Lecture 9: Matching, 1 of 2

March 19, 2025

Course Administration

1. Quantitative summaries should be in
2. I will guarantee a 2-week return; hope for a one-week return
3. No class next week
4. Consultations today and tomorrow instead

Course Administration

1. Quantitative summaries should be in
2. I will guarantee a 2-week return; hope for a one-week return
3. No class next week
4. Consultations today and tomorrow instead
5. Final paper instructions posted

Course Administration

1. Quantitative summaries should be in
2. I will guarantee a 2-week return; hope for a one-week return
3. No class next week
4. Consultations today and tomorrow instead
5. Final paper instructions posted
6. Workshop April 9
   - post draft Sunday April 5
   - post comments April 9 before class
7. Please come see me about your replication paper
8. Presentations April 16 & 23
9. Paper due April 28 by 5 pm
10. Any other issues?

Today

Matching framework

1. Why matching?
2. Propensity Score and Assumptions
3. How to implement

Examples

1. Common support example
2. Brooks on BIDs

Matching Motivation

1. What is matching and why matching?
2. Propensity score and assumptions
3. Estimation

Why Matching?

What is matching?

- Making a comparison group to treated group
- Based on observables
- Many different ways to construct this comparison group

What is matching?

- Making a comparison group to treated group
- Based on observables
- Many different ways to construct this comparison group

- Some say not a method for estimating treatment impact
- Just a method for generating credible sample

Where does matching fit in your toolkit?

- In my opinion, not usually as believable as the other strategies we've studied
- But easy to implement
- Sometimes as good as it will get
- A "control" strategy: just like regression, but with different weighting

Some notation, please

- Y_{1i} outcome for treated
- Y_{0i} outcome for untreated
- we just observe Y_i
- $D_i = 1$ is treated, 0 is otherwise
- X_i are covariates

Matching Overview

- Recall that the general problem is one of omitted variable bias:

$$Y = \beta_0 + \beta_1 D + \beta X + \epsilon,$$

but we fear that $\text{cov}(D, \epsilon) \neq 0$

Matching Overview

- Recall that the general problem is one of omitted variable bias:

$$Y = \beta_0 + \beta_1 D + \beta X + \epsilon,$$

  but we fear that $\text{cov}(D, \epsilon) \neq 0$
- We have explored a variety of techniques to conquer this problem
  - fixed effects
  - difference-in-difference
  - instrumental variables
  - regression discontinuity
- Matching has a similar flavor

Matching: pair a treated observation with the "most similar" non-treated observation(s), and use the difference in their outcomes as the treatment effect

Matching Intuition

- As per usual, we are concerned that

$$E(Y_0|X, D = 1) \neq E(Y_0|X, D = 0)$$

Matching Intuition

- As per usual, we are concerned that

$$E(Y_0|X, D = 1) \neq E(Y_0|X, D = 0)$$

- In words: unobservables may differ by treatment status
- Intuitionally, we are finding a Y_0 for the $D = 1$ guys
- Two empirical challenges:
  1. sufficiently explaining treatment
  2. finding the "most similar" observations
- Talk about cross-sectional matching, then re-visit for the diff-in-diff

Propensity Score and Assumptions

What to Match On?

- Match on things that determine treatment
- Don't match on things that treatment determines

Which Variables to Match On?

- Suppose you think that ten covariates are important determinants of treatment
- Hard to match on all ten
- $\rightarrow$ create a one-dimensional measure that uses all these covariates together

Propensity Score

- An index that describes likelihood of treatment
- Imagine ranking each observation in terms of likelihood of treatment
  - could you do this for an experiment?

