UNIVERSITY OF ILLINOIS

AT URBANA-CHAMPAIGN

Matching Estimators

1867

illinois.edu

Outline

- Why matching and what is needed?
- Types of matching
- How to choose *X*'s and run specification tests
- Matching and Difference-in-Difference
- Bounding Bias
- Practical Issues

Why Match and What is Needed?

Why Match?

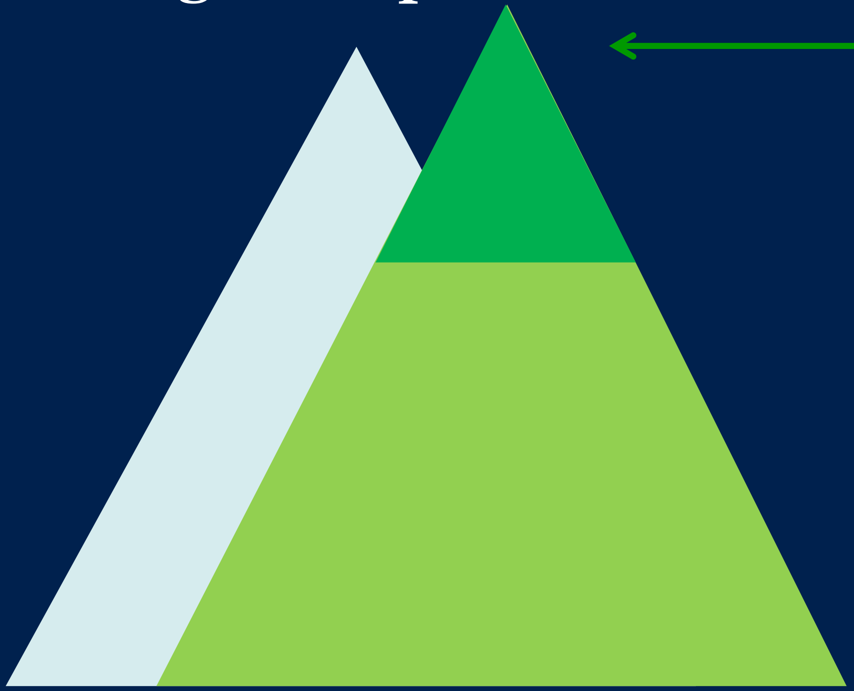
- Randomization is not always feasible

  …and doesn't always work
- Selection bias: program bias, administrative bias

- E.g. protected area established based on criteria on elevation, slope and species habitat
- E.g.2 Irrigation intervention targeted at vulnerable households in semi-arid areas

The Problem

- We want to estimate the effect of a policy that targets a specific area



Policy (Treatment)

Issue: *what would have happened without the policy?*

Matching

- Or, more likely…
- Find a set of 'matched' control (untreated) parcels with similar characteristics to the treated parcels
- If we do a good job finding control parcels that look like the treated parcels, we can replicate what would have happened without protection*

*Assumes unobservable characteristics are distributed in the same way as observables: i.e. Matching over observables = matching over unobservables

Why Not Just Run a Regression?

- Can still control for covariates – but some characteristics may have a non-linear effect
- Often is a matter of throwing away ‘bad’ controls

Assumptions Needed

- Conditional Independence Assumption (CIA)
  - Once you've controlled for the observable *X*'s, you've controlled for selection (either *X*'s explain everything, or *U*'s are distributed in the same way as the *X*'s)
- Common Support Assumption
  - Enough control observations have characteristics in the same range as the treated observations
  - E.g. if irrigation scheme targets all high elevation land in an eco-region, one might not be able to find good controls
  - E.g.2 if new sustainable intensification schemes target all farm households with less than 1 ha, might have difficulties finding controls

Types of Matching

Types of Matches

- Characteristic matching (nn matching; mahalanobis matching)
- Propensity Score Matching
- Exact matching (stratified matching)
- Caliper matching
- Kernel matching

i	T	Edu	Income
1	0	2	60
2	0	3	80
3	0	5	90
4	0	12	200
5	1	5	100
6	1	3	80
7	1	4	90
8	1	2	70

From Heinrich et al 2010

i	T	Edu	Income	Match	Yi	Yo	Diff
1	0	2	60				
2	0	3	80				
3	0	5	90				
4	0	12	200				
5	1	5	100	[3]	100	90	10
6	1	3	80	[2]	80	80	0
7	1	4	90	[2,3]	90	85	5
8	1	2	70	[1]	70	60	10

ATT = 6.25

Dimensionality Issue

- Can easily match over only one variable (dimension)
- Once one has multiple variables, one needs to weight them and create an index
- What is the appropriate weighting scheme?

