Propensity Score Analysis Using `teffects` in Stata

SOC 561

Programming for the Social Sciences

Hyungjun Suh

Apr. 25. 2016

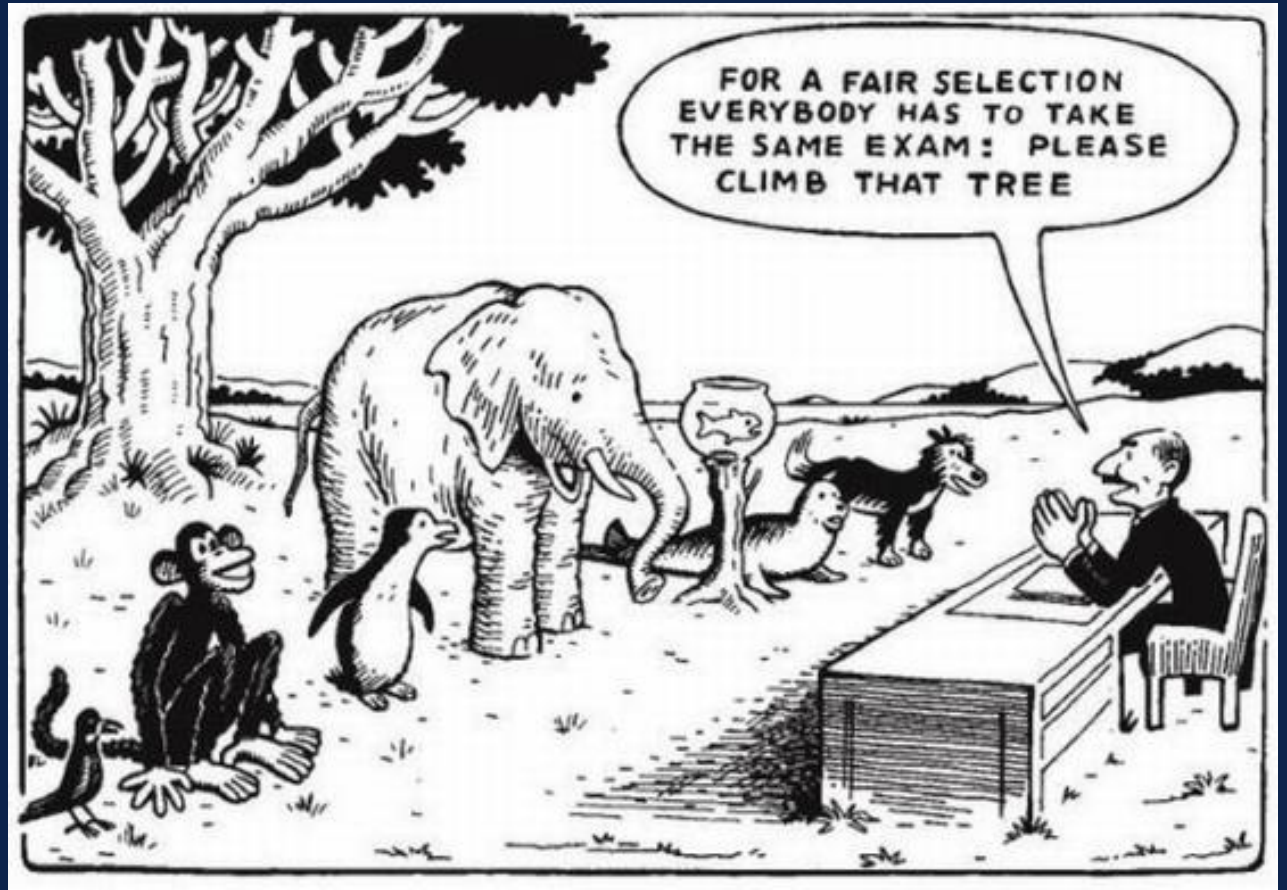
Overview

- Motivation
- Propensity Score Weighting
- Propensity Score Matching with `teffects`
- Example

Motivation: What is PSA?

