Problem Set 3 Solutions

Question 1. Return to the NELS-88 data used in Problem Set 2 to estimate the academic benefits to attending a Catholic high school. This time—rather than exact or nearest neighbor matching based on covariates—we will use propensity scores. (**30 points**)

- (a) For this analysis, use the "continuous" family income variable you created in Problem Set 2 and the dummy variables for parents' highest level of education. In addition, create z-scores for the 8th and 12th grade reading and math scores. (3 points)
 - See the attached log file for creation of additional variables.
- (b) Use teffects psmatch or the psmatch2 commands to estimate a propensity score model where the "treatment" is attending a Catholic high school. The goal will ultimately be (in part c) to estimate the ATT on 12th grade test scores, high school graduation, and post-secondary enrollment, using nearest neighbor matching (based on propensity scores). This will be an iterative process where you experiment with a propensity score model, check for balance, and then make adjustments to your propensity score model and matching rules as needed. I recommend using the quietly prefix with teffects psmatch, or omitting the outcome() option if you use the psmatch2 command. (You do not want treatment effect estimates to guide your specification search). Here are a few tips/requirements: (12 points)
 - Choose predictor variables that are likely to be associated with the "treatment" that you would ultimately like to see balanced between your treated and untreated group. You can use any variables in the dataset that you deem appropriate.
 - After fitting the propensity score model, check for balance on your predictor variables using tebalance summarize, box and density. (Or use pstest after psmatch2).
 - You may be able to improve balance by changing your propensity model specification—e.g., omitting or including variables (depending on how predictive or theoretically important they are), entering variables as continuous or categorical, adding interactions or nonlinear terms (e.g., quadratic)—and/or by tinkering with the number of nearest neighbors or caliper. Use your own judgment when deciding on "good enough" balance.
 - You do <u>not</u> need to provide results for all of the iterations you attempt. Just include a short written explanation justifying your choice, and provide balance tests/figures for your final specification. (You can refer to other specifications you tried in your write-up).

• Lastly, once you have settled on a propensity score model, check for overlap in the distribution of propensity scores between your treated and untreated group. (E.g., teoverlap or tebalance density after teffects psmatch, or using code provided in class). Include your graphical results with your output. Presuming you have good overlap, you can proceed to the next step. If there is poor overlap, you should revisit your propensity score model.

My final model is shown in the attached log (yours will likely differ). Of the confounders, I prioritized balance on baseline test scores, family income, and parental education, given the results of prior studies which demonstrate the importance of these in explaining later outcomes. The logit model results confirm these are important predictors of Catholic school enrollment as well. I experimented with several different specifications and variable definitions; these included models linear in prior test scores, models that used separate categorical variables for the highest education of each parent, models with interactions between gender and race, and models with a continuous measure of risk factors. The covariate balance was quite good for most specifications, but this one seemed especially good, as the balance test results show. (The log file shows the tebalance summarize for the same propensity score match and the tebalance box and tebalance density for select variables).

The teoverlap and tebalance density plots are also shown in the attached. Note that the former uses *all* observations, not just the matched sample. The latter shows both the full and matched sample.

(c) Once you are satisfied with your propensity score model, calculate the ATT for the four outcome variables: 12th grade reading and math z-scores, high school graduation, and post-secondary enrollment. Interpret the results in words. What assumption(s) are required for these estimates to be considered causal? (5 points)

The results are shown in the attached log and summarized in the table below. Attending a Catholic high school is estimated to have a 0.119 standard deviation effect on math scores and a 0.024 standard deviation effect on reading scores. Only the math effect is statistically significant. The ATT estimate is positive and marginally insignificant for high school graduation (3.0 percentage points) and positive and significant for post-secondary enrollment (7.3 ppts). Because these are ATT estimates they are interpreted as the average treatment effect for the population who attends Catholic high school. The key assumption is conditional independence—conditional on the covariates (here, factors used to create balance between the treated and untreated cases), treatment assignment is independent of potential outcomes.

(d) Rather than nearest neighbor matching, propensity scores can also be used in weighting estimators. Using the same propensity score model you settled on in part (b), estimate the ATT and ATE using inverse probability weighting (teffects ipw). How do your results differ from those in part (c), if at all? (5 points)

The results are shown in the attached log and summarized in the table below. The ATT results are very similar to those using propensity score matching.

(e) Imbens (2015) proposed an algorithm for estimating propensity score models. The algorithm iteratively adds linear and quadratic terms (and interactions), keeping terms that improve the predictive fit of the model. The user-written command psestimate executes this algorithm and outputs the resulting propensity scores to your dataset. (You will need to install it using ssc install psestimate). The syntax of the command is:

psestimate treat, totry(varnames) genpscore(newvarname)

where *treat* is the treatment variable, *varnames* is a list of predictor variables you would like to try in the algorithm, and *newvarname* will be the new variable containing the propensity score. Choose a set of variables for the algorithm to try. (Note if you choose a long list it make take quite awhile for the algorithm to run. If it takes an exceedingly long time, start with a shorter list). What specification did the algorithm end up with? (5 points)

Note: the resulting propensity scores from part (e) could then be used for matching, stratification, or weighting.

The resulting model specification is shown in the attached log. I had the algorithm try the same set of variables used in the earlier steps. The resulting specification included some nonlinear terms (e.g., a quadratic in prior math scores) and many interactions (e.g., Black and family income).

