Matching: An Overview

Charlie Gibbons
Political Science 236

October 11, 2008

Outline

Overview

General procedure

Match each observation in treatment to an observation in control as closely as possible on a set of covariates.

Overview

General procedure

Match each observation in treatment to an observation in control as closely as possible on a set of covariates.

General reasoning

If the two observations are alike in each of the covariates, then treatment and control groups will be alike in the distribution of those covariates. The selection on observables assumption implies that they will be alike in the distribution of outcomes as well. Then we are able to identify treatment effects.

Discarding data

Notice what this implies: Suppose that there is no observation in treatment like one of the controls; then that control is ignored.

Discarding data

Notice what this implies: Suppose that there is no observation in treatment like one of the controls; then that control is ignored.

Implication: Matching throws away data.

Discarding data

Notice what this implies: Suppose that there is no observation in treatment like one of the controls; then that control is ignored.

Implication: Matching throws away data.

Why is this bad?

Discarding data

Notice what this implies: Suppose that there is no observation in treatment like one of the controls; then that control is ignored.

Implication: Matching throws away data.

Why is this bad?

- If there is a heterogeneous treatment effect, then we limit the population for which we can make an inference.
- If there is a homogeneous treatment effect, then we lose precision (*i.e.*, our standard errors could be smaller).

Discarding data

Notice what this implies: Suppose that there is no observation in treatment like one of the controls; then that control is ignored.

Implication: Matching throws away data.

Why is this bad?

- If there is a heterogeneous treatment effect, then we limit the population for which we can make an inference.
- If there is a homogeneous treatment effect, then we lose precision (*i.e.*, our standard errors could be smaller).

Discarding data

Notice what this implies: Suppose that there is no observation in treatment like one of the controls; then that control is ignored.

Implication: Matching throws away data.

Why is this bad?

- If there is a heterogeneous treatment effect, then we limit the population for which we can make an inference.
- If there is a homogeneous treatment effect, then we lose precision (*i.e.*, our standard errors could be smaller).

Discarding data

Responses:

- If there is a heterogeneous treatment effect, then including the discarded data will produce a biased estimate.
- If there is a homogeneous treatment effect, then we wouldn't know that it is in fact homogeneous.

Discarding data

Responses:

- If there is a heterogeneous treatment effect, then including the discarded data will produce a biased estimate.
- If there is a homogeneous treatment effect, then we wouldn't know that it is in fact homogeneous.

Discarding data

Responses:

- If there is a heterogeneous treatment effect, then including the discarded data will produce a biased estimate.
- If there is a homogeneous treatment effect, then we wouldn't know that it is in fact homogeneous.

Discarding data

Responses:

- If there is a heterogeneous treatment effect, then including the discarded data will produce a biased estimate.
- If there is a homogeneous treatment effect, then we wouldn't know that it is in fact homogeneous.

How can we keep these data in OLS?

Discarding data

Responses:

- If there is a heterogeneous treatment effect, then including the discarded data will produce a biased estimate.
- If there is a homogeneous treatment effect, then we wouldn't know that it is in fact homogeneous.

How can we keep these data in OLS?

We *assume a model of how dissimilar observations relate.*

Discarding data

Responses:

- If there is a heterogeneous treatment effect, then including the discarded data will produce a biased estimate.
- If there is a homogeneous treatment effect, then we wouldn't know that it is in fact homogeneous.

How can we keep these data in OLS?

We *assume a model of how dissimilar observations relate.*

Goal of matching: Eliminate bias arising from having qualitatively different treatment and control groups.

Framework

Suppose that there is a treatment sample S_1 of size N and a control sample S_0 of size rN, $r \geq 1$.

Framework

Suppose that there is a treatment sample S_1 of size N and a control sample S_0 of size rN, $r \geq 1$.

If $r = 1$, then the observations will simply be shuffled by matching and no bias can be removed.

Framework

Suppose that there is a treatment sample S_1 of size N and a control sample S_0 of size rN, $r \geq 1$.

If $r = 1$, then the observations will simply be shuffled by matching and no bias can be removed.

A subset of N observations in S_0 will be matched to the treatment observations; the resulting subsample is S_0^*. Let $j = 1, \ldots, N$ index the pairs.

