Propensity Score Matching (PSM)

An introduction

Dr Ge Yu, ge.yu@ncl.ac.uk

Background

- In Public Health for many intervention/policy it is infeasible or unethical to assign subjects to different treatments
- There may be important differences in individual characteristics between treatment and control groups in non-experimental studies
- Thus, it is not straightforward to estimate the causal impact of an intervention/policy

What is Propensity Score Matching

- Quasi-experimental approach to estimate causal impact of an intervention/policy
- Uses observable characteristics to match people who were exposed to the intervention/policy to similar people who were not.
- Model assumes that these observable characteristics are the only factors that influence people’s outcomes from receiving the intervention/policy.

Propensity Score

- It is the probability of receiving treatment rather than control, given individual characteristics/covariates X.
- Define E=1 for treatment and E=0 for control
- We will denote the propensity score for subject i by ps_i

$$ps_i = P(E = 1|X_i)$$

Propensity Score

- Suppose age is the only X covariate, and older people were more likely to get treatment
- Then the PS would be larger for older ages:

$$ps_i > ps_j \; if \; age_i > age_j$$

- If subject i has a propensity score value of 0.4, that means that, given his/her particular variable values, there is a 40% chance he/she will be treated

Propensity Score

Suppose 2 subjects have the same value of the PS, but they slightly different (e.g. one lives in Newcastle and one lives in Birmingham). Despite the different covariate values, they were both equally likely to have been treated

- Both subjects' X is just as likely to be found in the treatment group
- If you restrict to a subpopulation of subjects who have the same value of the PS, there should be balance in the two treatment groups
- Implication: if we match on the PS, we should achieve balance

Estimated Propensity Score

- In an RCT, the PS is generally known.

  (e.g. $P(E=1|X)=P(E=1)=0.5$

- In an observational study, it will be unknown.
  - Notice, however, that the PS involves observed data: E and X
    - We therefore can estimate it
    - Typically when people talk about a PS, they are referring to the estimated PS.

Estimated Propensity Score

We need to estimate $P(E = 1|X)$:

- The outcome here is E, which is binary.
- Steps
  1. To select variables
     - ➢ All predictors of being treated
     - ➢ None of the effects of the treatment
  2. To select model for creating PS (e.g. logistic regression)
  3. To get the predicted probability (fitted value) for each subject

     It is NOT for inferential purposes

Propensity Score Matching (PSM)

- We previously noted that the PS is a balancing score.

  Matching on the PS should achieve balance.

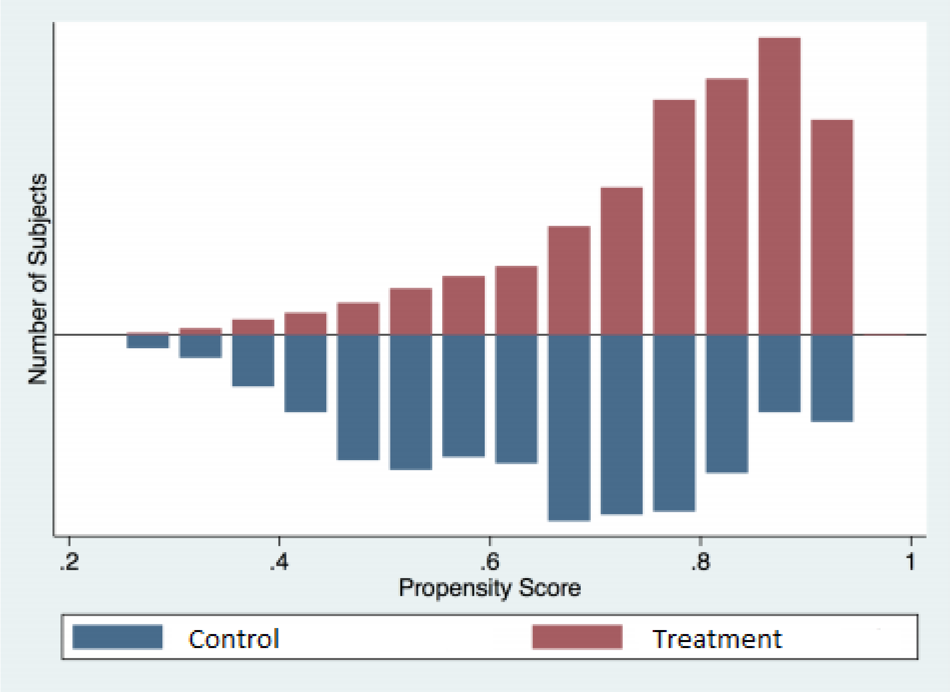
- The PS is a scalar – each subject will have exactly one value of the PS.