faminc8 -- total annual family income in 8th grade

			ļ	Freq.	Percent	Valid	Cum.
Valid	12	none <\$1000 \$1000-\$2999 \$3000-\$4999 \$5000-\$7499 7500-\$9999 \$10000-\$14999 \$15000-\$19999 \$20000-\$24999 \$25000-\$34999 35000-\$49999 50000-\$74999	+ 	18 42 84 85 144 175 447 441 655 1267 1419 894 5671	0.32 0.74 1.48 1.50 2.54 3.09 7.88 7.78 11.55 22.34 25.02 15.76 100.00	0.32 0.74 1.48 1.54 3.09 7.88 7.78 11.55 22.34 25.06 100.00	0.32 1.06 2.54 4.04 6.58 9.66 17.55 25.32 36.87 59.21 84.24 100.00
	10	Cul	1	3071	-00.00	100.00	

```
replace faminc8b = (15000+19999)/2 if faminc8==8
(441 real changes made)

replace faminc8b = (20000+24999)/2 if faminc8==9
(655 real changes made)

replace faminc8b = (25000+34999)/2 if faminc8==10
(1,267 real changes made)

replace faminc8b = (35000+49999)/2 if faminc8==11
(1,419 real changes made)

replace faminc8b = (50000+74999)/2 if faminc8==12
(894 real changes made)

label var faminc8b "family income in 8th grade (dollars)"

// father's highest education
fre fathed8
```

fathed8 -- father^s highest level of education

		Freq.	Percent	Valid	Cum.
Valid	1 not finish hs 2 hs grad 3 junior coll 4 coll <4 5 coll grad 6 masters 7 doctorate 8 dont know Total	873 1778 660 443 743 346 141 687 5671	15.39 31.35 11.64 7.81 13.10 6.10 2.49 12.11 100.00	15.39 31.35 11.64 7.81 13.10 6.10 2.49 12.11 100.00	15.39 46.75 58.38 66.20 79.30 85.40 87.89 100.00

gen fathed1 = fathed8==1 /* hs dropout */

gen fathed2 = fathed8==2 /* hs grad */

gen fathed3 = (fathed8 >= 3 & fathed8 <= 4) /* some college */

gen fathed4 = (fathed8>=5 & fathed8<=7) /* 4yr college or more */

label var fathed1 "father's highest ed: hs dropout"

label var fathed2 "father's highest ed: hs grad"

label var fathed3 "father's highest ed: some college"

label var fathed4 "father's highest ed: 4yr college or more"

// mother's highest education

fre mothed8

mothed8 -- mother's highest level of education

		Freq.	Percent	Valid	Cum.
Valid	1 not finish hs 2 hs grad 3 junior coll 4 coll <4 5 coll grad 6 masters 7 doctorate 8 dont know Total	815 2091 686 468 655 299 82 575	14.37 36.87 12.10 8.25 11.55 5.27 1.45 10.14	14.37 36.87 12.10 8.25 11.55 5.27 1.45 10.14	14.37 51.24 63.34 71.59 83.14 88.41 89.86 100.00

```
gen mothed1 = mothed8==1 /* hs dropout */
          gen mothed2 = mothed8==2 /* hs grad */
          gen mothed3 = (mothed8 \ge 3 \& mothed8 \le 4) /* some college */
          gen mothed4 = (mothed8>=5 & mothed8<=7) /* 4yr college or more */
          label var mothed1 "mother's highest ed: hs dropout"
          label var mothed2 "mother's highest ed: hs grad"
          label var mothed3 "mother's highest ed: some college"
          label var mothed4 "mother's highest ed: 4yr college or more"
          forvalues j=1/4 {
 2.
                replace fathed`j'=. if fathed8==.
                replace mothed`j'=. if mothed8==.
 3.
 4.
(0 real changes made)
          // highest education of two parents
          egen pared8=rowmax(fathed8 mothed8)
          gen pared1 = pared8==1 /* hs dropout */
          gen pared2 = pared8==2 /* hs grad */
          gen pared3 = (pared8>=3 & pared8<=4) /* some college */</pre>
          gen pared4 = (pared8>=5 & pared8<=7) /* 4yr college or more */</pre>
          label var pared1 "parent's highest ed: hs dropout"
          label var pared2 "parent's highest ed: hs grad"
          label var pared3 "parent's highest ed: some college"
          label var pared4 "parent's highest ed: 4yr college or more"
          label var pared8 "parent's highest education"
          // z-scores for 8th and 12th grade reading and math scores
          foreach g in 8 12 {
             foreach s in read math {
                egen `s'`g'z = std(`s'`g')
label var `s'`g'z "z-score for `g'th grade `s'"
 4.
 5.
  6.
                }
          // parent expects child to go to college
         codebook fhowfar mhowfar
fhowfar
                                                  how far in schl r's father wants r to go
                  Type: Numeric (byte)
                 Label: farcat2
```

Label: farcat2

Range: [1,6]
Units: 1
Unique values: 6

Missing .: 0/5,671

mhowfar how far in schl r^s mother wants r to go

Type: Numeric (byte)

Label: farcat2

Range: [1,6] Units: 1
Unique values: 6 Missing .: 0/5,671

Tabulation: Freq. Numeric Label
42 1 not fin
275 2 hs grad

42 1 not finish hs 275 2 hs grad 361 3 junior coll 577 4 coll <4 2,928 5 coll grad 1,488 6 postsec ed

gen collegexp=(fhowfar>=5 & fhowfar<.) | (mhowfar>=5 & mhowfar<.)

label var collegexp "=1 if mother or father expects college+"

. codebook riskdrop8

riskdrop8 # of risk factors for later dropout

Of Fish factors for fatter diopole

Type: Numeric (byte)

Range: [0,5] Units: 1

Unique values: 6 Missing .: 0/5,671

Tabulation: Freq. Value 3,369 0 1,406 1

1,406 1 623 2 214 3 52 4 7 5

gen riskdrop1=riskdrop8==1

gen riskdrop2=(riskdrop8>=2 & riskdrop8<=5)</pre>

. label var riskdrop1 "=1 if 1 dropout risk factor"

. label var riskdrop2 "=1 if 2-5 dropout risk factors"

