

# WWS 403 Methods Lab

## Lecture 3

---

Will Lowe

Department of Politics, Princeton University

## The Plan

Announcements: Brief debrief and extra sessions

Part 1: How to get people to vote?

Part 2: What should I put in my model?  $\beta$

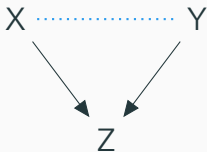
Lab: Because we love labs

## EXTRAS: SELECTION ON THE DEPENDENT VARIABLE

Even when there is no causal link between X and Y



conditioning on Z will create spurious association.



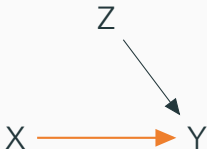
## SELECTION ON THE DEPENDENT VARIABLE

This is why selecting cases on the dependent variable is problematic.

## SELECTION ON THE DEPENDENT VARIABLE

This is why selecting cases on the dependent variable is problematic.

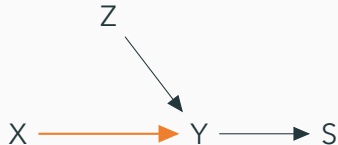
Assume Z raises Y but does not affect X



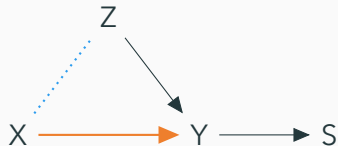
Z is not a confounder

## SELECTION ON THE DEPENDENT VARIABLE

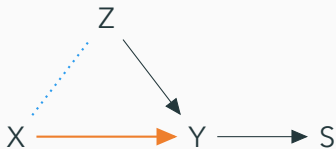
However, if we select (S) only units with, say, high Y



Then we will see spurious association between X and Z

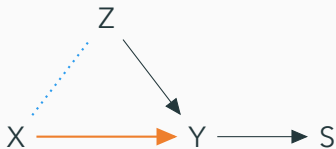


## SELECTION ON THE DEPENDENT VARIABLE



This subsample will have more high Z units with low X and more low Z units with high X than the population.

## SELECTION ON THE DEPENDENT VARIABLE



This subsample will have more high Z units with low X and more low Z units with high X than the population.

So effect of X on Y will seem to be *larger* than it really is.



## SELECTION ON THE DEPENDENT VARIABLE

This effect occurs regardless of whether  $Z$  is measured.

## SELECTION ON THE DEPENDENT VARIABLE

This effect occurs regardless of whether  $Z$  is measured.

Assume  $Z$  is *not* measured. Then it is part of the error term in the regression

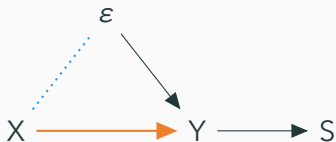
$$Y = \alpha + X\beta + \varepsilon$$

## SELECTION ON THE DEPENDENT VARIABLE

This effect occurs regardless of whether  $Z$  is measured.

Assume  $Z$  is *not* measured. Then it is part of the error term in the regression

$$Y = a + X\beta + \varepsilon$$



$\varepsilon$  is now correlated with  $X$  because of the selection, so  $\hat{\beta}$  no longer estimates  $\beta$ .