

# A First Course in Causal Inference, Ch14-15

Ruomeng Liu  
HKU CSS Reading Group

# Chapter 14. Using the Propensity Score in Regression for Causal Effects





Two simple methods to use the propensity score in regression model:

- Including propensity score as a covariate in regression.
- Running regression model weighted by the inverse of the propensity score.



## Regression with the propensity score as covariate

Quick recap:

- Treatment effect:  $\tau = E[Y(1) - Y(0)]$ ,  $Y(1)$  is the potential outcome when unit receive the treatment ( $Z = 1$ ), and  $Y(0)$  is the potential outcome when unit does not receive ( $Z=0$ )
- Propensity score is the sufficient statistic of the covariate, defined as:  $e(X) = \Pr(Z = 1 | X)$ ; when given the propensity score,  $Y(1), Y(0) \perp\!\!\!\perp Z | e(X)$ .
- Causal effect identification:  $\tau = E[E[Y | Z = 1, e(X)] - E[Y | Z = 0, e(X)]]$

## Regression with the propensity score as covariate

$$\arg \min_{a,b,c} E\{Y - a - bZ - ce(X)\}^2$$

Regression model:  $E\{Y \mid Z, e(X)\} = a + bZ + ce(X)$

**Theorem 14.1** *If  $Z \perp\!\!\!\perp \{Y(1), Y(0)\} \mid X$ , then the coefficient of  $Z$  in the population OLS fit of  $Y$  on  $\{1, Z, e(X)\}$  equals*

$$\tau_e = \tau_0 = \frac{E\{h_0(X)\tau(X)\}}{E\{h_0(X)\}},$$

*recalling that  $h_0(X) = e(X)\{1 - e(X)\}$  and  $\tau(X) = E\{Y(1) - Y(0) \mid X\}$ .*

$\tau_e = E\{Y(1) - Y(0)\} = \tau = E\{Y \mid Z = 1, X\} - E\{Y \mid Z = 0, X\}$ , given the ignorability assumption

the coefficient of  $Z$  in a population OLS regression of  $\{1, Z, e(x)\}$  equals the causal effect  $\tau$ .



## Regression weighted by the inverse of the propensity score

IPW adjust the sample distribution of the treatment group and the control group by assigning weights to each sample to make them more balanced in terms of the covariates.

For  $Z = 1$ , weight is  $1/e(X)$ ; For  $Z = 0$ , weight is  $1/(1-e(x))$

How to estimate  $\tau_0$  (ATE) by IPW?

$$\hat{\tau}^{\text{hajek}} = \frac{\sum_{i=1}^n \frac{Z_i Y_i}{\hat{e}(X_i)}}{\sum_{i=1}^n \frac{Z_i}{\hat{e}(X_i)}} - \frac{\sum_{i=1}^n \frac{(1-Z_i) Y_i}{1-\hat{e}(X_i)}}{\sum_{i=1}^n \frac{1-Z_i}{1-\hat{e}(X_i)}}$$

- The first part is the weighted mean of the treatment group results; the second part is the control group.
- Identically, WLS

## Regression weighted by the inverse of the propensity score

How to use WLS and IPW to estimate the causal effect

**Proposition 14.1**  $\hat{\tau}^{\text{hajek}}$  equals  $\hat{\beta}$  from the following WLS:

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{\alpha, \beta} \sum_{i=1}^n w_i (Y_i - \alpha - \beta Z_i)^2$$

with weights

$$w_i = \frac{Z_i}{\hat{e}(X_i)} + \frac{1 - Z_i}{1 - \hat{e}(X_i)} = \begin{cases} \frac{1}{\hat{e}(X_i)} & \text{if } Z_i = 1; \\ \frac{1}{1 - \hat{e}(X_i)} & \text{if } Z_i = 0. \end{cases} \quad (14.1)$$

$$e(X_i) = P(Z_i = 1 | X_i)$$

WLS gives a consistent estimator for the ATE  $\tau$  and also equivalent with the Hajek estimator.

Extension:

when the  $e(X)$  is an estimator rather than a true value, the estimator can still be consistent;

WLS can be extended to more generalized framework, e.g. nonlinear causal effect.

