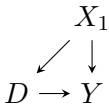
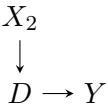
Roadmap

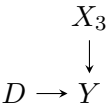
- What controls to include?
- Proof of Unconfoundedness given propensity score
- Why not simple linear covariate adjustment?

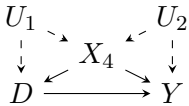
Good and bad controls

Which of these variables would you control for? Think identification and efficiency



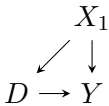




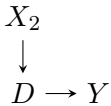


Good and bad controls

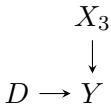
Which of these variables would you control for? Think identification and efficiency



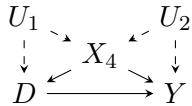
X_1 Necessary for identification



X_2 Bad for efficiency (but may be good for robustness)



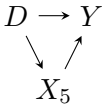
X_3 Good for efficiency

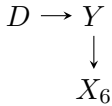


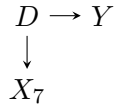
X_4 “Collider,” generates bias (but may serve as proxy for U_2 if U_2 affects D)

Good and bad controls

Which of these variables would you control for? Think identification and efficiency

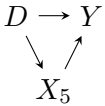




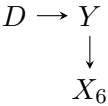


Good and bad controls

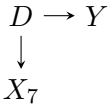
Which of these variables would you control for? Think identification and efficiency



X_5 Bad control, generate bias



X_6 Bad control, generate bias



X_7 Exercise for you

Roadmap

- What controls to include?
- Proof of Unconfoundedness given propensity score
- Why not simple linear covariate adjustment?

Proof

This proof is from Imbens (2005, RESTAT) (may skip in class).

Show $\mathbb{P}(D_i = 1 \mid Y_i(0), Y_i(1), \pi(\boldsymbol{X}_i)) = \mathbb{P}(D_i = 1 \mid \pi(\boldsymbol{X}_i)) = \pi(\boldsymbol{X})_i$ which implies that D_i and the potential outcomes are independent conditional on the propensity score

Proof

$$
\begin{aligned}
\mathbb{P}(D_i = 1 \mid Y_i(0), Y_i(1), \pi(\boldsymbol{X}_i)) &= \mathbb{E}[D_i = 1 \mid Y_i(0), Y_i(1), \pi(\boldsymbol{X}_i)] \\
&= \mathbb{E}[\,\mathbb{E}[D_i = 1 \mid Y_i(0), Y_i(1), \pi(\boldsymbol{X}_i), \boldsymbol{X}_i] \mid Y_i(0), Y_i(1), \pi(\boldsymbol{X}_i)] \\
&= \mathbb{E}[\,\mathbb{E}[D_i = 1 \mid Y_i(0), Y_i(1), \boldsymbol{X}_i] \mid Y_i(0), Y_i(1), \pi(\boldsymbol{X}_i)] \\
&= \mathbb{E}[\,\mathbb{E}[D_i = 1 \mid \boldsymbol{X}_i] \mid Y_i(0), Y_i(1), \pi(\boldsymbol{X}_i)] \\
&= \mathbb{E}[\pi(\boldsymbol{X}_i) \mid Y_i(0), Y_i(1), \pi(\boldsymbol{X}_i)] = \pi(\boldsymbol{X}_i),
\end{aligned}
$$

where 1st equality comes from D_i being a 0/1 variable, 2nd from LIE, 4th from CIA (conditional on $\boldsymbol{X}$), and fifth from definition of the propensity score

Roadmap

- What controls to include?
- Proof of Unconfoundedness given propensity score
- Why not simple linear covariate adjustment?

Simplest regression

Say the data-generating model is given by

$$y_i = D_i\tau_i + W_i'\beta + u_i$$

- Effects of treatment, τ_i, can vary arbitrarily by individuals
- D_i is a treatment variable and W_i is a vector of controls