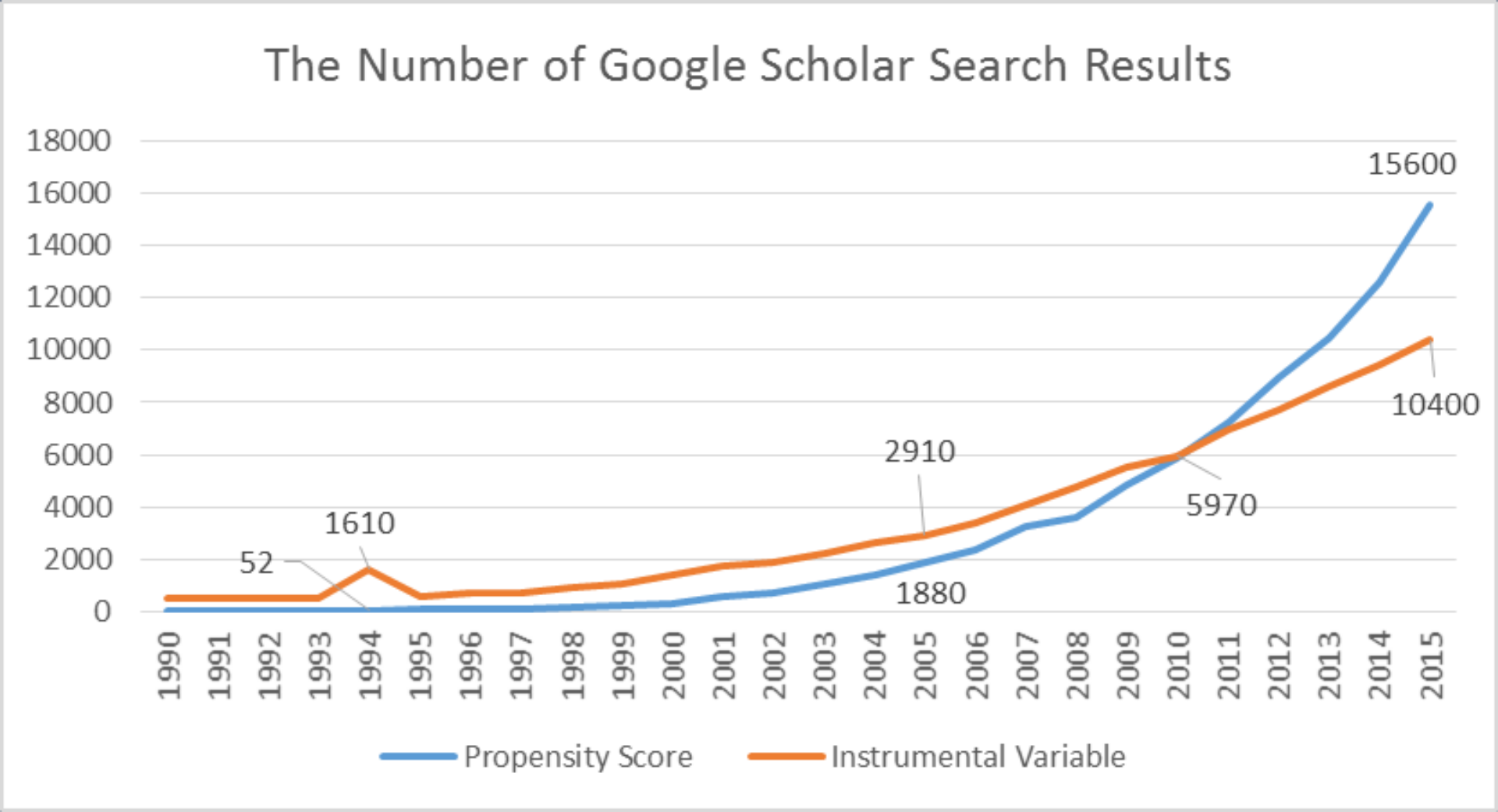
- Propensity Score Analysis (PSA): a method to estimate treatment effects with nonexperimental or observational data (Guo and Fraser 2015: 1)
- Propensity Score: the conditional probability of assignment to a particular treatment given a vector of covariates (Everitt & Skrondal 2010:345). -> **predicted probability.**

Motivation: Why is PSA important?



Motivation: Why is PSA important?

