Blinder Oaxaca Decomposition

Chungsang Tom Lam[1]

[1]Department of Economics
Clemson University

March 3, 2020

OLS estimation

- Consider the regression

$$y = D\gamma + X\beta + \epsilon \tag{1}$$

- where D is a dummy variable and X are the control variables.
- γ is the variable we are interested in
- An example would be an estimation of gender gap, where y is $ln(wages)$, D is the gender dummy and X contains other factors which explains y.

Blinder Oaxaca Decomposition

- We can run the regression separately

$$y_{D=1} = X_{D=1}\beta_{D=1} + \epsilon_{D=1}$$
$$y_{D=0} = X_{D=0}\beta_{D=0} + \epsilon_{D=0}$$

- Or simply

$$y_1 = X_1\beta_1 + \epsilon_1$$
$$y_0 = X_0\beta_0 + \epsilon_0$$

- By running OLS regression separately, we can get $\hat{\beta}_1$ and $\hat{\beta}_0$ as the estimators of β_1 and β_0

Blinder Oaxaca Decomposition

- Then we can examine the difference in the mean of y

$$mean(y_1) - mean(y_0)$$

- This is the average difference in y for individuals in group 1 and group 0.
- We can decompose this difference:

$$\begin{aligned}
& mean(y_1) - mean(y_0) \\
= & mean(X_1\hat{\beta}_1 + \epsilon_1) - mean(X_0\hat{\beta}_0 + \epsilon_0) \\
= & \hat{\beta}_1 mean(X_1) - \hat{\beta}_0 mean(X_0) \\
= & \hat{\beta}_1 mean(X_1) - \hat{\beta}_1 mean(X_0) + \hat{\beta}_1 mean(X_0) - mean(X_0)\hat{\beta}_0 \\
= & \hat{\beta}_1 (mean(X_1) - mean(X_0)) + mean(X_0)(\hat{\beta}_1 - \hat{\beta}_0)
\end{aligned}$$

Blinder Oaxaca Decomposition

- We can write the decomposition in a variety of ways:

$$
\begin{aligned}
&\hat{\beta}_1 mean(X_1) - \hat{\beta}_0 mean(X_0) \\
=&\frac{1}{2}\hat{\beta}_1 mean(X_1) + \frac{1}{2}\hat{\beta}_1 mean(X_1) - (\frac{1}{2}\hat{\beta}_0 mean(X_0) + \frac{1}{2}\hat{\beta}_0 mean(X_0)) \\
=&(\frac{1}{2}\hat{\beta}_1 + \frac{1}{2}\hat{\beta}_0)(mean(X_1) - mean(X_0)) + (\hat{\beta}_1 - \hat{\beta}_0)(\frac{1}{2} mean(X_1) + \frac{1}{2}
\end{aligned}
$$