Propensity Score

- An index that describes likelihood of treatment
- Imagine ranking each observation in terms of likelihood of treatment
  - could you do this for an experiment?
- Let's call some subset of X that can explain treatment D as Q

Propensity Score

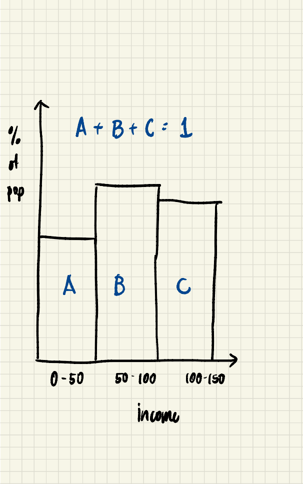
- An index that describes likelihood of treatment
- Imagine ranking each observation in terms of likelihood of treatment
    - could you do this for an experiment?
- Let's call some subset of X that can explain treatment D as Q
- We can estimate this likelihood of treatment as

$$D = \gamma Q + \eta,$$

  where D is treatment (can be discrete, but need not be) and we estimate with logit or probit
- We do need the regular OLS assumption here: $\text{cov}(Q, \eta) = 0$
- Propensity score for each obs is $\hat{D} = \hat{\gamma} Q$

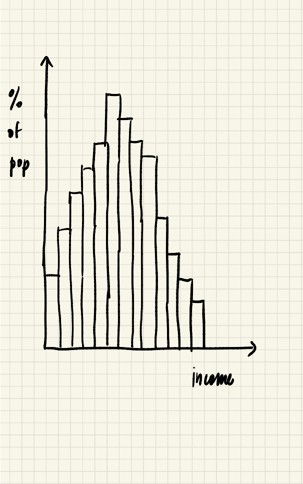
A Probability Density Function

- We want to know the distribution of a given variable
- Look at the probability density function. What is that?
- Imagine we're interested in the distribution of income. You could draw a figure like the one here
- Note that the areas $A + B + C = 1$



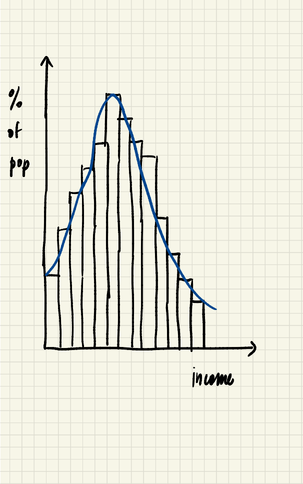
A Probability Density Function

- Now imagine that the boxes get very very narrow. The graph will start to look more like a curve.
- In the same way the area of the boxes equals 1, the area under the curve should sum to one (if you drew narrow boxes upward)
- In calculus $\int_0^I f(i)di = 1$



A Probability Density Function

- Now imagine that the boxes get very very narrow. The graph will start to look more like a curve.
- In the same way the area of the boxes equals 1, the area under the curve should sum to one (if you drew narrow boxes upward)
- In calculus $\int_0^I f(i)di = 1$

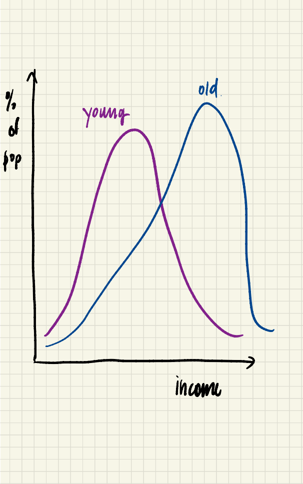


Probability Density Function by Group

- We can also draw two income pdfs
- Maybe one for those under age 45 and one for those over
- The area under each curve should still sum to one

Probability Density Function by Group

- We can also draw two income pdfs
- Maybe one for those under age 45 and one for those over
- The area under each curve should still sum to one



Visualizing the Propensity Score

- All observations have binary treament: $D = 0, 1$
- But a continuous propensity score $0 \leq \hat{D} \leq 1$
- Here are 40 observations

Visualizing the Propensity Score

- All observations have binary treament: $D = 0, 1$
- But a continuous propensity score $0 \leq \hat{D} \leq 1$
- Here are 40 observations, blue for treated, red for comparison

Two Key Assumptions for Matching to Yield Causal Estimates

1. Overlap, or Common Support
2. Unconfoundedness

1. Overlap = Common Support

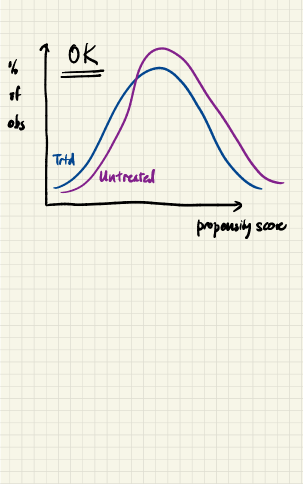
Overlap $\equiv$ treated and untreated observations with similar $\hat{D}$

- We need some treated and untreated for any $\hat{D}$
- Think about the distribution of the propensity score
- What does an ok picture look like?
- What does a bad picture look like?