- PSA can address a selection bias of an independent variable.
- PSA demonstrates a causation, while covariate control remains an association.
- Thus, practically, it will give you an edge in publication ☺


The Number of Google Scholar Search Results
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
15600
10400
2910
5970
1610
52
1880
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
Propensity Score
Instrumental Variable

Glossary

- Treatment group: a group which is assigned a certain condition.
- Control group: a group which is not assigned the certain condition.
- eg1) new drug test: those who took a drug (the treated) / otherwise (the controlled).
- eg2) homeownership: homeowners (the treated) / non-owners (the controlled).
- Propensity score: predicted probability

Glossary

- Potential-outcome means (POMs): the means of Y_1 and Y_0 in the **population**. Here, Y_1 refers to the outcome of the treatment group, and Y_0 refers to the outcome of the control group.
- Average Treatment Effect (ATE): the mean of the difference (Y_1 - Y_0). The average effect of a binary independent variable on an outcome.
- Average Treatment Effect for the Treated (ATT or ATET): the means of the difference (Y_1 - Y_0) among the treated.

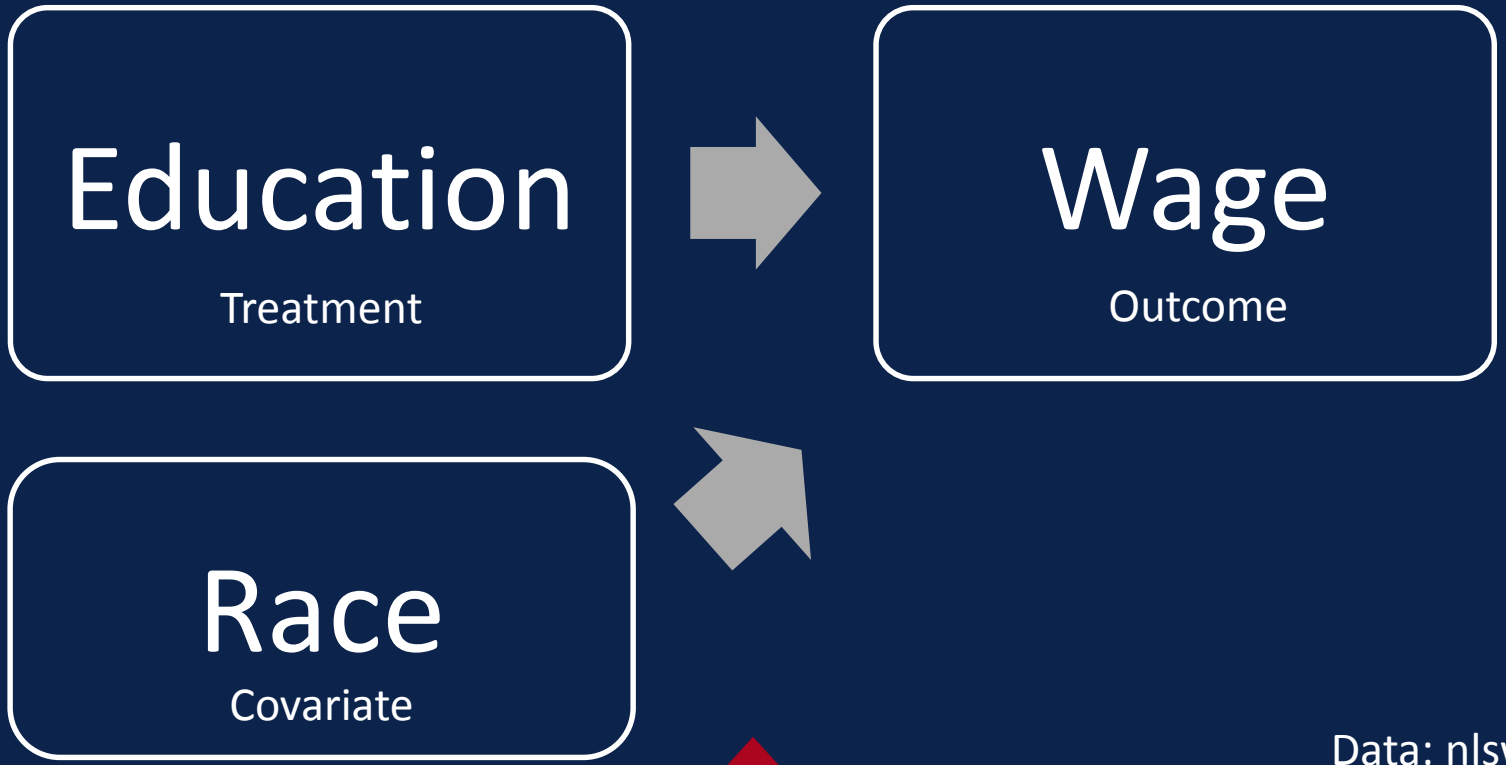
General Comments

- Various kinds of PSA
- Propensity score matching, weighting, subclassification, among others.
- This presentation focuses on a binary treatment: multinomial and continuous treatments are also possible.
- This presentation focuses on application perspective. For theoretical and mathematical explanation, see Guo and Fraser (2015).

Propensity Score Weighting

- Strengths of PSW
- It allows to use most types of multivariate analysis.
- It allows to use most observations unlike propensity score matching.

Running Example: Education and Wage



Data: nlsw88.dta

Running Example: Education and Wage

```
. sum wage

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        wage |      2,246    7.766949    5.755523    1.004952   40.74659

. tab1 collgrad white

-> tabulation of collgrad

college graduate |      Freq.     Percent        Cum.
-----------------+-----------------------------------
not college grad |      1,714       76.31       76.31
    college grad |        532       23.69      100.00
-----------------+-----------------------------------
           Total |      2,246      100.00

-> tabulation of white

 race white |      Freq.     Percent        Cum.
------------+-----------------------------------
  non-white |        609       27.11       27.11
      white |      1,637       72.89      100.00
------------+-----------------------------------
      Total |      2,246      100.00
```

Propensity Score Weighting

- Step1: check data balance with T-tests or bivariate linear regressions.

$$X = \beta_1 Treatment + \beta_0$$

where X is equal to each covariate (logit(X) if the covariate is binary, X if the covariate is continuous).

A significant coefficient (β_1) means that there is a significant difference between the treated and the controlled. Consequently, **to be balanced, β_1 should be insignificant.**

Propensity Score Weighting

