(Mis)use of matching techniques

Paweł Strawiński

University of Warsaw

5th Polish Stata Users Meeting,
Warsaw, 27th November 2017

Research financed under National Science Center, Poland grant
2015/19/B/HS4/03231

Outline

Motivation

- Matching techniques became very popular among researchers
  Search on `ideas.repec.org` for documents created since 2007 gives result:
  "propensity: 11794,
  score : 14188,
  matching : 23245"
- However, matching is often overlooked as a "magic bullet" that solves all statistical problems
- The problem is that practitioners either are not aware or ignore shortcomings of matching procedures

The idea of matching

- Lets assume we observe two groups.
- One is exposed to experimental treatment and the other to control treatment
- We would like to compare the values of the outcome variable in two groups
- If a study is non-randomized simple difference in outcome variable may be biased estimator of the effect of treatment
- Matching is statistical technique that allows to transform non-randomised data to randomised one.

How it works

- Assume that each observation is described by pair of random variables (Y, X). We call Y outcome variable(s), and X object characteristic(s)
  1. For each observation in the treatment group, find observation in the control group with the same (or at least very similar) X value(s).
  2. The Y values of these matching observations from the control group are then used to compute the counterfactual outcome without treatment for the observation from treatment group.
  3. An estimate for the average treatment effect can be obtained as the mean of the differences between the observed values from the treatment and the matched counterfactual values from the control group.

Outcome of interest

- The Average Treatment on the Treated (ATT)

$$ATT = \sum_i Y_i^1 - E(Y_i^0 | T = 1)$$

- The Average Treatment Effect (ATE)

$$ATE = \frac{\#T = 1}{N} \hat{ATT} + \frac{\#T = 0}{N} \left(\sum_i E(Y_i^1 | T = 0) - Y_i^0 \right)$$

- Those expected values are not directly observed. They are retrieved from observed data by reweighting procedure. Different algorithms uses different reweighting.

Matching methods

- Exact matching
- Distance matching
- Propensity Score Matching

Exact Matching

- Exact matching is matching on discrete metric
- Observations are matched if and only if $X_i = X_j$
- The result is perfect matching on covariate values
- Is the problem is multivariate the result could be an empty set

Distance Matching

- Matching on distance metric that measures proximity of observations
- The idea then is to use close observations, but not necessarily ideally matched
- The most popular algorithm is Mahalanobis distance matching

$$MD = (X_i - X_j)'\Sigma(X_i - X_j)$$

where Σ is empirical covariance matrix of X

- Performs well when X are discrete
- When X are continuous can be computationally burdensome

Propensity Score Matching

- Method proposed by Rosenbaum and Rubin (1983). Instead matching on multidimensional X matching is done on propensity score $\pi(X)$ which is $E(T = 1 \mid X)$
- It requires estimation of the propensity score $\pi(X)$ usually by logit or probit model. Then observations with closest values of $\pi(X)$ are matched
- Performs well when X are continuous
- When X are discrete it often is difficult to choose best match

