Week 3: Matching and PSM

JJ Chen

January 30, 2015

Today's Tasks

- ▶ Homework Questions
- ▶ Matching and Propensity Score
  - ▶ Understand functional form and weight of regressions

Recap

- ► Week 1: RCM and RCTs
  - ► RCTs give IA: $Y_{0i}, Y_{1i} \perp\!\!\!\perp D_i$
  - ► Extend to CIA: $Y_{0i}, Y_{1i} \perp\!\!\!\perp D_i \mid X_i$

- ► Week 2: Reg and OVB
  - ► If CIA is believeable, regression recovers causal effect
  - ► Understand Reg through RCM
  - ► Understand OVB through CIA

- ► Week 3: Mat and PSM
  - ► If CIA + CSA are believeable, matching gives ATET (or ATE)
  - ► Understand Reg through Mat

Intuition: Understand Reg through Mat

- ▶ We will still want to run regressions
  - ▶ The matching idea helps to simplify complicated matrices from last semester, and focus on identification issues

- ▶ Idea: match silimar units, and produce a mean comparison
  - ▶ OLS gives conditional mean comparison
  - ▶ FE compares means of each unit to their average
  - ▶ DID compares difference in mean across locations
  - ▶ RD compares means around the cutoff
  - ▶ IV compares means of instrumented and non-instrumented

- ▶ Goal: give a believable mean comparison
  - ▶ Still need to work out the details

Matching: Idea

- ▶ Idea: recreate a control group based on the treatment group (preprocess the data)
    - ▶ Based on treatment unit's covariate values, match each unit in the treatment group to an untreated unit based on
- ▶ Drop people in the original control group (bias?)
    - ▶ If we drop obs based on a rule that does not depend on the outcome variable, then it should be good
- ▶ Invoke CIA, and compare mean

Exact Matching

▶ Today we will just talk about exact matching

  ▶ There are many others: stratification, nearest neighbor, optimal, . . .

▶ Two Assumptions for Identification

  ▶ CIA: $Y_{0i}, Y_{1i} \perp\!\!\!\perp D_i \mid X_i$
  ▶ CSA: $0 < \Pr(D_i = 1 \mid X_i) < 1$
  ▶ If you just need, say ATET, notations can be simplified

    ▶ MIA:
      $Y_{0i} \perp\!\!\!\perp D_i \mid X_i \implies E(Y_{0i} \mid D_i = 1, X_i) = E(Y_0 i \mid D_i = 0, X_i)$
    ▶ CSA: $\Pr(D_i = 1 \mid X_i) < 1 \implies$ at least there's a control people

Exact Matching Example: Original Data

Unit	X	D	Y
1	1	0	10
2	1	1	15
3	1	1	20
4	2	0	25
5	2	1	30
6	2	0	30
7	3	1	25
8	3	1	35
9	4	0	50
10	5	0	55

Exact Matching Example: Balanced Data

Unit	X_i	Y_{1i}	Y_{0i}	Effect
2	1	15	10	5
3	1	20	10	10
5	2	30	27.5	2.5
7	3	25	NA	NA
8	3	35	NA	NA

- ATET: $(5 + 10 + 2.5) / 3 = 5.8$
- What if we want ATEN?

Exact Matching: Comments

▶ Construct counterfactual using untreated units
  ▶ rely on CIA, CIA, CIA, CIA. . . .
  ▶ again, hard to justify

▶ Drop units whose $\Pr\left(D_i = 1 \mid X_i\right) = 1$
  ▶ No variation in treatment status
  ▶ Matching only estimate the effect when there is CS

▶ Matching only estimates ATET in the CS region
  ▶ This is why we need to show the CS region

What if I just run a regression?

```
X <- c(1, 1, 1, 2, 2, 2, 3, 3, 4, 5)
D <- c(0, 1, 1, 0, 1, 0, 1, 1, 0, 0)
Y <- c(10, 15, 20, 25, 30, 30, 25, 35, 50, 55)
lm(Y ~ D + X)
```

```
##
## Call:
## lm(formula = Y ~ D + X)
##
## Coefficients:
## (Intercept)            D              X
##       6.757        -1.216          9.730
```

I thought regression would give similar result. . .

- ▶ Turned out we need to worry about functional form
    - ▶ In this case, the relationship between Y and X is not quite linear
    - ▶ And the relationship between X and D is also not quite linear
- ▶ Regresion doesn't care about common support
    - ▶ a treated unit with $x_i = 2$ is compared with a control unit with $x_i = 5$
    - ▶ because we assume linearity in x_i
- ▶ Pros: Extrapolation uses more data, gives small SE
- ▶ Cons: Linearity might be problematic

What if I allow x_i to be more flexible?

```
out <- lm(Y ~ D + factor(X))
knitr::kable(summary(out)$coef, digits=2)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	11.67	3.73	3.13	0.04
D	5.00	3.95	1.26	0.27
factor(X)2	15.00	3.95	3.79	0.02
factor(X)3	13.33	4.37	3.05	0.04
factor(X)4	38.33	5.89	6.51	0.00
factor(X)5	43.33	5.89	7.35	0.00

What if I allow x_i to be more flexible?

- ▶ Great, now we got 5, which is much closer to 5.8
  - ▶ Reg actually pick only the common support region if you do it in this way
- ▶ Matching is a nonparametric way to get the effect
  - ▶ Reg can do the same job, provided you know how to add controls
  - ▶ And they both rely on CIA, CIA, CIA, CIA...

Why doesn't Reg give me a 5.8 but 5?

- ▶ Weights:
  - ▶ Mat weights the cell average based on the fraction of the treated units
  - ▶ Reg weights the cell average based on the variance of the treated units
- ▶ Mat: $2/3$ of the treated units have $x_i = 1$ and the average treatment effect is 7.5; $1/3$ of the treated observations have $x_i = 2$ and the average treatment effect is 2.5
  - ▶ 7.5 * $(2/3)$ + 2.5 * $(1/3)$ = 5.8
- ▶ Reg: the variance in cell $x_i = 1$ and $x_i = 2$ are both $p \times (1 - p) = 1/3 * 2/3$ and both cells have 3 observations, so the weights are the same
  - ▶ 7.5 * $(1/2)$ + 2.5 * $(1/2)$ = 5

Motivation and steps

- ▶ Motivation: reduce dimensions
    - ▶ What matters is the design, not a looooong controls
- ▶ The PS theorem looks like CIA
- ▶ Steps:
    - ▶ Choose variables that affect the selection into treatment, run a logit/probit/LPM, predict the scores, match a subsample, check for balacnce, estimate ATET/ATEN/ATE among the common support
    - ▶ Best practice according to Gary King: choose n-match-check, tweak-match-check, tweak-match-check, tweak-match-check. . . . . .
    - ▶ Checking for balance: $X_i \perp\!\!\!\perp D_i \mid \Pr(D_i = 1 \mid X_i)$