Covariate Adjustment

Ec142 Applied Econometrics

January 23rd, 2018

Bryan S. Graham

University of California - Berkeley

Potential Outcomes Notation

Let $W \in \{0,1\}$ be a binary treatment or policy.

Let $Y(1)$ and $Y(0)$ denote a unit's *potential* outcome under active treatment ($W = 1$) and control ($W = 0$) respectively.

The *observed* outcome, Y, equals

$$\boxed{Y = (1 - W)\,Y(0) + WY(1)}$$

For each unit we observe either $Y(1)$ or $Y(0)$, but never both.

Causal Estimands

For a given unit, the difference $Y(1) - Y(0)$ is the *causal effect* of treatment.

Since only $Y(1)$ *or* $Y(0)$ is observed, it is impossible to learn this effect.

We either observe our earnings at age 25 conditional on completing college or not.

The hope, however, is that *averages* of unit-specific causal effects are identified (i.e., learnable).

<u>Average Treatment Effects</u>

1. The *Average Treatment Effect* (ATE) equals

$$\boxed{\beta^{\mathsf{ATE}} = \mathbb{E}\left[Y(1) - Y(0)\right]}$$

   Equals the expected benefit (in terms of the outcome Y) for a randomly sampled unit from the population of interest.

2. The *Average Treatment Effect on the Treated* (ATT) equals

$$\beta^{\mathsf{ATT}} = \mathbb{E}\left[Y(1) - Y(0) \middle| W = 1\right]$$

   The average benefit among those units actually treated.

Random Assignment

If units are *randomly assigned* to either the active treatment ($W = 1$) or control ($W = 0$) then

$$(Y(0), Y(1)) \perp W$$

The distribution of potential outcomes is the same across the treated and controls so that

$$\begin{aligned} b &= \mathbb{E}[Y|W=1] - \mathbb{E}[Y|W=0] \\ &= \mathbb{E}[Y(1)|W=1] - \mathbb{E}[Y(0)|W=0] \\ &= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \\ &= \beta^{\mathrm{ATE}} \end{aligned}$$

A simple treatment-control comparison identifies the ATE (and ATT).

Selection Bias

If units are not randomly assigned to treatment or control, then

$$\begin{aligned} b &= \mathbb{E}[Y|W=1] - \mathbb{E}[Y|W=0] \\ &= \mathbb{E}[Y(1)|W=1] - \mathbb{E}[Y(0)|W=0] \\ &\neq \beta^{\text{ATE}} \end{aligned}$$

The distribution of $Y(0)$ among those who, for example, actually complete college may differ from that across the entire population.

$$b = \underbrace{\mathbb{E}[Y(1) - Y(0)|W=1]}_{\text{ATT}} + \underbrace{\mathbb{E}[Y(0)|W=1] - \mathbb{E}[Y(0)|W=0]}_{\text{Selection Bias}}$$

Selection on Observables

A.1 Exogeneity:

$$(Y(0), Y(1)) \perp W | X = x, \ x \in \mathbb{X}$$

For every subpopulation defined in terms of X, treatment varies independently of the potential outcomes within that subpopulation.

Under A.1 X contains all covariates which predict *both* treatment assignment and the potential outcomes.

Overlap

Let

$$e(x) = \Pr(W = 1 | X = x)$$

denote the *propensity score*.

$e(x)$ *may* vary with x so that different subpopulations more systematically take up the active or control treatment...

...but we require that for all subpopulations a positive fraction of units are treated and a positive fraction are not treated.

Overlap (continued)

A.2 Overlap:

$$0 < \kappa \leq e\left(x\right) \leq 1 - \kappa < 1, \; x \in \mathbb{X}$$

A.2 ensures we can make treatment vs. control comparisons in *all* subpopulations defined in terms of $X = x$

Identification of the ATE

Let $\beta(x) = \mathbb{E}\left[\left. Y(1) - Y(0) \right| X = x\right]$ be the *conditional average treatment effect* (CATE).

The CATE equals the average causal effect of treatment in a specific subpopulation.

