Some regression results

Regression as a matching estimator

Suppose that conditional independence holds given a set of discrete covariates x:

$$(y_1, y_0) \perp d|x$$

As we know, in this case the x-specific ATT $E(y_1 - y_0|x)$ can be identified using

$$\delta_x = E(y|x, d = 1) - E(y|x, d = 0)$$

Then the ATT can be identified as

$$\delta_{ATT} = E(y_1 - y_0|d = 1) = \sum_x \delta_x P(x|d = 1),$$

where the sum runs over all combinations of values of the discrete covariates

Similarly, the ATT can be identified as

$$\delta_{ATE} = E(y_1 - y_0) = \sum_x \delta_x P(x)$$

A simple matching estimator would replace δ_x, $P(x|d=1)$ and $P(x)$ with their sample analogs to estimate δ_{ATT} and δ_{ATE}

It turns out that we can also interpret a fairly ordinary regression as a particular type of matching estimator

Suppose that we ran the regression

$$y = \sum_x d_x \alpha_x + \delta_R d + e_i,$$

where d is treatment status and the d_x are dummies for every possible combination of the discrete covariates

A regression that includes every possible combination of a set of discrete covariates is called *saturated*

By the FWL theorem,

$$\delta_R = \frac{Cov(y, \tilde{d})}{V(\tilde{d})} = \frac{E\left\{[d - E(d|x)]y\right\}}{E\left\{d - E(d|x)\right]^2\}}$$
$$= \frac{E\left\{[d - E(d|x)]E(y|d, x)\right\}}{E\left\{d - E(d|x)\right]^2\}}$$

The second equality uses the fact that a saturated regression of d on x identifies $E(d|x)$ and the third uses the fact that regressing y on d and x is the same as regressing y on $E(y|d, x)$

Now we can write

$$E(y|d,x) = E(y|d=0,x) + \delta_x d$$

so the numerator of δ_R becomes

$$\begin{aligned} E\left\{[d - E(d|x)]E(y|d=0,x)\right\} + E\left\{[d - E(d|x)]\delta_x d\right\} \\ = E\left\{[d - E(d|x)]\delta_x d\right\} = E\left\{[d - E(d|x)]^2 \delta_x\right\} \end{aligned}$$

The first equality holds because $E(y|d=0,x)$ is a function of x, and hence uncorrelated with $d - E(d|x)$[1]

The second is an idempotency property (we only need to express one of the terms as deviations from means)

---

[1] $E\{f(x)[d - E(d|x)]\} = E\left[E\left\{f(x)[d - E(d|x)|x\right\}\right] = E\left[f(x)E\left\{[d - E(d|x)|x\right\}\right] = 0$

Now we've shown that

$$
\begin{aligned}
\delta_R &= \frac{E\left\{[d - E(d|x)]^2 \delta_x\right\}}{E\left\{[d - E(d|x)]^2\right\}} \\
&= \frac{E\left\{E\left[\left(d - E(d|x)\right)^2 |x\right] \delta_x\right\}}{E\left\{E\left[\left(d - E(d|x)\right)^2 |x\right]\right\}} \\
&= \frac{E\left[\sigma_d^2(x)\delta_x\right]}{E\left[\sigma_d^2(x)\right]},
\end{aligned}
$$

where

$$
\sigma_d^2 = E\left[\left(d - E(d|x)\right)^2 |x\right]
$$

is the variance of d conditional on x

Finally, we can write this as

$$\delta_R = \frac{\sum_x \delta_x P(d=1|x)[1-P(d=1|x)]P(x)}{\sum_x P(d=1|x)[1-P(d=1|x)]P(x)}$$

For comparison, the matching estimator can be written

$$\delta_{ATT} = \sum_x \delta_x P(x|d=1) = \frac{\sum_x \delta_x P(d=1|x)P(x)}{\sum_x P(d=1|x)P(x)}$$

Thus, the matching estimator puts more weight on values of x that are more likely to be treated, while the regression estimator puts more weight on values where the *variance* of treatment status is greater

Regression and nonlinearity

- Consider a regression of y on s
- We know that if $E(y|s)$ is a linear function of s, then this regression will identify the CEF
- But how can we interpret the coefficients from a regression of y on s when $E(y|s)$ is nonlinear?
- Suppose that the conditional expectation function $h(t) = E(y|s = t)$ is nonlinear

Yitzhaki (1996) showed that a regression of y on s identifies a weighted average of $h'(t)$

$$\frac{E\{y[s-E(s)]\}}{E\{s[s-E(s)]\}} = \frac{\int h'(t)\mu_t dt}{\int \mu_t dt},$$

where

$$\mu_t = [E(s|s \geq t) - E(s|s < t)]P(s \geq t)[1 - P(s \geq t)]$$

Hence, the regression puts more weight on points where the difference in conditional means above/below that value is greater, and on points closer to the conditional median, where $P(s|s \geq t)[1 - P(s|s \geq t)]$ is maximized Derivation

