Classwork 16: Omitted Variable Bias (part 1: analytical)

1. Show that $\hat{\beta}_1$ is an unbiased estimator of β_1 when the model is correctly specified as $y_i = \beta_0 + \beta_1 x_i + u_i$ and exogeneity holds.

You can start with the formula $\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$. I think a good first step is to substitute in for y_i and $\bar{y}$.

2. Signing the Bias (omitted variable bias)

2a) Suppose the true data generating process for y is $y = \beta_0 + \beta_1 x + \beta_2 z + u$ but we omit z by estimating the incorrect model $y = \beta_0 + \beta_1 x + u$. Show that $\hat{\beta}_1 = \beta_1 + \beta_2 \frac{\sum_i (x_i - \bar{x})(z_i - \bar{z})}{\sum_i (x_i - \bar{x})^2} + \frac{\sum_i (x_i - \bar{x}) u_i}{\sum_i (x_i - \bar{x})^2}$.

You can start with a similar procedure as question 1: take $\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$ and plug in true formulas for y and $\bar{y}$.

2b) The first term is the true value β_1. The third term is the sampling error, which is 0 in expectation if exogeneity holds with respect to x and u. The second term $\beta_2 \frac{\sum_i (x_i - \bar{x})(z_i - \bar{z})}{\sum_i (x_i - \bar{x})^2}$ is the omitted variable bias. Fill in the blanks:

β_2 is the true effect of z on ____________.

$\sum_i (x_i - \bar{x})(z_i - \bar{z})$ is $(n-1)$____________.

$\sum_i (x_i - \bar{x})^2$ is $(n-1)$____________.

2c) In the workbook chapter 15, you learned that if an omitted variable affects both x and y, it causes omitted variable bias and confounds the relationship you're trying to measure between x and y.

Considering the OVB term in 2b, what would the OVB be if z did not affect x? What about if z did not affect y? Explain.

2d) We can also use the OVB term in 2b to sign the bias. That is, we can predict whether $\hat{\beta}_1$ will be greater than or less than the true β_1 in the presence of bias from some omitted variable.

If z positively affects y and z positively affects x, will the OVB bias term be positive or negative? Will $\hat{\beta}_1$ be greater than or less than the true β_1?

2e) If z negatively affects y and z negatively affects x, will the OVB bias term be positive or negative? Will $\hat{\beta}_1$ be greater than or less than the true β_1?

2f) If z positively affects y but z negatively affects x, will the OVB bias term be positive or negative? Will $\hat{\beta}_1$ be greater than or less than the true β_1?

3) Consider the model Earnings$_i = \beta_0 + \beta_1$Years of Education$_i + u_i$. Answer the questions below.

3a) When we have to omit ability from this model, does it make it look like education is a better or worse investment than it actually is? Explain.

3b) When we have to omit conscientiousness from this model, does it make it look like education is a better or worse investment than it actually is? Explain.

3c) When we have to omit conformity from this model, does it make it look like education is a better or worse investment than it actually is? Explain.

4) Consider the model High School GPA$_i = \beta_0 + \beta_1$Number of Opposite Sex Friends$_i + u_i$.

When we have to omit **Strict Parents** from this model, does it make it look like having friends of the opposite sex is better or worse for your GPA than it actually is? Explain.