1. Overlap = Common Support

Overlap ≡ treated and untreated observations with similar $\hat{D}$

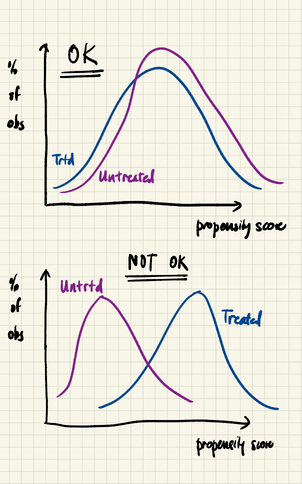
- We need some treated and untreated for any $\hat{D}$
- Think about the distribution of the propensity score
- What does an ok picture look like?
- What does a bad picture look like?



1. Overlap = Common Support

Overlap ≡ treated and untreated observations with similar $\hat{D}$

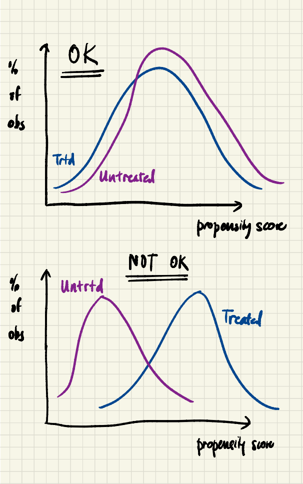
- We need some treated and untreated for any $\hat{D}$
- Think about the distribution of the propensity score
- What does an ok picture look like?
- What does a bad picture look like?



1. Overlap = Common Support

We need both treated and untreated observations with similar $\hat{D}$

- What does this mean, empirically?
- Is this true in an experiment?
- Estimates are only valid where there is common support



2. Unconfoundedness

- In the cross-section

$$E(Y_0|\hat{D}, D = 1) = E(Y_0|\hat{D}, D = 0)$$

- In words, for a given propensity score, same untreated outcome
- Or, the matched observation is a good control

2. Unconfoundedness

- In the cross-section

$$E(Y_0|\hat{D}, D = 1) = E(Y_0|\hat{D}, D = 0)$$

- In words, for a given propensity score, same untreated outcome
- Or, the matched observation is a good control
- Also known as
  - ignorable treatment assignment
  - conditional independence
  - selection on observables

2. Unconfoundedness

- In the cross-section

$$E(Y_0|\hat{D}, D = 1) = E(Y_0|\hat{D}, D = 0)$$

- In words, for a given propensity score, same untreated outcome
- Or, the matched observation is a good control
- Also known as
  - ignorable treatment assignment
  - conditional independence
  - selection on observables
- Treatment is “ignorable” given Q

Restating Unconfoundedness Assumptions

- You cannot select into treatment based on anticipated impact
- Not as tough as an instrument, since Q could still affect Y
- But you do need to be able to entirely explain treatment so that the remaining variation is random
- Age, gender and race are ok, since treatment probably doesn't affect them
- But they may not be sufficient

Implementation and Estimation

Finding "Most Similar" Observations and Estimation

1. One nearest neighbor
2. x nearest neighbors
3. Kernel matching estimator
4. Local linear regression
5. Coarsened exact matching
6. Propensity score weighting
7. Literally zillions of others

1. One Nearest Neighbor

- For each treated observation, find closest observation in propensity score space
- For each person i and another person j, calculate difference in propensity score:

$$|\Pr(D=1|X_i) - \Pr(D=1|X_j)| = |\hat{D}_i - \hat{D}_j|$$

1. One Nearest Neighbor

- For each treated observation, find closest observation in propensity score space
- For each person i and another person j, calculate difference in propensity score:

$$|\Pr(D = 1|X_i) - \Pr(D = 1|X_j)| = |\hat{D}_i - \hat{D}_j|$$

- Let
  - $A_i = 1$ if $|\hat{D}_i - \hat{D}_j| = \min_j(|\hat{D}_i - \hat{D}_j|)$
  - $A_i = 0$ otherwise