Characteristic Matching

- Based on *X*'s, where they are weighted by inverse variance

  X_1 slope, $\sigma^2 = 5$

  X_2 wealth, $\sigma^2 = 700$

Characteristic Matching

- Based on *X*'s, where they are weighted by inverse variance

  X_1 slope, $\sigma^2 = 5$

  X_2 wealth, $\sigma^2 = 700$

- Malahanobis: Proximity to the center of the mass of the treated

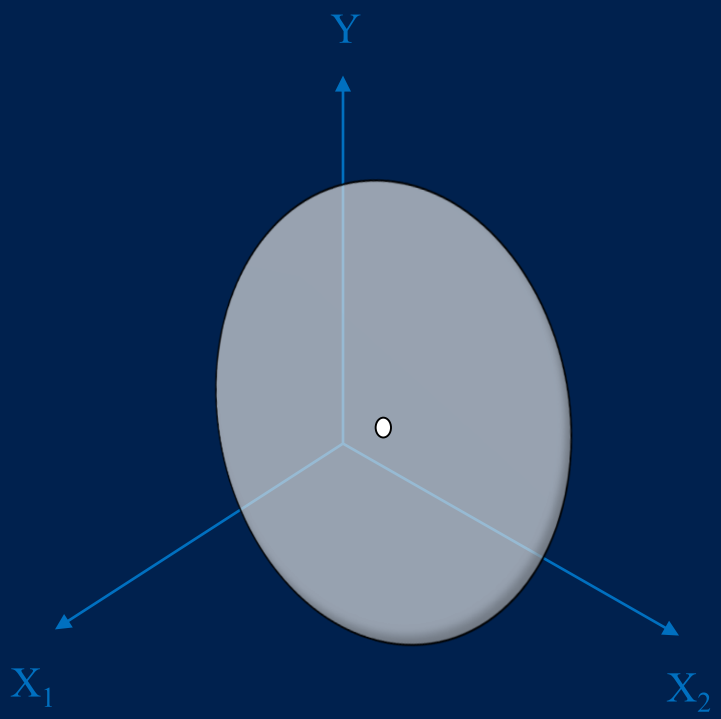
Characteristic Matching

- Based on X's, where they are weighted by inverse variance

  X_1 slope, $\sigma^2 = 5$

  X_2 wealth, $\sigma^2 = 700$

- Malahanobis: Proximity to the center of the mass of the treated



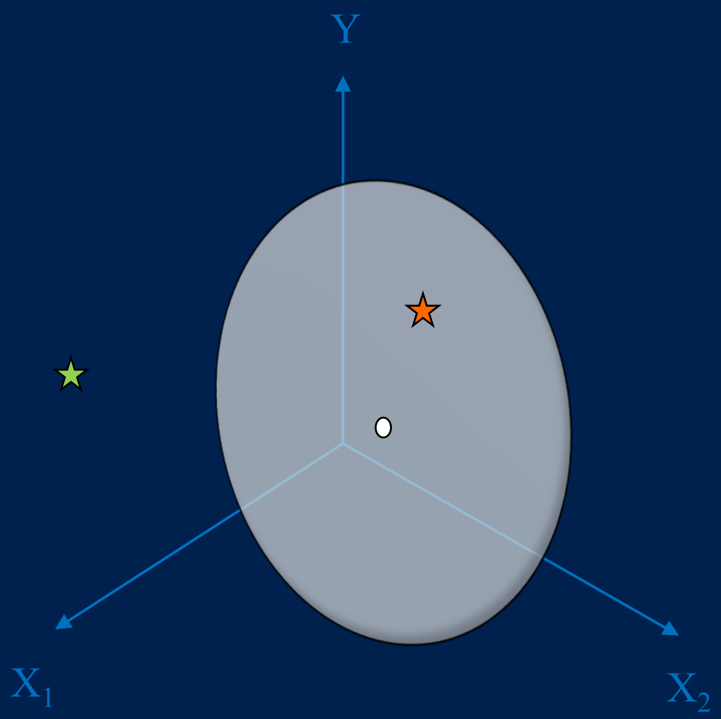
Characteristic Matching

- Based on *X*'s, where they are weighted by inverse variance

  X_1 slope, $\sigma^2 = 5$

  X_2 wealth, $\sigma^2 = 700$

- Malahanobis: Proximity to the center of the mass of the treated



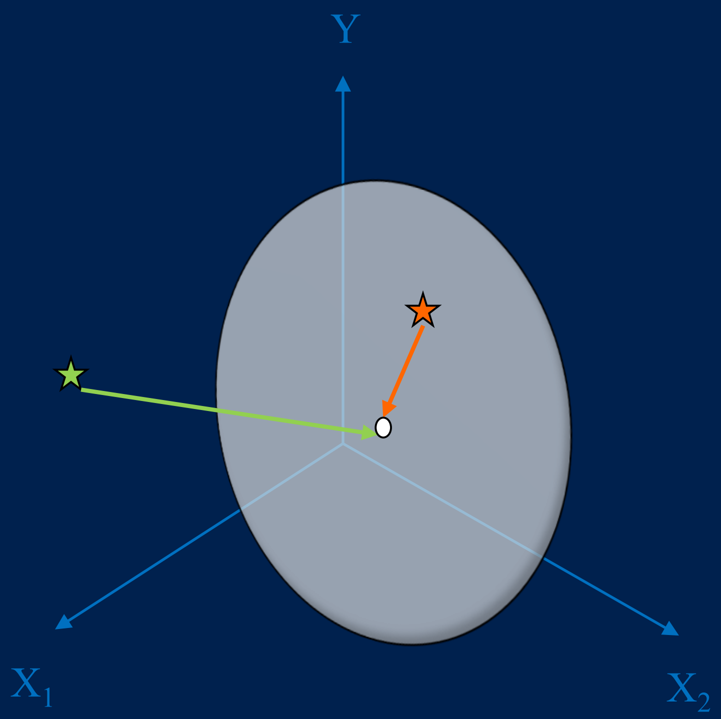
Characteristic Matching

- Based on *X*'s, where they are weighted by inverse variance

  X_1 slope, $\sigma^2 = 5$

  X_2 wealth, $\sigma^2 = 700$

- Malahanobis: Proximity to the center of the mass of the treated



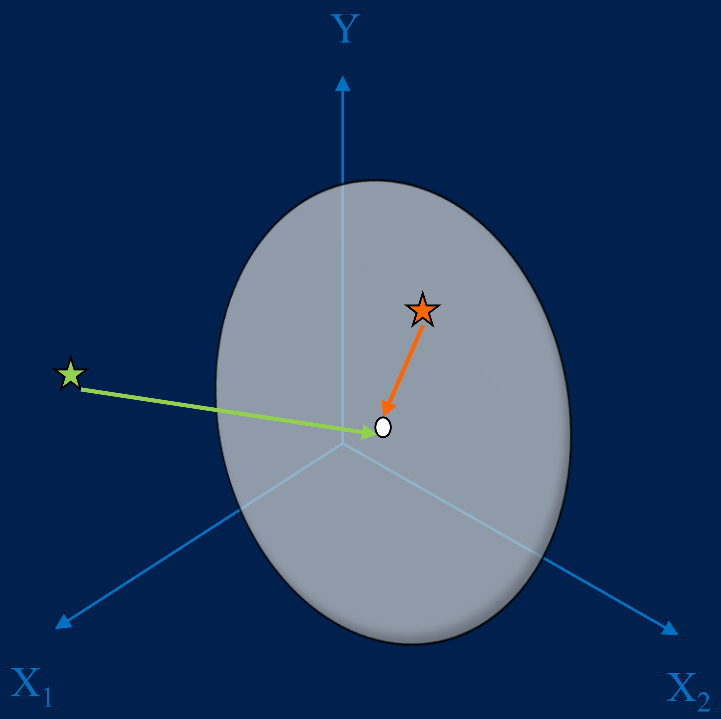