## Regression weighted by the inverse of the propensity score

### Two estimators

#### 1. Regression Estimator

Treatment group prediction results

$$\hat{\tau}_{\text{wls}}^{\text{reg}} = \frac{1}{n} \sum_{i=1}^n \mu_1(X_i, \hat{\beta}_1) - \frac{1}{n} \sum_{i=1}^n \mu_0(X_i, \hat{\beta}_0)$$

#### 2. Doubly Robust Estimator

$$\hat{\tau}_{\text{wls}}^{\text{dr}} = \hat{\tau}_{\text{wls}}^{\text{reg}} + \frac{1}{n} \sum_{i=1}^n \frac{Z_i \{Y_i - \mu_1(X_i, \hat{\beta}_1)\}}{e(X_i, \hat{\alpha})} - \frac{1}{n} \sum_{i=1}^n \frac{(1 - Z_i) \{Y_i - \mu_0(X_i, \hat{\beta}_0)\}}{1 - e(X_i, \hat{\alpha})}$$

corrects the observation bias by weighting it to a propensity score weighted average

## Regression weighted by the inverse of the propensity score

**Theorem 14.2** If  $\bar{X} = 0$  and  $(\mu_1(X_i, \hat{\beta}_1), \mu_0(X_i, \hat{\beta}_0)) = (\hat{\beta}_{10} + \hat{\beta}_{1x}^T X_i, \hat{\beta}_{00} + \hat{\beta}_{0x}^T X_i)$  based on the WLS fit of  $Y_i$  on  $(1, Z_i, X_i, Z_i X_i)$  with weights (14.1), then

$$\hat{\tau}_{\text{wls}}^{\text{dr}} = \hat{\tau}_{\text{wls}}^{\text{reg}} = \hat{\beta}_{10} - \hat{\beta}_{00},$$

which is the coefficient of  $Z_i$  in the WLS fit.

Intercept  
difference

## Regression weighted by the inverse of the propensity score

**Theorem 14.2** If  $\bar{X} = 0$  and  $(\mu_1(X_i, \hat{\beta}_1), \mu_0(X_i, \hat{\beta}_0)) = (\hat{\beta}_{10} + \hat{\beta}_{1x}^T X_i, \hat{\beta}_{00} + \hat{\beta}_{0x}^T X_i)$  based on the WLS fit of  $Y_i$  on  $(1, Z_i, X_i, Z_i X_i)$  with weights (14.1), then

$$\hat{\tau}_{\text{wls}}^{\text{dr}} = \hat{\tau}_{\text{wls}}^{\text{reg}} = \hat{\beta}_{10} - \hat{\beta}_{00},$$

which is the coefficient of  $Z_i$  in the WLS fit.

Intercept  
difference

Use:

- WLS can effectively adjust the imbalance of covariate distribution and is particularly suitable for observational data.
- WLS may perform worse when the model is wrong or heteroskedasticity exists.

## Regression weighted by the inverse of the propensity score

TABLE 14.1: Regression estimators in CREs and unconfounded observational studies. The weights  $w_i$ 's are defined in (14.1). Assume covariates are centered at  $\bar{X} = 0$ .

	CRE	unconfounded observational studies
without $X$	$Y_i \sim (1, Z_i)$	$Y_i \sim (1, Z_i)$ with weights $w_i$
with $X$	$Y_i \sim (1, Z_i, X_i, Z_i X_i)$	$Y_i \sim (1, Z_i, X_i, Z_i X_i)$ with weights $w_i$

## Regression weighted by the inverse of the propensity score

ATT

**Proposition 14.2**  $\hat{\tau}_T^{\text{hajek}}$  is numerically identical to  $\hat{\beta}$  in the following WLS:

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{\alpha, \beta} \sum_{i=1}^n w_{Ti} (Y_i - \alpha - \beta Z_i)^2$$