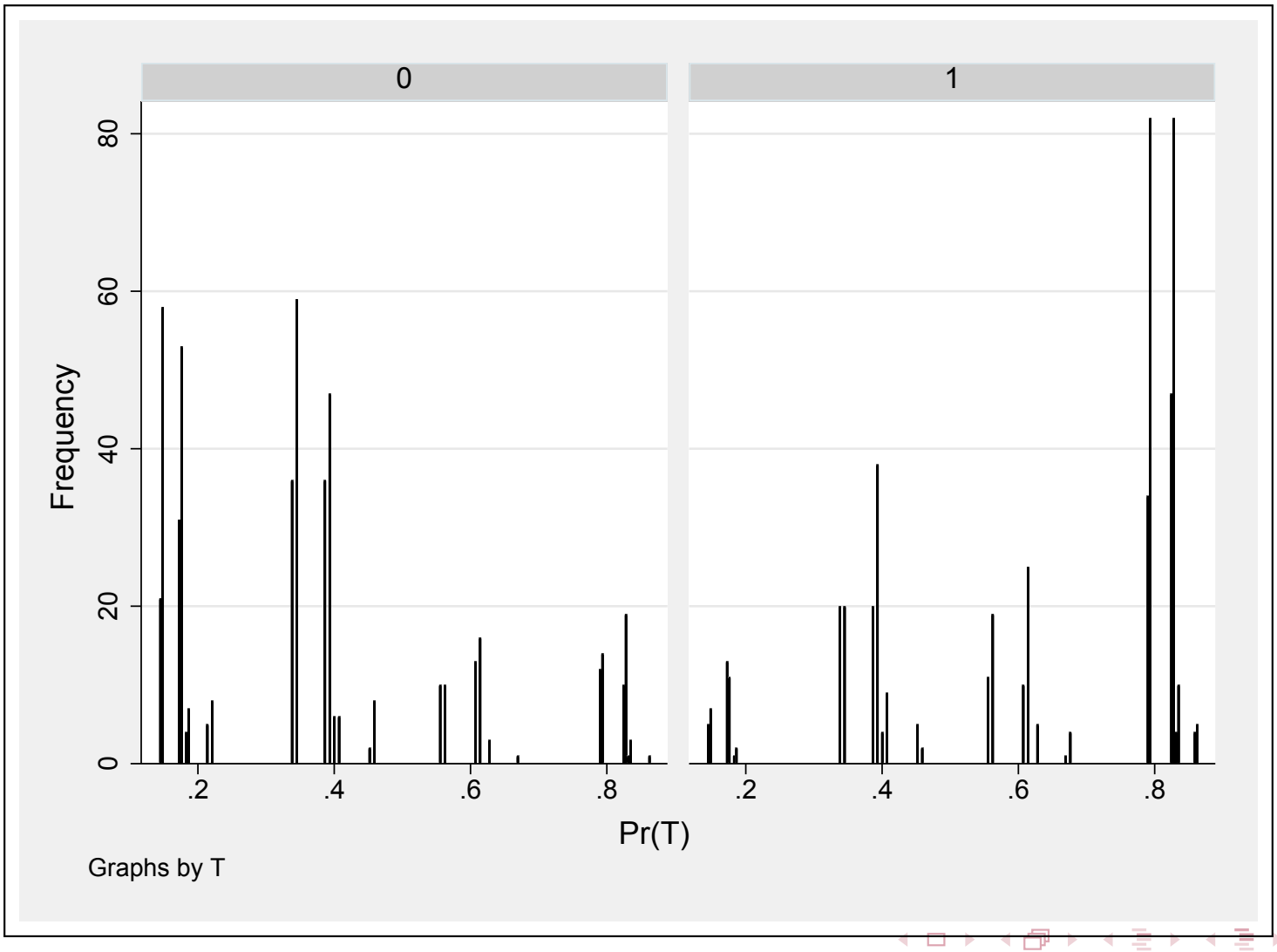
- Pair matching (no replacement)
- Nearest neighbour match (with replacement)
- Nearest k-neighbour match (with replacement)
- Caliper matching
- Kernel matching

- The aim of examples is to show in which circumstances exact matching and matching on the propensity score leads to poor quality results