Characteristic Matching

- Based on *X*'s, where they are weighted by inverse variance

  X_1 slope, $\sigma^2 = 5$

  X_2 wealth, $\sigma^2 = 700$

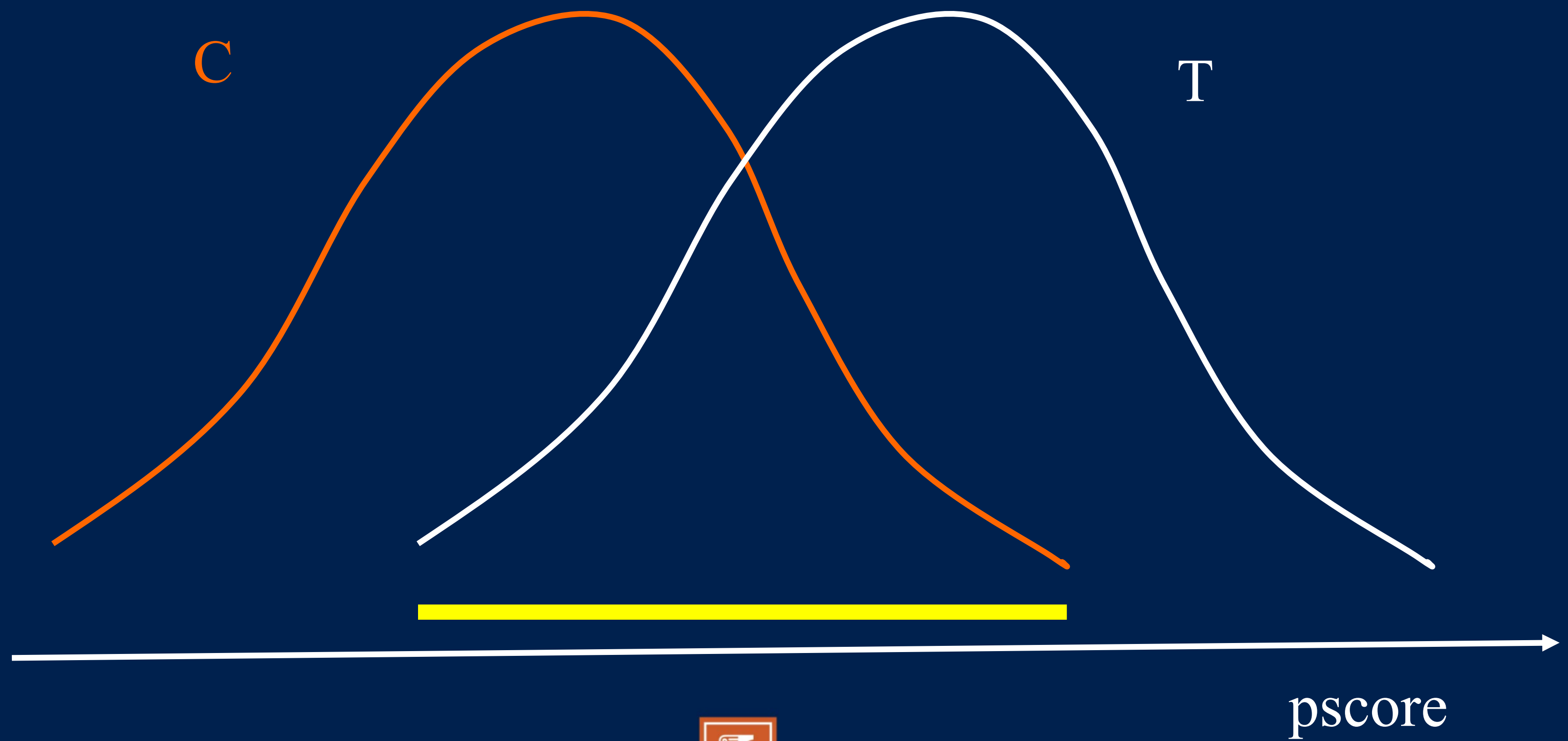
- Malahanobis: Proximity to the center of the mass of the treated
- Different ways of measuring 'distance'



Propensity Score Matching

- Measure ‘distance’ by probability of treatment
- First estimate the probability of treatment given observable characteristics using a probit or logit
- Then predict that probability for all treated and control observations
- Use these predicted probabilities as ‘coordinates’ to allocate which controls are near to which treatment observations

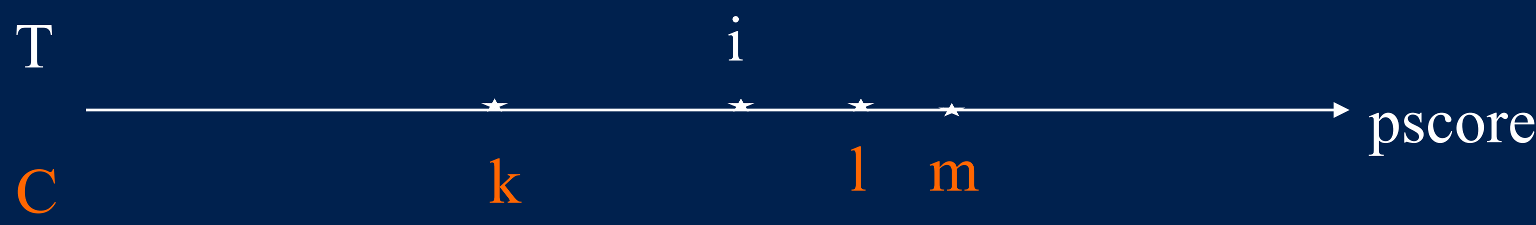
Common support



Exact Matching

- Forcing the control observations to be exactly the same over some dimension
  - E.g. sex; ecoregion

How to define a ‘nearby’ control?

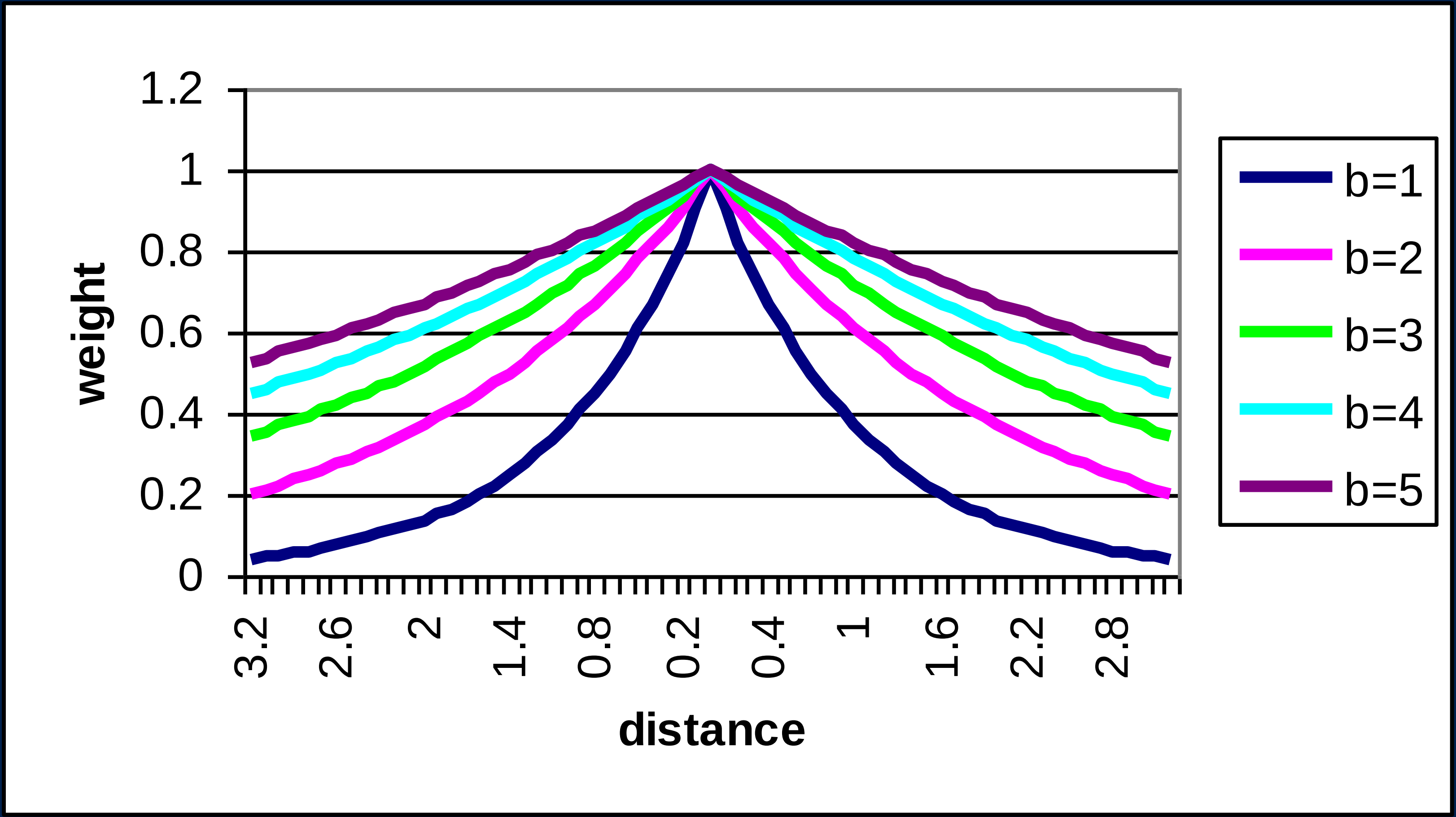


1) Pick ‘n’ closest observations in terms of propensity score, weighted equally: $\frac{1}{n}$
2) Pick a certain distance and include all observations within that distance
3) Weight observations using a kernel

