

### OVB and Instrumental Variables

Due Friday July 29th, 8pm on Gradescope

The first question is worth 40% and the second question is worth 60%. Each sub question will be weighted equally and assessed according to the following criteria: No Answer = 0, incorrect or unclear = 1pt, correct but incomplete = 2pt, correct and complete = 3pt.

1. In a video watched on lecture one, an interviewer brings up the point that the gender pay gap in the UK is 9% as evidence that modern society is still primarily dominated by men. Her point is that gender causes a wage differential.
  - a. Write down the underlying regression that corresponds to interviewers' claim that women earn 9% less than men on average in the UK. How large would the standard errors need to be in this estimation for the gap to be indistinguishable from 0 (i.e. we cannot reject the null hypothesis of no difference)?

To this claim, the interviewee responds: "Multivariate analysis of the pay gap indicates that it doesn't exist ... If you are a social scientist worth your salt, you never do a univariate analysis [...]. You break it down by age, by occupation, interest, personality [and the gap disappears]"

- b. Write down the underlying regression that corresponds to the commentator's response.
  - c. Using concepts from class, particularly of regression, what is the commentator implying regarding the first estimate (in a)?
  - d. Discuss how the OVB formula could help us understand the effect of including a variable like "personality" on the effect of gender on wages. Write down the long, short and auxiliary equations and interpret the OVB formula (the right hand side, not the  $\beta_L - \beta_S$  part)

[help: here you should define the variable personality and interpret its meaning. For example following the (overly simplified) logic of the commentator you can think of personality measuring agreeableness with 0 being not agreeable at all and 100 being highly agreeable. Also following his arguments, jobs with more agreeable people tend to have lower salaries and women tend to be more agreeable. Use this in the OVB formula. This is the section where the note below applies the most]

(Note: remember we are just half way through uncovering the BS in the commentator's argument, first we need to understand his point, so then we can expose the flaws in it. If you are frustrated with the state of the debate up to here, use that frustration as a motivation for learning the remaining material of the course!)

2. Watch [this video](#) from the great Econimate, and answer the following questions:

- a. What is the main causal question of interest? And what would a simple difference in groups be measuring? [hint: define the outcome variable for conflict as a binary variable, 0 no conflict in a given location and time, 1 yes conflict]

In b-e you are asked to use the OVB formula to describe the identification problem in this case

- b. Write down a regression equation that would estimate the simple difference in groups and label it the short equation.
- c. Think of an omitted variable that could be affecting beta\_s and write down the long equation (tip: here you have to speculate about a variable that you don't have, so to make it concrete, think of the omitted factor and then turn it into some type of score. For example, in a different analysis we could suspect that motivation is an omitted factor. To study its effect on the OVB formula, we just assume that we have something like a motivation score where 0 is the least and 100 is the most motivated that an individual can be). [hint: in the 1min 30sec you see suggestions of other factors, turn one of those into a omitted variable, for example a variable could be "positive attitude towards leaders in time of peace"]
- d. Write down the auxiliary equation
- e. Discuss the OVB formula in this case (again, both sides of the equation)
- f. Describe one of the instruments that are used in the IV approach. Define the variable. [hint: this are describe starting at 1min55sec, the key word here "arbitrary" used in as a synonym of random, there are two instruments]
- g. Write down the first stage equation for this IV.
- h. Describe the exclusion restriction in this case.