

Regression $\{0, 1\}$
↓

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

↓

$$E[y_{0i}] + (y_{1i} - y_{0i})x_i + y_{0i} - E[y_{0i}]$$

$$E[y_i | x_i = 1] = \alpha + \beta + E[\varepsilon_i | x_i = 1]$$

$$E[y_i | x_i = 0] = \alpha + E[\varepsilon_i | x_i = 0]$$

$$E[y_i | x_i = 1] - E[y_i | x_i = 0]$$

$$= \beta + E[\varepsilon_i | x_i = 1] - E[\varepsilon_i | x_i = 0]$$

$$\varepsilon_i = y_{0i} - E[y_{0i}]$$

$$= \beta + E[y_{0i} | x_i = 1] - E[y_{0i} | x_i = 0]$$

selection bias.