Kernel matching

- weight each observation based on its ‘distance’
- $\omega_{ij} = \left[1 - \left(\frac{\delta_{ij}}{d_i}\right)^3\right]^3 I\left(\delta_{ij} < d_i\right)$
- Where δ_{ij} is the distance between i and j
- d_i is the distance to the q^{th} nearest neighbor
- q is often referred to as the ‘window size’
- Alternatively, $\omega_{ij} = \Phi\left[\frac{\delta_{ij}}{s_i b}\right]$
- where s is the std. dev. of distances between i and all others, and b is the bandwidth

Effect of bandwidth on distance weights



Matching Options

- With or without replacement
- Weighted
- Efficiency vs bias trade-off

How to Choose X's and Run Specification Tests

How to Choose Covariates?

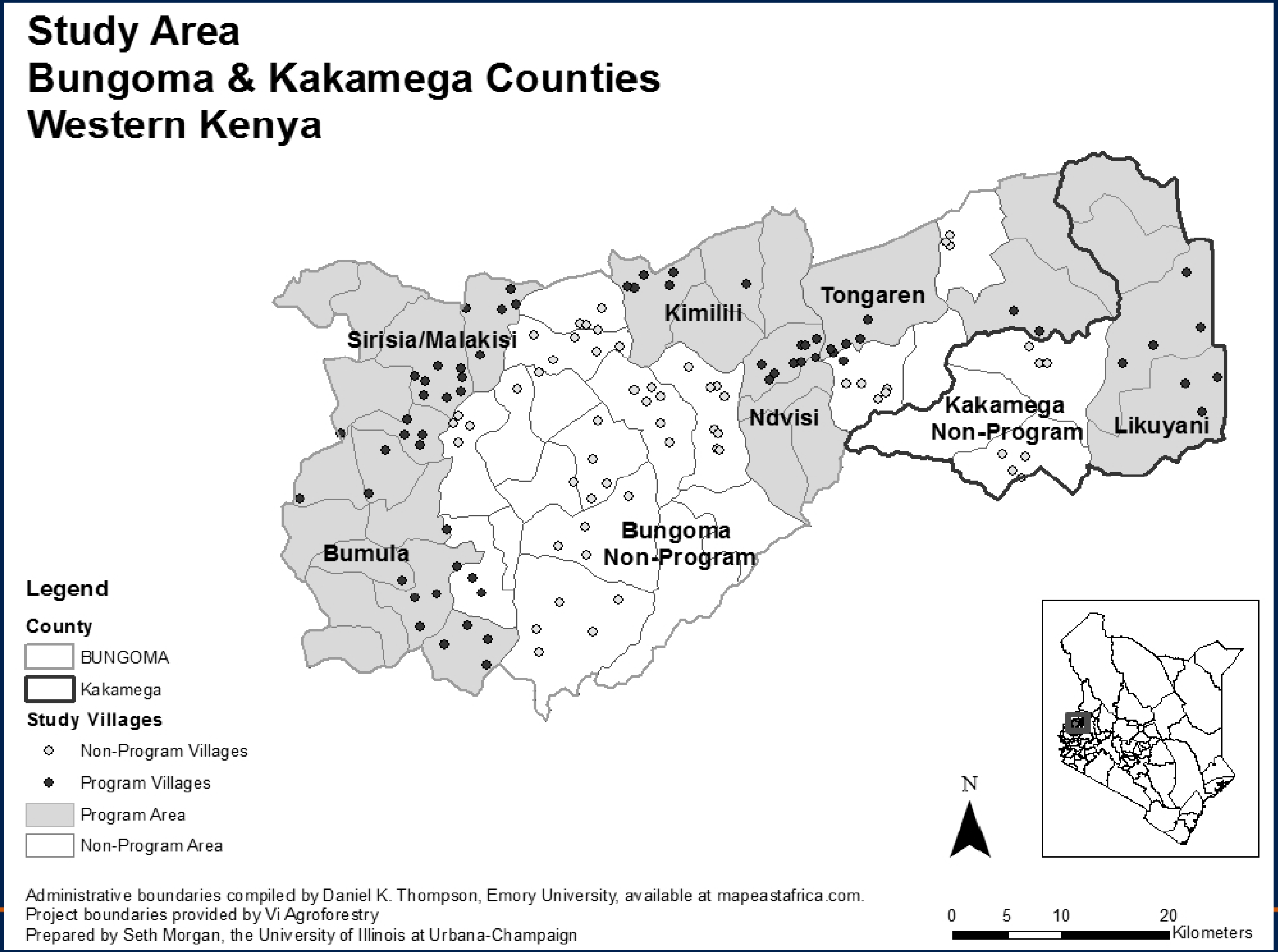
1. Things that affect selection into treatment
2. Things that affect outcome
   - If don't condition on them in the estimation

- Regressions on treatment and on outcome
- Do not use characteristics that might be affected by treatment
  - Consider spillovers

Things to Keep in Mind…

- Unlike a regression, adding in a covariate that doesn't affect selection and/or outcome can matter
- Ensure data for controls and treatment are from the same source (same frequency/granularity; same probability of missing etc)
- Heterogeneity analysis – may want to match over subsets (exact matching)
  - E.g. gender in PSM