Framework

Suppose that there is a treatment sample S_1 of size N and a control sample S_0 of size rN, $r \geq 1$.

If $r = 1$, then the observations will simply be shuffled by matching and no bias can be removed.

A subset of N observations in S_0 will be matched to the treatment observations; the resulting subsample is S_0^*. Let $j = 1, \ldots, N$ index the pairs.

The conditional expectation of Y given X is the *response surface*. Treatment and control groups each have their own response surfaces; $R_1(x)$ and $R_0(x)$ respectively.

Framework

Hence our model becomes:

$$\begin{aligned} y_{1j} &= R_1(x) + \epsilon_{1j} \\ y_{0j} &= R_0(x) + \epsilon_{0j} \\ \mathbb{E}\left[\epsilon_{ij} | x_{ij}\right] &= 0 \quad \forall\, i = 1, 2; \;\; j = 1, \ldots, N \end{aligned}$$

Assumptions

Note that, unlike OLS, we are *not assuming anything about the relationship between* Y *and* X; specifically, we are not making a linearity assumption; $R_i(x)$ can take any form.

Assumptions

Note that, unlike OLS, we are *not assuming anything about the relationship between Y and X*; specifically, we are not making a linearity assumption; $R_i(x)$ can take any form.

What assumptions are we making?

- Selection on observables: $(Y_i(0), Y_i(1)) \perp T_i | X_i$
- Common support on covariates: $0 < \Pr(T = 1|X = x) < 1 \; \forall \, x \in X$
- Stable Unit Treatment Value Assumption (SUTVA): $(Y_i(0), Y_i(1)) \perp T_j \; \forall \, j \neq i$.

Assumptions

Note that, unlike OLS, we are *not assuming anything about the relationship between Y and X*; specifically, we are not making a linearity assumption; $R_i(x)$ can take any form.

What assumptions are we making?

- Selection on observables:
  $(Y_i(0), Y_i(1)) \perp T_i | X_i$
- Common support on covariates:
  $0 < \Pr(T = 1 | X = x) < 1 \; \forall \; x \in X$
- Stable Unit Treatment Value Assumption (SUTVA):
  $(Y_i(0), Y_i(1)) \perp T_j \; \forall \; j \neq i$.

Assumptions

Note that, unlike OLS, we are *not assuming anything about the relationship between Y and X*; specifically, we are not making a linearity assumption; $R_i(x)$ can take any form.

What assumptions are we making?

- Selection on observables:
  $(Y_i(0), Y_i(1)) \perp T_i | X_i$
- Common support on covariates:
  $0 < \Pr(T = 1 | X = x) < 1 \; \forall \; x \in X$
- Stable Unit Treatment Value Assumption (SUTVA):
  $(Y_i(0), Y_i(1)) \perp T_j \; \forall \; j \neq i$.

Assumptions

Note that, unlike OLS, we are *not assuming anything about the relationship between Y and X*; specifically, we are not making a linearity assumption; $R_i(x)$ can take any form.

What assumptions are we making?

- Selection on observables:
  $(Y_i(0), Y_i(1)) \perp T_i | X_i$
- Common support on covariates:
  $0 < \Pr(T = 1 | X = x) < 1 \; \forall \; x \in X$
- Stable Unit Treatment Value Assumption (SUTVA):
  $(Y_i(0), Y_i(1)) \perp T_j \; \forall \; j \neq i$.

Estimator

The average treatment effect for the treated τ is defined by

$$\tau \equiv \mathbb{E}_1[R_1(x) - R_0(x)]$$

Estimator

The average treatment effect for the treated τ is defined by

$$\tau \equiv \mathbb{E}_1[R_1(x) - R_0(x)]$$

We estimate τ by finding its sample analogue:

$$\hat{\tau} = \frac{1}{N}\sum_{j=1}^{N} y_{1j} - \frac{1}{N}\sum_{j=1}^{N} y_{0j}^{*} = \bar{y}_1 - \bar{y}_0^{*}$$