  The matching problem is simplified, in that we are only matching on one variable

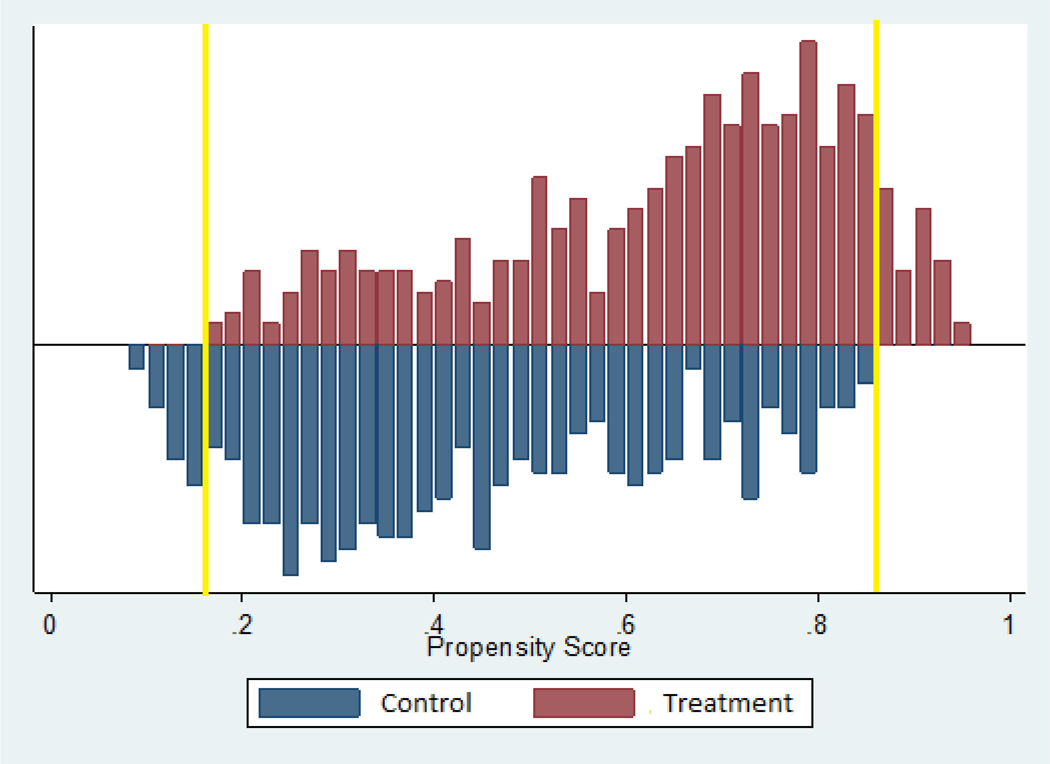
- Overlap

  To compare the distribution of the PS for treatment and control groups in a plot

PSM: overlap example



PSM: overlap example



PSM: matching

- There are two broad groups of matching estimators:
  - Individual matching: is a one to one match of the control and treatment group (e.g. nearest neighbour matching, caliper & radius matching)
  - Weighted matching: is where a weighted average is estimated for the control group. The weight given to a control individual, is in proportion to the closeness of observable characteristics in the control and treatment groups (e.g. kernel based matching, Mahalaobis metric matching)