1. One Nearest Neighbor

- For each treated observation, find closest observation in propensity score space
- For each person i and another person j, calculate difference in propensity score:

$$|\Pr(D=1|X_i) - \Pr(D=1|X_j)| = |\hat{D}_i - \hat{D}_j|$$

- Let
  - $A_i = 1$ if $|\hat{D}_i - \hat{D}_j| = \min_j(|\hat{D}_i - \hat{D}_j|)$
  - $A_i = 0$ otherwise
- When you have a pair-wise match, you can simply average the difference between the treated and the matched N observations:

$$\beta = \frac{1}{n}\sum_{n \in N}(Y_{1i} - \sum_j (A_i Y_{0j})) = \sum_{n \in N}(Y_i - \sum_j A_i Y_j)$$

Why Is Nearest Neighbor Matching Better Than a Regression?

- Why is this better than a regular old regression where we just control for Q and X?

Why Is Nearest Neighbor Matching Better Than a Regression?

- Why is this better than a regular old regression where we just control for Q and X?
- Not obviously better
- Here we allow the covariates to have a very non-linear effect on treatment
- Regular regression can control only for a linear effect of X.

2. Multiple (x) Nearest Neighbors

- For each treated observation, find the nearest x observations, similarly to finding the closest one
- Let A_{xi} be 1 for all x of them
- You can do a simple estimation as before

$$\beta = \frac{1}{n} \sum_{n \in N} (Y_{1i} - \sum_{j} (\frac{1}{x} A_{xi} Y_{0j}))$$

- Again, take the average difference of treated and control observations

3. Kernel Matching Estimator

- Nearest neighbor matching gives an equal weight to all the neighbors
- Suppose we'd like to be a little more sophisticated and weight observations that are nearby in $\hat{D}$ more heavily

3. Kernel Matching Estimator

- Nearest neighbor matching gives an equal weight to all the neighbors
- Suppose we'd like to be a little more sophisticated and weight observations that are nearby in $\hat{D}$ more heavily
- In general, we want the weights for the matches for each i to sum to one
- There are two key choices: bandwidth and functional form
    - Bandwidth is the width of the kernel
    - Functional form is the type of the kernel
- x nearest neighbors is actually a uniform kernel, but with a varying bandwidth
- A big literature on how to choose the bandwidth

4. Local Linear Regression in Particular

- You may see estimates with a local linear regression (not always in the matching context, either)
- Lowess without matching: for each bandwidth, find $\hat{\beta}$, which yields $\hat{Y}$. Plot $\hat{Y}$
- Lowess with matching: for each bandwidth,

$$\beta = \frac{1}{n}\sum_{n \in N}(\hat{Y}_{1i} - \sum_{j}(W(i,j)A_{xi}\hat{Y}_{0j})),$$

where $W(i,j)$ are the lowess weights

5. Coarsened Exact Matching

Iacus, King and Porro, 2011

- Instead of matching with a propensity score
- Match to a comparison group by bins of characteristics
- For example, one bin could be
  - males 41-42
  - income $50-$52k
  - 5 feet 8 inches to 5 feet 10 inches

⋮

5. Coarsened Exact Matching

Iacus, King and Porro, 2011

- Instead of matching with a propensity score
- Match to a comparison group by bins of characteristics
- For example, one bin could be
  - males 41-42
  - income $50-$52k
  - 5 feet 8 inches to 5 feet 10 inches

  ⋮
- This requires a lot of data!
- Good if you need to match on only a few covariates (Ripollone et al, 2020)

6. Using the Propensity Score Weight

- This is what I do in my paper
- General idea: weight by probability of treatment, so that weights are

$$\sqrt{\frac{D_i}{p(X_i)} + \frac{(1-D_i)}{(1-p(X_i))}}$$

Treated observation

- weight by 1/propensity score
- ≥ 1
- more likely you are to be treated the less weight you get
- bad control for the untreated

