Econometrics 2 – Part 1

Arieda Muço

Central European University

Winter 2023

Bad Controls

- Assume that C_i is randomly assigned, so it is independent of all potential outcomes
- Compare mean earnings for white collar workers with or without college

$$\begin{aligned} Y_i &= C_i Y_{1i} + (1 - C_i) Y_{0i} \\ W_i &= C_i W_{1i} + (1 - C_i) W_{0i} \end{aligned}$$

$$\begin{aligned} E[Y_i|C_i = 1] - E[Y_i|C_i = 0] &= E[Y_{1i} - Y_{0i}] \\ E[W_i|C_i = 1] - E[W_i|C_i = 0] &= E[W_{1i} - W_{0i}] \end{aligned}$$

Bad Controls

- By the joint independence of $(Y_{1i}, Y_{0i}, W_{1i}, W_{0i}) \amalg C_i$ we have

$$\begin{aligned} E[Y_{1i}|W_{1i}=1, C_{1i}=1] - E[Y_{0i}|W_{0i}=1, C_{0i}=0] &= \\ E[Y_{1i}|W_{1i}=1] - E[Y_{0i}|W_{0i}=1] \end{aligned}$$

- This expression illustrates the apples-to-oranges nature of the bad-control problem

$$E[Y_{1i}|W_{1i}=1] - E[Y_{0i}|W_{1i}=1] + E[Y_{0i}|W_{1i}=1] - E[Y_{0i}|W_{0i}=1]$$

Bad Controls

$$\underbrace{E[Y_{1i}|W_{1i}=1]-E[Y_{0i}|W_{1i}=1]}_{\text{causal effect on college grads}}+\underbrace{E[Y_{0i}|W_{1i}=1]-E[Y_{0i}|W_{0i}=1]}_{\text{selection bias}}$$

$$\underbrace{E[Y_{1i}-Y_{0i}|W_{1i}=1]}_{\text{causal effect on college grads}}+\underbrace{E[Y_{0i}|W_{1i}=1]-E[Y_{0i}|W_{0i}=1]}_{\text{selection bias}}$$

How bad control creates selection bias

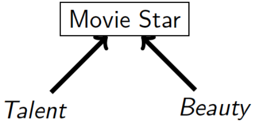
<table>
<tr><th rowspan="2">Type of worker</th><th colspan="2">Potential occupation</th><th colspan="2">Potential earnings</th><th colspan="2">Average earnings by occupation</th></tr>
<tr><th>Without college (1)</th><th>With college (2)</th><th>Without college (3)</th><th>With college (4)</th><th>Without college (5)</th><th>With college (6)</th></tr>
<tr><td>Always Blue (AB)</td><td>Blue</td><td>Blue</td><td>1,000</td><td>1,500</td><td rowspan="2">Blue 1,500</td><td>Blue 1,500</td></tr>
<tr><td>Blue White (BW)</td><td>Blue</td><td>White</td><td>2,000</td><td>2,500</td><td rowspan="2">White 3,000</td></tr>
<tr><td>Always White (AW)</td><td>White</td><td>White</td><td>3,000</td><td>3,500</td><td>White 3,000</td></tr>
</table>

Collider Bias

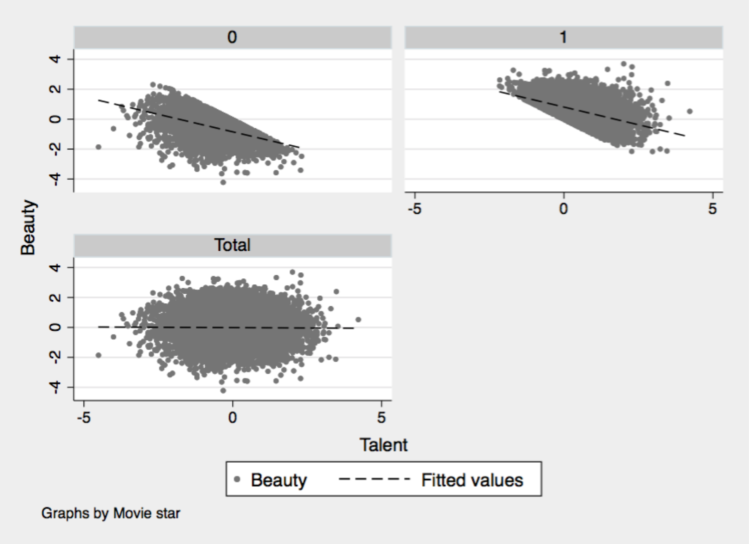
- CNN.com headline: Megan Fox voted worst but sexiest actress of 2009
- Assume talent and beauty are independent, but each causes someone to become a movie star. What's the correlation between talent and beauty for a sample of movie stars compared to the population as a whole (stars and non-stars)?

Collider Bias Exapmle

- What if the sample consists *only* of movie stars?



Collider Bias Examle (see STATA code for replication)



Proxy Variables

- Sometimes you will see cases when OVB is attempted to be solved via obtaining a proxy variable
- In the wage equation, one possibility is to use the intelligence quotient, or IQ, as a proxy for ability
- This does not require IQ to be the same thing as ability; what we need is for IQ to be correlated with ability

- If A_i is observed:

$$Y_i = \alpha + \rho s_i + A_i' \gamma + \nu_i \Rightarrow \text{Long Regression}$$

- A_i is unobserved, how can we estimate ρ?
- Call the proxy variable IQ_i. What do we require from IQ_i?

$$A_i = \delta_0 + \delta_1 IQ_i + u_i$$

- where u_i is an error due to the fact that A_i and IQ_i are not exactly related
- typically, we think of $\delta_1 > 0$. If $\delta_1 = 0$, then it's not a suitable proxy
- The intercept δ_0, which can be positive or negative, simply allows A_i and IQ_i to be measured on different scales. (For example, unobserved ability is certainly not required to have the same average value as IQ in the U.S. population.)

- Assumptions:

$$
\begin{array}{ll}
1) & E[\nu_i|s_i, A_i] = 0 \\
2) & E[u_i|IQ_i] = 0
\end{array}
$$

- IQ_i is irrelevant in the population model, once A_i has been included
- Proxy variables lead to bias if assumptions not satisfied.