PSM: individual matching

- Nearest neighbour matching
  - An individual in the treatment group is matched to an individual in the control group
- Caliper matching
  - For a pre-specified maximum tolerated difference between the treatment and control groups (specified as a standard deviation, $\delta>0$), an individual in the treatment group is matched to an individual in the control group
  - If the treated individual is not within the SD of a control individual, then this treated individual is left unmatched

PSM: weight matching

- Kernel based matching
    - A method is to give larger weight to controls with smaller distances using some kernel function with specified bandwidth.
- Mahalanobis metric matching
    - A method is to match based on a distance metric that measures the proximity between observations in the multivariate space of covariates.
    - The idea then is to use observations that are “close”, but not necessarily equal, as matches.

PSM: transformation

- In practice, logit (log-odds) of the PS is often used, rather than the PS itself
    - The PS is bounded between 0 and 1, making many values seem similar.
    - Logit of the PS is unbounded, this transformation essentially stretches the distribution, while preserving ranks.
    - Match on $logit(\pi)$ rather than π.

PSM: after matching

- The outcome analysis methods can be the same as would be used in any other matched analysis
  - Randomisation tests
  - Conditional logistic regression, GEE, stratified Cox model

PSM: matching in Stata

- Install ‘psmatch2’ package

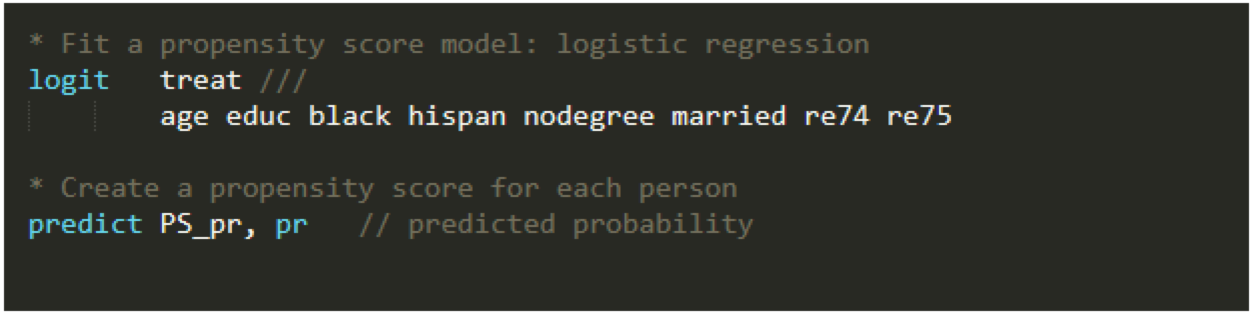
  ‘ssc install psmatch2’

- ‘psmatch2’: implements various types of PSM estimators
  - One-to-one, k-nearest neighbours, radius, kernel, local linear regression, spline, Mahalanobis