Pre-Survey Matching

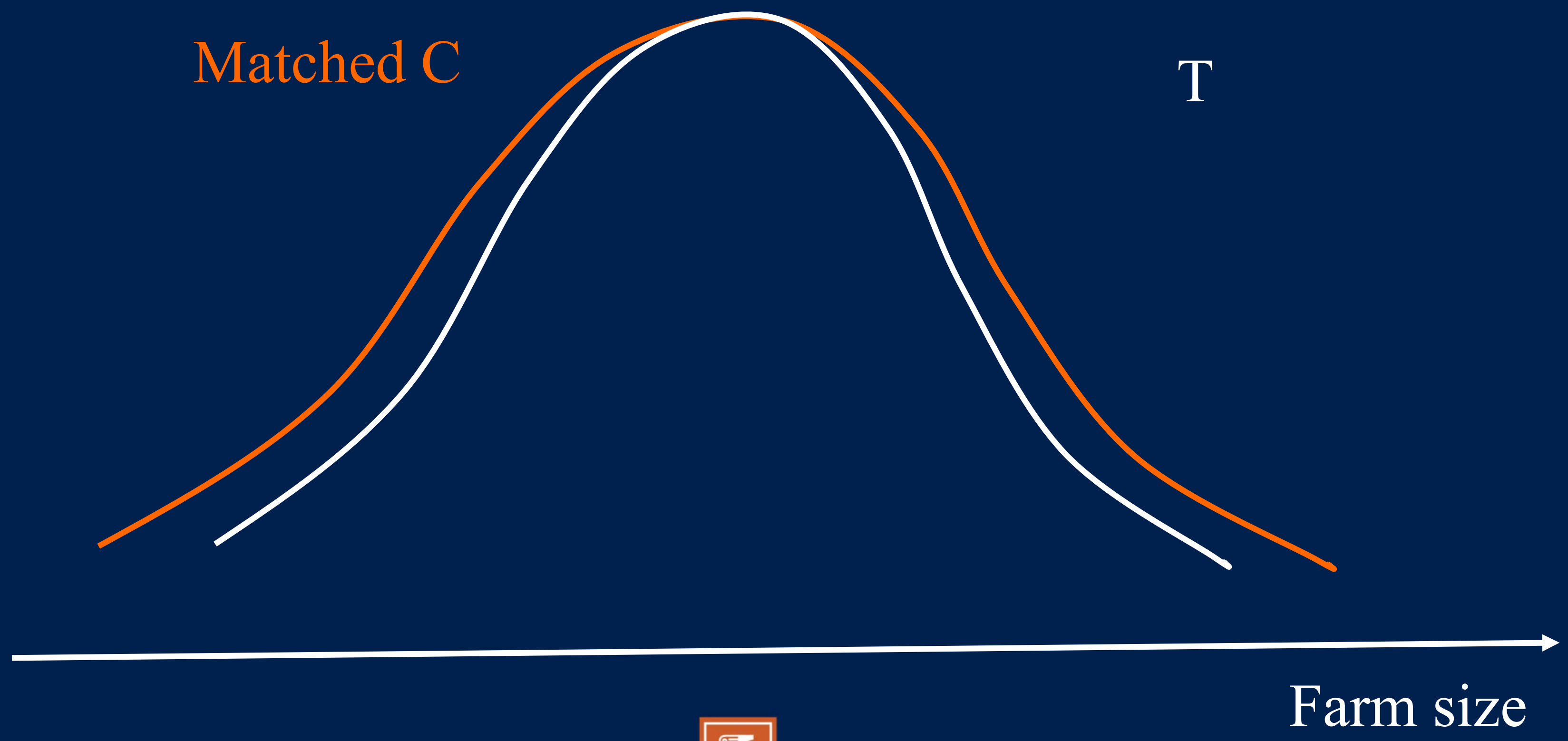


What if we don't want to throw away data?

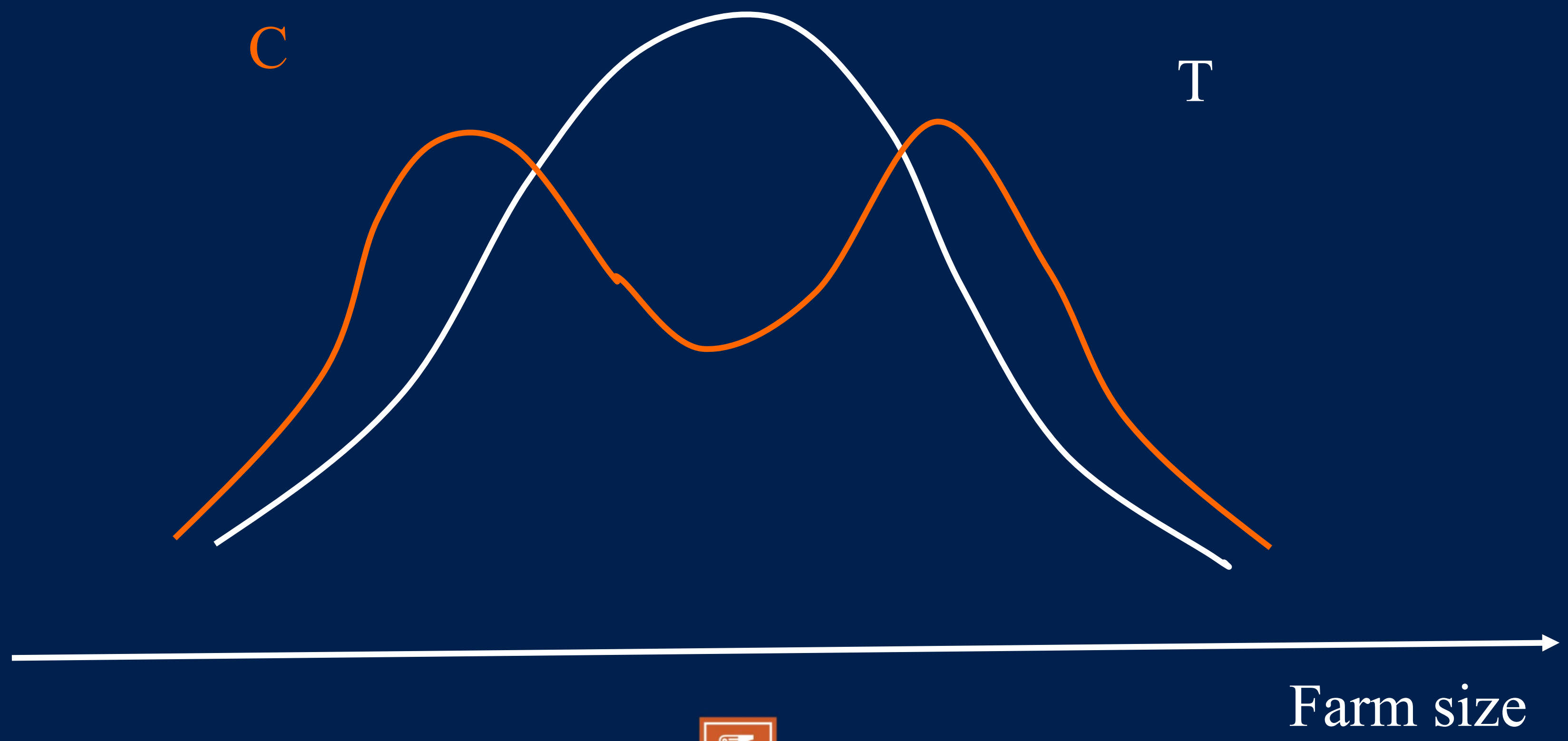
Specification Tests

- Covariate balance test
  - Including distribution
- Multiple algorithms for robustness
  - Test to examine if results are sensitive to a few bad matches
- Visual inspection of propensity scores (before and after)
- Mapping

Checking X's



But maybe...



Matching and Difference-in-Difference

DiD Matching

- Matching can be used in a DiD framework
- Control for time-invariant unobservables
- (Heckman Ichimura and Todd 1997; Heckman et al 1998)

Other options

- Matching with Continuous Treatment (Imbens 2000)
- Matching with a roll-out design