Observe that

$$\begin{aligned} b(x) &= \mathbb{E}\left[\left. Y \right| W = 1, X = x\right] - \mathbb{E}\left[\left. Y \right| W = 0, X = x\right] \\ &= \mathbb{E}\left[\left. Y(1) \right| W = 1, X = x\right] - \mathbb{E}\left[\left. Y(0) \right| W = 0, X = x\right] \\ [\mathbf{A.1}] &= \mathbb{E}\left[\left. Y(1) \right| X = x\right] - \mathbb{E}\left[\left. Y(0) \right| X = x\right] \\ &= \beta(x) \end{aligned}$$

Identification of the ATE (continued)

Conditioning on covariates allows us to make “apples-to-apples” comparisons.

Selection on observables (A.1) implies that, within subpopulations with the same covariate value, treatment and control units are comparable (i.e., have the same distribution of potential outcomes).

Identification of the ATE (continued)

Under overlap (A.2) we can average over the marginal distribution of covariates to recover the ATE.

$$\begin{aligned} b &= \mathbb{E}_X\left[\mathbb{E}\left[Y \mid W=1, X\right] - \mathbb{E}\left[Y \mid W=0, X\right]\right] \\ &= \mathbb{E}_X\left[\mathbb{E}\left[Y(1) \mid X\right] - \mathbb{E}\left[Y(0) \mid X\right]\right] \\ [\mathbf{A.2}] &= \mathbb{E}_X\left[\beta(X)\right] \\ &= \beta^{\text{ATE}} \end{aligned}$$

We find the CATE for every subgroup defined in terms of $X = x$.

Under A.2 all such CATEs are identified.

We recover the ATE by averaging these CATEs.

We will introduce several approaches to operationalizing this idea with real data later.

Optimizing Agents

In many social science settings agents choose, or exert partial control over, their treatment status.

Agents may also know (or partially know) their potential outcomes and hence the benefits of treatment.

Purposeful treatment selection by agents can render our selection on observables assumption (A.1) implausible.

Note: Our overlap assumption is testable and we will discuss methods of testing it later.

Optimizing Agents (continued)

Let $Y(w) = g(w, \epsilon)$ for $w \in \{0, 1\}$. We can think of $g(w, \epsilon)$ as a *production function* with *input* w and *productivity* ϵ.

A firm does not know their own productivity ϵ, but observes the productivity *signal* X.

A firm also knows the cost, C, of using the input.

Maximizing expected profits yields

$$W = \arg\max_{w \in \{0,1\}} \mathbb{E}\left[g(w, \epsilon) - wC \middle| X, C \right]$$

Optimizing Agents (continued)

Maximizing expected profits yields

$$\begin{aligned} W &= \arg\max_{w \in \{0,1\}} \mathbb{E}\left[g(w,\epsilon) - wC \,|\, X, C \right] \\ &= \mathbf{1}\left(\mathbb{E}\left[g(1,\epsilon) - g(0,\epsilon) \,|\, X, C \right] \geq C \right) \end{aligned}$$

If $\epsilon \perp C | X = x,\ x \in \mathbb{X}$ then:

1. Cost heterogeneity across firms does not predict productivity, ϵ, conditional on the signal, X.
2. Firms with identical signals will choose different inputs (i.e., both $W = 0$ and $W = 1$) due to cost heterogeneity.

Optimizing Agents (continued)

We have, under these assumptions,

$$W = \mathbf{1}\left(\mathbb{E}\left[g\left(1,\epsilon\right) - g\left(0,\epsilon\right)\right| X\right] \geq C\right)$$

and hence that

$$\left(g\left(1,\epsilon\right), g\left(0,\epsilon\right)\right) \perp W \right| X = x,\ x \in \mathbb{X}.$$

This gives A.1 (Selection on Observables).

Observationally identical firms choose different treatments because they face different costs.

These costs are unrelated to unobserved productivity.

Since X coincides with the firm's signal, firms homogenous in X have identical expectations about their unobserved productivity.

Optimizing Agents (continued)

The propensity score equals

$$e\left(x\right) = F_{C|X}\left(\mathbb{E}\left[g\left(1,\epsilon\right) - g\left(0,\epsilon\right)\middle| x\right]\middle| X = x\right).$$

As long as costs vary sufficiently, the overlap (A.2) condition will also hold.

Optimizing Agents (continued)

Our two key covariate adjustment assumptions are consistent with agents optimally choosing their treatment...

...but we have made very specific assumptions regarding agent utilities, expectations and input costs.

These assumptions will be convincing in some settings and not in others.

It is always important to think through the economics of a problem.

Next Time

We will introduce the inverse probability weighting (IPW) estimate of the average treatment effect.