Example: data

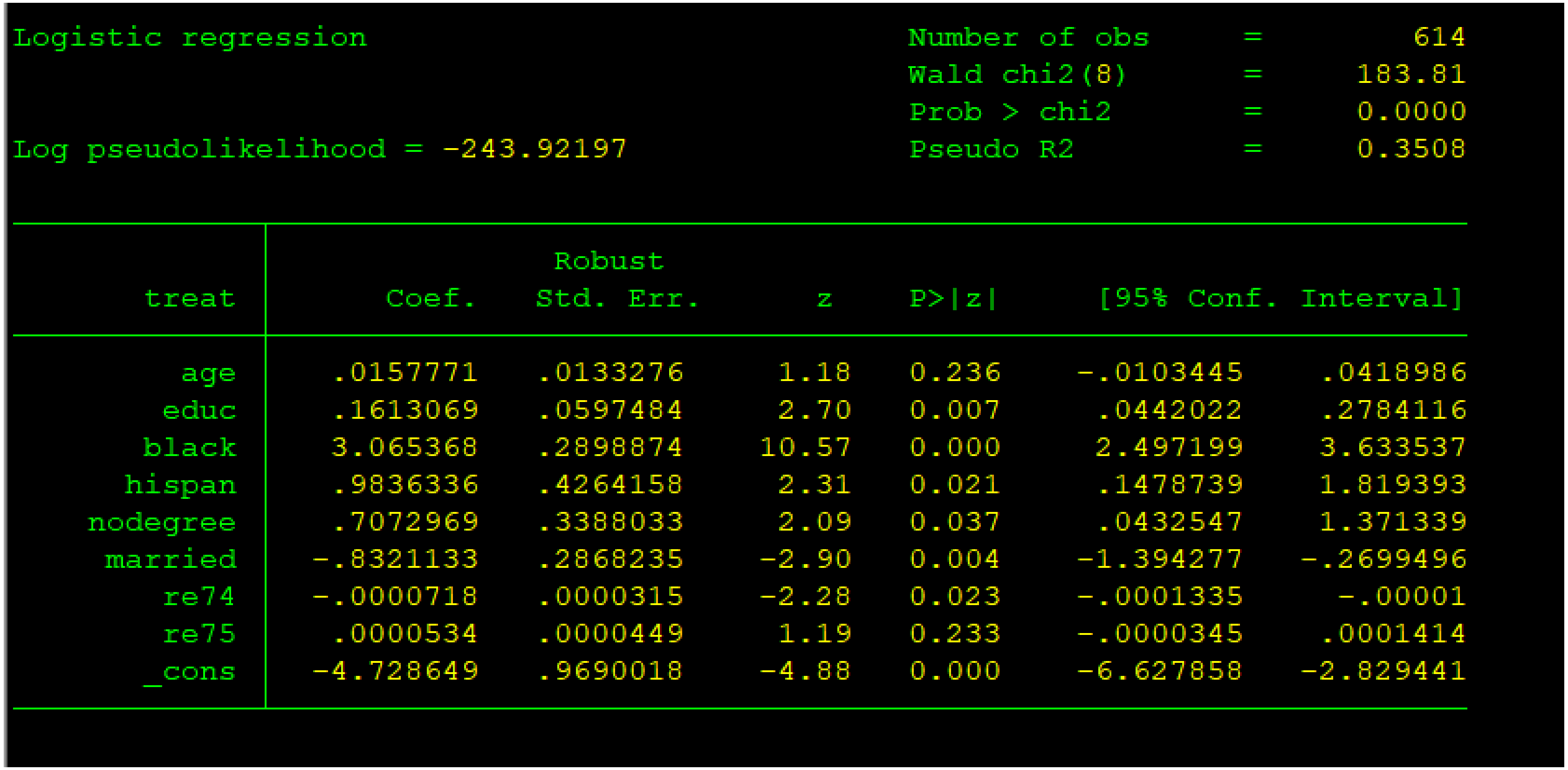
- **Dataset**: from study “causal effects in non-experimental studies: the evaluation of community physical activity promotion program”
- **Objective**: to estimate the average treatment effect (ATE) of the program on steps a day among people who participated in the program
- **Program:** a temporary community program that promoted active travel such as walking and cycling to encourage individuals to pursue physically active lifestyles
- **Comparator**: a national survey for the same years as the program
- **Matching covariates**: age, education, race, ethnicity, marital status, pre-program daily steps