Measuring the degree of potential bias

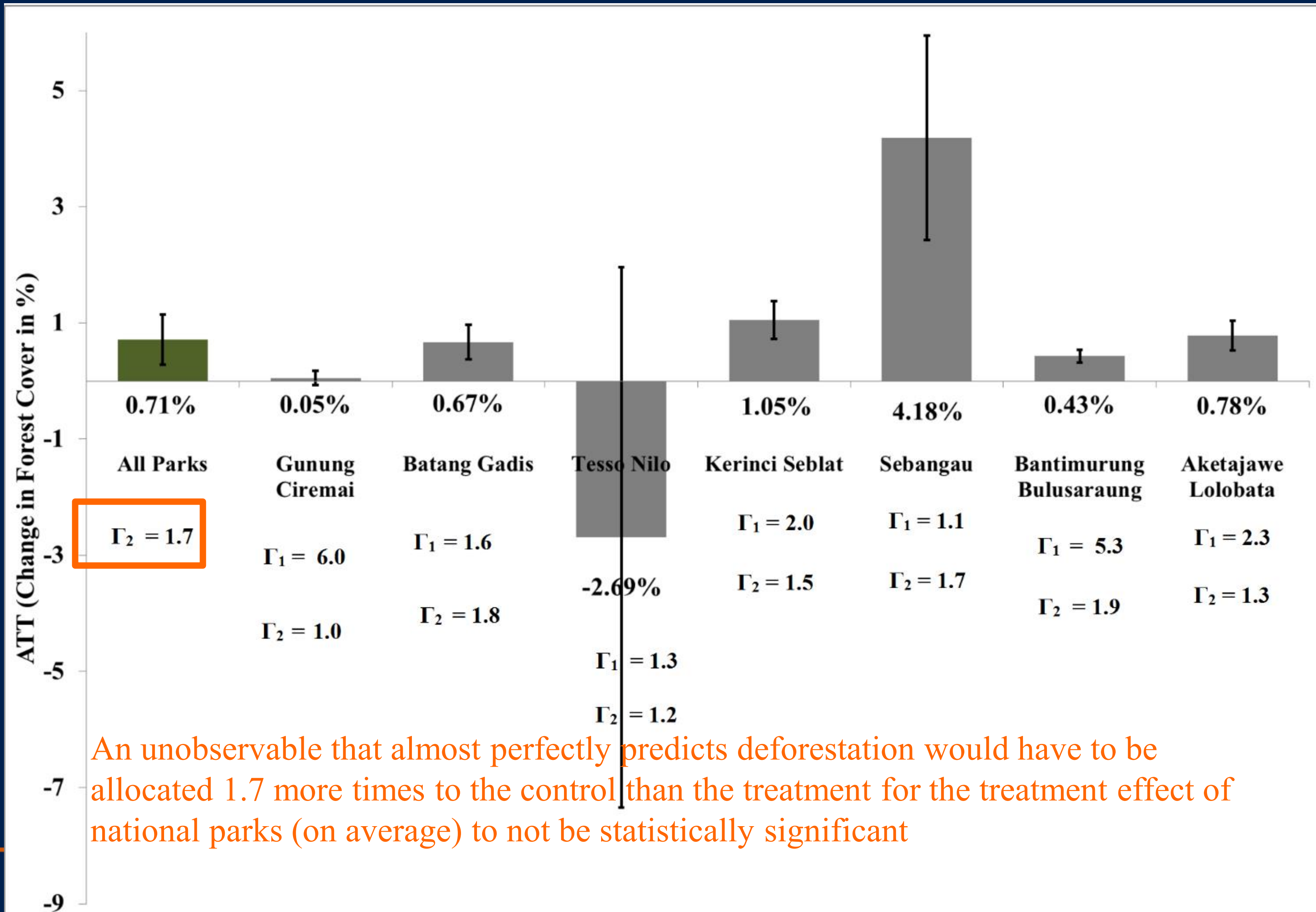
Rosenbaum Bounds

- Rosenbaum (2002): Identify "hidden bias" from unobservable covariates
- Ask how much unobservables might affect results (make the ATT insignificant)
- Specifically, estimates an odds ratio of how much could an unobserved variable bias outcome by affecting selection

e.g. Shah and Baylis

- Comparing effect of parks across Indonesia
  - Do unobservable covariates affect whether individual park ATT is different from the national ATT (Γ_1)
  - Do unobservable covariates affect the park level ATT estimates (Γ_2)

Test for Hidden Bias



An unobservable that almost perfectly predicts deforestation would have to be allocated 1.7 more times to the control than the treatment for the treatment effect of national parks (on average) to not be statistically significant

Interpretation of the Γ

- Lalonde (1985): effect of job training on wages
- match on age, education, race, marital status, high school degree, earnings for the two years before the training program and unemployment before the training program.

ATT: 1767.7 (830.85); p.stat = 0.033; $\Gamma = 1.05$

Interpretation of the Γ

- Lalonde (1985): effect of job training on wages
- match on age, education, race, marital status, high school degree, earnings for the two years before the training program and unemployment before the training program.

ATT: 1767.7 (830.85); p.stat = 0.033; $\Gamma = 1.05$

If the odds of a person being in the program are 1.05 times higher than in the control due to an unobservable, the estimated treatment effect is no longer significant at the 0.05 level.

Interpretation of the Γ

- Lalonde (1985): effect of job training on wages
- match on age, education, race, marital status, high school degree, earnings for the two years before the training program and unemployment before the training program.

ATT: 1767.7 (830.85); p.stat = 0.033; $\Gamma = 1.05$

If the odds of a person being in the program are 1.05 times higher than in the control due to an unobservable, the estimated treatment effect is no longer significant at the 0.05 level.

```
> #Sensitivity Tests
> psens(mgen1, Gamma=1.5, GammaInc=.1)
Rosenbaum Sensitivity Test for Wilcoxon Signed Rank P-Value
     Gamma L. Bound P-Value U. Bound P-Value
[1,]   1.0           0.0346           0.0346
[2,]   1.1           0.0062           0.1271
[3,]   1.2           0.0009           0.3000
[4,]   1.3           0.0001           0.5164
[5,]   1.4           0.0000           0.7139
[6,]   1.5           0.0000           0.8539
```

If the odds of a person being in the training program are only 1.1 times higher b/c of an unobservable that affects income, the p-value as high as 0.127

Can also bound the estimated treatment effect

Rosenbaum Sensitivity Test for Hodges-Lehmann Point Estimate

	Gamma	L. Bound HL Est.	U. Bound HL Est.
[1,]	1.0	1194.000000	1194.0
[2,]	1.1	560.780000	1231.2
[3,]	1.2	274.080000	1598.4
[4,]	1.3	-0.015006	1944.9
[5,]	1.4	-113.220000	2218.3
[6,]	1.5	-333.220000	2424.8

Median effect size if no difference in unobservables

If people are 1.5 times more likely to be in treatment due to an unobservable that affects income, the median treatment effect could be as high as $2424.80 or as low as -333.22

Can also bound the estimated treatment effect

Rosenbaum Sensitivity Test for Hodges-Lehmann Point Estimate

	Gamma	L. Bound HL Est.	U. Bound HL Est.
[1,]	1.0	1194.000000	1194.0
[2,]	1.1	560.780000	1231.2
[3,]	1.2	274.080000	1598.4
[4,]	1.3	-0.015006	1944.9
[5,]	1.4	-113.220000	2218.3
[6,]	1.5	-333.220000	2424.8

Median effect size if no difference in unobservables

If people are 1.5 times more likely to be in treatment due to an unobservable that affects income, the median treatment effect could be as high as $2424.80 or as low as -333.22

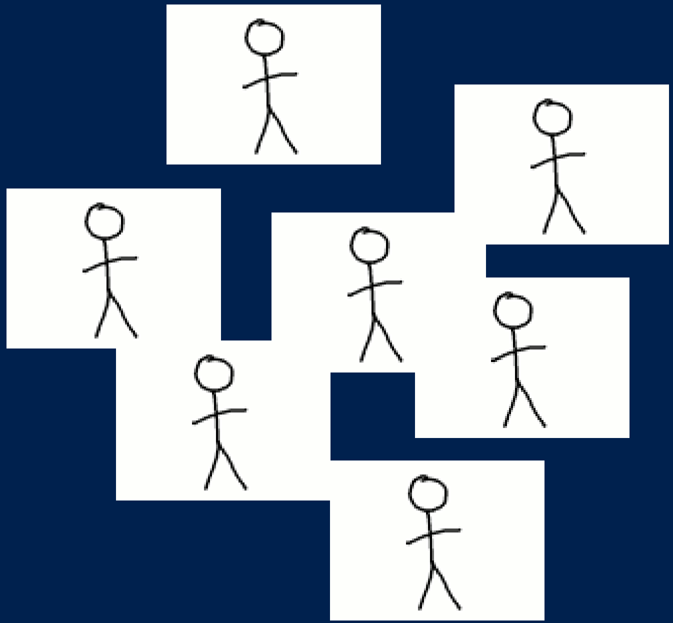
Practical note: most socio-economic studies have $\Gamma < 2$.