Example 1

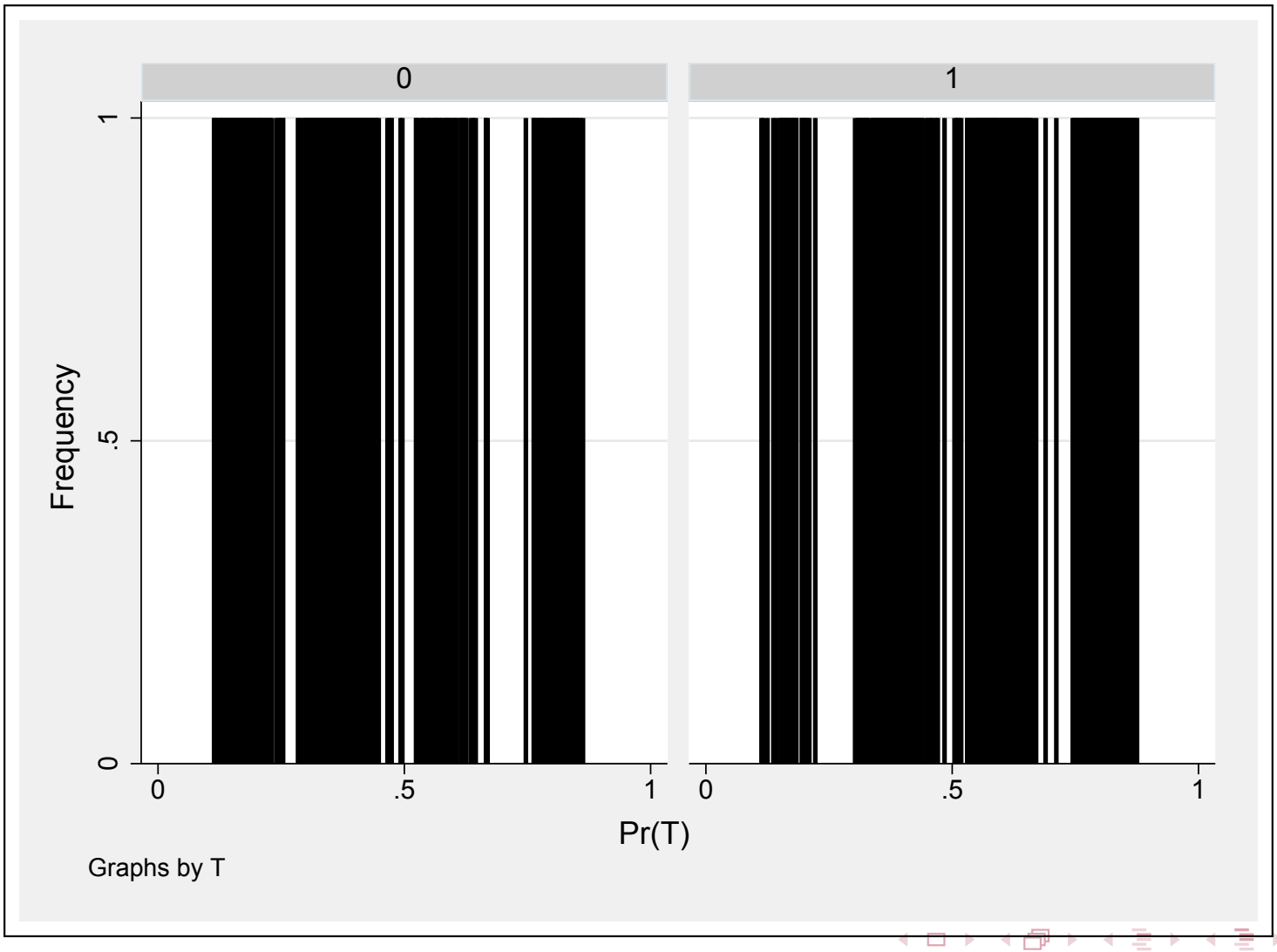
- We analyse sample of 1000 observations
- There are 500 treated objects and 500 non treated object
- Objects are described by 5 dummy variables and 5 variables with continuous distribution
- In each model for propensity score 5 dummy variables are used and different number of continuously distributed variables from 0 to 5