If the regression includes covariates x, this becomes

$$\frac{E[\int h'_x(t)\mu_{tx}dt]}{E[\int \mu_{tx}dt]}$$

where $h'_x = \partial E(y|x, s = t)/\partial t$ and

$$\mu_{tx} = [E(s|x, s \geq t) - E(s|x, s < t)]P(s \geq t, x)[1 - P(s \geq t, x)]$$

Here, we average along the nonlinear CEF for each value of x, then we average across values of x

The preceding results show that a linear regression recovers a weighted average of the nonlinear CEF

However, this weighted average might not seem particularly intuitive

To get more intuition, suppose that s is normally distributed, and define $z = [s - E(s)]/\sigma_s$ and $t^* = [t - E(s)]/\sigma_s$

Then

$$E(s|s \geq t) = E(s) + \sigma_s E(z|z \geq t^*)$$

and, using results on truncated normal rvs,

$$\mu_t = \sigma_s \left(\frac{\phi(t^*)}{1 - \Phi(t^*)} - \frac{-\phi(t^*)}{\Phi(t^*)} \right) [1 - \Phi(t^*)]\Phi(t^*) = \sigma_s \phi(t^*)$$

In this case, a regression of y on s identifies the average derivative $E[h'(s)]$

Extending LATE

- We previously showed that, with heterogeneous treatment effects, the Wald estimator identifies the local average treatment effect
- However, this result assumes that there are no covariates in the regression
- Researchers often include covariates in IV regressions when they think the IV is only exogeneous conditional on some covariates
- How can we incorporate covariates into LATE-type results?

Angrist and Imbens (1995) showed that if the conditions of the LATE theorem hold conditional on x, then 2SLS based on the first-stage equation

$$d = \pi_x + \pi_{1x} z + \xi$$

and the second-stage equation

$$y = \alpha_x + \rho_c d + \eta,$$

where π_x and α_x denote a saturated model for x (dummies for every combination of values of discrete covariates) identifies a weighted average of x-specific LATEs, or

$$\rho_c = E[\omega(x)\lambda(x)]$$

In this weighted average,

$$\lambda(x) = E(y_1 - y_0 | x, d_1 > d_0)$$

and

$$\omega(x) = \frac{V\left[E(d|x,z)|x\right]}{E\left\{V\left[E(d|x,z)|x\right]\right\}}$$

The weights on each x-specific LATE are proportional to the average conditional variance of the population first-stage fitted values $E(d|x,z)$

Although this shows that 2SLS with covariates can still have a "LATEy" interpretation, it isn't common to run fully saturated regressions

Abadie (2003) showed that if $g(y, d, x)$ is a function of y, d and x, then

$$E[g(y, d, x)|d_1 > d_0] = \frac{E[\kappa g(y, d, x)]}{E(\kappa)},$$

where

$$\kappa = 1 - \frac{d(1-z)}{1 - P(z=1|x)} - \frac{(1-d)z}{P(z=1|x)}$$

Thus, we can identify the expected value of any function in the compliant subpopulation as a weighted average of that function

One implication of this is that if we wanted a *linear approximation* to the CEF of y among compliers (which would have a causal interpretation), we could use

$$\arg\min_{a,b} E\left[\kappa\left(y - ad - x'b\right)\right]$$

In addition, if $P(z = 1|x)$ is linear, this minimization problem results in the usual 2SLS estimator

Thus, 2SLS with covariates identifies the best approximation to the CEF among compliers under the assumption that $P(z = 1|x)$ is linear in x

While we might not strictly believe this assumption, it gives us a LATE-like interpretation of the usual 2SLS estimator

Deriving Yitzhaki Back

We can write the regression of y on s as

$$\frac{Cov(y,s)}{Var(s)} = \frac{E\{h(s)[s-E(s)]}{E\{s[s-E(s)]\}}$$

and write

$$h(s) = \left[\lim_{t\to-\infty} h(t)\right] + \int_{-\infty}^{s} h'(t)dt$$

The numerator of the regression can be written

$$\int_{-\infty}^{\infty}\int_{-\infty}^{u} h'(t)[u-E(s)]g(u)dtdu,$$

where $g(u)$ is the density of s at u

Switching the order of integration gives

$$\int_{-\infty}^{\infty} h'(t) \int_{t}^{\infty} [u - E(s)]g(u)dudt$$

Some manipulation shows that the inner integral equals
$\mu_t = [E(s|s \geq t) - E(s|s < t)]P(s \geq t)[1 - P(s \geq t)]$