

# EC203 – Applied Econometrics

Term 2, Week 7

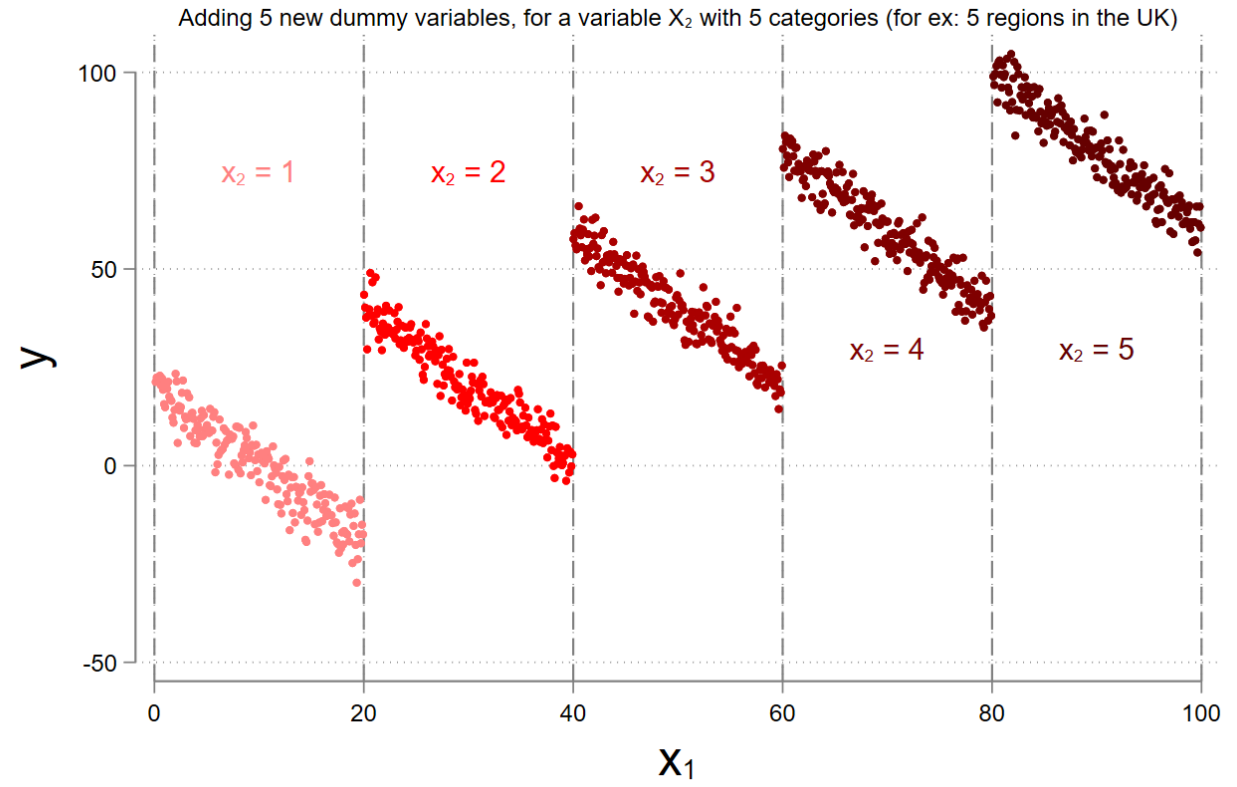
