Take-Home Assignment - Part 2

Theory

OLS Proof:

After reviewing it, wait fifteen minutes and try to write down the OLS Proof of Unbiasedness for yourself (and annotate each transformation – i.e. jot down, what this step does). Highlight where you got stuck and had to revisit the proof; mark in different color where you feel like you don't fully understand the step; mark in a third color where you use E(u|X)=0 (or E(uX)=0)

Endogeneity

A.) Now redo the proof, but assume E(u|X) = 0.65
⇨ What bias do you get?
⇨ Proof you answer by showing it in the proof.

B.) Write down all 4 sources of endogeneity.

C.) Below is a Haiku that describes a situation where endogeneity biased the estimation result. Which source of endogeneity does it refer to?

Haiku: Bias

Regressed hospital visits on mortality.

Found that hospitals kill people.

But Healthy don't go to hospital.

D.) Try to think of 2 (or up to 4 if you want to do one for each source) examples/regressions where an important source of endogeneity is likely to bias an OLS-estimation.
Briefly describe it in a few lines. (hint: a Haiku (3 lines with 3-7 words; See below for an example) that describes it).

Regression Exercises

Question E1: OmittedVariables

Generate a 5-dimensional multivariate normal (Y, X1, X2, Z1, Z2). 5347 observations (alternatively use your own data), with bilateral correlations as follows:

Correlations	Y	x1	x2	Z1	Z2
Y	1.0	0.2	0.1	0.35	0.0
x1	0.2	1.0	0.0	0.4	0.0
x2	0.1	0.0	1.0	0.0	0.4
Z1	0.35	0.4	0.0	1.0	0.6
Z2	0.0	0.0	0.4	0.6	1.0

Add a vector of ones (or any other constant)

1.) Manually compute the sample covariances for all pairs and the sample variances. Compare the difference between the sample covariances (the matrix you get from your data) and the theoretical Var-Cov Matrix (the matrix above). Is it exactly the same? Is it almost the same?
2.) Univariate Regression:
    a. Use these to obtain the regression coefficient when regressing Y only on X1 (slide set from Intro)
    b. Use standard regression commands to regress y only on x1. Compare the coefficient.
3.) Multivariate Regression on foot
    a. Next define your data (ones and the two X columns as Matrix X [ignore the Z variables for now])
    b. Compute X´X
    c. Compute beta-hat for x1 and x2
    d. Now run the standard regression command
    e. Compare.
4.) Compare your univariate to you multivariate regression
    a. What would theory predict regarding the difference of the coefficients between between uni and multivariate regression?
    b. Comparing your results, is the estimated coefficient for X1 the same or different between uni and multivariate?
    c. Is that exactly what theory predicted? (Why or why not?)
5.) Finally: Compute the CEF for Y|x1 (discretize x1 as you see fit).
    a. Draw and report.
    b. b. When you compare points 2, 3 and 4, what do you observe?

P.S. Please try to set a "seed" at 420711 (cf. googling):