

Omitted Variable & Diff-in-Diff

Difference of two differences.

Assume treatment is taking a drug or going to school.

$T_i = 1$ if take magic drug

X_i = hours of exercise you do

Y_i = health (0-100)

$$Y_i = a + bT_i + cX_i + e_i$$

So now we can take conditional expectations

$$\mathbb{E}[Y_i|T_i = 1] - \mathbb{E}[Y_i|T_i = 0] = ATE$$

Let's do it:

$$\mathbb{E}[Y_i|T_i = 1] = a + b + c\mathbb{E}[X_i|T_i = 1] + \mathbb{E}[e_i|T_i = 1]$$

$$\mathbb{E}[Y_i|T_i = 0] = a + 0 + c\mathbb{E}[X_i|T_i = 0] + \mathbb{E}[e_i|T_i = 0]$$

$$\text{Let's assume } \mathbb{E}[e_i|T_i = 1] = \mathbb{E}[e_i|T_i = 0] = 0$$

We are left with

$$b + c(\mathbb{E}[X_i|T_i = 1] - \mathbb{E}[X_i|T_i = 0])$$

What's with this last term. If people who exercise (and care more about their health) also are more likely to take the pill, then

$$\mathbb{E}[X_i|T_i = 1] - \mathbb{E}[X_i|T_i = 0] > 0$$

Our goal is get rid of this term. Let's assume there are two time periods.
0=1980 1=1985

$$\mathbb{E}[Y_{it}|T_{i1}]$$

$$\mathbb{E}[Y_{i0}|T_i = 1] - \mathbb{E}[Y_{i0}|T_i = 0]$$

$$\mathbb{E}[Y_{i0}|T_i = 1] = a + b(0) + c\mathbb{E}[X_{i0}|T_{i1} = 1]$$

$$\mathbb{E}[Y_{i0}|T_i = 0] = a + 0 + c\mathbb{E}[X_{i0}|T_{i1} = 0]$$

Subtract and get

$$c(\mathbb{E}[X_{i0}|T_{i1} = 1] - \mathbb{E}[X_{i0}|T_{i1} = 0])$$

Now the big diff-in-diff

$$\begin{aligned} \mathbb{E}[Y_i|T_i = 1] - \mathbb{E}[Y_i|T_i = 0] &= (\mathbb{E}[Y_{i0}|T_i = 1] - \mathbb{E}[Y_{i0}|T_i = 0]) \\ &= b + c(\mathbb{E}[X_i|T_i = 1] - \mathbb{E}[X_i|T_i = 0]) - c(\mathbb{E}[X_{i0}|T_{i1} = 1] - \mathbb{E}[X_{i0}|T_{i1} = 0]) \\ &= b \end{aligned}$$