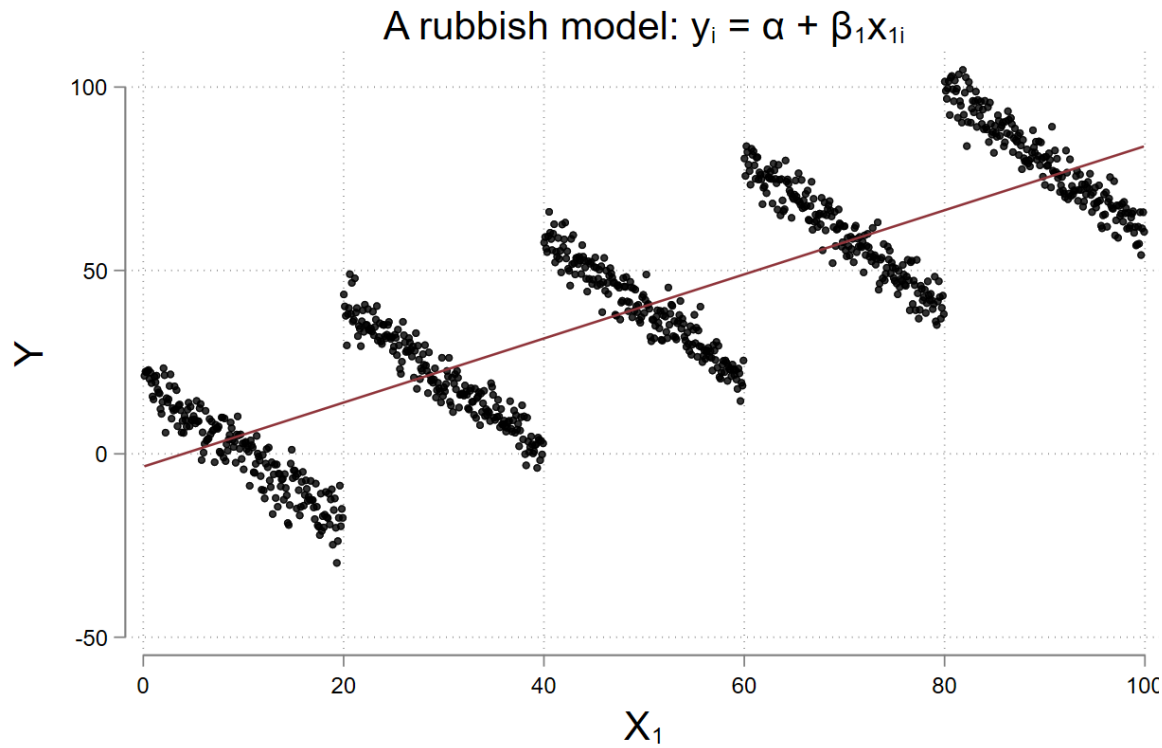
*Sushil Mathew*



