

The Evaluation Problem

$$Y = \begin{cases} Y_0 & \text{if } D = 0 \\ Y_1 & \text{if } D = 1 \end{cases}$$

or

$$Y = (1 - D)Y_0 + DY_1$$

D : enrollment in a program.

Y_j : outcome.

Econometric Specification

Linear Model:

$$Y_0 = \mu_0(X) + U_0$$

$$Y_1 = \mu_1(X) + U_1$$

U_0, U_1 : error terms $E[U_0|X] = 0, E[U_1|X] = 0$

The mean treatment effect:

$$E[Y_1 - Y_0] = E_X[\mu_1(X) - \mu_0(X)]$$

The mean treatment of the treated effect:

$$E[Y_1 - Y_0|D = 1] = E_X[\mu_1(X) - \mu_0(X) + E[U_1 - U_0|X, D = 1]]$$

Program Evaluation Problem

- ▶ Suppose we want to evaluate the Mean Treatment of the Treated. $E[Y_0|D = 1]$ is not known. We do not know the outcome of participants if they did not participate and the outcome of nonparticipants if they participated.
- ▶ If the assignment to the program is random, and (if the random variable determining assignment does not directly affect the outcome), then the participation status should be independent to the outcome Y_0 . That is,

$$Y_0 \perp D$$

where \perp denotes independence. This also implies

$$E[Y_0|D = 0] = E[Y_0|D = 1]$$

Then, because of the random assignment, Mean Treatment of the Treated is

$$\begin{aligned} E[Y_1 - Y_0|D = 1] &= E[Y_1|D = 1] - E[Y_0|D = 1] \\ &= E[Y_1|D = 1] - E[Y_0|D = 0] \end{aligned}$$

both $E[Y_0|D = 0]$ and $E[Y_1|D = 1]$ can be estimated from the data as follows

$$\begin{aligned} \hat{E}[Y_0|D = 0] &= \frac{\sum_{it} (1 - D_{it}) Y_{(it)}}{\sum_i (1 - D_{it})} \\ \hat{E}[Y_1|D = 1] &= \frac{\sum_{it} D_{it} Y_{(it)}}{\sum_i D_{it}} \end{aligned}$$

In many actual policy cases, it is very rare that random assignment was implemented. Then, one can still think that the following assumption is roughly satisfied.

Assumptions for Matching:

Assumption A1

$$E[Y_0|X, D = 0] = E[Y_0|X, D = 1]$$

$$E[Y_1|X, D = 0] = E[Y_1|X, D = 1]$$

Assumption A2

$$0 < P(D = 1|X) < 1$$

such that

$$Supp(X|D = 0) = Supp(X|D = 1)$$

Then, Mean Treatment of the Treated conditional on X is

$$\begin{aligned} E[Y_1 - Y_0|X, D = 1] &= E[Y_1|X, D = 1] - E[Y_0|X, D = 1] \\ &= E[Y_1|X, D = 1] - E[Y_0|X, D = 0] \end{aligned}$$

both $E[Y_0|X, D = 0]$ and $E[Y_1|X, D = 1]$ can be estimated from the data as follows

$$\begin{aligned} \hat{E}[Y_0|X, D = 0] &= \frac{\sum_{it} (1 - D_{it}) D_{it}(X_{it} = X) Y_{(it)}}{\sum_i (1 - D_{it}) D_{it}(X_{it} = X)} \\ \hat{E}[Y_1|X, D = 1] &= \frac{\sum_{it} D_{it} D_{it}(X_{it} = X) Y_{(it)}}{\sum_{it} D_{it} D_{it}(X_{it} = X)} \end{aligned}$$

- ▶ But this is very demanding to evaluate because as you increase the number of variables in X , the dimensionality of X increases and it becomes more difficult to get enough sample size for each X bin.
- ▶ Similar problems arise even if we would evaluate the expectation conditional on X using kernels by matching the outcomes of treatment and control group individuals with X_{it} 's that are close to X .

Result by Rosenbaum and Rubin

Suppose the following Assumptions hold

Assumption A1

$$Y_0 \perp D|X$$

That is, Y_0 is conditionally independent to treatment D given X .

Assumption A2

$$0 < P(D = 1|X) < 1$$

Theorem

Then,

$$Y_0 \perp D|P(D = 1|X)$$

Notice that they are essentially the same as the Assumptions about for matching, except that we need conditional independence rather than the assumption for matching, which is about conditional mean being the same for $D = 0$ and $D = 1$.

Implications of the Result

- ▶ Eliminates the curse of dimensionality problem in matching estimation. For $j = 0, 1$

$$E[Y_j|X, D = 0] = E[Y_j|P(D = 1|X), D = 0]$$

$$E[Y_j|X, D = 1] = E[Y_j|P(D = 1|X), D = 1]$$

- ▶ Derivation of the matching estimator of the treatment effect becomes easier because we just need to derive

$$\widehat{E}[Y_1|P(D = 1|X), D = 1] - \widehat{E}[Y_0|P(D = 1|X), D = 0]$$

, which only involves matching of one dimensional $P(D = 1|X)$, instead of

$$\widehat{E}[Y_1|X, D = 1] - \widehat{E}[Y_0|X, D = 0]$$

Two Step Estimation

Step 1

Estimate the propensity score using probit or logit.

$$D = 1 (X\beta + \epsilon > 0)$$

or

$$D^* = X\beta + \epsilon$$

$$D = \begin{cases} 0 & \text{if } D^* \leq 0 \\ 1 & \text{if } D^* > 0 \end{cases}$$

If the error term is i.i.d. extreme value distributed, then the estimated propensity score is

$$\hat{P}(D = 1|X) = \frac{\exp(X\hat{\beta})}{1 + \exp(X\hat{\beta})}$$

Step 2

Derive the estimate of the Mean Treatment of the Treated conditional on $\hat{P}(D = 1|X)$

$$\begin{aligned} & \hat{E} \left[Y_1 - Y_0 | D = 1, \hat{P}(D = 1|X) \right] \\ = & \hat{E} \left[Y_1 | D = 1, \hat{P}(D = 1|X) \right] - \hat{E} \left[Y_0 | D = 0, \hat{P}(D = 1|X) \right] \end{aligned}$$

Why Does it Work?

- ▶ Suppose you have a sample where $P(D = 1|X) = 0.6$
- ▶ Then, given X , 60% are program participants and 40% are nonparticipants. This is true for any X with $P(D = 1|X) = 0.6$. That is, for those sample, participation is assigned evenly across all X .
- ▶ Then, the distribution of X for participants and the distribution of X for nonparticipants are the same.
- ▶ Under that environment, mean nonparticipation outcome is the same for participants and nonparticipants. This comes from the assumption that mean nonparticipation outcome is the same for participants and nonparticipants given X .