Example: fit PS model



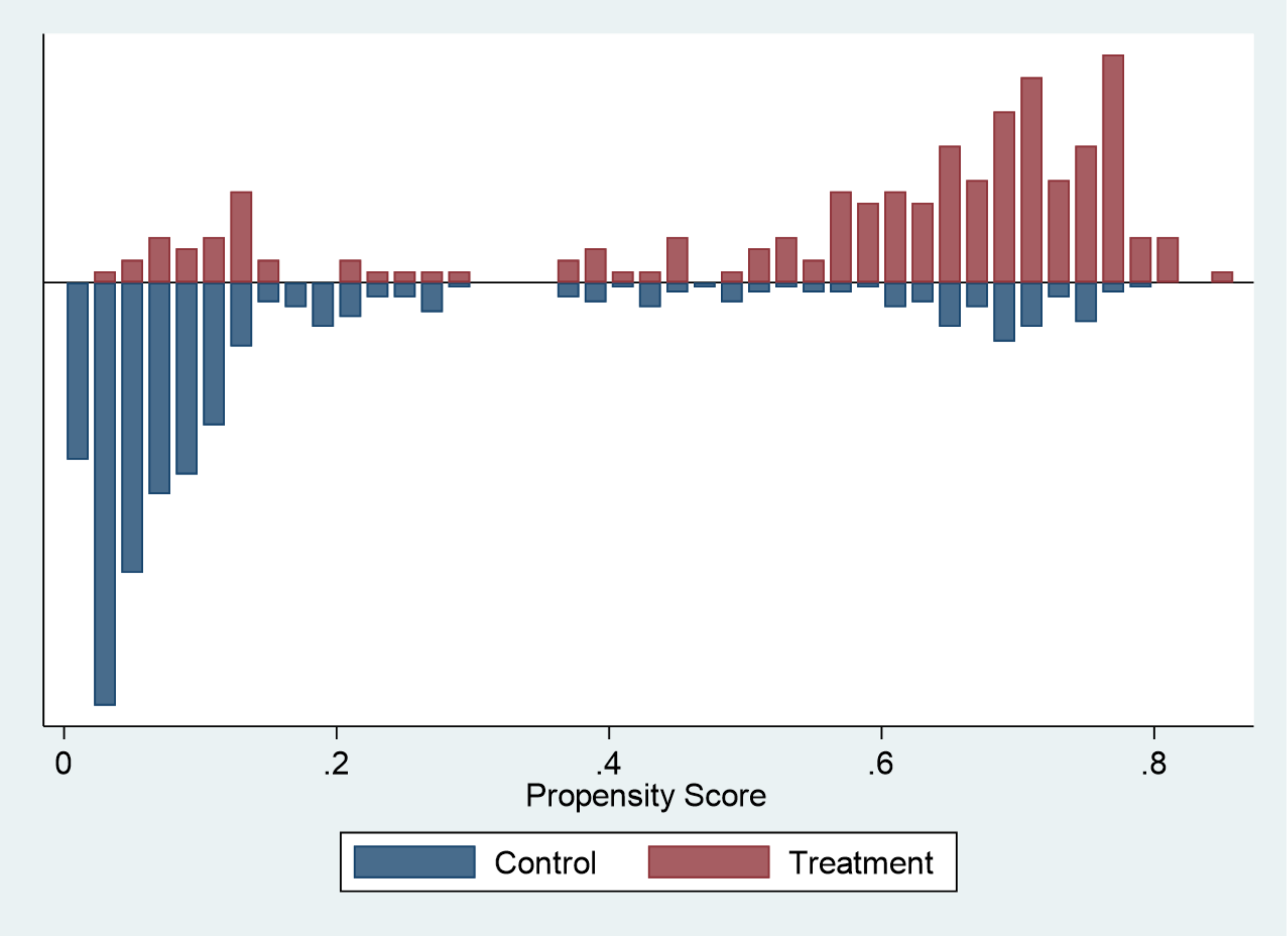
```
* Fit a propensity score model: logistic regression
logit   treat ///
        age educ black hispan nodegree married re74 re75

* Create a propensity score for each person
predict PS_pr, pr   // predicted probability
```

