# Regression in Action: Bias, squares, and measurement error

Econ 140, Section 3

Jonathan Old

- 1. Regression
- 2. Recap
- 3. Selection bias in action: Omitted Variable Bias
- 4. Measurement Error

Any questions?

## ... Remember - Every question is useful!

## Regression

### Regression: Basics, Interpretation, Quadratic Terms

See in Datahub

## Recap

#### **Recap: Selection bias**

We saw that whenever we do a difference-in-means comparison (or a regression), we get:



## Causal Effect +

Selection Bias

- We cannot observe them so we can never be sure!
- Econometrics is all about uncertainty. You can always state that there are different possibilities and you cannot know for sure

#### Recap: How to think about Selection Bias



Confounder / Omitted Variable

Figure 1: Selection bias

## Selection bias in action: Omitted Variable Bias

## The most important slide of this course (until the midterm)

We can summarize everything of OVB in three equations. Let *Y<sub>i</sub>* be the outcome variable, *X<sub>i</sub>* our regressor of interest, and *Z<sub>i</sub>* the "omitted" variable.

[Long regression]  $Y_i = c_1 + \beta_L X_i + \delta Z_i + e_i$ [Short regression]  $Y_i = c_2 + \beta_S X_i + u_i$ [Auxiliary regression]  $Z_i = c_3 + \gamma X_i + v_i$ 

Then, the Omitted variable bias formula states that:



We call  $\delta\gamma$  the **omitted variable bias**. We can appraise the direction of the bias by multiplying our guesses for the signs of  $\delta$  and  $\gamma$ .

#### Wait a minute!

We saw that the OLS coefficient in the short regression is:



#### This looks awfully much like:

#### Estimating the effect of iPads on grades

Let us start with a difference-in-means comparison:

$$\Delta = E[\text{Grade}_i | i\text{Pad}_i = 1] - E[\text{Grade}_i | i\text{Pad}_i = 0]$$

Add and subtract  $E[Grade_i(0)|iPad_i = 1]$ 

 $= E[Grade_i(1)|iPad_i = 1] - E[Grade_i(0)|iPad_i = 1] +$ 

 $E[Grade_i(0)|iPad_i = 1] - E[Grade_i(0)|iPad_i = 0]$ 

Use properties of expectations

 $= E[Grade_i(1) - Grade_i(0)|iPad_i = 1] +$ 

 $E[Grade_i(0)|iPad_i = 1] - E[Grade_i(0)|iPad_i = 0]$ 

= ATT + Selection bias

Selection bias: Students with and without iPad have different potential grades: even if they both had iPads, they would be different.

6

Control variables are additional variables (or covariates) included in a regression. We do this for various reasons (in decreasing order of importance):

- To remove selection bias / omitted variable bias
- To increase precision of our estimates
- To know about the (conditional/partial) correlation of other variables
- To better predict the outcome

Let's see graphically how control variables work.

Any questions?

## ... Remember - Every question is useful!

## **Measurement Error**

See on Blackboard and in **RStudio**.