- If all covariates are already balanced, the researcher may move on to main analyses (albeit it is a rare case in social sciences). In this case, the researcher do not have to control covariates (actually, insignificant variables are not even covariates!).
- Otherwise, significant coefficients mean there is a selection bias, and the researcher needs to address it.

Running Example: Education and Wage

```
. logit white collgrad

Iteration 0:   log likelihood =  -1312.558
Iteration 1:   log likelihood = -1305.8609
Iteration 2:   log likelihood = -1305.8297
Iteration 3:   log likelihood = -1305.8297

Logistic regression                             Number of obs     =      2,246
                                                LR chi2(1)        =      13.46
                                                Prob > chi2       =     0.0002
Log likelihood = -1305.8297                     Pseudo R2         =     0.0051

------------------------------------------------------------------------------
       white |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    collgrad |   .4262018   .1189256     3.58   0.000     .1931119    .6592916
       _cons |   .8955541   .0532331    16.82   0.000     .7912191    .9998891
------------------------------------------------------------------------------
```

Propensity Score Weighting

- Step2: obtain a propensity score.

$$logit(\widehat{Treatment}) = \boldsymbol{XB} + \epsilon$$

where **X** is a covariate vector and **B** is a vector of coefficients.

"a careful selection of conditioning variables and a correct specification of the logistic regression are crucial to propensity score matching" (Guo and Fraser 2015:141).

```
. logit collgrad white // DV: treatment

Iteration 0:   log likelihood =   -1229.55
Iteration 1:   log likelihood = -1222.8571
Iteration 2:   log likelihood = -1222.8218
Iteration 3:   log likelihood = -1222.8218

Logistic regression                             Number of obs     =      2,246
                                                LR chi2(1)        =      13.46
                                                Prob > chi2       =     0.0002
Log likelihood = -1222.8218                     Pseudo R2         =     0.0055

------------------------------------------------------------------------------
    collgrad |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       white |   .4262018   .1189256     3.58   0.000     .1931119    .6592916
       _cons |  -1.490091   .1045975   -14.25   0.000    -1.695099   -1.285084
------------------------------------------------------------------------------

. predict preprob, pr

. list white collgrad preprob in 1/10

     +-------------------------------------------+
     |     white          collgrad    preprob |
     |-------------------------------------------|
  1. | non-white   not college grad   .183908 |
  2. | non-white   not college grad   .183908 |
  3. | non-white   not college grad   .183908 |
  4. |     white       college grad  .2565669 |
  5. |     white   not college grad  .2565669 |
```

Propensity Score Weighting

- Step3: calculate propensity score weights with a formula (Guo and Fraser 2015: 245).

$$Weight = \frac{T}{P} + \frac{1 - T}{(1 - P)}$$

T = a binary treatment. The treated (T=1) and the controlled (T=0).