# Types of endogeneity

...and ways to fix them

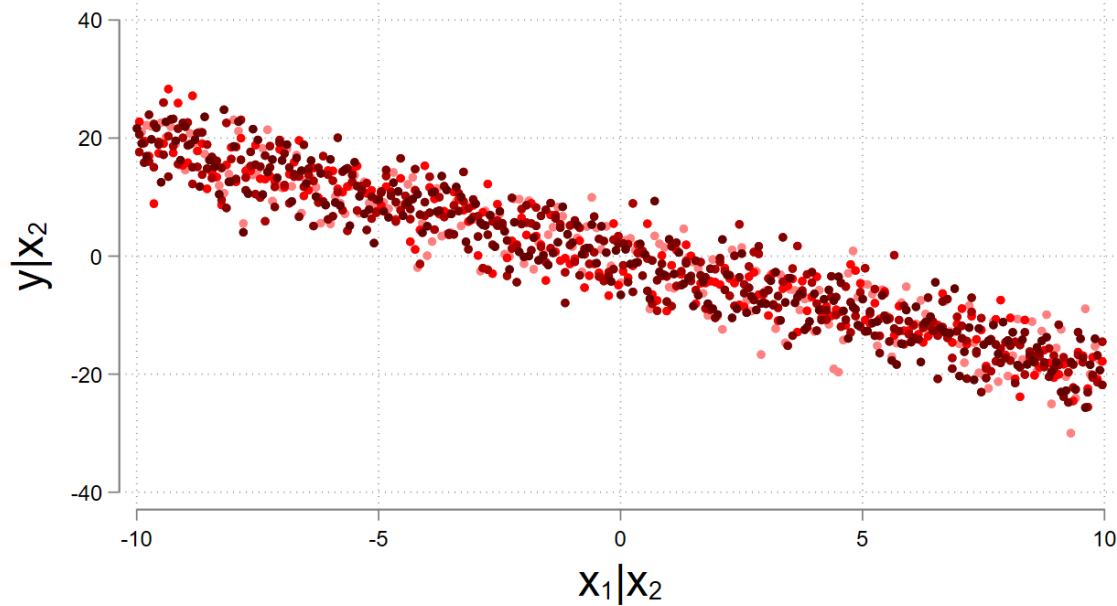
# Problem 1: Omitted variable bias



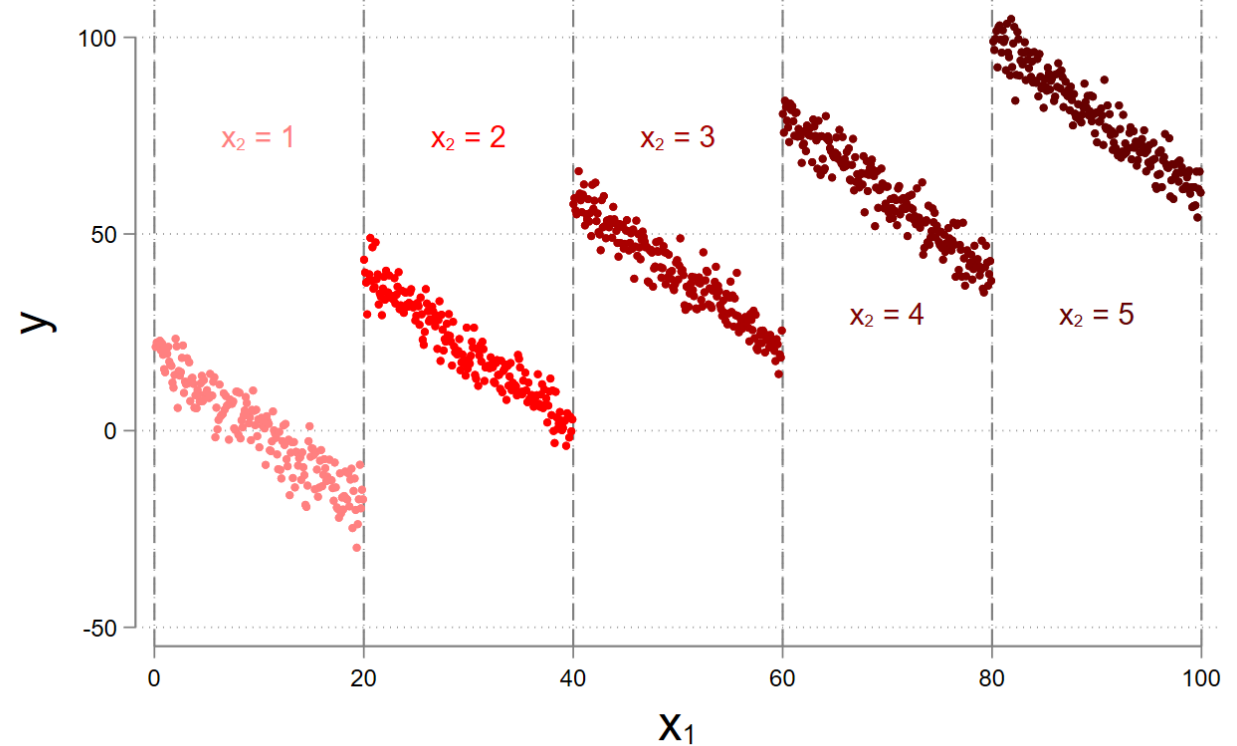
# Solution 1.1: Adding variables to a regression is good