Example 1. 5 dummy

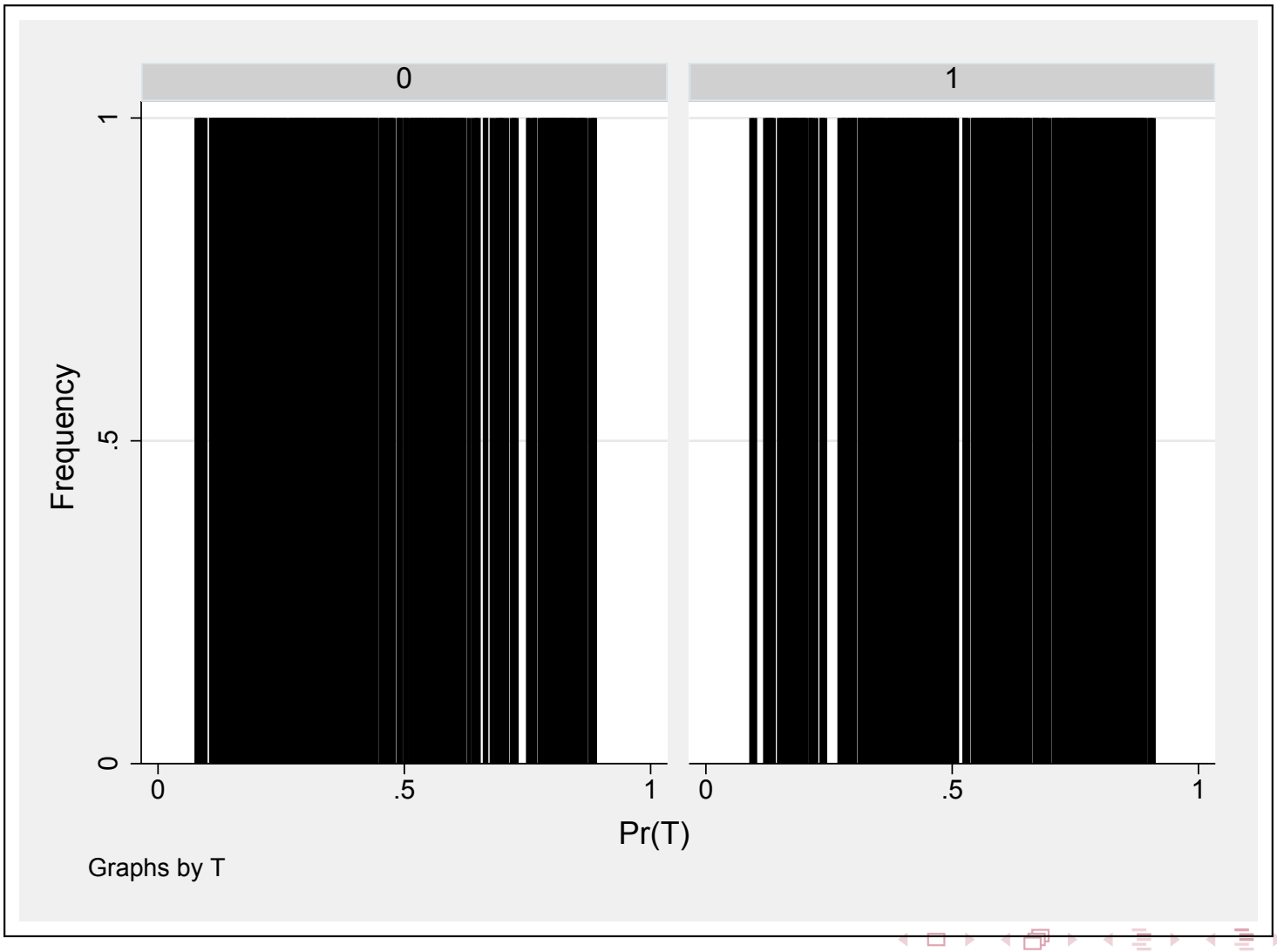


Example 1. 5 dummy and 1 continuous covariate

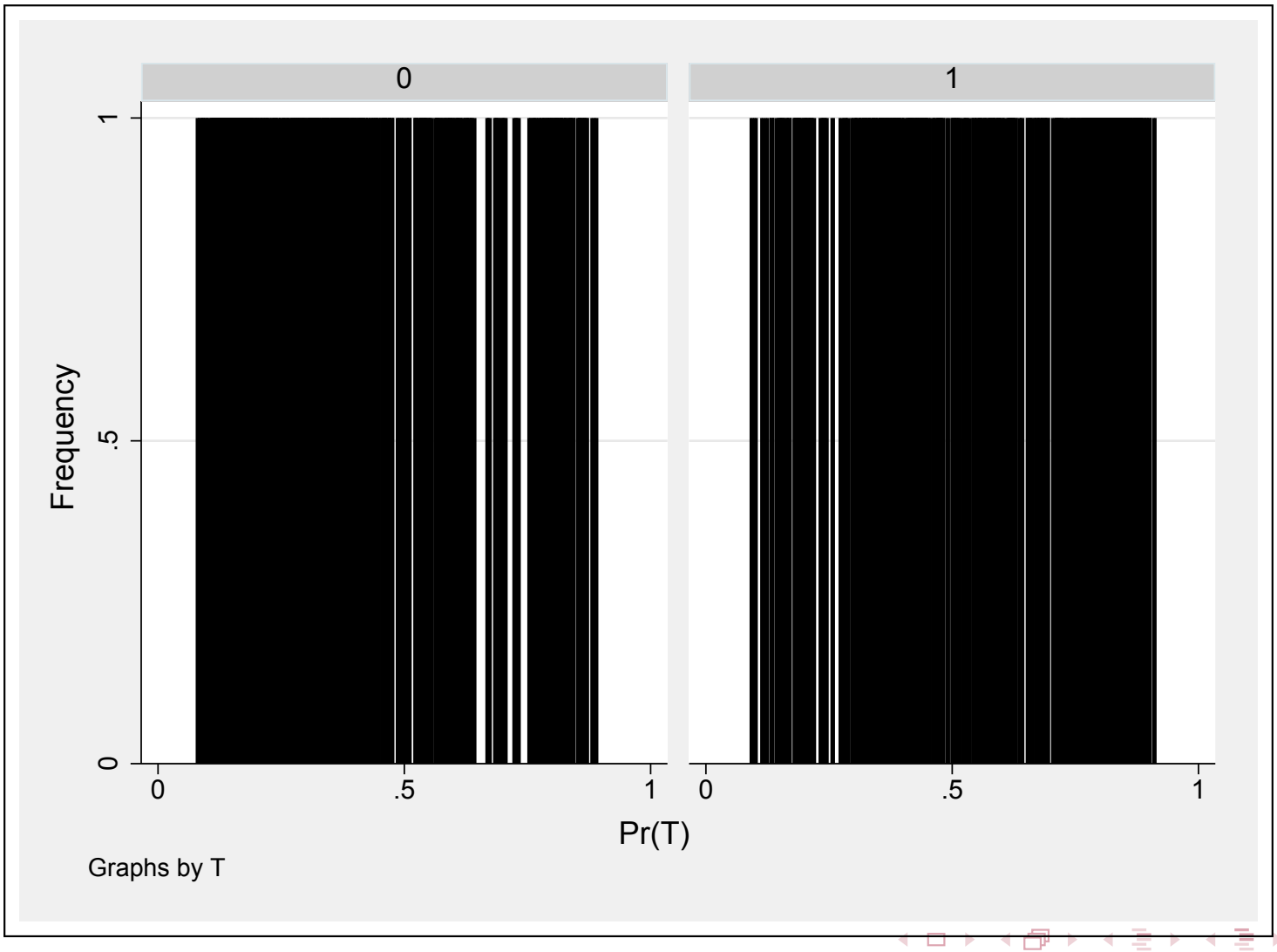