Example: PS model output



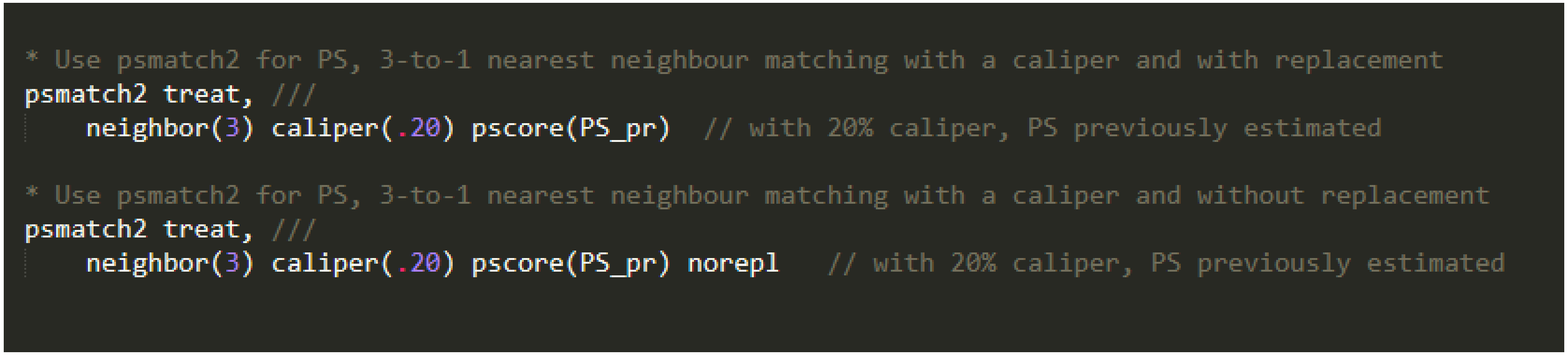
```
Logistic regression                               Number of obs   =        614
                                                  Wald chi2(8)    =     183.81
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -243.92197                 Pseudo R2       =     0.3508

------------------------------------------------------------------------------
             |               Robust
       treat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0157771   .0133276     1.18   0.236    -.0103445    .0418986
        educ |   .1613069   .0597484     2.70   0.007     .0442022    .2784116
       black |   3.065368   .2898874    10.57   0.000     2.497199    3.633537
      hispan |   .9836336   .4264158     2.31   0.021     .1478739    1.819393
    nodegree |   .7072969   .3388033     2.09   0.037     .0432547    1.371339
     married |  -.8321133   .2868235    -2.90   0.004    -1.394277   -.2699496
        re74 |  -.0000718   .0000315    -2.28   0.023    -.0001335     -.00001
        re75 |   .0000534   .0000449     1.19   0.233    -.0000345    .0001414
       _cons |  -4.728649   .9690018    -4.88   0.000    -6.627858   -2.829441
------------------------------------------------------------------------------
```

Example: plot of PS, pre-matching



Example: PS caliper matching

- Previously estimated PS
- Three to one matching
- With caliper of 0.2
- With replacement (by default)
- Without replacement



```
* Use psmatch2 for PS, 3-to-1 nearest neighbour matching with a caliper and with replacement
psmatch2 treat, ///
    neighbor(3) caliper(.20) pscore(PS_pr)  // with 20% caliper, PS previously estimated

* Use psmatch2 for PS, 3-to-1 nearest neighbour matching with a caliper and without replacement
psmatch2 treat, ///
    neighbor(3) caliper(.20) pscore(PS_pr) norepl   // with 20% caliper, PS previously estimated
```