What controlling for a variable looks like visually

Notice the axes scale and colour of dots.  
Compare this to Fig2 colours



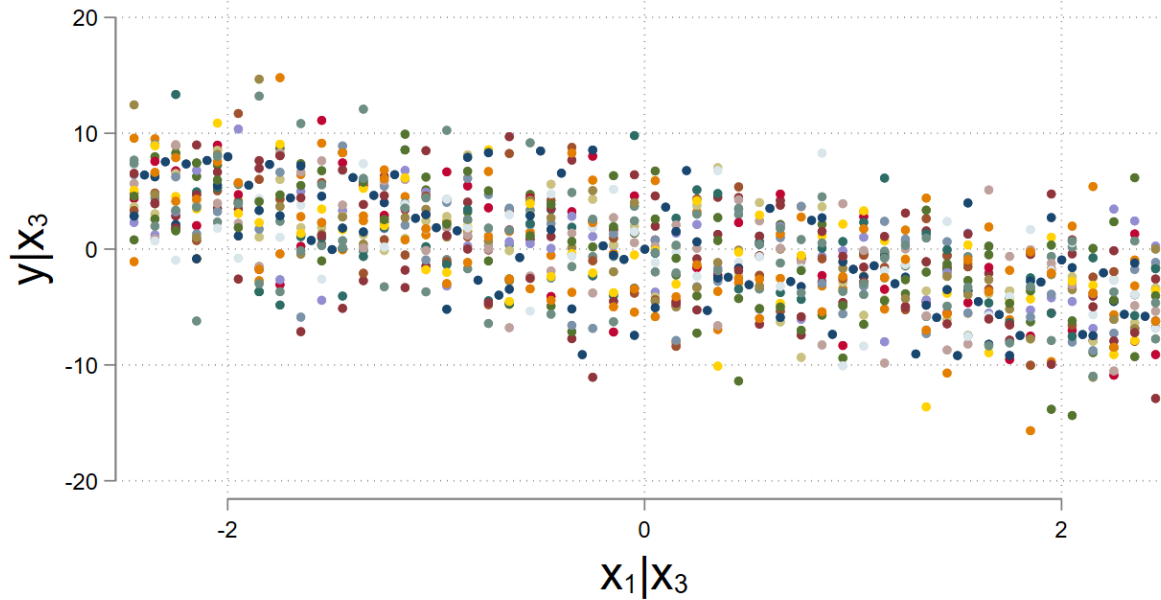
Adding 5 new dummy variables, for a variable  $X_2$  with 5 categories (for ex: 5 regions in the UK)



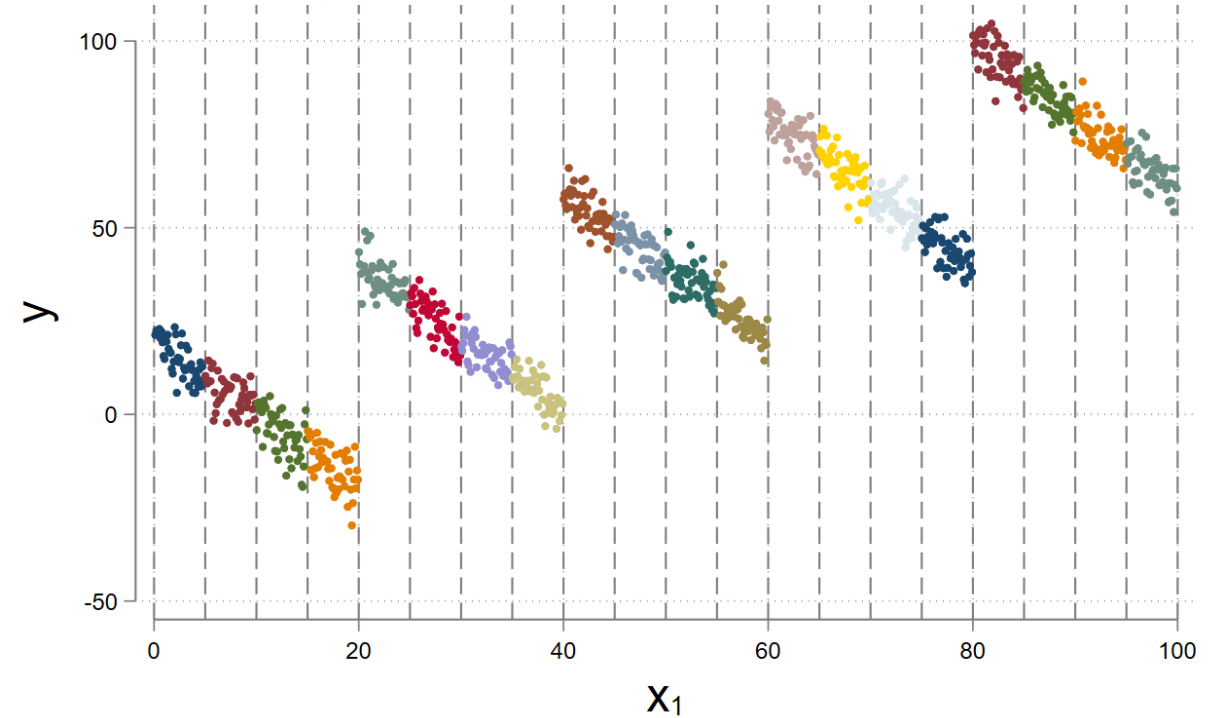
# Solution 1.1: Not perfect solution

What controlling for an irrelevant variable looks like visually

The slope is still the same  
But the uncertainty in the slope has increased  
Also compare scale of y and x axis with fig3