with weights

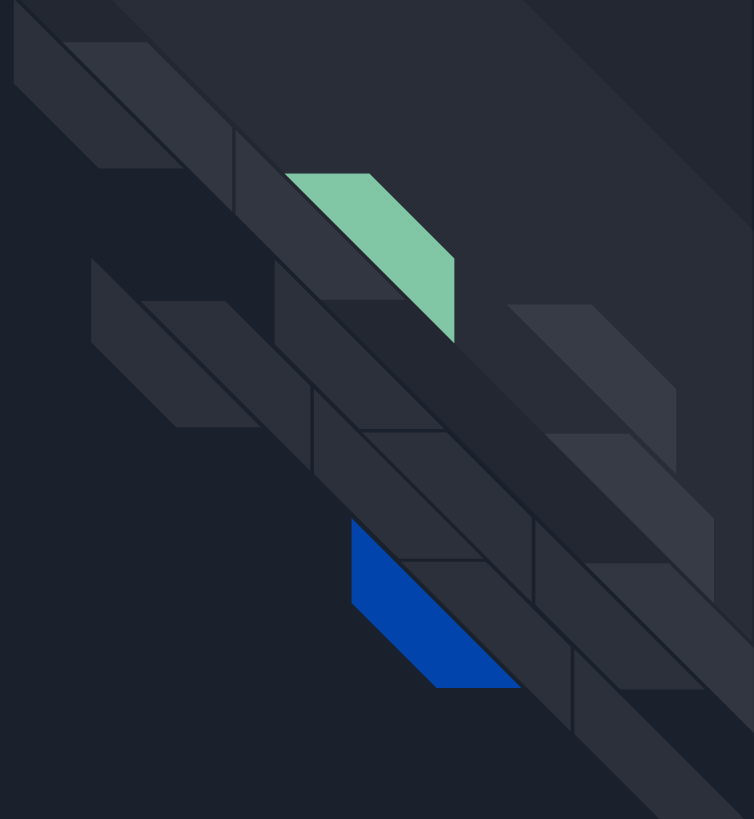
$$w_{Ti} = Z_i + (1 - Z_i)\hat{\delta}(X_i) = \begin{cases} 1 & \text{if } Z_i = 1; \\ \hat{\delta}(X_i) & \text{if } Z_i = 0. \end{cases} \quad (14.2)$$

**Theorem 14.3** If  $\hat{X}(1) = 0$  and  $\mu_0(X_i, \hat{\beta}_0) = \hat{\beta}_{00} + \hat{\beta}_{0x}^T X_i$  based on the WLS fit of  $Y_i$  on  $(1, Z_i, X_i, Z_i X_i)$  with weights (14.2), then

$$\hat{\tau}_{T, \text{wls}}^{\text{dr}} = \hat{\tau}_{T, \text{wls}}^{\text{reg}} = \hat{\beta}_{10} - \hat{\beta}_{00},$$

which is the coefficient of  $Z_i$  in the WLS fit.

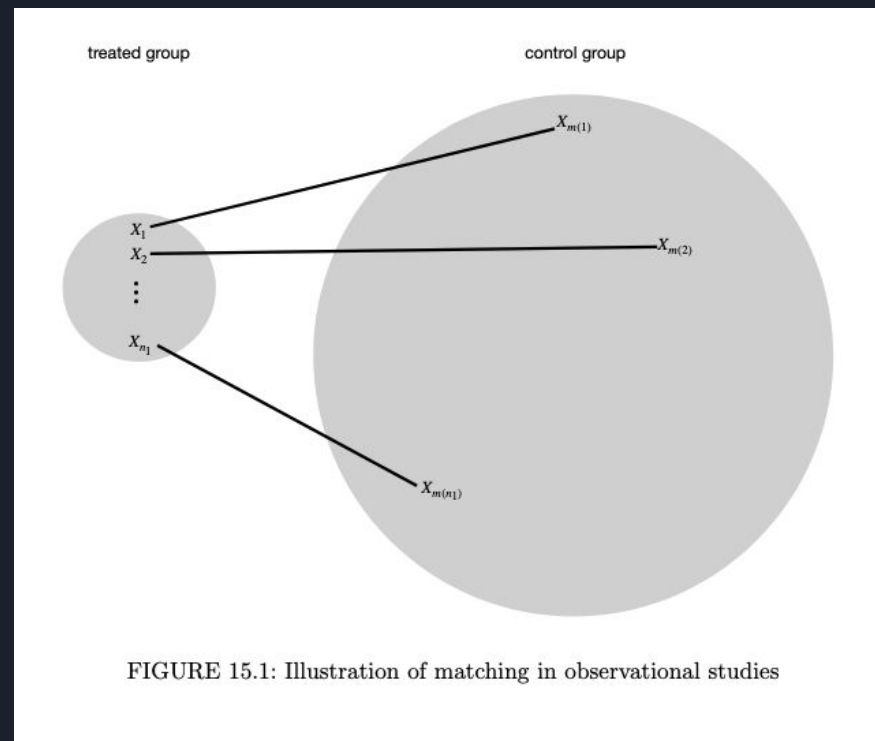
# Chapter 15. Matching in Observational Studies