. fre parmar8

parmar8 -- parents marital status in 8th grade

		 F	req.	Percent	Valid	Cum.
Valid	1 divorced 2 widowed 3 separated 4 never married 5 not married but cohabit 6 married Total		543 77 140 91 93 4727 5671	9.58 1.36 2.47 1.60 1.64 83.35	9.58 1.36 2.47 1.60 1.64 83.35	9.58 10.93 13.40 15.01 16.65 100.00

```
gen byte parmarried=parmar8==6
          label var parmarried "=1 if parents married in 8th grade"
. // ****************
. // Part b - finding a propensity score
. // model for later use
          // Here I use psmatch2 (with logit, 1 nn match). I use the "ties"
          // option to line up with teffects psmatch (used later in part c). If
          // there are multiple nearest neighbors with the same pscore, all of
// them will be matched.
          psmatch2 catholic c.math8z##c.math8z c.read8z##c.read8z female ///
             black hisp api nativam riskdrop1 riskdrop2 disrupt nohw8 ///
             c.faminc8b##c.faminc8b pared1-pared4 parmarried collegexp, ties logit
Logistic regression
                                                         Number of obs = 5,671
                                                         LR chi2(20) = 269.04
Prob > chi2 = 0.0000
                                                         Pseudo R2 = 0.0709
Log likelihood = -1763.1344
```

catholic	Coefficient	Std. err.	z	P> z	[95% conf.	interval]
math8z	0605341	.0732046	-0.83	0.408	2040125	.0829443
c.math8z#c.math8z	1680258	.0495161	-3.39	0.001	2650755	070976
read8z	.3255739	.0679117	4.79	0.000	.1924695	.4586783
c.read8z#c.read8z	.0100256	.0484394	0.21	0.836	0849138	.104965
female black hisp api nativam riskdrop1 riskdrop2 disrupt8 nohw8 faminc8b	.0419112 2319429 0828599 -1.165599 -4855327 .0232422 4050387 .218962 6282156 .0000303	.0950359 .1917082 .1589371 .2812792 .2864356 .1225025 .2595379 .1285199 .1773147	0.44 -1.21 -0.52 -4.14 -1.70 0.19 -1.56 1.70 -3.54 2.09	0.659 0.226 0.602 0.000 0.090 0.850 0.119 0.088 0.000 0.036	1443558 6076841 3943709 -1.716896 -1.046936 2168583 9137236 0329324 975746 1.93e-06	.2281781 .1437983 .2286512 6143023 .0758707 .2633426 .1036463 .4708563 2806852 .0000587
c.faminc8b#c.faminc8b	 -2.12e-10	1.74e-10	-1.22	0.221	-5.52e-10	1.28e-10
pared1 pared2 pared3 pared4 parmarried collegexp _cons	7369748 .1017425 .0949015 .3071211 .2471734 .2901846 -3.252608	.3315306 .156011 .158119 .1506848 .1803656 .1432064 .377756	-2.22 0.65 0.60 2.04 1.37 2.03 -8.61	0.026 0.514 0.548 0.042 0.171 0.043 0.000	-1.386763 2040334 215006 .0117842 1063367 .0095053 -3.992996	0871868 .4075184 .4048091 .6024579 .6006836 .5708639 -2.51222