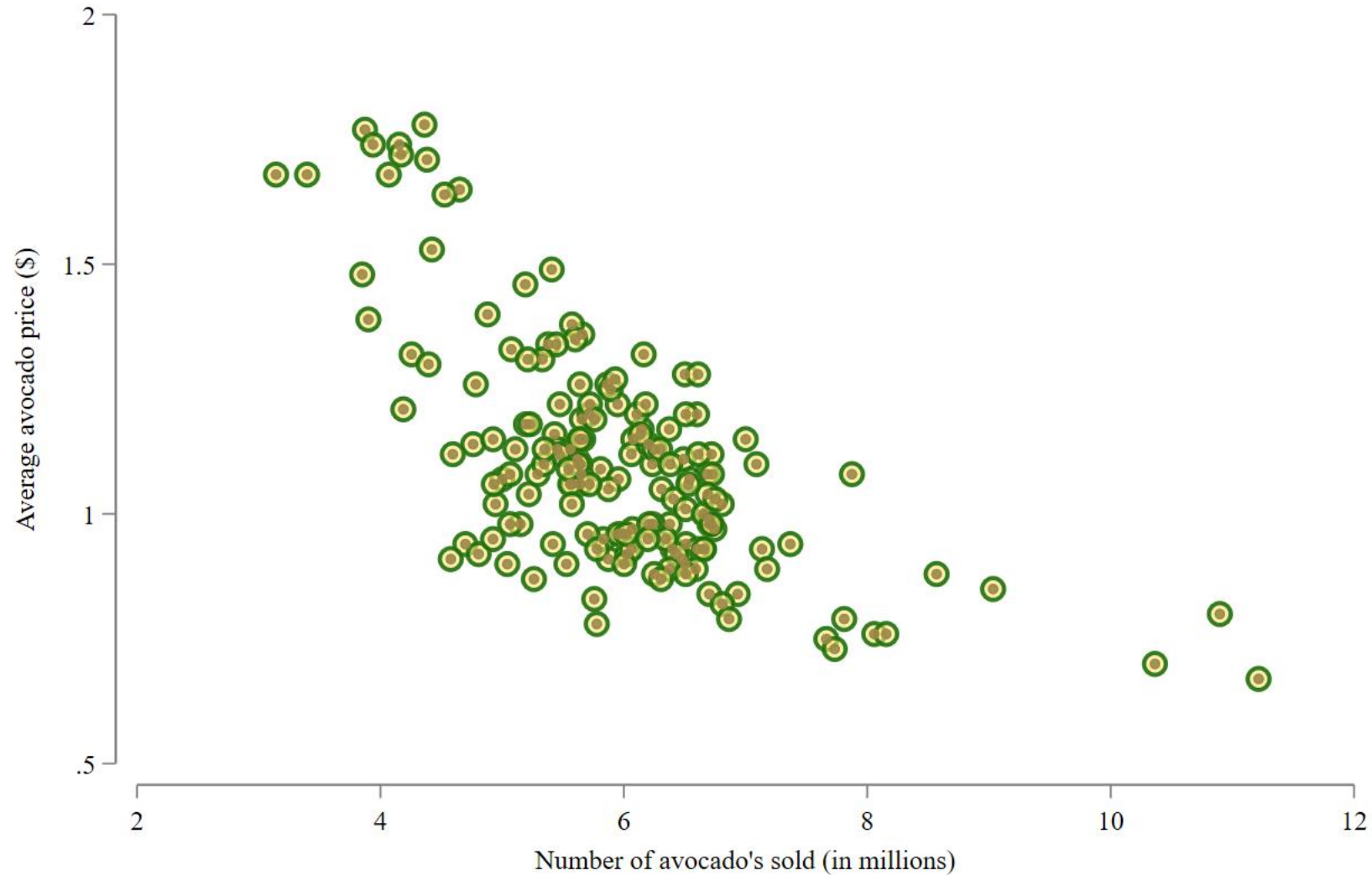
Adding a variable  $X_3$  with 20 (irrelevant) categories



