Ec533: Labour Economics for Research Students

Regression versus matching

Jörn-Steffen Pischke

LSE

October 29, 2009

Covariate matching

You want to estimate the effect of treatment on the treated

$$\delta_{TOT} \equiv E[\mathrm{Y}_{1i} - \mathrm{Y}_{0i} | \mathrm{D}_i = 1].$$

Using the law of iterated expectations

$$\delta_{TOT} = E\{E[\mathrm{Y}_{1i} | \mathrm{X}_i, \mathrm{D}_i = 1] - E[\mathrm{Y}_{0i} | \mathrm{X}_i, \mathrm{D}_i = 1] | \mathrm{D}_i = 1\}$$

Under conditional independence (CIA):

$$E[\mathrm{Y}_{0i} | \mathrm{X}_i, \mathrm{D}_i = 0] = E[\mathrm{Y}_{0i} | \mathrm{X}_i, \mathrm{D}_i = 1].$$

The matching estimand

Using this we get

$$\begin{aligned}\delta_{TOT} &= E\left\{E\left[\mathrm{Y}_{1i}|\mathrm{X}_i,\mathrm{D}_i=1\right]-E\left[\mathrm{Y}_{0i}|\mathrm{X}_i,\mathrm{D}_i=0\right]|\mathrm{D}_i=1\right\}\\ &= E\left\{E\left[y_i|\mathrm{X}_i,\mathrm{D}_i=1\right]-E\left[y_i|\mathrm{X}_i,\mathrm{D}_i=0\right]|\mathrm{D}_i=1\right\}\\ &= E[\delta_X|\mathrm{D}_i=1],\end{aligned}$$

where

$$\delta_X \equiv E\left[y_i|\mathrm{X}_i,\mathrm{D}_i=1\right]-E\left[y_i|\mathrm{X}_i,\mathrm{D}_i=0\right]$$

is an X-specific difference in means at covariate value X_i.
With discrete X_i, the matching estimand is

$$\delta_M = \sum_x \delta_x P(\mathrm{X}_i = x|\mathrm{D}_i=1),$$

where $P(\mathrm{X}_i = x|\mathrm{D}_i=1)$ is the probability mass function for X_i given $\mathrm{D}_i=1$.

Rewriting the matching estimand

Using Bayes Rule

$$P(\mathrm{X}_i = x | \mathrm{D}_i = 1) = \frac{P(\mathrm{D}_i = 1 | \mathrm{X}_i = x) \cdot P(\mathrm{X}_i = x)}{P(\mathrm{D}_i = 1)}.$$

the matching estimand can be written as

$$\begin{aligned} \delta_M &= \sum_x \delta_x P(\mathrm{X}_i = x | \mathrm{D}_i = 1) \\ &= \frac{\sum_x \delta_x P(\mathrm{D}_i = 1 | \mathrm{X}_i = x) P(\mathrm{X}_i = x)}{\sum_x P(\mathrm{D}_i = 1 | \mathrm{X}_i = x) P(\mathrm{X}_i = x)} \end{aligned}$$

Regression

Suppose we run instead the regression

$$y_i = \sum_x d_{ix}\beta_x + \delta_R \mathrm{D}_i + \varepsilon_i,$$

where d_{ix} is a dummy that indicates $\mathrm{X}_i = x$, β_x is a regression-effect for $\mathrm{X}_i = x$, and δ_R is the regression estimand. Note that this regression model allows a separate parameter for every value taken on by the covariates. It can be shown that δ_R can be written as

$$\delta_R = \frac{\sum_x \delta_x \left[P(\mathrm{D}_i = 1|\mathrm{X}_i = x)(1 - P(\mathrm{D}_i = 1|\mathrm{X}_i = x))\right] P\left(\mathrm{X}_i = x\right)}{\sum_x \left[P(\mathrm{D}_i = 1|\mathrm{X}_i = x)(1 - P(\mathrm{D}_i = 1|\mathrm{X}_i = x))\right] P\left(\mathrm{X}_i = x\right)}$$

Regression vs. matching

So the matching estimand

$$\delta_M = \sum_x \delta_x \left[\frac{P(\mathrm{D}_i = 1|\mathrm{X}_i = x)P(\mathrm{X}_i = x)}{\sum_x P(\mathrm{D}_i = 1|\mathrm{X}_i = x)P(\mathrm{X}_i = x)} \right]$$

and the regression estimand

$$\delta_R = \sum_x \delta_x \left[\frac{[P(\mathrm{D}_i = 1|\mathrm{X}_i = x)(1 - P(\mathrm{D}_i = 1|\mathrm{X}_i = x))]\, P\left(\mathrm{X}_i = x\right)}{\sum_x [P(\mathrm{D}_i = 1|\mathrm{X}_i = x)(1 - P(\mathrm{D}_i = 1|\mathrm{X}_i = x))]\, P\left(\mathrm{X}_i = x\right)} \right]$$

differ by the weights they use to combine the covariate specific treatment effects δ_x.

Regression vs. matching in words

- Matching uses weights which depend on

$$P(\mathrm{D}_i = 1|\mathrm{X}_i = x)$$

i.e. the fraction of treated observations in a covariate cell (or the mean of D_i). This is larger in cells where there are many treated observations (makes sense as we want the effect of treatment on the treated).

- Regression uses weights which depend on

$$P(\mathrm{D}_i = 1|\mathrm{X}_i = x)(1 - P(\mathrm{D}_i = 1|\mathrm{X}_i = x))$$

i.e. the variance of D_i in the covariate cell. This weight is largest in cells where there are half treated and half untreated observations. This also makes sense because these cells will produce the lowest variance estimates of δ_x. If all the δ_x are the same, the most efficient estimand uses the lowest variance cells most heavily. This is what regression does.