6. Using the Propensity Score Weight

- This is what I do in my paper
- General idea: weight by probability of treatment, so that weights are

$$\sqrt{\frac{D_i}{p(X_i)} + \frac{(1 - D_i)}{(1 - p(X_i))}}$$

Treated observation

- weight by 1/propensity score
- ≥ 1
- more likely you are to be treated the less weight you get
- bad control for the untreated

Untreated observation

- weight by 1/(1 - propensity score)
- ≥ 1
- bigger $p(X_i) \rightarrow$ smaller $(1 - p(X_i)) \rightarrow$ bigger fraction
- more likely you are to be treated the more weight you get

6., Alternatives: Use Machine Learning

- So far we used a logit or probit estimation to generate a propensity score
- Machine learning options exist
  - random forest
  - XGboost
  - genetic matching
- Not clear when these other methods outperform simpler ones

Diff-in-diff Extension to This Framework

- Let t be the time before treatment, and t' be after
- General idea: match on pre-treatment characteristics
- More plausible to think that these are independent of treatment

Diff-in-diff Extension to This Framework

- Let t be the time before treatment, and t' be after
- General idea: match on pre-treatment characteristics
- More plausible to think that these are independent of treatment
- The common support assumption remains the same
- Unconfoundedness is now: $E(Y_{0t} - Y_{0t'}|\hat{D}, D = 1) = E(Y_{0t} - Y_{0t'}|\hat{D}, D = 0)$
- Any examples where you could use this?

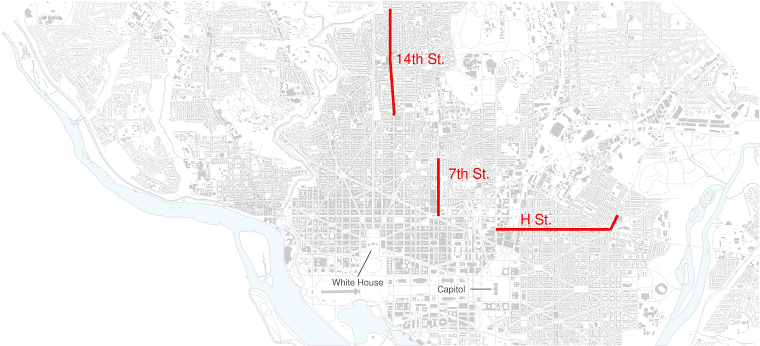
Lecture 9: Matching Example

Denoument of Larger Project

- Do neighborhoods substantially harmed in DC's 1968 civil disturbance have greater variation in property value today?
- Use lot-level data
- Calculate coefficient of variation by block

$$CV_s = \frac{\text{sd(improvements per lot square foot}_i)}{\text{mean(improvements per lot square foot}_i)}$$