// check for balance
pstest

Variable	Mean Treated Control	 %bias	t-te	est p> t	V(T)/ V(C)
math8z c.math8z#c.math8z read8z c.read8z#c.read8z female black hisp api nativam riskdrop1 riskdrop2 disrupt8 nohw8	.22418	-4.4 -7.4 -2.0 -0.5 2.0 -2.5 -5.0 0.0 -1.0 4.4 1.7 -2.2 3.8	-0.79 -1.28 -0.35 -0.08 0.35 -0.48 -0.88 0.00 -0.19 0.80 0.42 -0.39 0.86	0.429 0.200 0.729 0.937 0.727 0.631 0.378 1.000 0.851 0.426 0.674 0.697 0.392	0.93 0.96 1.01 0.92
faminc8b	39534 40008	-2.8	-0.50	0.615	1.02

```
      c.faminc8b#c.faminc8b
      | 1.8e+09
      1.9e+09
      -2.6 | -0.43 | 0.670 | 0.97

      pared1
      | .02027 | .02196 | -0.8 | -0.20 | 0.840 | .
      | .

      pared2
      | .25169 | .24155 | 2.3 | 0.40 | 0.686 | .
      | .

      pared3
      | .22804 | .23311 | -1.2 | -0.21 | 0.836 | .
      | .

      pared4 | .38007 | .375 | 1.1 | 0.18 | 0.857 | .
      | .

      parmarried | .91216 | .91385 | -0.5 | -0.10 | 0.918 | .
      | .

      collegexp | .88514 | .88176 | 0.9 | 0.18 | 0.856 | .
      | .

* if variance ratio outside [0.85; 1.18]
Ps R2 LR chi2 p>chi2 MeanBias MedBias B R %Var
0.004 6.95 0.997 2.3 2.0 15.3 0.91 0
______
* if B>25%, R outside [0.5; 2]
            // treated and matched untreated cases
            tabulate treated weight
 psmatch2: |
Treatment | psmatch2: weight of matched controls assignment | 1 2 3
                                                               4 | Total
-----
 Untreated | 442 58 10 1 | 511
Treated | 592 0 0 0 | 592
 Total | 1,034 58 10 1 | 1,103
           count if _treated==0 & _weight~=.
  511
            // propensity scores (pre-match)
            psgraph, name (psg, replace) ///
>
                    title (Propensity score distributions: full sample)
            // overlapping histograms of propensity scores, matched sample
            twoway (histogram _pscore if catholic==0 [fweight=_weight], bin(20) /// fcolor(none) lcolor(blue)) (histogram _pscore_if catholic==1, ///
               bin(20) bin(20) fcolor(none) lcolor(red)), ///
               legend(label(1 Untreated) label(2 Treated)) ///
               name(psg2, replace) ///
               title (Propensity score distributions: matched sample)
           graph combine psg psg2, row(1) xsize(8) ysize(4) ///
                     name(graph1, replace) ///
                      title(Propensity score distributions: after psmatch2)
            graph export graph1.pdf, name(graph1) as(pdf) replace
file graph1.pdf saved as PDF format
            // now show tebalance and overlap after teffects psmatch--
            // same model specification as above
            capture drop *
            capture drop nn*
            quietly teffects psmatch (math12z) (catholic c.math8z##c.math8z ///
               c.read8z##c.read8z female black hisp api nativam riskdrop1 ///
               riskdrop2 disrupt nohw8 c.faminc8b##c.faminc8b pared1-pared4 ///
               parmarried collegexp, logit), atet gen(nn)
```

```
// check for balance (now using tebalance)
       tebalance summarize math8z read8z female black hisp api nativam ///
            riskdrop1 riskdrop2 disrupt nohw8 faminc8b pared1-pared4 ///
            parmarried collegexp
Covariate balance summary
                 Raw
                       Matched
______
Number of obs = 5,671 1,184
Treated obs = 592 592
Control obs = 5,079 592
          |Standardized differences | Variance ratio
           | Raw Matched
                                   Raw Matched
```

```
tebalance box math8z, name(balmath, replace)
         tebalance box read8z, name(balread, replace)
         tebalance density faminc8b, name(balinc, replace)
          graph combine balmath balread balinc, col(1) xsize(4) ysize(8) ///
>
                 name(graph2, replace) ///
>
                  title (Covariate balance: after teffects psmatch, size (medsmall))
          graph export graph2.pdf, as(pdf) replace
file graph2.pdf saved as PDF format
          // propensity scores (all observations)
          teffects overlap, ptlevel(1) name(graph3a, replace) ///
             title (Propensity score distribution: all obs, size (medsmall))
         capture drop ps0 ps1
         predict ps0 ps1, ps
          // propensity scores (all observations + matched sample)
          tebalance density, name(graph3b, replace) ///
                  byopts(title(Propensity score distribution: all and matched sample, size
 (medsmall)))
```

```
// NOTE: the above tebalance density command shows the densities of
          // pscore for the matched sample. FYI the code below will manually
          // take the estimate pscores and produce overlapping densities. This
          // is a bit cumbersome since it requires identifying the matches
          // manually using the nn1 id
          capture drop ob
         gen ob=_n
          preserve
          tempfile treated
          keep if catholic == 1
(5,079 observations deleted)
          keep nn1
          bysort nn1: gen weight= N
          by nn1: keep if _n==1
(81 observations deleted)
         rename nn1 ob
         save `treated'
file C:\Users\corcorsp\AppData\Local\Temp\ST ec14 000002.tmp saved as .dta format
         restore
          merge m:1 ob using `treated'
(variable ob was float, now double to accommodate using data's values)
                                Number of obs
    Not matched
                                         5,160
                                           ,160 (_merge==1)
0 (_merge==2)
        from master
                                          5,160
        from using
                                           511 (_merge==3)
    Matched
    _____
         replace weight=1 if catholic==1
(592 real changes made)
         tabstat ps1 if weight~=., by(catholic)
Summary for variables: ps1
Group variable: catholic (attended catholic hs?)
catholic | Mean
-----
     no | .1408498
    yes |
            .145605
  Total | .143402
          twoway (kdensity ps1 if catholic==0 & weight~=.) ///
  (kdensity ps1 if catholic==1) , name(graph3c, replace) ///
  legend(label(1 Untreated) label(2 Treated)) xtitle(Propensity score) ///
>
>
             title (Propensity score distribution: matched sample - manual, size (medsmall))
          graph combine graph3a graph3b graph3c, col(1) xsize(4) ysize(8) ///
                   name(graph3, replace) ///
                   title (Propensity score distributions: after teffects psmatch, size (mediu
> m))
```

```
graph export graph3.pdf, as(pdf) replace
file graph3.pdf saved as PDF format
       graph close all
       drop nn1 ps0 ps1 ob weight merge
 // *********
 // Part c - estimate ATT using model
. // decided above
. // ****************
. // show using both teffects psmatch and psmatch2
       foreach j in math12z read12z hsgrad inpse {
 2.
           display in red "Outcome: `j'"
 3.
         _eststo psm`j': teffects psmatch (`j') (catholic ///
>
           c.math8z##c.math8z c.read8z##c.read8z female black ///
>
              hisp api nativam riskdrop1 riskdrop2 disrupt nohw8 ///
              c.faminc8b##c.faminc8b pared1-pared4 ///
           parmarried collegexp, logit), atet
            psmatch2 catholic c.math8z##c.math8z c.read8z##c.read8z ///
>
              female black hisp api nativam riskdrop1 riskdrop2 disrupt ///
              nohw8 c.faminc8b##c.faminc8b pared1-pared4 ///
           parmarried collegexp, logit outcome(`j')
>
 5.
Outcome: math12z
Treatment-effects estimation
                                 Number of obs
                                                    5,671
1
                                     min =
Treatment model: logit
                                             max =
______
math12z | Coefficient std.err. z P>
                             z P>|z| [95% conf. interval]
ATET
  catholic |
(yes vs no) | .1190682 .0423528 2.81 0.005 .0360583
                                                  .2020782
                                         Number of obs = 5,671
Logistic regression
                                        LR chi2(20) = 269.04
Prob > chi2 = 0.0000
Pseudo R2 = 0.0709
Log likelihood = -1763.1344
        catholic | Coefficient Std. err. z P>|z| [95% conf. interval]
          math8z | -.0605341 .0732046 -0.83 0.408 -.2040125 .0829443
                          .0495161
  c.math8z#c.math8z | -.1680258
                                   -3.39 0.001
                                                -.2650755
                                                         - 070976
           read8z | .3255739 .0679117 4.79 0.000
                                                .1924695 .4586783
  c.read8z#c.read8z |
                  .0100256 .0484394
                                   0.21 0.836
                                                -.0849138
                                                          .104965
        faminc8b | .0000303 .0000145
                                   2.09 0.036
                                                1.93e-06 .0000587
c.faminc8b#c.faminc8b | -2.12e-10 1.74e-10 -1.22 0.221 -5.52e-10 1.28e-10
```

Number of obs = 5,671

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
math12z			04278779 .248024987		.043090321 .056728857	9.51 2.10

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support On suppor	Total
Untreated Treated	5 , 079 592	5,079 592
Total	5,671	5,671

Outcome: read12z

Treatment-effect	s estimation	Number of obs	=	5 , 671
Estimator :	propensity-score matching	Matches: requested	=	1
Outcome model :	matching	min	=	1
Treatment model:	logit	max	=	1

read12z	Coefficient		z	P> z	[95% conf.	interval]
ATET catholic (yes vs no)	.0242824	.0426474	0.57	0.569	0593049	.1078698

