ECON 20: Lecture #4
Practice Problems

1 Omitted variables bias

Here we use the same geenral framework as in class. We are interested in the causal effect of a variable x on the variable y. Suppose the true model for y is:

$$y_i = \alpha + \beta x_i + \gamma z_i + \epsilon_i \tag{1}$$

Further, assume that $E[\epsilon|x, z] = 0$, meaning that if we were to estimate equation (1), the *long regression*, our estimate of β would be an unbiased estimate of the true causal effect.

Now suppose you do not have data on z and so you instead estimate the short *short regression*, $y_i = \alpha + \tilde{\beta} x_i + \tilde{\epsilon}_i$. In class, we showed that when we do so, we obtain:

$$E\left[\tilde{\beta}\right] = \beta + \gamma \frac{Cov(x_i, z_i)}{Var(x_i)} \tag{2}$$

We say that an estimate is positively biased (or *biased upward*) when the estimate is larger than the truth and negatively biased (or *biased downward*) when the estimate is smaller than the truth.

1. When is our estimate of β from the short regression positively biased?
2. When is our estimate of β from the short regression negatively biased?
3. When is our estimate of β from the short regression unbiased?
4. You may find it useful to write the sign of bias in table form.
5. In our STATA exercise from last class, we regressed the natural log of the wage on years of education and obtained:

$$\hat{lnwage} = \beta_0 + .12 \cdot yrsed$$

Now suppose that we add the variable *age* to our regression and we we obtain:

$$\hat{lnwage} = \beta_0 + .12 \cdot yrsed + 0.091 \cdot age$$

What can we conclude by comparing the two regressions? Be as specific as possible.

6. Now suppose you are able to get a new variable called IQ, which measures an individual's IQ score. You regress the IQ score on years of education and get:

$$\hat{IQ} = \beta_0 + 0.4 \cdot yrsed$$

Interpret the 0.4 in the above regression.

7. Given the above regression, under what condition is our estimate of β *not biased* by excluding IQ from the short regression of $lnwage$ on $yrsed$? Express your answer in words and be as specific as possible.

8. Assume that the condition above (your answer to part 7) is not satisfied. What do you expect the direction of the OVB from the exclusion of IQ from the short regression to be?

9. Suppose I told you that if we add IQ to the regression, that is if we estimate $lnwage = \alpha + \beta \cdot yrsed + \gamma \cdot IQ + u$, the estimated coefficient on IQ is $\hat{\gamma} = 0.15$. What is the estimated coefficient on $yrsed$ ($\hat{\beta}$) in this regression?

10. Make a subjective judgement about the severity of the bias. In other words, how severe was the bias in our estimate of β from the exclusion of IQ from the regression?