Unbiasedness: Proof

Is the estimator unbiased? By the law of iterated expectations:

$$\begin{aligned}
\mathbb{E}\left[\hat{\tau}-\tau\right] &= \mathbb{E}_X\left[\mathbb{E}[\hat{\tau}-\tau|X=x]\right] \\
&= \mathbb{E}_X\left[\mathbb{E}[\bar{y}_1-\bar{y}_0^*-\mathbb{E}_1[R_1(x)-R_0(x)]|X=x]\right] \\
&= \mathbb{E}_X\left[\mathbb{E}[\bar{y}_1|X=x]\right]-\mathbb{E}_X\left[\mathbb{E}[\bar{y}_0^*|X=x]\right] \\
&\quad -\mathbb{E}_X\left[\mathbb{E}\left[\mathbb{E}_1[R_1(x)]|X=x\right]\right]+\mathbb{E}_X\left[\mathbb{E}\left[\mathbb{E}_1[R_0(x)]|X=x\right]\right]
\end{aligned}$$

Unbiasedness: Proof

$$\begin{aligned}
\mathbb{E}_X\left[\mathbb{E}[\bar{y}_1|X=x]\right] &= \mathbb{E}_X\left[\mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N} y_{1j}\middle| X=x\right]\right] \\
&= \mathbb{E}_X\left[\mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N} R_1\left(x_{1j}\right)+\epsilon_{1j}\middle| X=x\right]\right] \\
&= \mathbb{E}_X\left[\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}[R_1\left(x_{1j}\right)|X=x]\right] \\
&\quad +\mathbb{E}_X\left[\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}[\epsilon_{1j}|X=x]\right]
\end{aligned}$$

Unbiasedness: Proof

Continuing,

$$
\begin{aligned}
&= \frac{1}{N}\sum_{j=1}^{N} \mathbb{E}_X[R_1(x)] \\
&= \mathbb{E}_X[R_1(x)] \\
&= \mathbb{E}_1[R_1(x)]
\end{aligned}
$$

Unbiasedness: Summary

Thus we have

$$\mathbb{E}_X\left[\mathbb{E}[\bar{y}_1|X=x]\right] = \mathbb{E}_1[R_1(x)].$$

By similar logic, we have

$$\mathbb{E}_X\left[\mathbb{E}[\bar{y}_0^*|X=x]\right] = \mathbb{E}_{0*}[R_0(x)].$$

Also, notice that

$$\begin{aligned}
\mathbb{E}_X\left[\mathbb{E}\left[\mathbb{E}_1[R_1(x)]|X=x\right]\right] &= \mathbb{E}_1[R_1(x)] \\
\mathbb{E}_X\left[\mathbb{E}\left[\mathbb{E}_1[R_0(x)]|X=x\right]\right] &= \mathbb{E}_1[R_0(x)]
\end{aligned}$$

Unbiasedness: Summary

Thus we have:

$$
\begin{aligned}
\mathbb{E}\left[\hat{\tau} - \tau\right] &= \mathbb{E}_1[R_1(x)] - \mathbb{E}_{0*}[R_0(x)] - \mathbb{E}_1[R_1(x)] + \mathbb{E}_1[R_0(x)] \\
&= \mathbb{E}_1[R_0(x)] - \mathbb{E}_{0*}[R_0(x)]
\end{aligned}
$$

which is 0 ($i.e.$, $\hat{\tau}$ is unbiased) when

$$\mathbb{E}_1[R_0(x)] = \mathbb{E}_{0*}[R_0(x)]$$

Unbiasedness: Summary

Thus we have:

$$\begin{aligned}\mathbb{E}\left[\hat{\tau} - \tau\right] &= \mathbb{E}_1[R_1(x)] - \mathbb{E}_{0*}[R_0(x)] - \mathbb{E}_1[R_1(x)] + \mathbb{E}_1[R_0(x)] \\ &= \mathbb{E}_1[R_0(x)] - \mathbb{E}_{0*}[R_0(x)]\end{aligned}$$

which is 0 (*i.e.*, $\hat{\tau}$ is unbiased) when

$$\mathbb{E}_1[R_0(x)] = \mathbb{E}_{0*}[R_0(x)]$$

Matching attempts to equate the distribution of X in the matched subsample S_0^* and in the treated sample S_1. Selection on observables implies that the distribution of potential outcomes in the two samples would be equal. Then, our estimator $\hat{\tau}$ would be unbiased.