Practical Issues

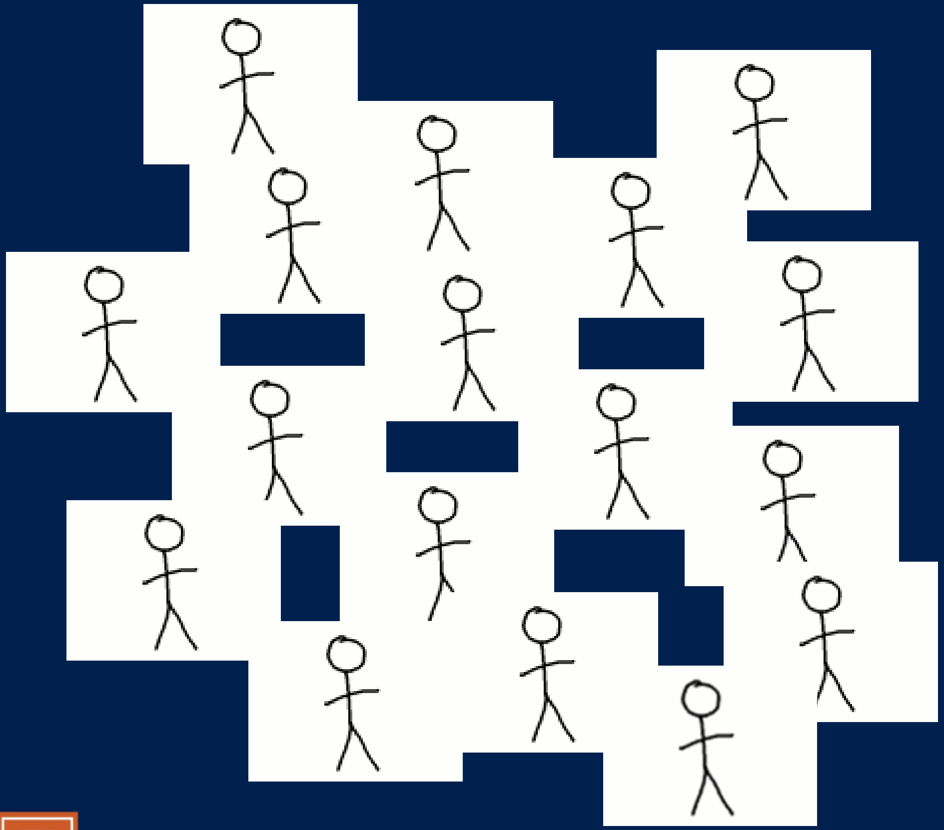
- Unit of observation
- Number of treated vs control
  - Can we afford to throw out controls?
- Unobservables
- Spillovers (contaminated controls)
- Different data sources (what happens when you observe treated obs at greater granularity than controls?)