Example 1. 5 dummy and 2 continuous covariates



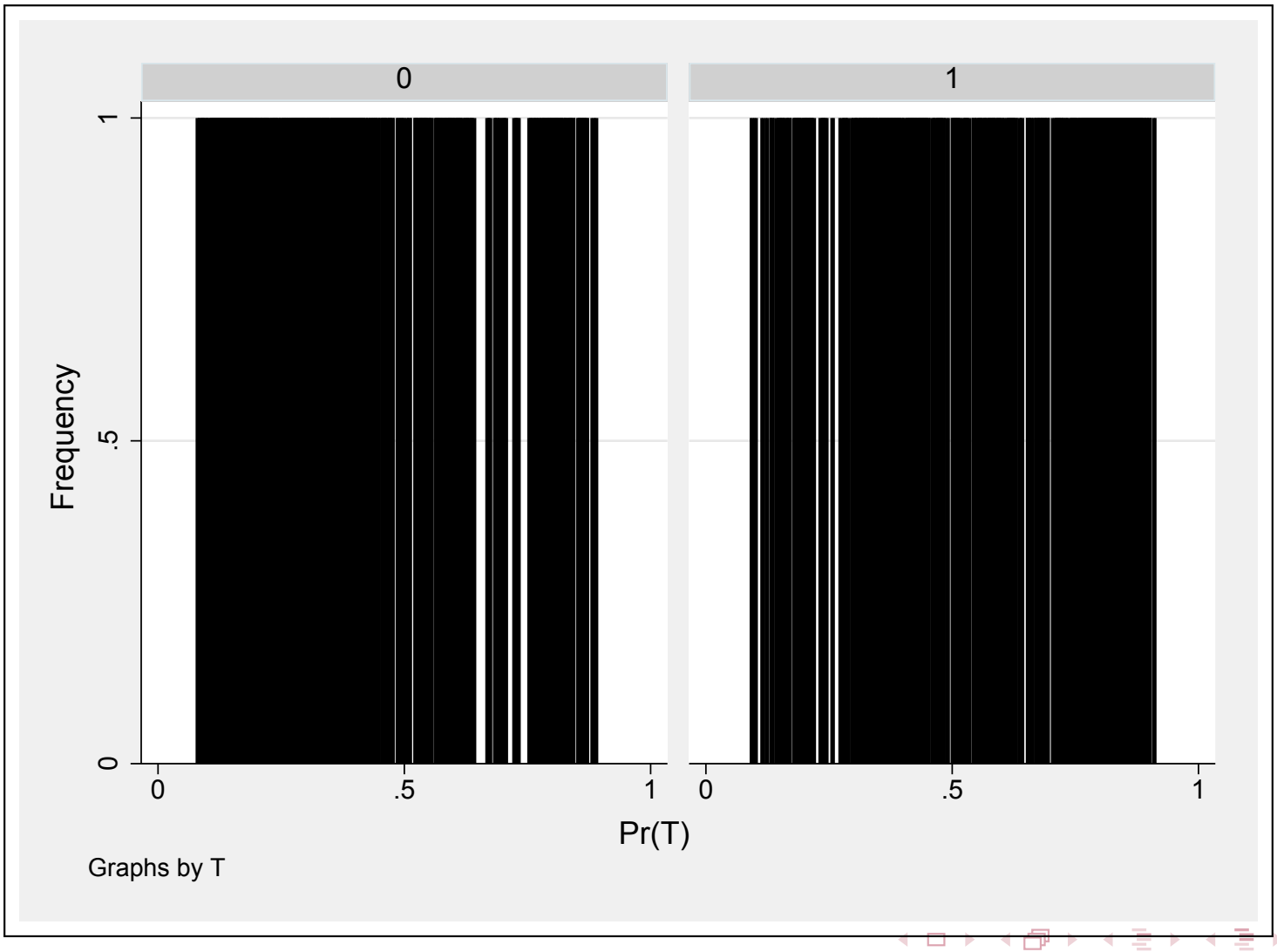
Example 1. 5 dummy and 3 continuous covariates