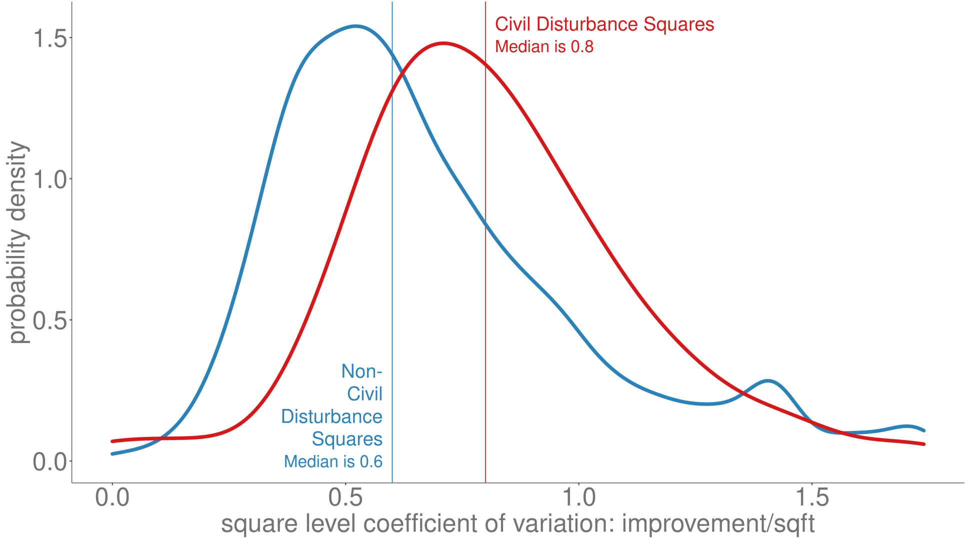
Neighborhoods Harmed



Comparing Squares with and without Civil Disturbance

- For each 2019 DC square
- Find coeff. of variation
- In improvements per sq ft
- Compare civil disturbance areas to the city at large

Distribution of Log of Coefficient of Variation



Why Should You Be Suspicious of This Comparison?

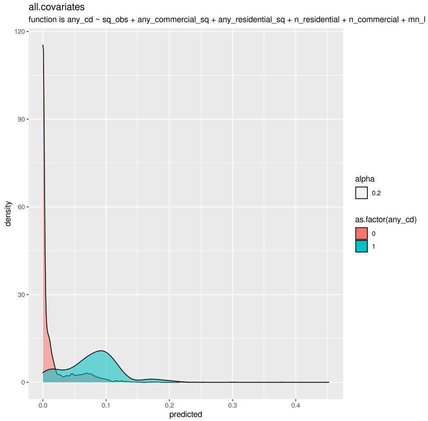
Why Should You Be Suspicious of This Comparison?

Destroyed areas

- are more commercial
- are more dense
- are more central
- have more individual lots (?)
- have more retail (?)

So comparing them a neighborhood of single-family homes would not be compelling.

My First Try in Matching



- First try at kitchen sink propensity function
- Picture is not compelling!

Two Matching Strategies

	Full Sample			Restricted Sample			Restricted Sample + Wtd		
	Mean			Mean			Mean		
	CD	Others	t	CD	Others	t	CD	Others	t
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log(CV of imp/sq ft)	-0.22	-0.43		-0.24	-0.29		-0.22	-0.29	
Matching Variables									
Log of mean land value/sq ft	5.6	4.79	-18.06	5.6	5.54	-1.34	5.56	5.54	-0.38
Lots in square	44.42	33.15	-3.39	45.87	35.53	-2.77	39.8	36.16	-1.02
Share square lots comm.	0.21	0.15	-2.21	0.22	0.27	1.86	0.26	0.27	0.37
Sample restriction variables									
1{All lots residential}	0.04	0.23	7.39	0	0				
1{Land value per sqft in CD range}	0	0.42	53.75	0	0				
1{Zero com. lots}	0.1	0.5	10.95	0	0				
Squares	71	4042		63	1213		63	1213	

Bottom Line: Differences Persist Even With Matching

	Sample and Weights					
	Full	Restricted	Restricted & Weighted			
	(1)	(2)	(3)	(4)	(5)	(6)
1{CD square}	0.210***	0.051	0.072***	0.080***	0.074***	0.074***
	(0.075)	(0.082)	(0.03)	(0.029)	(0.03)	(0.03)
Log(mean land val / sq ft)				-0.042	-0.026	-0.027
				(0.03)	(0.034)	(0.035)
N lots in square				-0.001	-0.002	-0.002
				(0.001)	(0.001)	(0.001)
Share lots comm.				0.325***	0.229***	0.229***
				(0.063)	(0.076)	(0.076)
Log(mean impv / sq ft)					-0.019	-0.019
					(0.017)	(0.018)
N comm. lots					0.007**	0.007**
					(0.003)	(0.003)
Mean imp/land value						-0.001
						(0.006)
R^2	0.002	0	0.005	0.045	0.05	0.05
Observations	4,112	1,276	1,276	1,276	1,276	1,276

Lecture 9: Brooks on BIDs

Setting the Stage for the Paper

- What is a BID?
- What is a collective action problem?

Setting the Stage for the Paper

- What is a BID?
- What is a collective action problem?
- What's the unit of analysis in this paper?