Logistic regression

LR chi2(20) = 269.04 Prob > chi2 = 0.0000 Pseudo R2 = 0.0709 Log likelihood = -1763.1344

catholic	Coefficient	Std. err.	z	P> z	[95% conf.	interval]
math8z	0605341	.0732046	-0.83	0.408	2040125	.0829443
c.math8z#c.math8z	1680258	.0495161	-3.39	0.001	2650755	070976
read8z	.3255739	.0679117	4.79	0.000	.1924695	.4586783
c.read8z#c.read8z	.0100256	.0484394	0.21	0.836	0849138	.104965
female black hisp api nativam riskdrop1 riskdrop2 disrupt8 nohw8 faminc8b	.0419112 2319429 0828599 -1.165599 4855327 .0232422 4050387 .218962 6282156 .0000303	.0950359 .1917082 .1589371 .2812792 .2864356 .1225025 .2595379 .1285199 .1773147 .0000145	0.44 -1.21 -0.52 -4.14 -1.70 0.19 -1.56 1.70 -3.54 2.09	0.659 0.226 0.602 0.000 0.090 0.850 0.119 0.088 0.000 0.036	1443558 6076841 3943709 -1.716896 -1.046936 2168583 9137236 0329324 975746 1.93e-06	.2281781 .1437983 .2286512 6143023 .0758707 .2633426 .1036463 .4708563 2806852 .0000587
pared1 pared2 pared3 pared4 parmarried collegexpcons	7369748 .1017425 .0949015 .3071211 .2471734 .2901846	.3315306 .156011 .158119 .1506848 .1803656 .1432064 .377756	-2.22 0.65 0.60 2.04 1.37 2.03 -8.61	0.221 0.026 0.514 0.548 0.042 0.171 0.043 0.000	-1.386763 2040334 215006 .0117842 1063367 .0095053 -3.992996	0871868 .4075184 .4048091 .6024579 .6006836 .5708639 -2.51222

Variable Sample | Treated Controls Difference S.E. T-stat read12z Unmatched | .328277963 -.038263548 .366541511 .043159142 8.49
ATT | .328277963 .303995523 .02428244 .057104586 0.43

psmatch2: Treatment assignment	support	I Tota	1				
Untreated Treated		1					
+		-+	-				
Total Outcome: hsg		5,671	L				
Treatment-ef Estimator Outcome mode Treatment mo	: proper	nsity-score ing		Matches	of obs : requested min max	= 1	
		AI rol	oust.				
hsgrad	Coeffic	ient std. 6	err. z	P> z	[95% con	f. interval]	
ATET catholic	 					.0541486	
Logistic reg					LR chi2(20 Prob > chi	obs = 5,671) = 269.04 2 = 0.0000	
Log likeliho	od = -1763	.1344			Pseudo R2	= 0.0709	
					P> z	[95% conf.	interval]
	math8z	+ 0605341	.073204	16 -0.83	0.408		.0829443
c.math8z	#c.math8z	 1680258	3 .049516	51 -3.39	0.001	2650755	070976
	read8z	 .3255739	.067911	.7 4.79	0.000	.1924695	.4586783
c.read8z	#c.read8z	.010025	.048439	0.21	0.836	0849138	.104965
c.faminc8b#c	black hisp api nativam riskdrop1 riskdrop2 disrupt8 nohw8 faminc8b	.0000303	9 .191708 9 .158937 9 .281279 7 .286435 122502 7 .259537 2 .128519 5 .177314 8 .000014	1.70 17 -3.54 15 2.09 .0 -1.22	0.602 0.000 0.090 0.850 0.119 0.088 0.000 0.036	1443558 6076841 3943709 -1.716896 -1.046936 2168583 9137236 0329324 975746 1.93e-06 -5.52e-10 -1.386763 2040334 215006 .0117842 1063367 .0095053 -3.992996	.2286512 6143023 .0758707 .2633426 .1036463 .4708563 2806852 .0000587
		tched .974	4662162 .	910218547	.064443615	S.E .01195636 .012694853	5.39

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment	psmatch2: Common support	
assignment	On suppor	Total
Untreated Treated	5 , 079 592	5,079 592
Total Outcome: inp	5,671 se	5,671

Treatment-effects estimation Estimator : propensity-score matching Outcome model : matching Treatment model: logit	Number of obs Matches: requested min max	=	5,	,671 1 1 1
AI robust				

inpse	 Coefficient	std. err.	Z	P> z	[95% conf.	interval]
ATET catholic		0210202	2 21	0 001	0206544	1156150
(yes vs no)	.0726351 	.0219293	3.31	0.001	.0296544	.1156158

Number of obs = 5,671 LR chi2(20) = 269.04 Prob > chi2 = 0.0000 Pseudo R2 = 0.0709 Logistic regression

Log likelihood = -1763.1344

catholic	Coefficient	Std. err.	Z	P> z	[95% conf.	interval]
math8z	0605341	.0732046	-0.83	0.408	2040125	.0829443
c.math8z#c.math8z	1680258	.0495161	-3.39	0.001	2650755	070976
read8z	.3255739	.0679117	4.79	0.000	.1924695	.4586783
c.read8z#c.read8z	.0100256	.0484394	0.21	0.836	0849138	.104965
female black hisp api nativam riskdrop1 riskdrop2 disrupt8 nohw8 faminc8b	.0419112 2319429 0828599 -1.165599 4855327 .0232422 4050387 .218962 6282156	.0950359 .1917082 .1589371 .2812792 .2864356 .1225025 .2595379 .1285199 .1773147	0.44 -1.21 -0.52 -4.14 -1.70 0.19 -1.56 1.70 -3.54 2.09	0.659 0.226 0.602 0.000 0.090 0.850 0.119 0.088 0.000 0.036	1443558 6076841 3943709 -1.716896 -1.046936 2168583 9137236 0329324 975746 1.93e-06	.2281781 .1437983 .2286512 6143023 .0758707 .2633426 .1036463 .4708563 2806852 .0000587
c.faminc8b#c.faminc8b	 -2.12e-10	1.74e-10	-1.22	0.221	-5.52e-10	1.28e-10
pared1 pared2 pared3 pared4 parmarried collegexp _cons	7369748 .1017425 .0949015 .3071211 .2471734 .2901846 -3.252608	.3315306 .156011 .158119 .1506848 .1803656 .1432064 .377756	-2.22 0.65 0.60 2.04 1.37 2.03 -8.61	0.026 0.514 0.548 0.042 0.171 0.043 0.000	-1.386763 2040334 215006 .0117842 1063367 .0095053 -3.992996	0871868 .4075184 .4048091 .6024579 .6006836 .5708639 -2.51222

