Matching Estimators
EC455 - Quantitative Approaches and Policy Analysis

Greg Fischer

Lent Term 2014
Outline

1. Introduction: the intuition
2. Matching in practice
3. Matching vs. Regression
4. Application: Piped water and diarrhea in India
5. Summary
Outline

1 Introduction
2 Matching in practice
3 Matching vs. Regression
4 Application: Piped water and diarrhea in India
5 Summary
Suppose you hear about an employability program designed to get people back to work

- People who sign up for the program get five weeks of job skills training and job search support
- You want to evaluate whether the program actually increased earnings
- You have a single cross section of administrative for a large number of individuals
- Some signed up for the program; some did not
- There are no arbitrary assignment rules to exploit (true randomization or otherwise)
The problem of simply comparing treated and untreated means is well known

\[
E \left[ Y_i^T \mid T \right] - E \left[ Y_i^C \mid C \right] = E \left[ Y_i^T - Y_i^C \mid T \right] + E \left[ Y_i^C \mid T \right] - E \left[ Y_i^C \mid C \right]
\]

- Observed Earnings Diff.
- Avg. Treatment on Treated
- Selection Bias

- Those who sign up for program would have different earnings than those who did not even in the absence of the program
- In other words, you are concerned that the treated individuals are inherently different from the untreated
- They were selected into the program
The intuition behind matching estimators is simple

- For each person who is enrolled in the program, match them with someone who is as similar as possible and not enrolled
- Compute the difference in outcomes for each match
- The treatment effect is the weighted average of these differences
Matching estimators are widely used in program evaluation

- Matching estimators appeared in the 1990s in the literature on evaluating labor market programs.
- In contrast to the techniques we have covered this term, it does not rely on natural experiments.
- The basic idea originated and remains heavily used in medical studies.
- It is sometimes viewed as a “magic bullet” to establish causality in the absence of experimental data.
- It is not
- We will discuss its strengths and limitations.
Stock and Watson do not cover matching estimators

A fairly non-technical summary of the matching approach is in:


As an application of this approach we are going to look at:

Outline

1. Introduction
2. Matching in practice
3. Matching vs. Regression
4. Application: Piped water and diarrhea in India
5. Summary
Matching relies on a key identifying assumption

- Conditional on observable variables, treatment is as good as random
- More formally, recall the problem that got us here:

\[ E \left[ Y_i^C \mid T \right] \neq E \left[ Y_i^C \mid C \right] \]

- Matching relies on the assumption that we can find a set of variables \( X \) such that

\[ E \left[ Y_i^C \mid X, T \right] = E \left[ Y_i^C \mid X, C \right] \]
A good match involves finding an untreated subject as similar as possible in determinants of potential outcomes.

<table>
<thead>
<tr>
<th>Treated Subjects</th>
<th>Untreated Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Gender</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
</tr>
</tbody>
</table>
Exact matching finds someone identical in relevant observables

<table>
<thead>
<tr>
<th>Treated Subjects</th>
<th>Untreated Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td><strong>Gender</strong></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td><strong>Gender</strong></td>
</tr>
<tr>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>38</td>
<td>0</td>
</tr>
<tr>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
</tr>
</tbody>
</table>

This still leaves a lot of room for omitted variables
So we may want to expand our set of controls

<table>
<thead>
<tr>
<th>Treated Subjects</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Gender</td>
<td>Months Unemp.</td>
<td>Univ. Grad</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>55</td>
<td>0</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Untreated “Control” Subjects</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Gender</td>
<td>Months Unemp.</td>
<td>Univ. Grad</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>38</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>55</td>
<td>0</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
But expansion comes at a cost

<table>
<thead>
<tr>
<th>Treated Subjects</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Gender</td>
<td>Months Unemp.</td>
<td>Univ. Grad</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>55</td>
<td>0</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Untreated “Control” Subjects</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Gender</td>
<td>Months Unemp.</td>
<td>Univ. Grad</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>38</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>55</td>
<td>0</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
It may be hard to find an exact match

- When we attempt to match on many characteristics
- Or a characteristics that can take on many values (e.g., a continuous variable)
- We are unlikely to find matches for all treated individuals in our pool of the untreated

This is known as "The Curse of Dimensionality"
Fortunately, propensity score matching solves "The Curse of Dimensionality"

- We use a technique called propensity score matching developed by Rosenbaum and Rubin (1983)
- Forget about finding an exact match on each relevant characteristic
- Use observable characteristics to compute probability that individual will enroll in treatment
- This value, \( p(X) \in [0, 1] \), is called the **propensity score**
- The propensity score summarizes all the observed characteristics that influence likelihood of being treated
Now we match on the propensity score rather than individual characteristics

- We match treated individuals with untreated individuals that have the closest propensity score
- These “closest” individuals generate our counterfactual:
  \[ E \left[ Y_i^C \middle| p(X), T \right] \]
- We are trying to mimic randomized assignment by choosing for the comparison group individuals who were just as likely to be treated
Rosenbaum and Rubin (1983) show that if:

- $0 < \text{Prob}(T = 1 | X) < 1$, i.e., there are individuals with the set of characteristics $x = X$ in both the treatment and control groups, and
- $E \left[ Y_i^C | X, T \right] = E \left[ Y_i^C | X, C \right]$, the treatment assignment is as good as random conditional on $X$

Then

- $E \left[ Y_i^C | p(X), T \right] = E \left[ Y_i^C | p(X), C \right]$
- Matching on the propensity score, $p(X)$, is as good as matching on each individual $X$
The propensity score works as well as X itself, but how good is that?