## Regression weighted by the inverse of the propensity score

Exact match

$$\text{pr}(Z_i = 1, Z_{m(i)} = 0 \mid Z_i + Z_{m(i)} = 1, X_i, X_{m(i)}) = 1/2$$





## 15.2 A more complicated but realistic scenario

Perfect matching can be difficult, then approximate matching can be achieved.

Distance Metric:

- 1) Euclidean Distance

$$d(X_i, X_k) = (X_i - X_k)^T (X_i - X_k),$$

- 2) Mahalanobis Distance

$$d(X_i, X_k) = (X_i - X_k)^T \Omega^{-1} (X_i - X_k)$$



## 15.2 A more complicated but realistic scenario

### Subtle Issues

- 1) One to One matching and One to M matching
- 2) Matching with replacement
  - a) With replacement is more computationally convenient; gives matches higher quality
  - b) Without replacement involves computationally intensive discrete optimization
- 3) External validity
- 4) High dimensionality
  - a) Propensity score

## 15.3 Matching estimator for the average causal effect

Point estimation

$$\hat{Y}_i(0) = M^{-1} \sum_{k \in J_i} Y_k,$$

$$\hat{Y}_i(1) = M^{-1} \sum_{k \in J_i} Y_k,$$

Bias correction

$$\hat{B}_i = (2Z_i - 1)M^{-1} \sum_{k \in J_i} \{\hat{\mu}_{1-Z_i}(X_i) - \hat{\mu}_{1-Z_i}(X_k)\}$$

$$\hat{\tau}^{\text{mbc}} = \hat{\tau}^{\text{m}} - \hat{B},$$

Treatment  
group

## 15.3 Matching estimator for the average causal effect

**Proposition 15.1** *We have*

$$\hat{\tau}^{\text{mbc}} = n^{-1} \sum_{i=1}^n \hat{\psi}_i \quad (15.1)$$

where

$$\hat{\psi}_i = \hat{\mu}_1(X_i) - \hat{\mu}_0(X_i) + (2Z_i - 1)(1 + K_i/M)\{Y_i - \hat{\mu}_{Z_i}(X_i)\}$$

with  $K_i$  being the times that unit  $i$  is used as a match.

## 15.5 Case study

### Experimental data

```
> library("car")
> library("Matching")
>
> ## Chapter 15.5.1
> ## experimental data
> data("lalonge")
> y = lalonge$re78
> z = lalonge$treat
> x = as.matrix(lalonge[, c("age", "educ", "black",
+                           "hisp", "married", "nodegr",
+                           "re74", "re75")])
>
```

```
>
> ## analysis the randomized experiment
> neymanols = lm(y ~ z)
> fisherols = lm(y ~ z + x)
> xc = scale(x)
> linols = lm(y ~ z*xc)
> resols = c(neymanols$coef[2],
+            fisherols$coef[2],
+            linols$coef[2],
+            sqrt(hccm(neymanols, type = "hc2")[2, 2]),
+            sqrt(hccm(fisherols, type = "hc2")[2, 2]),
+            sqrt(hccm(linols, type = "hc2")[2, 2]))
> resols = matrix(resols, 3, 2)
> rownames(resols) = c("neyman", "fisher", "lin")
> colnames(resols) = c("est", "se")
> resols
      est      se
neyman 1794.343 670.9967
fisher 1676.343 677.0493
lin     1621.584 694.7217
```