Example 1. 5 dummy and 4 continuous covariates



Example 1. 5 dummy and 5 continuous covariates



Example 1. Number of distinct pscore values

- With 0 continuous variables 31 distinct values
- With at least 1 continuous variable 1000 distinct values
- Therefore, distinct values of the propensity score are important but not sufficient to find good matches
- Its distribution also matters

Example 1. What I do not like about teffects

```
. teffects psmatch (y) (T d1 d2 d3 d4 d5), atet

Treatment-effects estimation                    Number of obs      =      1,000
Estimator      : propensity-score matching      Matches: requested =          1
Outcome model  : matching                                       min =          1
Treatment model: logit                                          max =         82
------------------------------------------------------------------------------
             |               AI Robust
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ATET         |
           T |
   (1 vs 0)  |   46.52442   105.3741     0.44   0.659     -160.005    253.0538
------------------------------------------------------------------------------
```

Example 1. What I do not like about teffects

```
. psmatch2 T d1 d2 d3 d4 d5, outcome(y) logit quietly
There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.
------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference         S.E.   T-stat
-----------------------------+------------------------------------------------------------
             y1  Unmatched | 2107.59819   2074.21094   33.3872521   71.9959108     0.46
                       ATT | 2107.59819   1879.20512   228.393068    330.29759     0.69
-----------------------------+------------------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.
```

Example 1. What I do not like about teffects

```
. teffects psmatch (y) (T x1 d1 d2 d3 d4 d5), atet

Treatment-effects estimation                   Number of obs      =      1,000
Estimator      : propensity-score matching     Matches: requested =          1
Outcome model  : matching                                   min =          1
Treatment model: logit                                      max =          1
------------------------------------------------------------------------------
             |               AI Robust
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ATET         |
           T |
   (1 vs 0)  |   33.21557     145.23     0.23   0.819      -251.43    317.8611
------------------------------------------------------------------------------

. psmatch2 T x1 d1 d2 d3 d4 d5, outcome(y) logit qui
------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference          S.E.   T-stat
-----------------------------+------------------------------------------------------------
              y1  Unmatched | 2107.59819   2074.21094   33.3872521   71.9959108     0.46
                        ATT | 2107.59819   2074.38262   33.2155662   127.669286     0.26
-----------------------------+------------------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.
```

