ,                                  ///
151      legend(order(1 "Income Quant 1" 2 "Income Quant 2" 3 "Income Quant 3" 4 "Income Quant 4"
      5 "Income Quant 5")) ///
152      ytit("Density") xline(0)
153
154
155    * save plot
156    graph export "$outdir\3_plot.png" , replace
157
158
159   *************************
160   * 4 replicate reg tables *
161   *************************
162
163
164   eststo clear
165
166   * do the regressions the way they did them
167   eststo: reg enrolt1  pass_1 r_score_1 pass_score_1 if qqt1<=4 & abs(r_score_1)<=44, r
168   estadd scalar Bandwidth = 44
169
170   eststo: reg enrolt1  pass 1 r score 1 pass score 1 if pre sel==0 & abs(r score 1)<=44, r
171
172   estadd scalar Bandwidth = 44
173
174
175
176   esttab using "$outdir\ps2_table_4.tex", ///
177   mtitles("(1)" "(2)") nonumbers replace label stats(Bandwidth) se
178
179   eststo clear
180
181
182
183   * this is an extension of the Imbens and Kalyanaraman approach. It give similar results
184   * but is more robust and bias corrected
185   eststo: rdrobust enrolt1 r_score_1 if qqt1 <= 4
186   display e(h_l)
187   display e(h_r)
188
189   estadd scalar Bandwidth = e(h_l)
190
191   eststo: rdrobust enrolt1 r_score_1 if qqt1 > 4
192   display e(h_l)
193   display e(h_r)
194   estadd scalar Bandwidth = e(h_l)
195
196
197
198   esttab using "$outdir\ps2_table_4ii.tex", ///
199   mtitles("(1b)" "(2b)") nonumbers replace label stats(Bandwidth) se
200
201
202   eststo clear
203
204
205   ***************
206   *5  Replicate IV *
207   ***************
208
209   * dod table 4 regeressions
```

```
eststo: reg everelig1 pass_1 r_score_1 pass_score_1 if qqt1<=4 & abs(r_score_1)<=44, r
eststo: ivreg everenroll1  (everelig1=pass_1) r_score_1 pass_score_1 if qqt1<=4 & abs(r_score_1)<=44, r
eststo: reg everenroll1  pass_1 r_score_1 pass_score_1 if pre_sel==0 & abs(r_score_1)<=44, r

esttab using "$outdir\ps2_table_5.tex", ///
mtitles("(1 FS)" "(1)" "(2)") nonumbers replace label stats(Bandwidth) se


***************
* 6 make fig 1 *
***************



rdplot enrolt1 psut1 if qqt1<=4 , c(475) shade ci(95) binselect(espr) graph_options(title(RD Plot Pre-Selected))

graph export "$outdir\rdplot_1.png" , replace



rdplot enrolt1 psut1 if pre_sel==0, c(475) shade ci(95) binselect(espr) graph_options(title(RD Plot Not Pre-Selected))
graph export "$outdir\rdplot_2.png" , replace



*********************
* q 7 placebo tests *
*********************

* set up a matrix for the results
matrix Res = J(89,3,.)

forvalues i = 431/519{

* create variables for regression
gen pass_i = (psut1 >= `i' & psut1 != .)
gen r_score_i= psut1 - `i'
gen pass_score_i = pass_i * r_score_i


* get matrix position
local mat_post = `i' - 430

* store cutoff in matrix
matrix Res[`mat_post',1] =  `i'

* run regression a with these vars
reg enrolt1  pass_i r_score_i pass_score_i if qqt1<=4 & abs(r_score_i)<=44, r

* store result in matrix
matrix Res[`mat_post',2] = abs(_b[pass_i])

* run regression b with these vars
reg enrolt1  pass_i r_score_i pass_score_i if pre_sel==0 & abs(r_score_i)<=44, r

* store result in other matrix column
matrix Res[`mat_post',3] = abs(_b[pass_i])

* drop variables for next iteration
drop pass_i
drop r_score_i
drop pass_score_i

}
```

```
277
278   * make the results the data set
279   drop _all
280   svmat float Res
281
282   count if Res2 < .175
283   count if Res3 < .002727
284   * make a histogram of each
285   hist Res2, bin(15) kdens  addplot(pci 0 .175 25 .175, lcolor(black) lwidth(1)) ///
286      legend(order(1 "Density" 2 "Kernal Density" 3 "Coefficient at 475")) freq ///
287      title(Placebo Test Pre-Selected Students) xtitle(Absolute Vlaue of Coefficient for above
      score i)
288
289      graph export "$outdir\placebo_1.png" , replace
290
291
292
293   hist Res3, bin(15) kdens  addplot(pci 0 .003 15 .003, lcolor(black) lwidth(1)) freq  ///
294      legend(order(1 "Density" 2 "Kernal Density" 3 "Coefficient at 475")) ///
295         title(Placebo Test Non Pre-Selected Students) xtitle(Absolute Vlaue of Coefficient for
       above score i)
296
297
298      graph export "$outdir\placebo_2.png" , replace
299
300
301
```

References

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295-2326. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA11757`