## Solution 1.2: Instrumental variables

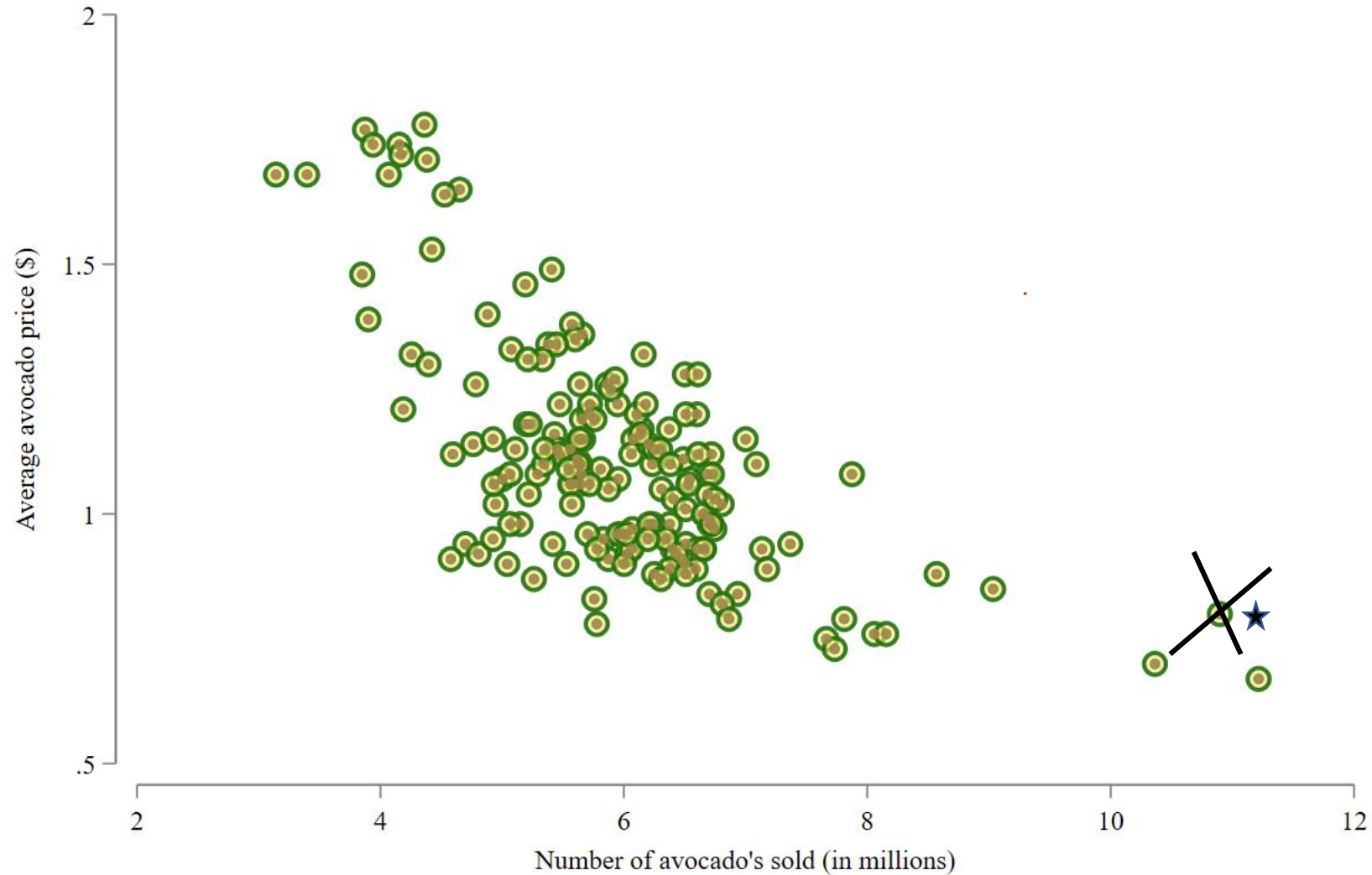
## Problem 2: Reverse causality or simultaneity

## Problem 2: Reverse causality or simultaneity





# Problem 2: Reverse causality or simultaneity



## Solution 2.1: Instrumental variables

# Problem 3: Measurement Error

<https://twitter.com/simonhhess/status/1590366800687992832>

## Solution 3.1: Instrumental variables

# Endogeneity can be due to

- Omitted variable bias
- Reverse causality/simultaneity
- Measurement error in the x variable.