Setting the Stage for the Paper

- What is a BID?
- What is a collective action problem?
- What's the unit of analysis in this paper? Reporting district in a given year

Setting the Stage for the Paper

- What is a BID?
- What is a collective action problem?
- What's the unit of analysis in this paper? Reporting district in a given year
- What's the identification problem?

Estimation

What is the estimating equation?

Estimation

What is the estimating equation?

$$\text{crime} = \beta_0 + \beta_1 \text{BID}_i * \text{after}_{it} + \beta_2 \text{year}_t * \text{division}_a + \beta_3 \text{rd}_i + \epsilon_{iat}$$

Estimation

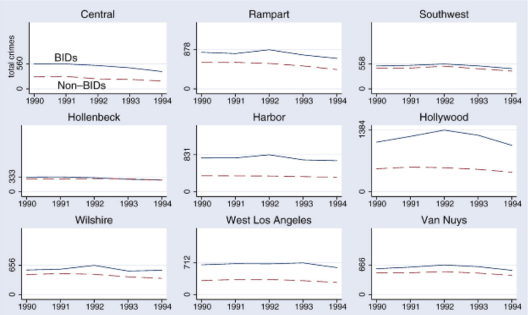
What is the estimating equation?

$$\text{crime} = \beta_0 + \beta_1 \text{BID}_i * \text{after}_{it} + \beta_2 \text{year}_t * \text{division}_a + \beta_3 \text{rd}_i + \epsilon_{iat}$$

- What is the coefficient of interest here?

Identifying Assumptions

$$\text{crime} = \beta_0 + \beta_1 \text{BID}_i * \text{after}_{it} + \beta_2 \text{year}_t * \text{division}_a + \beta_3 \text{rd}_i + \epsilon_{iat}$$



- Why do I have figure A2?
- Why do I have year*division fixed effects?

Propensity Score Matching and BIDs

1. Calculate a propensity score

Propensity Score Matching and BIDs

1. Calculate a propensity score

$$\Pr(\text{BID}_i = 1) = e(X_i) = f(\text{serious crime 1990 - 1994, less serious crime 1990-1994, era of development, census variables})$$

- why does this have only pre-BID crime?

Propensity Score Matching and BIDs

1. Calculate a propensity score

$$\Pr(\text{BID}_i = 1) = e(X_i) = f(\text{serious crime 1990 - 1994, less serious crime 1990-1994, era of development, census variables})$$

   - why does this have only pre-BID crime?

2. Make weights

$$\lambda_i = \sqrt{\frac{\text{BID}_i}{e(X_i)} + \frac{(1 - \text{BID}_i)}{(1 - e(X_i))}}$$

   - What does this equation mean?

Propensity Score Matching and BIDs

1. Calculate a propensity score

$$\Pr(\text{BID}_i = 1) = e(X_i) = f(\text{serious crime 1990 - 1994, less serious crime 1990-1994,} \\ \text{era of development, census variables})$$

   - why does this have only pre-BID crime?

2. Make weights

$$\lambda_i = \sqrt{\frac{\text{BID}_i}{e(X_i)} + \frac{(1 - \text{BID}_i)}{(1 - e(X_i))}}$$

   - What does this equation mean?
     - if a BID obs, $\text{BID}_i = 1$, more weight if propensity score is low
     - if not a BID obs, $\text{BID}_i = 0$, more weight if propensity score is high

Propensity Score Matching and BIDs

1. Calculate a propensity score

$$\Pr(\text{BID}_i = 1) = e(X_i) = f(\text{serious crime 1990 - 1994, less serious crime 1990-1994,} \\ \text{era of development, census variables})$$

   - why does this have only pre-BID crime?