## 15.5 Case study

### Experimental data

```
> matchest.adj = Match(Y = y, Tr = z, X = x, BiasAdjust = TRUE)
> summary(matchest.adj)
```

```
Estimate... 2119.7
AI SE..... 876.42
T-stat..... 2.4185
p.val..... 0.015583
```

```
Original number of observations..... 445
Original number of treated obs..... 185
Matched number of observations..... 185
Matched number of observations (unweighted). 268
```

## 15.5 Case study

### Observational data

```
> neymanols = lm(y ~ z)
> fisherols = lm(y ~ z + x)
> xc = scale(x)
> linols = lm(y ~ z*xc)
> resols = c(neymanols$coef[2],
+           fisherols$coef[2],
+           linols$coef[2],
+           sqrt(hccm(neymanols, type = "hc2")[2, 2]),
+           sqrt(hccm(fisherols, type = "hc2")[2, 2]),
+           sqrt(hccm(linols, type = "hc2")[2, 2]))
> resols = matrix(resols, 3, 2)
> rownames(resols) = c("neyman", "fisher", "lin")
> colnames(resols) = c("est", "se")
> resols
```

	est	se
neyman	-8506.495	583.4426
fisher	1067.546	628.4389
lin	-4265.801	3211.7718

```
> matchest = Match(Y = y, Tr = z, X = x, BiasAdjust = TRUE)
> summary(matchest)
```

Estimate...	1747.8
AI SE.....	916.59
T-stat.....	1.9068
p.val.....	0.056543
Original number of observations.....	16177
Original number of treated obs.....	185
Matched number of observations.....	185
Matched number of observations (unweighted).	248

# 15.5 Case study

## Covariate balance check

```
> lm.before = lm(z ~ x)  
> summary(lm.before)
```

```
Residuals:  
    Min       1Q   Median       3Q      Max   
-0.18508 -0.01057  0.00303  0.01018  1.01355  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)      
(Intercept)  1.404e-03  6.326e-03  0.222  0.8243      
xage         -4.043e-04  8.512e-05 -4.750 2.05e-06 ***  
xeduc        3.220e-04  4.073e-04  0.790  0.4293      
xblack       1.070e-01  2.902e-03 36.871 < 2e-16 ***  
xhispan      6.377e-03  3.103e-03  2.055  0.0399 *     
xmarried    -1.525e-02  2.023e-03 -7.537 5.06e-14 ***  
xnodegree   1.345e-02  2.523e-03  5.331 9.89e-08 ***  
xre74       7.601e-07  1.806e-07  4.208 2.59e-05 ***  
xre75      -1.231e-07  1.829e-07 -0.673  0.5011      
xu74        4.224e-02  3.271e-03 12.914 < 2e-16 ***  
xu75        2.424e-02  3.399e-03  7.133 1.02e-12 ***
```

```
Residuals:  
    Min       1Q   Median       3Q      Max   
-0.66864 -0.49161 -0.03679  0.50378  0.65122  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)      
(Intercept)  6.003e-01  2.427e-01  2.474  0.0137 *     
xage         3.199e-03  3.427e-03  0.933  0.3511      
xeduc       -1.501e-02  1.634e-02 -0.918  0.3590      
xblack       6.141e-05  7.408e-02  0.001  0.9993      
xhispan      1.391e-02  1.208e-01  0.115  0.9084      
xmarried    -1.328e-02  6.729e-02 -0.197  0.8437      
xnodegree   -3.023e-02  7.144e-02 -0.423  0.6723      
xre74       6.754e-06  9.864e-06  0.685  0.4939      
xre75      -9.848e-06  1.279e-05 -0.770  0.4417      
xu74        2.179e-02  1.027e-01  0.212  0.8321      
xu75       -2.642e-02  8.327e-02 -0.317  0.7512
```



# summary

## Chapter 14

- Including PS as covariate
  - Adjust for confounding in regression
  - Relies on regression adjustment
  - Simple: add PS as covariate in regression
- IPW
  - By reweighting to balance covariate
  - Weights directly balance covariates
  - Require reweighting and diagnostics for weights

## Chapter 15

- Matching in observational studies
  - estimator