# Endogeneity can be due to

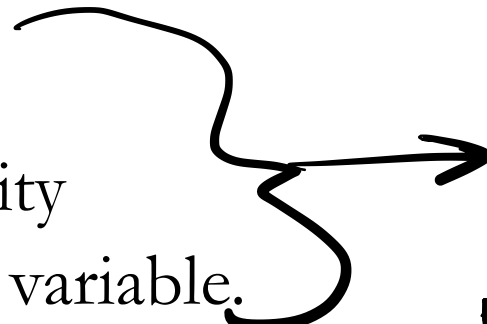
- Omitted relevant variable
- Reverse causality/simultaneity
- Measurement error in the x variable.

## Consequence of endogeneity

Bias in the estimated parameters.

# Endogeneity can be due to

- Omitted relevant variable
- Reverse causality/simultaneity
- Measurement error in the x variable.



$E(\epsilon | x) \neq 0 \leftarrow$

$y = \alpha + \beta x + \epsilon$

$x$  is  
ENDOGEN.

## Consequence of endogeneity

Bias in the estimated parameters.

## One solution for all these problems

Instrumental variables

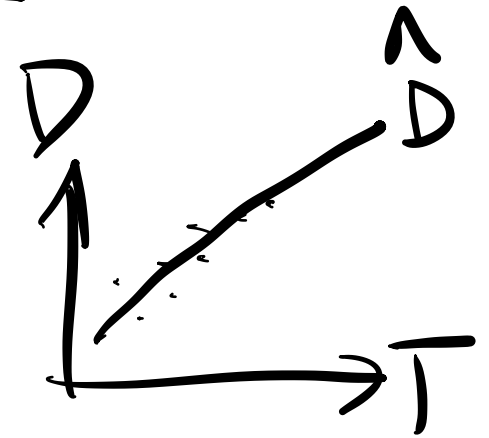
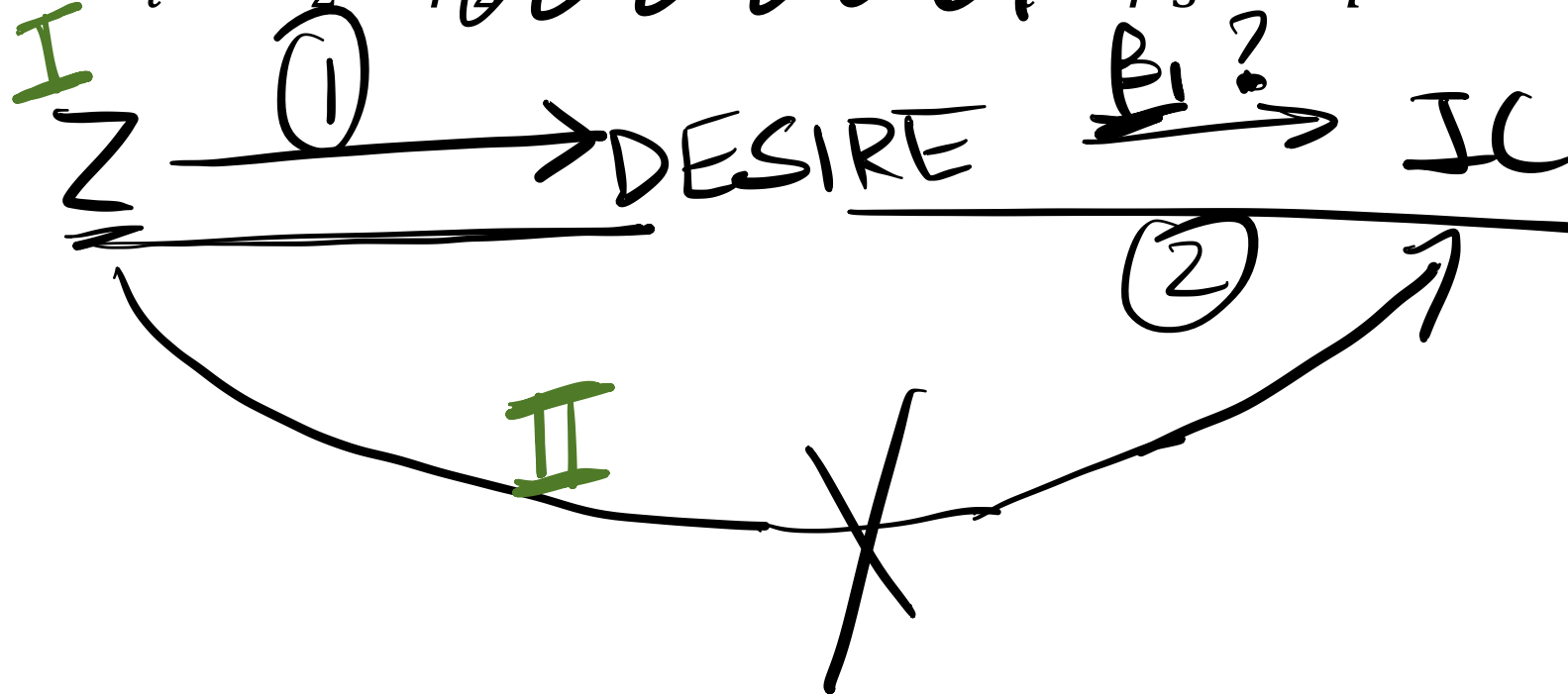
Q1 - 2



