

Ph.D. Econometrics II  
Heinz School, Carnegie Mellon University  
90-907, Fall 2000

Final

Instructions You may use any books, notes, calculators, and other aids you like. You may not converse, nor may you cooperate.

Please complete all questions.

Each question is worth 33 points (+1 free).

Please show all relevant work.

Please interpret your results in plain English.

1. Consider the classical bivariate regression model:

$$Y = \beta_1 + \beta_2 X + \epsilon \quad (1)$$

But you try to estimate this equation:

$$X = \alpha_1 + \alpha_2 Y + \epsilon \quad (2)$$

- (a) What happens if you run OLS on equation 2?
- (b) Suppose you have a good instrument for  $X$  ( $X$ , not a typo). What happens if you run IV on equation 2 using this good instrument?

2. Consider the attached output from the hospital cost dataset we have used previously. Recall, we are estimating the equation:

$$\ln C_{it} = \beta_1 + \beta_2 \ln v_{it} + \beta_3 \ln d_{it} + \epsilon_{it} \quad (3)$$

$C$  is cost.  $v$  is visits.  $d$  is days.

Estimates of this equation appear in the output. For a hospital with  $\epsilon = 0$  and days = 1 and visits at its sample mean, please construct an estimate and 95% confidence interval for the marginal cost of a visit (ie  $\frac{\partial C}{\partial v}$ ) — again no typo the  $C$  is not logged (who cares what the marginal log costs of a visit are).

3. Consider a wage equation like:

$$W = \dots + \beta \text{COL} + \dots + \epsilon \quad (4)$$

$W$  is wages, and COL is a dummy for attended college.

You run OLS, but someone (righteously) objects that, since IQ does not appear in the equation,  $\text{Cov}(\text{COL}, \epsilon) > 0$ .

We know of three ways to deal with this problem.

- Selection Model
- Fixed Effects
- Instrumental Variables

Tell me how each one fixes the problem. Notice, for the selection model, I am asking you to go beyond what we have learned in the course (we don't know a model which applies directly), but it is a pretty clear how to build the model.

Tell me the pros and cons of each approach and make sure you mention data requirements and the assumptions you must make for each one.