Variable Sample | Treated Controls Difference S.E. T-stat

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support On suppor	Total
Untreated Treated	5,079 592	5,079 592
Total	5,671	5,671

```
table _treated _weight
```

```
psmatch2: weight of matched controls
                                      1 2 3 4 Total
psmatch2: Treatment assignment |
                                       442
                                                58
                                                       10
                                                                1
                                                                          511
  Untreated
  Treated
                                       592
                                                                         592
                                    1,034 58 10 1 1,103
  Total
        count if _treated==0 & _weight~=.
 511
. // Part d - estimate ATT and ATE using IPW and
foreach j in math12z read12z hsgrad inpse {
    display in red "Outcome: `j'"
  3.
             _eststo ipwatt`j': teffects ipw (`j') (catholic ///
                    c.math8z##c.math8z c.read8z##c.read8z female black ///
>
                   hisp api nativam riskdrop1 riskdrop2 disrupt nohw8 ///
>
                   c.faminc8b##c.faminc8b pared1-pared4 ///
>
                parmarried collegexp, logit), atet
             _eststo ipwate`j': teffects ipw (`j') (catholic ///
>
                    c.math8z##c.math8z c.read8z##c.read8z female black ///
>
                   hisp api nativam riskdrop1 riskdrop2 disrupt nohw8 ///
>
                   c.faminc8b##c.faminc8b pared1-pared4 ///
               parmarried collegexp, logit), ate
Outcome: math12z
Iteration 0: EE criterion = 7.819e-21
Iteration 1: EE criterion = 1.322e-32
    The Gauss-Newton stopping criterion has been met but missing standard errors
    indicate some of the parameters are not identified.
                                               Number of obs = 5,671
Treatment-effects estimation
Estimator : inverse-probability weights
Outcome model : weighted mean
Treatment model: logit
                                     Robust
            math12z | Coefficient std. err.
                                                z P>|z| [95% conf. interval]
ATET
            catholic |
        (yes vs no)
                                                                  .0755148
                         .1208028 .0231066
                                                5.23 0.000
                                                                              .1660909
POmean
            catholic |
             no | .2462904 .0319048 7.72 0.000 .1837582 .3088226
Warning: Convergence not achieved.
Iteration 0: EE criterion = 7.819e-21
Iteration 1: EE criterion = 3.839e-34
    The Gauss-Newton stopping criterion has been met but missing standard errors
    indicate some of the parameters are not identified.
Treatment-effects estimation
                                              Number of obs = 5,671
Estimator : inverse-probability weights
Outcome model : weighted mean
Treatment model: logit
                                    Robust
            math12z | Coefficient std. err.
                                                                [95% conf. interval]
                                                  z P>|z|
ATE
        catholic | (yes vs no) | .0983067 .0546021 1.80 0.072 -.0087113
```

POmean

catholic |

no | -.0126136 .0134597 -0.94 0.349 -.0389941 .0137669

Warning: Convergence not achieved.

Outcome: read12z

Iteration 0: EE criterion = 7.819e-21
Iteration 1: EE criterion = 6.888e-34

The Gauss-Newton stopping criterion has been met but missing standard errors

indicate some of the parameters are not identified.

Treatment-effects estimation Number of obs = 5,671

Estimator : inverse-probability weights
Outcome model : weighted mean

Treatment model: logit

_____ Robust read12z | Coefficient std. err. z P>|z| [95% conf. interval] ATET catholic | (yes vs no) | .0412662 .0277719 1.49 0.137 -.0131658 .0956982 catholic L no i .2870117 .0293521 9.78 0.000 .2294827 .3445408

Warning: Convergence not achieved.

Iteration 0: EE criterion = 7.819e-21
Iteration 1: EE criterion = 1.166e-33

The Gauss-Newton stopping criterion has been met but missing standard errors indicate some of the parameters are not identified.

Treatment-effects estimation Number of obs = 5,671

Estimator : inverse-probability weights
Outcome model : weighted mean
Treatment model: logit

Robust read12z | Coefficient std. err. z P>|z| [95% conf. interval] ATE catholic | (yes vs no) | .0210077 .0602739 0.35 0.727 -.0971271 .1391425 POmean catholic | no | -.0043111 .0135662 -0.32 0.751 -.0309003 .0222782 _____

Warning: Convergence not achieved.

Outcome: hsgrad

Iteration 0: EE criterion = 7.819e-21
Iteration 1: EE criterion = 7.463e-34

The Gauss-Newton stopping criterion has been met but missing standard errors indicate some of the parameters are not identified.

Treatment-effects estimation Number of obs = 5,671

Estimator : inverse-probability weights
Outcome model : weighted mean

Treatment model: logit

______ Robust hsgrad | Coefficient std. err. z P>|z| [95% conf. interval] _____ ATET catholic | (yes vs no) | .0221956 .0071377 3.11 0.002 .0082059 . 0361853 POmean catholic | no | .9524665 .003826 248.95 0.000 .9449677 .9599654

Warning: Convergence not achieved.

Iteration 0: EE criterion = 7.819e-21
Iteration 1: EE criterion = 4.915e-34

The Gauss-Newton stopping criterion has been met but missing standard errors indicate some of the parameters are not identified.

Number of obs = 5,671 Treatment-effects estimation

Estimator : inverse-probability weights
Outcome model : weighted mean

Treatment model: logit

______ Robust hsgrad | Coefficient std. err. z P>|z| [95% conf. interval] ATE catholic | (yes vs no) | .0290058 .0209966 1.38 0.167 -.0121467 .0701584 POmean catholic | no | .9146284 .0038238 239.19 0.000 .9071339 .922123

Warning: Convergence not achieved.