Example: baseline characteristics

Covariates		Unmatched data				Matched data			
		Control *N*= 429	Treatment *N*= 185	*P* -Value	*Standardised differences**	Control *N*= 555	Treatment *N*= 185	*P* -Value	*Standardised differences**
Age (years)		28.0 (10.8)	25.8 (7.2)	0.010	-0.242	25.0 (10.4)	25.8 (7.2)	0.303	0.095
Length of schooling (years)		10.2 (2.9)	10.3 (2.0)	0.633	0.045	10.5 (2.5)	10.3 (2.0)	0.548	-0.054
Black				0.000	1.668			0.732	0.029
	No	342 79.7%	29 15.7%			93 16.8%	29 15.7%		
	Yes	87 20.3%	156 84.3%			462 83.2%	156 84.3%		
Hispanic				0.003	-0.277			0.861	-0.015
	No	368 85.8%	174 94.1%			520 93.7%	174 94.1%		
	Yes	61 14.2%	11 5.9%			35 6.3%	11 5.9%		
No high school degree				0.008	0.235			0.493	0.059
	No	173 40.3%	54 29.2%			177 31.9%	54 29.2%		
	Yes	256 59.7%	131 70.8%			378 68.1%	131 70.8%		
Married				0.000	-0.719			0.181	0.111
	No	209 48.7%	150 81.1%			473 85.2%	150 81.1%		
	Yes	220 51.3%	35 18.9%			82 14.8%	35 18.9%		
Daily steps in 1974		5619.2 (6788.8)	2095.6 (4886.6)	0.000	-0.596	2217.6 (4362.3)	2095.6 (4886.6)	0.749	-0.026
Daily steps in 1975		2466.5 (3292.0)	1532.1 (3219.3)	0.001	-0.287	1641.6 (2483.6)	1532.1 (3219.3)	0.631	-0.038

Note: Continuous variables are reported as mean (standard deviation), while categorical variables are reported as number(percent)

* The standardised difference in percent is the the mean difference as a percentage of the average standard deviation

Example: bias reduction

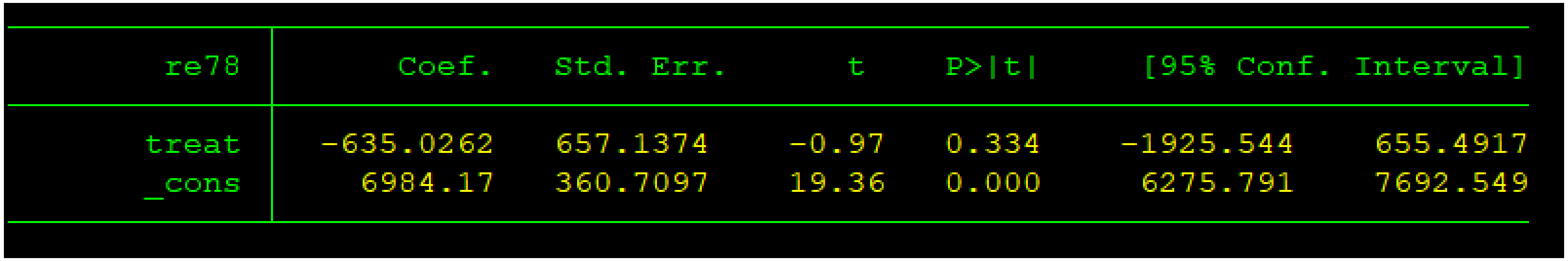
- Bias reduction

$$bias\ reduction = 1 - \left(\frac{|standardised\ differences\ matched|}{|standardised\ differences\ unmatched|}\right)$$

  - Age: $1 - \left(\frac{0.095}{0.242}\right)$=60.7%
  - Black: $1 - \left(\frac{0.029}{1.668}\right)$=98.3%
  - Hispanic: $1 - \left(\frac{0.015}{0.277}\right)$=93.4%
  - No degree: $1 - \left(\frac{0.059}{0.235}\right)$=74.9%
  - Married: $1 - \left(\frac{0.111}{0.719}\right)$=84.6%
  - Daily steps in 1974: $1 - \left(\frac{0.026}{0.596}\right)$=95.6%
  - Daily steps in 1975: $1 - \left(\frac{0.038}{0.287}\right)$=86.8%