Simultaneity/reverse causality  $\uparrow \uparrow$  DES  $\xrightarrow{\text{ICE}} \uparrow \uparrow$

$$IceCreamShops_i = \alpha_1 + \beta_1 \hat{Desire}_i + \varepsilon_{1i} \quad (A)$$

$$Desire_i = \alpha_2 + \beta_2 \cancel{IceCreamShops_i} + \beta_3 Temperature_i + \varepsilon_{2i} \quad (B)$$



# Simultaneity/reverse causality

$$IceCreamShops_i = \alpha_1 + \beta_1 Desire_i + \varepsilon_{1i}$$

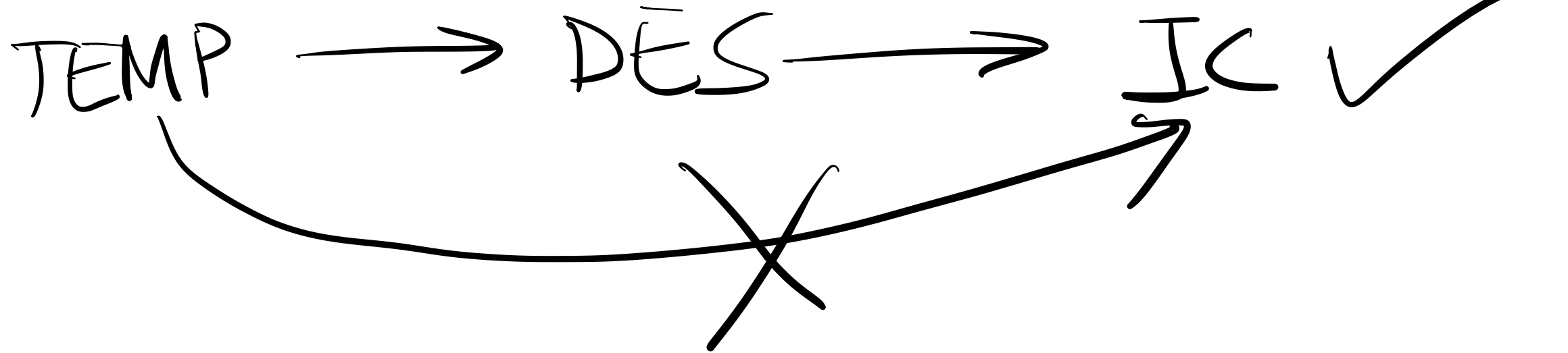
$$Desire_i = \alpha_2 + \beta_2 IceCreamShops_i + \beta_3 Temperature_i + \varepsilon_{2i}$$

I: RELEVANCE  
:  $COV(Z, D) \neq 0$

II: IV EXOG.  
 $COV(Z, \varepsilon_1) = 0$

Q1, Q2

$$\left\{ \begin{array}{l} E(\varepsilon_1 \mid Desire) \neq 0 \\ E(\varepsilon_2 \mid Ice\ Cream\ Shops) \neq 0 \end{array} \right\}$$



Q3



# Simultaneity/reverse causality fix

$$\text{IceCreamShops}_i = \alpha_1 + \beta_1 \text{Desire}_i + \varepsilon_{1i}$$

$$\text{Desire}_i = \alpha_2 + \beta_2 \text{IceCreamShops}_i + \beta_3 \text{Temperature}_i + \varepsilon_{2i}$$

THEOR.  
MODEL

WE

BELIEVE  
THIS IS TRUE

fit (fit)