example

- High-yielding seed variety introduced to a limited area



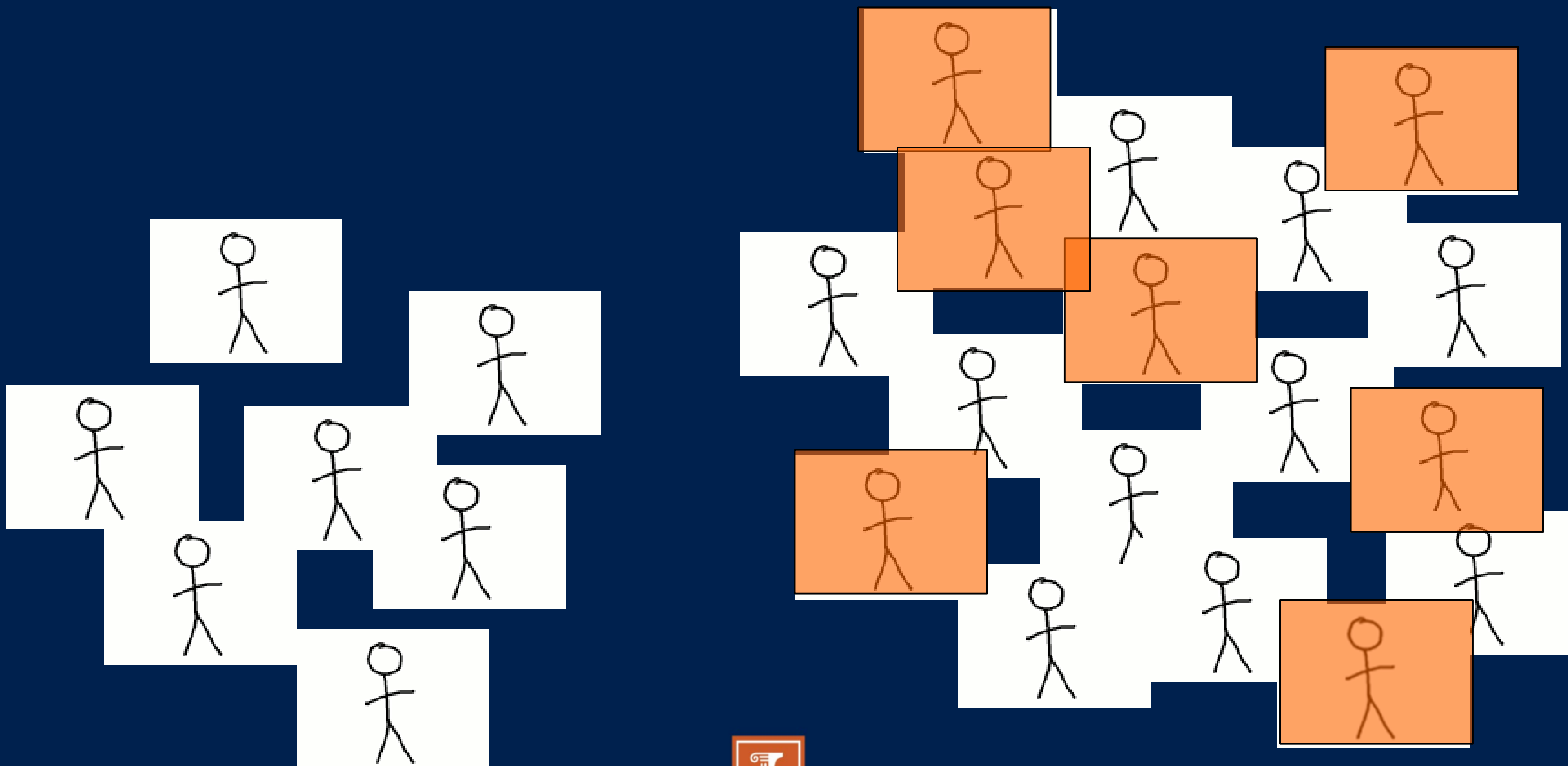
Treated Village



Potential Controls

example

- 1 to 1 matching

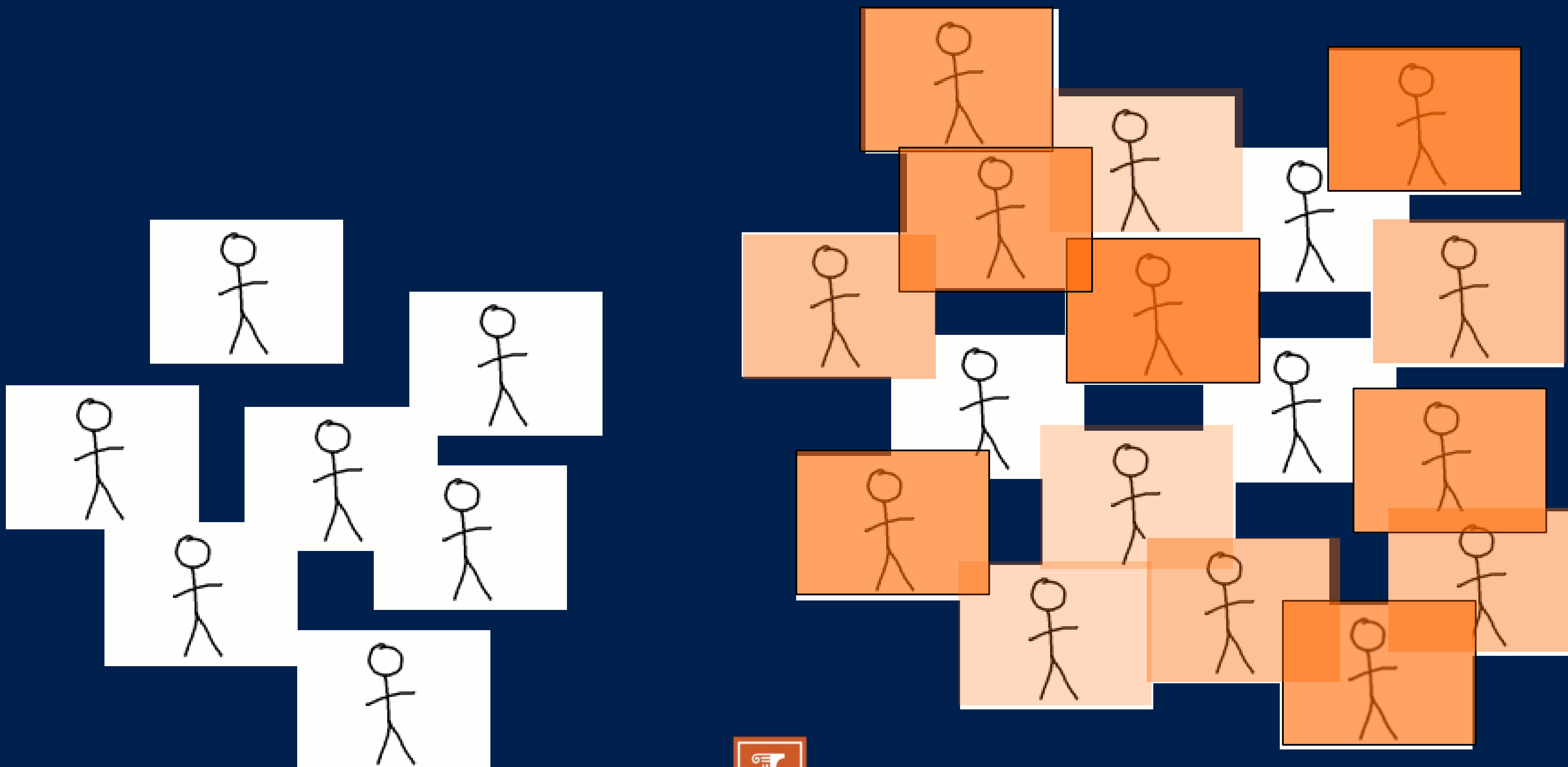


Treated Village

Matched Controls

example

- Kernel matching

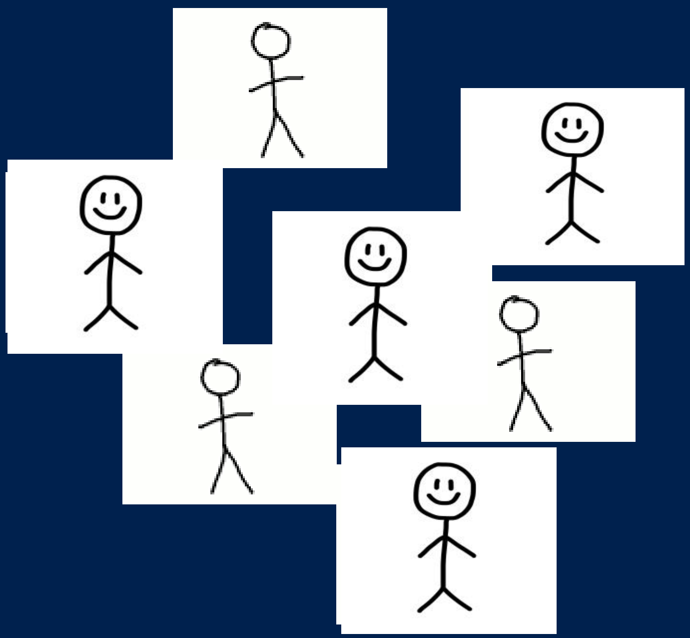


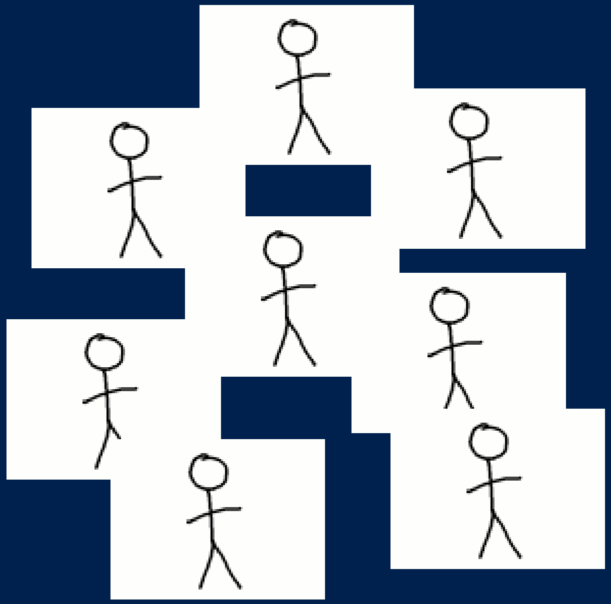
Treated Village

Weighted Matched Controls

example

- Now assume only some people adopt





example

- PSM