February 3rd Lecture Notes

Matching

Causal Inference with matching.
(For the moment assume we have matched one control to each treatment unit without replacement)

ATT:

$$\tau_t = \frac{1}{N_t}\sum_{i=1}^{N} W_i[Y_i(1) - Y_i(0)]$$

- Once you match the data, you can perform any analysis on it (matched data). Eg. regress Y on W, X's.
- If you have perfect matches then one can think of " matched obs data " as randomized experiment (unconfounded).

Paired Analysis

Notation:
$K_i^t = \textit{index of } i^{th} \textit{ treatment unit.}$
$K_i^c = \textit{index of matching control unit.}$
$= \arg\min_{\boldsymbol{j \in Ctrl}} \parallel X_{k_i}^t - X_j \parallel$

$$\hat{\boldsymbol{\tau}}_{\boldsymbol{i}} = Y_{k_i^t}^{obs} - Y_{k_i^c}^{obs} = Y_{k_i^t}(1) - Y_{k_i^c}(0)$$

$$\hat{\boldsymbol{\tau}}_{\boldsymbol{ATT}} = \frac{1}{N_T}\sum_{i=1}^{N} W_i \hat{\tau}_i$$

$$\hat{V}(\hat{\tau}_{ATT}) = \frac{1}{N_T}\frac{1}{N_T - 1}\sum_{i=1}^{N_T}(\hat{\tau}_i - \hat{\tau}_{ATT})^2$$

NOTE: If you don't match perfectly estimators are biased and variance formula is only approximate. However, in case of a perfect match variance formula is exact.

Comments

- If we allow replacement, then controls are used more than once.

- ATT estimate ($\hat{\tau}_{ATT}$) is still ok. i.e.

$$\hat{\tau}_{ATT} = \frac{1}{N_T} \sum_{i=1}^{N} W_i Y_i - (1 - W_i) K(i) Y_i$$

  where k(i) = number of times unit i is used as a control matched to a treatment unit.
- variance estimator in this case is not obvious.

- If we allow multiple matches i.e. m controls to each treatment then

$$k(i) = \frac{1}{m}(\# \text{ trt units for which is a match})$$

- $$\text{ATE} = \tau = \frac{1}{N} \sum_{i=1}^{N} (Y_i(1) - Y_i(0))$$
  - we need to match but also need to match controls.
  - like imputation strategy

$$\hat{Y}_i(0) \begin{cases} Y_i & \text{if } W_i = 0 \\ Y_{match} & \text{if } W_i = 1 \end{cases}$$

$$\hat{Y}_i(1) \begin{cases} Y_{match} & \text{if } W_i = 0 \\ Y_i & \text{if } W_i = 1 \end{cases}$$

  - $\hat{\tau}_{ATE}$ is similar to above sums while accounting for number of times each unit appears as a match (or a target)

Propensity Score (Ch12(Theory), 13(Estimation), 17(Analysis), Imbens (2004))

$$e(x) = \frac{1}{N_x} \sum_{i=1}^{N} Pr(W_i = 1 | Y(0), Y(1), X)$$

= empirical average probability of getting treated.

Under regular assignment mechanism = $\frac{1}{N_x} \sum_{i=1}^{N} q(x) = q(x)$ = "actual" probability of getting treatment under regular assignment.

We use e(x) to be the propensity score.

Theory

- e(x) is a balancing score.
  $W_i \perp X_i | e(x_i)$

- e(x) is a function of covariates such that W is independent of all elements of X.
- e(x) is lowest dimension balancing score.

Proof

$$P(W_i|X_i, e(x_i)) = P(W_i = 1|e(x_i))$$

Since $e(x_i)$ is a function of X_i

$$P(W_i|X_i, e(x_i)) = P(W_i = 1|X_i) = e(x_i)$$

From Law of Iterated Conditional Expectation,

$$\begin{aligned} P(W_i = 1|e(x_i) &= E[W_i|e(x_i)] = E[E(W_i|X_i, e(x_i))|e(x_i)] \\ &= E[P(W_i = 1|X_i, e(x_i))|e(x_i)] \\ &= E[e(x_i)|e(x_i)] = e(x_i) \end{aligned}$$

- If we have an uncounfounded design given X then it is also unconfounded given e(x).
  If $W_i \perp Y_i(0), Y_i(1)|X$ then $W_i \perp Y_i(0), Y_i(1)|e(x)$.

- *Superpopulation.* If e(x) is known then we can obtain causal inference via weighting. Notice that

$$\begin{aligned} E\left[\frac{Y_i^{obs} W_i}{e(x_i)}\right] &= E\left[E\left(\frac{Y_i^{obs} W_i}{e(x_i)}\middle| X_i\right)\right] \\ &= E\left[\frac{1}{e(x_i)} E(Y_i(1)|X_i)\ E(W_i|X_i)\right] \\ &= E[E(Y_i(1)|X_i)] = E[Y_i(1)] \end{aligned}$$

$$E\left[\frac{Y_i^{obs}(1 - W_i)}{1 - e(x_i)}\right] = E[Y_i(0)]$$

$$\begin{aligned} \Rightarrow \boldsymbol{\tau}(ATE) &= E[Y_i(1) - Y_i(0)] \\ &= E\left[\frac{Y_i^{obs} W_i}{e(x_i)}\right] - E\left[\frac{Y_i^{obs}(1 - W_i)}{1 - e(x_i)}\right] \end{aligned}$$

$$\Rightarrow \hat{\boldsymbol{\tau}} = \frac{1}{N}\sum_{i=1}^{N} \frac{W_i Y_i^{obs}}{e(x_i)} - \frac{1}{N}\sum_{i=1}^{N} \frac{(1 - W_i) Y_i^{obs}}{1 - e(x_i)}$$

Note: The weighting makes good sense ... if $e(x) = 0.9$, then we expect 9 times as many treated units as controls for covariates like these. Each treated unit receives weight $1/0.9 = 1.11$ and each control unit receives weight $1/0.1 = 10$.