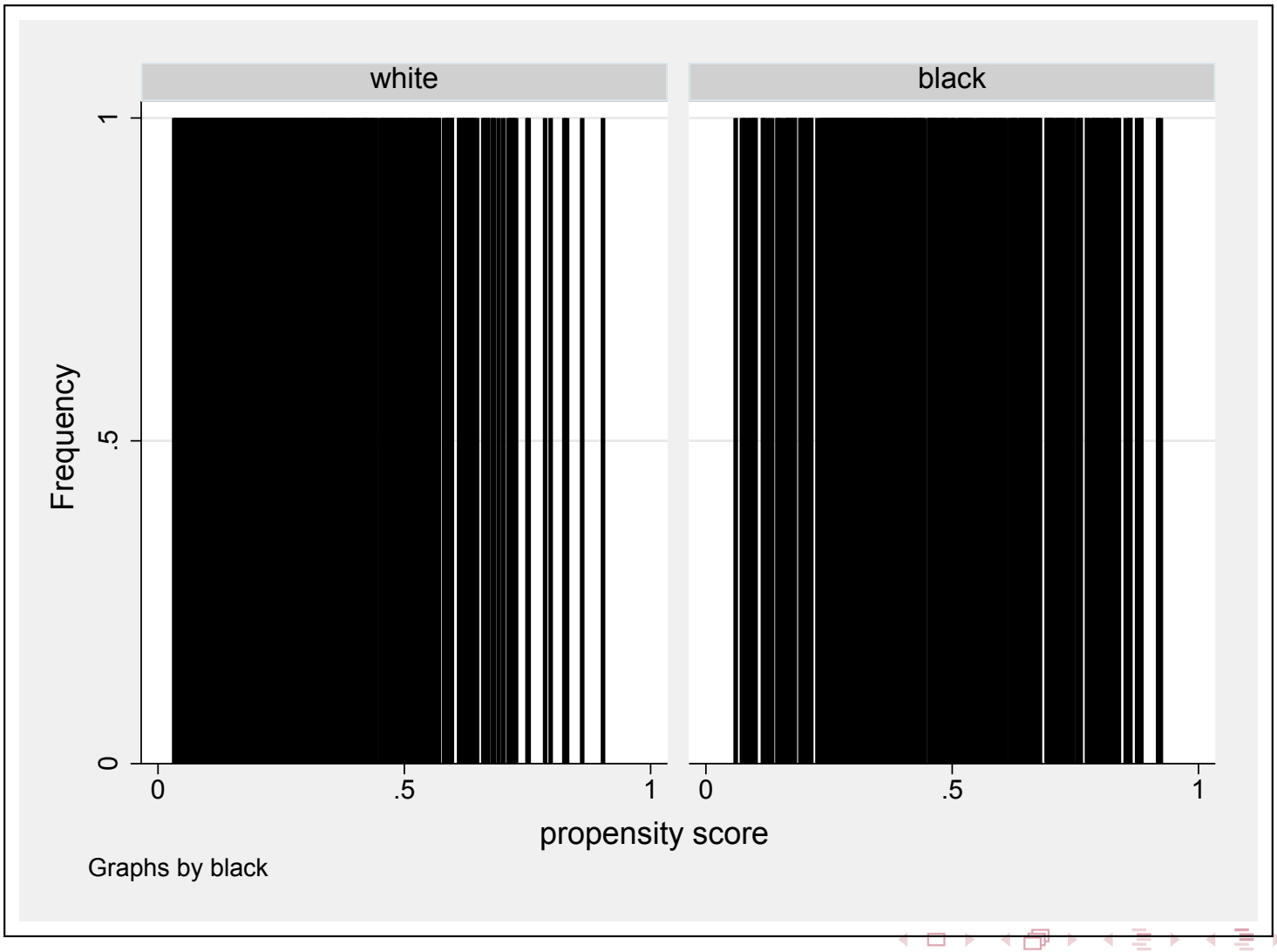
Example 2.

- We analyse National Longitudinal Surveys 1988, Women samples data

```
sysuse nlsw88.dta, clear
(NLSW, 1988 extract)

drop if race==3
(26 observations deleted)

keep if hours==40
(1,142 observations deleted)

gen black=(race==2)
```

- We restrict sample to white and black woman
- And consider only those who work for 40 hours a week
- In model for propensity score we use 8 dummy variables are used and 3 continuously distributed variables

Example 2. 8 dummy and 3 continuous covariates



Conclusions and Recomendations

- The Propensity Score Matching and other matching techniques should be used with caution
- If You lacking continuously distributed covariates the PSM matching is bad idea. Distance matching is better choice but remember about problem of identical distances
- Three continuously distributed covariates are usually enough to receive close to continuous distribution for the estimated propensity score