- Matching is not a randomized experiment!
- A randomized experiment ensures that the treatment is uncorrelated with both observable and unobservable determinants of the outcome.
- Just like simple OLS, matching controls for observables.
- But it cannot control for unobservable determinants of the outcome.
- How important such unobservables are will vary from application to application.
- Diagnosing this importance is an active area of research.
Implementation starts with matchmaking and matchmaking starts with data

- You want to match on characteristics that have not been affected by the treatment
  - A baseline is helpful for this
- Often studies will have different data for treated and untreated individuals
- The variables used to match should be as comparable as possible

This highlights importance of planning evaluations before implementation wherever possible
Next you generate your propensity scores

- **Step 1:** Regress the treatment dummy, $T$, on the set of available controls $X$
- **Step 2:** For each observation, record the predicted probability of treatment, $\hat{T}$, this is the propensity score
- **Step 3:** Restrict the sample to observations for which there is common support in the propensity score distribution
A hypothetical distribution of propensity scores
In the middle of the distribution, it is easy to find an untreated individual with a similar propensity score.

But we may not see untreated individuals with high propensity scores.

Intuitively, individuals who were very likely to enroll in the program are so dissimilar to untreated individuals that it is hard to find a valid counterfactual for them.

A lack of common support tends to appear in the tails of the distribution.
Step 4: For each treated individual, locate an untreated individuals or subgroup of untreated individuals with similar propensity scores.

There are many ways to generate these subgroups, e.g.,

- Nearest-neighbor picks the control observation with the closest propensity score
- Radius matching picks all control observations within a certain radius
- You might also hear terms like caliper matching, interval matching and kernel matching

Despite fancy names, they are all just different ways to select comparison group.

Step 4': Check to see if Xs are balanced after matching.
Estimate your treatment effects

- **Step 5:** Compare outcomes for each treatment observation and its matched subgroup
  - Under the identifying assumptions, this is the treatment effect for that individual
- **Step 6:** Average these measured effects across all treated observations
  - The result is our estimate of the treatment effect on the treated
How do you actually implement the analysis

- Stata has a convenient add-in for implementing propensity score matching
  - Usually Google is your best path to finding such tools
  - In Stata, type “search pscore” and install the latest version of the pscore suite
- Warning: calculating standard errors can be a little complicated
  - Intuitively: you have to account for the randomness involved in selecting your match as well as that from estimating the treatment effect
Outline

1. Introduction
2. Matching in practice
3. Matching vs. Regression
4. Application: Piped water and diarrhea in India
5. Summary
Let’s revisit the program evaluation with which we started

What is the difference between matching on gender and graduation status and running the regression

\[ Income_i = \alpha + \beta_1 Gender_i + \beta_2 Grad_i + \beta_3 (Gender_i \times Grad_i) + \gamma T_i + \varepsilon_i? \]

They both rely on unconfoundedness conditional on the controls

\[ E \left[ Y_i^C \mid T = 1, Gender_i, Grad_i \right] = E \left[ Y_i^C \mid T = 0, Gender_i, Grad_i \right] \]
The matching estimator and OLS estimator weight the observations differently.

- The matching estimator puts more weight on observations with values of $X$ for which $p(X)$ is large, conditional on having common support.
  - This comes immediately from taking the average of estimated treatment effects for each treated observation.
- OLS puts the most weight on observations with values of $X$ for which there is an equal number of treated and control individuals.
  - OLS minimizes the sum of squared errors.
There are other subtle differences between matching and regression

- The matching estimator does not impose a specific functional form on the relationship between $X$ and outcomes.
- Matching forces the researcher to clarify if “comparable” control observations are available.
  - You can and should do this with regression, but it is also easy to cheat.
- Matching makes it harder to extrapolate, perhaps erroneously, into regions outside the data range.
- **Note that careful use of regression techniques can capture many of the benefits of matching.**
- But calculating standard errors for matching estimators can be tricky.
Outline

1. Introduction
2. Matching in practice
3. Matching vs. Regression
4. Application: Piped water and diarrhea in India
5. Summary
Jalan and Ravallion (2003) want to evaluate the impact of having piped water on diarrhea for children in rural India.

- Would it be possible to evaluate this question with a randomized experiment?
- Jalan and Ravallion (2003) use cross-section data from a household survey to answer this question.
- They recognize that houses with and without piped water may differ and estimate a propensity score matching model to try to account for this.
- You will look at their findings and potential weaknesses of their approach in the class exercise.
Outline

1. Introduction
2. Matching in practice
3. Matching vs. Regression
4. Application: Piped water and diarrhea in India
5. Summary
Matching rests on the same core assumption as OLS: controlling for observables, the expected potential outcomes of treated and untreated observations are the same.

Fancy terminology and arcane techniques do not get around this fact.

Matching does have some advantages, and you should know how to interpret results from it.

But the importance of “unobservables” is hard to gauge and will vary from application to application.

As always: interpret results critically.