P = the obtained propensity score.

In sum, weights are 1/P for the treated and 1/(1-P) for the controlled.

Running Example: Education and Wage

```
. gen psweight=.
(2,246 missing values generated)

. replace psweight=(1/preprob) if collgrad==1 // the treated = college graduate
(532 real changes made)

. replace psweight=(1/(1-preprob)) if collgrad==0 // the controlled == non-college graduate
(1,714 real changes made)

. list white preprob psweight in 1/10
```

	white	preprob	psweight
1.	non-white	.183908	1.225352
2.	non-white	.183908	1.225352
3.	non-white	.183908	1.225352
4.	white	.2565669	3.897619
5.	white	.2565669	1.345111

Propensity Score Weighting

- Step4: check data balance applying the obtained propensity score weights.

Running Example: Education and Wage

```
. logit white collgrad [pweight=psweight]

Iteration 0:   log pseudolikelihood =  -2625.116
Iteration 1:   log pseudolikelihood =  -2625.116  (backed up)

Logistic regression                             Number of obs     =      2,246
                                                Wald chi2(1)      =       0.00
                                                Prob > chi2       =     1.0000
Log pseudolikelihood =  -2625.116               Pseudo R2         =     0.0000

------------------------------------------------------------------------------
             |               Robust
       white |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    collgrad |   3.66e-15   .118952     0.00   1.000    -.2331417    .2331417
       _cons |   .9888023   .053245    18.57   0.000     .8844441    1.093161
------------------------------------------------------------------------------
```

Propensity Score Weighting

- Step5: perform main analyses applying the obtained weights.

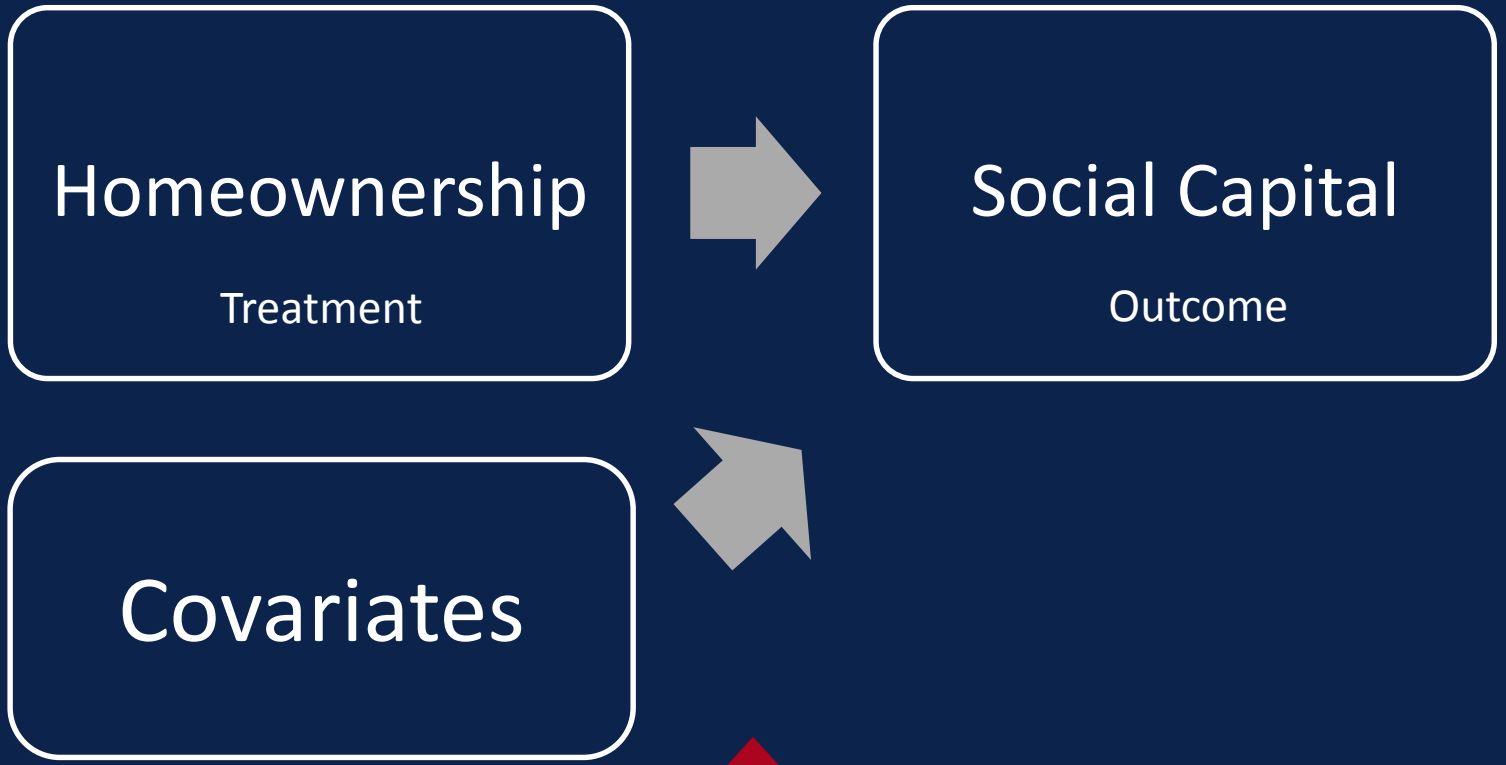
Running Example: Education and Wage

```
. reg wage collgrad [pweight=psweight]
(sum of wgt is   4.4920e+03)

Linear regression                               Number of obs     =      2,246
                                                F(1, 2244)        =     135.38
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0877
                                                Root MSE          =     5.8941

------------------------------------------------------------------------------
             |               Robust
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    collgrad |   3.654446   .3140864    11.64   0.000     3.038515    4.270376
       _cons |   6.937017   .1292918    53.65   0.000     6.683473    7.190561
------------------------------------------------------------------------------
```