Outcome: inpse

Iteration 0: EE criterion = 7.819e-21
Iteration 1: EE criterion = 1.478e-32

The Gauss-Newton stopping criterion has been met but missing standard errors indicate some of the parameters are not identified.

Treatment-effects estimation Number of obs = 5,671

Estimator : inverse-probability weights
Outcome model : weighted mean

Treatment model: logit

______ Robust inpse | Coefficient std. err. z P>|z| [95% conf. interval] ______ ATET catholic | (yes vs no) | .0636729 .0148925 4.28 0.000 .0344843 .0928616 catholic | no | .7978135 .008995 88.70 0.000 .7801837 .8154434 -----

Warning: Convergence not achieved.

Iteration 0: EE criterion = 7.819e-21
Iteration 1: EE criterion = 2.078e-32

The Gauss-Newton stopping criterion has been met but missing standard errors indicate some of the parameters are not identified.

Number of obs = 5,671 Treatment-effects estimation

Estimator : inverse-probability weights
Outcome model : weighted mean

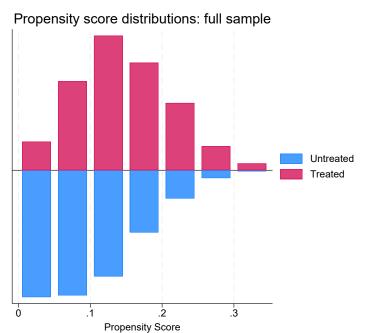
Treatment model: logit

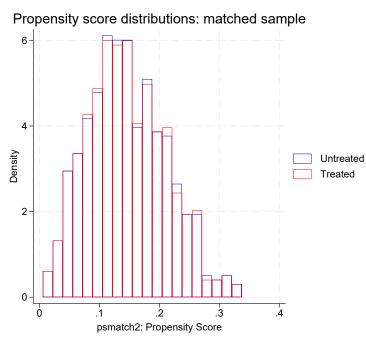
-----Robust inpse | Coefficient std. err. z P>|z| [95% conf. interval] ATE catholic | (yes vs no) | .1301173 .0200468 6.49 0.000 .0908263 .1694082 POmean catholic | no | .7025744 .0062853 111.78 0.000 .6902554 .7148934

Warning: Convergence not achieved.

```
esttab all using PS3estimates.csv, se paren csv replace
(output written to PS3estimates.csv)
. // ****************
. // Part e - try psestimate
. // ***************
      ssc install psestimate
checking psestimate consistency and verifying not already installed...
all files already exist and are up to date.
       psestimate catholic, totry(math8z read8z female black hisp ///
>
             api riskdrop1 riskdrop2 disrupt nohw8 faminc8b pared1-pared4 ///
             parmarried collegexp) genpscore(pscore)
Selecting first order covariates... (153)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5 .....s...s...s..
.....s....s....s....s....s...
Selected first order covariates are: faminc8b read8z api riskdrop2 nohw8 pared1 math8z col
> legexp pared4 black parmarried disrupt8
Selecting second order covariates... (3,081)
1 ---+-- 1 ---+-- 2 ---+-- 3 ---+-- 4 ---+-- 5
100
150
....S......
                                      200
.....s....s.....
                                      250
                                      300
350
....s...s...s...
450
                                      500
..s......
                                      550
....s...s...
                                      600
                                      650
700
Selected second order covariates are: c.math8z#c.math8z c.disrupt8#c.collegexp c.black#c.c
> ollegexp c.pared1#c.read8z c.black#c.faminc8b c.pared1#c.faminc8b c.math8z#c.api c.parma
> rried#c.riskdrop2 c.disrupt8#c.black
Final model is: faminc8b read8z api riskdrop2 nohw8 pared1 math8z collegexp pared4 black p
> armarried disrupt8 c.math8z#c.math8z c.disrupt8#c.collegexp c.black#c.collegexp c.pared1
> #c.read8z c.black#c.faminc8b c.pared1#c.faminc8b c.math8z#c.api c.parmarried#c.riskdrop2
> c.disrupt8#c.black
       \ensuremath{//} Close log and convert to PDF
       capture log close
```

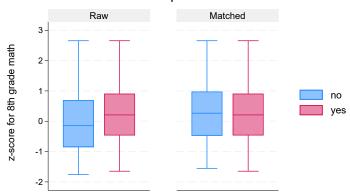
Propensity score distributions: after psmatch2