2. Make weights

$$\lambda_i = \sqrt{\frac{\text{BID}_i}{e(X_i)} + \frac{(1 - \text{BID}_i)}{(1 - e(X_i))}}$$

   - What does this equation mean?
     - if a BID obs, $\text{BID}_i = 1$, more weight if propensity score is low
     - if not a BID obs, $\text{BID}_i = 0$, more weight if propensity score is high

3. Run a weighted regression

Interpreting the Results

	Observations	Totals		
		Serious	Less serious	Overall
No fixed effects	13,117	51.07	88.55	139.62
		6.53**	6.26**	6.55**
Fixed effects	13,117	−44.46	−12.69	−57.15
		7.91**	5.46*	11.18**
Almost BIDs	3250	−31.64	−23.12	−54.76
		12.47*	8.07**	17.98**
Matching	12,831	−25.84	−10.04	−35.79
		4.36**	4.67*	7.48**
Neighbors	5434	−39.91	−5.35	−45.26
		8.02**	6.61	12.49**

**Significant at the 1% level. *Significant at the 5% level.

- What does it mean that the first row in the table has positive coefficients and the second row has negative ones?

Interpreting the Results

	Observations	Totals		
		Serious	Less serious	Overall
No fixed effects	13,117	51.07	88.55	139.62
		6.53**	6.26**	6.55**
Fixed effects	13,117	−44.46	−12.69	−57.15
		7.91**	5.46*	11.18**
Almost BIDs	3250	−31.64	−23.12	−54.76
		12.47*	8.07**	17.98**
Matching	12,831	−25.84	−10.04	−35.79
		4.36**	4.67*	7.48**
Neighbors	5434	−39.91	−5.35	−45.26
		8.02**	6.61	12.49**

**Significant at the 1% level. *Significant at the 5% level.

- What does it mean that the first row in the table has positive coefficients and the second row has negative ones?
- How do we interpret -44.5?

Interpreting the Results

	Observations	Totals		
		Serious	Less serious	Overall
No fixed effects	13,117	51.07	88.55	139.62
		6.53**	6.26**	6.55**
Fixed effects	13,117	−44.46	−12.69	−57.15
		7.91**	5.46*	11.18**
Almost BIDs	3250	−31.64	−23.12	−54.76
		12.47*	8.07**	17.98**
Matching	12,831	−25.84	−10.04	−35.79
		4.36**	4.67*	7.48**
Neighbors	5434	−39.91	−5.35	−45.26
		8.02**	6.61	12.49**

**Significant at the 1% level. *Significant at the 5% level.

- What does it mean that the first row in the table has positive coefficients and the second row has negative ones?
- How do we interpret -44.5?
- What does the matching coefficient mean?

Interpreting the Results

	Observations	Totals		
		Serious	Less serious	Overall
No fixed effects	13,117	51.07	88.55	139.62
		6.53**	6.26**	6.55**
Fixed effects	13,117	−44.46	−12.69	−57.15
		7.91**	5.46*	11.18**
Almost BIDs	3250	−31.64	−23.12	−54.76
		12.47*	8.07**	17.98**
Matching	12,831	−25.84	−10.04	−35.79
		4.36**	4.67*	7.48**
Neighbors	5434	−39.91	−5.35	−45.26
		8.02**	6.61	12.49**

**Significant at the 1% level. *Significant at the 5% level.

- What does it mean that the first row in the table has positive coefficients and the second row has negative ones?
- How do we interpret -44.5?
- What does the matching coefficient mean?
- Questions about any of the strategies?

Interpreting the Results

	Observations	Totals		
		Serious	Less serious	Overall
No fixed effects	13,117	51.07	88.55	139.62
		6.53**	6.26**	6.55**
Fixed effects	13,117	−44.46	−12.69	−57.15
		7.91**	5.46*	11.18**
Almost BIDs	3250	−31.64	−23.12	−54.76
		12.47*	8.07**	17.98**
Matching	12,831	−25.84	−10.04	−35.79
		4.36**	4.67*	7.48**
Neighbors	5434	−39.91	−5.35	−45.26
		8.02**	6.61	12.49**

**Significant at the 1% level. *Significant at the 5% level.

- What does it mean that the first row in the table has positive coefficients and the second row has negative ones?
- How do we interpret -44.5?
- What does the matching coefficient mean?
- Questions about any of the strategies?
- Which strategy did you prefer and why?

Next Lecture

- Next week: no class
- Next next week: Lecture 10
  - Matching II: synthetic controls
  - Read selected bits of paper
  - I'll return comments on quantitative summaries
- Lecture 11: Half-class with request and in-class workshop