More Complex Example: Homeownership and Social Capital



Covariates

- Sex
- Age
- Employment status
- Employment sector
- Material status
- Education
- Settlement type
- Household income
- Ethnicity
- Religion
- Housing amenities
- Housing environment
- Living with extended family

Example: Homeownership and Social Capital

```
* data balance check with the derived propensity score
* each covariate becomes a dependent variable for regression. (Guo and Fraser 2015: 247)
* make matrix in which results are stored.
matrix res = J(16, 2, .)
* column = p-value of coefficients
matrix colnames res = unweighted weighted
matrix rownames res = extfamB female educB singleB sepaB employed urban ethnicB ///
    islamB otherrelB workorgG qualamen qualenvir satnbhdR age re_hhinc_r
local irow = 0
* binary variables
foreach i in extfamB female educB singleB sepaB employed urban ethnicB islamB otherrelB workorgG {
    local ++irow
    logit `i' hhowner
    qui test hhowner
    matrix res[`irow',1] = r(p)
    logit `i' hhowner [pweight=psweight]
    qui test hhowner
    matrix res[`irow',2] = r(p)
}
```

Table 5 Data Balance Before and After Propensity Score Weight Adjustment

	Homeowners vs. Non-owners		Respondent Owners vs. Family Owners	
Covariate	Unweighted P-value	Weighted P-value	Unweighted P-value	Weighted P-value
Living w/ Extended Family	0.000	0.037	0.000	0.350
Female	0.176	0.067	0.442	0.648
Bachelor's Degree	0.006	0.691	0.002	0.762
Single	0.122	0.867	0.000	0.249
Separated/Divorced/ Widowed	0.022	0.951	0.000	0.837
Unemployed	0.971	0.775	0.003	0.758
Other Jobs	0.007	0.274	0.000	0.298
Urban	0.004	0.376	0.003	0.874
Ethnic Russian	0.808	0.298	0.000	0.570
Islam	0.179	0.334	0.000	0.566
Other Religions	0.017	0.819	0.037	0.738
Working at Gvm't-related Org.	0.001	0.864	0.089	0.621
Housing Amenities	0.005	0.336	0.068	0.665
Housing Environment	0.981	0.494	0.249	0.747
Neighborhood Satisfaction	0.108	0.678	0.130	0.756
Age	0.000	0.366	0.000	0.298
Monthly Household Income	0.384	0.126	0.006	0.658

Note: P-values of the coefficient of homeownership and property rights are displayed.