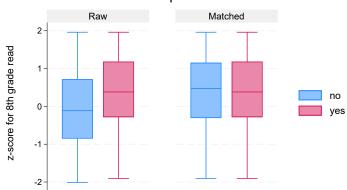


Covariate balance: after teffects psmatch

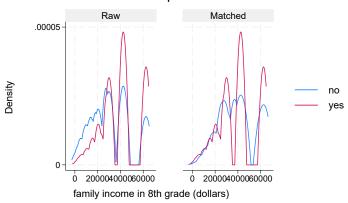
Balance plot



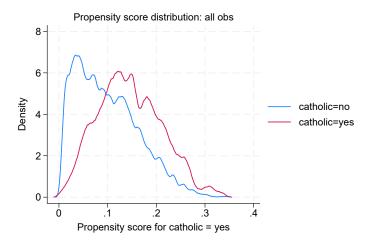
Balance plot



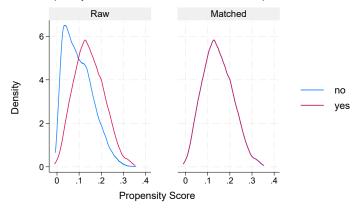
Balance plot

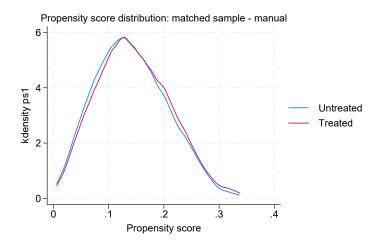


Propensity score distributions: after teffects psmatch



Propensity score distribution: all and matched sample





	PSM (1) math12z	PSM (2) read12z	PSM (3) hsgrad	PSM (4) inpse	(5)	IPW-ATE (6) math12z	(7)	(8)	IPW-ATT (9) hsgrad	IPW-ATE (10) hsgrad	IPW-ATT (11) inpse	IPW-ATE (12) inpse
main r1vs0.catholic	0.119** (0.0424)	0.0243 (0.0426)	0.0304* (0.0121)		0.121*** (0.0231)	0.0983 (0.0546)	0.0413 (0.0278)	0.0210 (0.0603)	0.0222** (0.00714)	0.0290 (0.0210)	0.0637*** (0.0149)	
POmean 0.catholic					0.246*** (0.0319)	-0.0126 (0.0135)	0.287*** (0.0294)		0.952*** (0.00383)			
TME1 math8z					-0.0605 (0.0720)	-0.0605 (0.0720)	-0.0605 (0.0720)	-0.0605 (0.0720)	-0.0605 (0.0720)	-0.0605 (0.0720)	-0.0605 (0.0720)	-0.0605 (0.0720)
c.math8z#c.math8z					-0.168*** (0.0504)	-0.168*** (0.0504)	-0.168*** (0.0504)	-0.168*** (0.0504)		-0.168*** (0.0504)	-0.168*** (0.0504)	-0.168*** (0.0504)
read8z					0.326*** (0.0681)	0.326*** (0.0681)	0.326*** (0.0681)	0.326*** (0.0681)	0.326*** (0.0681)	0.326*** (0.0681)	0.326*** (0.0681)	0.326*** (0.0681)
c.read8z#c.read8z					0.0100 (0.0491)	0.0100 (0.0491)	0.0100 (0.0491)	0.0100 (0.0491)	0.0100 (0.0491)	0.0100 (0.0491)	0.0100 (0.0491)	0.0100 (0.0491)
female					0.0419 (0.0939)	0.0419 (0.0939)	0.0419 (0.0939)	0.0419 (0.0939)	0.0419 (0.0939)	0.0419 (0.0939)	0.0419 (0.0939)	0.0419 (0.0939)
black					-0.232 (0.190)	-0.232 (0.190)	-0.232 (0.190)	-0.232 (0.190)	-0.232 (0.190)	-0.232 (0.190)	-0.232 (0.190)	-0.232 (0.190)
hisp					-0.0829 (0.157)	-0.0829 (0.157)	-0.0829 (0.157)	-0.0829 (0.157)	-0.0829 (0.157)	-0.0829 (0.157)	-0.0829 (0.157)	-0.0829 (0.157)
api					-1.166*** (0.282)	-1.166*** (0.282)	-1.166*** (0.282)	-1.166*** (0.282)	-1.166*** (0.282)	-1.166*** (0.282)	-1.166*** (0.282)	-1.166*** (0.282)
nativam					-0.486 (0.283)	-0.486 (0.283)	-0.486 (0.283)	-0.486 (0.283)	-0.486 (0.283)	-0.486 (0.283)	-0.486 (0.283)	-0.486 (0.283)
riskdrop1					0.0232 (0.117)	0.0232 (0.117)	0.0232 (0.117)	0.0232 (0.117)	0.0232 (0.117)	0.0232 (0.117)	0.0232 (0.117)	0.0232 (0.117)
riskdrop2					-0.405 (0.242)	-0.405 (0.242)	-0.405 (0.242)	-0.405 (0.242)	-0.405 (0.242)	-0.405 (0.242)	-0.405 (0.242)	-0.405 (0.242)
disrupt8					0.219 (0.127)	0.219 (0.127)	0.219 (0.127)	0.219 (0.127)	0.219 (0.127)	0.219 (0.127)	0.219 (0.127)	0.219 (0.127)
nohw8					-0.628*** (0.177)	-0.628*** (0.177)	-0.628*** (0.177)	-0.628*** (0.177)	-0.628*** (0.177)	-0.628*** (0.177)	-0.628*** (0.177)	-0.628*** (0.177)
faminc8b												0.0000303° 0.0000151
c.faminc8b#c.famino									-2.12e-10 (1.80e-10)			-2.12e-10 (1.80e-10)
pared1					-0.737* (0.323)	-0.737* (0.323)	-0.737* (0.323)	-0.737* (0.323)	-0.737* (0.323)	-0.737* (0.323)	-0.737* (0.323)	-0.737* (0.323)
pared2					0.102 (0.156)	0.102 (0.156)	0.102 (0.156)	0.102 (0.156)	0.102 (0.156)	0.102 (0.156)	0.102 (0.156)	0.102 (0.156)
pared3					0.0949 (0.158)	0.0949 (0.158)	0.0949 (0.158)	0.0949 (0.158)	0.0949 (0.158)	0.0949 (0.158)	0.0949 (0.158)	0.0949 (0.158)

	PSM (1) math12z	PSM (2) read12z	PSM (3) hsgrad	PSM (4) inpse	IPW-ATT (5) math12z	IPW-ATE (6) math12z	IPW-ATT (7) read12z	IPW-ATE (8) read12z	IPW-ATT (9) hsgrad	IPW-ATE (10) hsgrad	IPW-ATT (11) inpse	IPW-ATE (12) inpse
pared4					0.307* (0.151)	0.307* (0.151)	0.307* (0.151)	0.307* (0.151)	0.307* (0.151)	0.307* (0.151)	0.307* (0.151)	0.307* (0.151)
parmarried					0.247 (0.170)	0.247 (0.170)	0.247 (0.170)	0.247 (0.170)	0.247 (0.170)	0.247 (0.170)	0.247 (0.170)	0.247 (0.170)
collegexp					0.290* (0.147)	0.290* (0.147)	0.290* (0.147)	0.290* (0.147)	0.290* (0.147)	0.290* (0.147)	0.290* (0.147)	0.290* (0.147)
_cons					-3.253*** (0.362)	-3.253*** (0.362)	-3.253*** (0.362)	-3.253*** (0.362)	-3.253*** (0.362)	-3.253*** (0.362)	-3.253*** (0.362)	-3.253*** (0.362)
N	5671	5671	5671	5671	5671	5671	5671	5671	5671	5671	5671	5671

** p<0.01'** p<0.001"