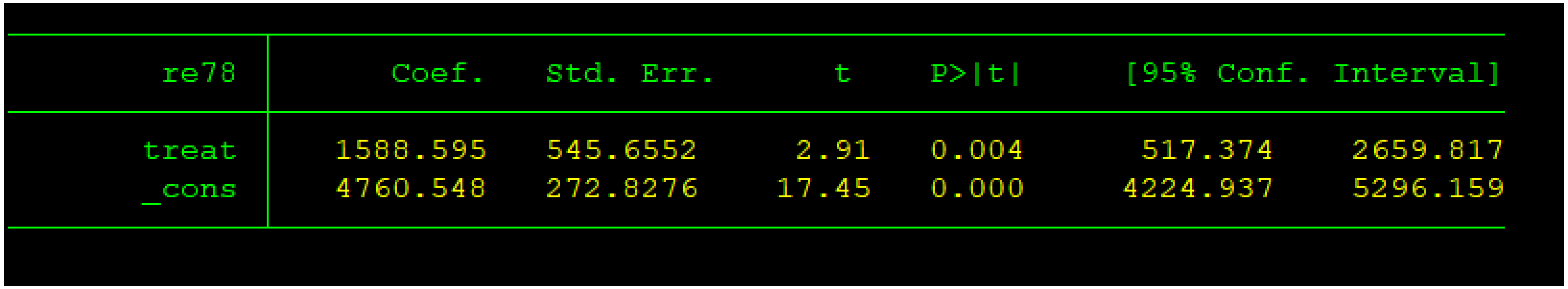
Example: ATE analysis

- Unmatched dataset



```
        re78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |  -635.0262   657.1374    -0.97   0.334    -1925.544    655.4917
       _cons |    6984.17   360.7097    19.36   0.000     6275.791    7692.549
------------------------------------------------------------------------------
```

- Matched dataset



```
        re78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |   1588.595   545.6552     2.91   0.004      517.374    2659.817
       _cons |   4760.548   272.8276    17.45   0.000     4224.937    5296.159
------------------------------------------------------------------------------
```

Conclusions

- Strengths
  - Useful when adjusting for a large number of covariates
  - Useful for both categorical and continuous variables
  - No PS analysis will bias effect estimation
- Limitations
  - Can only adjust for observed covariates
  - Work better in larger samples to attain distributional balance of observed covariates
  - Exacerbate hidden bias
  - Including irrelevant covariates in PS model may reduce efficiency

Thank you for your attention!

References

1. Leuven, Edwin, and Barbara Sianesi. "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing." (2018).
2. Austin, P.C. and Mamdani, M.M. (2006). A comparison of propensity score methods: A case-study illustrating the effectiveness of post-AMI statin use. *Statistics in Medicine* 25: 2084-2106.
3. Cook, T.D., Shadish, W.R., and Wong, V.C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management* 27(4): 724-750.
4. Dehejia, R.H. and Wahba, S. (1999). Causal effects in non-experimental studies: Re-evaluating the evaluation of training programs. *Journal of the American Statistical Association* 94: 1053-1062.
5. Ho, D., Imai, K., King, G., and Stuart, E.A. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* 15(3): 199-236.
6. Little, R.J. and Rubin, D.B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytic approaches. *Annual Review of Public Health*, 21: 121-145.
7. Liu, W., Kuramoto, S.K., and Stuart, E.A. (2013). An Introduction to Sensitivity Analysis for Unobserved Confounding in Non-Experimental Prevention Research. Forthcoming in *Prevention Science*.
8. Rubin, D.B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology* 2: 169-188.
9. Rubin, D.B. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine*, 26: 20-36.
10. Stuart, E.A. and Green, K. (2008). Using full matching to estimate causal effects in non-experimental studies: Examining the relationship between adolescent marijuana use and adult outcomes. *Developmental Psychology* 44(2): 395-406.
11. Stuart, E.A. (2010). Matching Methods for Causal Inference: A review and a look forward. *Statistical Science* 25(1): 1-21.