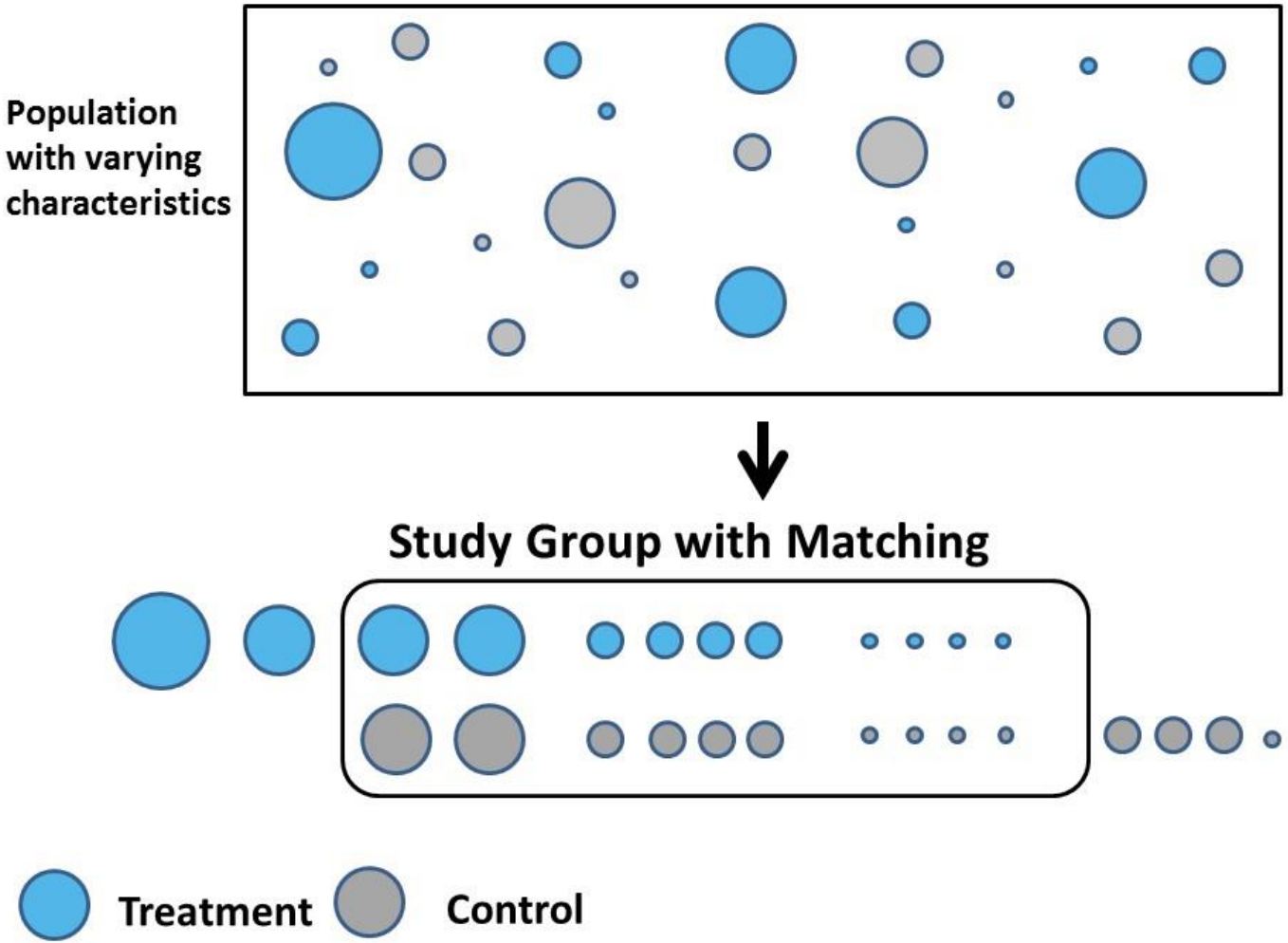
Results after PSW Adjustment

```
. reg netO hhowner [pweight=psweight]
(sum of wgt is   4.2790e+03)

Linear regression                               Number of obs     =      2,207
                                                F(1, 2205)        =       6.69
                                                Prob > F          =     0.0098
                                                R-squared         =     0.0071
                                                Root MSE          =      1.565
```

netO	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
hhowner	.2650957	.1024964	2.59	0.010	.0640962	.4660953
_cons	2.528809	.0949365	26.64	0.000	2.342635	2.714984

Understanding Matching



Source: http://cdn2.hubspot.net/hubfs/355318/images/Propensity-Score-Graphic.jpg?t=1461337172841

Propensity Score Matching with `teffects`

1) How to install

- It is a built-in command. **Stata 13.1 or greater is required.**

2) Basic Syntax

- **`teffects psmatch`** `(outcome) (treatment covariates),` **`nn(#) gen(`**`match`**`)`

teffects

3) Examples

- **teffects psmatch** (wage) (collgrad white)

Outcome — Treatment — Covariate

```
Treatment-effects estimation                Number of obs      =      2,246
Estimator      : propensity-score matching  Matches: requested =          1
Outcome model  : matching                                  min =        112
Treatment model: logit                                     max =       1217
```

wage	Coef.	AI Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE						
collgrad						
(college grad vs not college grad)	3.654446	.3142033	11.63	0.000	3.038619	4.270273

teffects

3) Examples: matching with multiple neighbors

```
. teffects psmatch (wage) (collgrad white age south smsa c_city), nn(4) // 'nn' means 'nearest neighbor.'

Treatment-effects estimation                   Number of obs      =      2,246
Estimator      : propensity-score matching     Matches: requested =          4
Outcome model  : matching                                  min =          4
Treatment model: logit                                     max =         54
```

	wage	AI Robust Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE							
	collgrad						
(college grad vs not college grad)		3.299533	.3118173	10.58	0.000	2.688382	3.910683

teffects

3) Examples: Post Estimation

- (after a teffects psmatch analysis)
- **Predict** ps0 ps1, **ps** // get propensity scores
- **Predict** y0 y1, **po** // get potential outcomes
- **Predict** te, **te** // get treatment effects, $(Y_1 - Y_0)$.

```
. list white collgrad wage match1 ps0 ps1 y0 y1 te if ob==1 | ob==2 | ob==1671 | ob==441, nolabel
```

	white	collgrad	wage	match1	ps0	ps1	y0	y1	te
1.	0	0	11.73913	1671	.8291792	.1708208	11.73913	12.77777	1.03865
2.	0	0	6.400963	441	.7862242	.2137758	6.400963	12.89855	6.497582
441.	0	1	11.22383	2	.7862242	.2137758	6.830397	11.22383	4.393431
1671.	0	1	12.77777	1	.8291792	.1708208	9.384411	12.77777	3.393364

Other Methods of `teffects`

- teffects ra (Regression Adjustment)
- teffects ipw (Inverse Probability Weighting)
- teffects aipw (Augmented Inverse Probability Weighting)
- teffects ipwra (Inverse Probability Weighted Regression Adjustment)
- teffects nnmatch (Nearest Neighbor Mathcing)

See Stata Treatment-Effects Reference Manual for more information.

References

- Everitt, B.S. & A. Skrondal. 2010. The Cambridge Dictionary of Statistics. 4th edition. Cambridge University Press.
- Guo, S. & M. Fraser. 2015. Propensity Score Analysis. 2nd edition. Sage.
- Propensity Score Matching in Stata using teffects (https://www.ssc.wisc.edu/sscc/pubs/stata_psmatch.htm)
- Stata Treatment-Effects Reference Manual Release 14.

- Questions and Comments to

  suhhyungjun@email.arizona.edu


תודה
Dankie
Gracias
Спасибо
شكراً
Merci
Takk
Köszönjük
Terima kasih
Grazie
Dziękujemy
Dėkojame
Ďakujeme
Vielen Dank
Paldies
Kiitos
Täname teid
谢谢
Thank You
Tak
感謝您
Obrigado
Teşekkür Ederiz
감사합니다
Σας Ευχαριστούμ
ขอบคุณ
Bedankt
Děkujeme vám
ありがとうございます
Tack