Homework #3 – Answers
Economics 113
Introduction to Econometrics
Professor Spearot
Due Wednesday, October 29th, 2008 – Beginning of class

1. Please interpret the slope coefficient in each of the following four specifications:

$$y = \beta_0 + \beta_1 x + \varepsilon$$
$$y = \beta_0 + \beta_1 \log(x) + \varepsilon$$
$$\log(y) = \beta_0 + \beta_1 x + \varepsilon$$
$$\log(y) = \beta_0 + \beta_1 \log(x) + \varepsilon$$

For the first specification:

$$\frac{\partial y}{\partial x} = \beta_1$$

Thus, the slope coefficient equals the unit change in *y* after a unit change in <u>*x*</u>.

$$\partial y = \beta_1 \frac{\partial x}{x} = \left(\frac{\beta_1}{100}\right)\left(\frac{\partial x}{x} * 100\right) = \left(\frac{\beta_1}{100}\right)(\text{percentage change in x})$$

Thus, the slope coefficient divided by 100 equals the unit change in *y* after a percentage change in <u>*x*</u>.

$$\frac{\partial y}{y} = \beta_1 \partial x$$
$$\frac{\partial y}{y} * 100 = (\beta_1 * 100)\partial x$$

Thus, the slope coefficient multiplied by 100 equals the percentage change in *y* resulting from a unit change in *x*.

$$\frac{\partial y}{y} = \beta_1 \frac{\partial x}{x}$$
$$\frac{\partial y}{y} * 100 = \beta_1 \left(\frac{\partial x}{x} * 100\right)$$

The slope coefficient gives the percentage change in *y* as a function of a percentage change in *x*.

2. For the following examples, discuss whether each satisfy the four assumptions we use for linear regression. If not, which assumptions are violated?

a) To examine the link between attendance and grades, I construct an indicator variable, *PickUp*, which takes on the value of 1 if a student picked-up his or her exam on Friday, October 17th, and 0 otherwise. I then run the following regression using data for the *entire class*:

$$ExamGrade = \beta_0 + \beta_1 PickUp + \varepsilon$$

Here the problem is one of omitted variables. The indicator PickUp is likely correlated with a variable in the error term. For example, since studying probably improves your grade, and studying is likely correlated with your probability of attending class, there is bias in our estimates. In this case the assumption of zero conditional mean is violated.

b) To examine the link between studying and grades, I construct a new variable, *Study*, which is the self-reported hours-studied prior to the exam. The sampling was done on Friday, October 17th. Using this sample, I run the following regression:

$$ExamGrade = \beta_0 + \beta_1 Study + \varepsilon$$

You could make the same argument as in (a), instead using average attendance as the omitted variable. However, the bigger issue is that we are using a selected sample, so the random sampling assumption is violated. That is, we are only using the sample of those who attend class, which is likely not representative of the entire class.

3. Taken from Problem 3.7 in Wooldridge
Which of the following can cause OLS estimators to be biased?

(i) Heteroskedasticity
(ii) Omitting an important variable
(iii) A high correlation coefficient (say .95) between two random variables

Only number two causes the estimators to be biased. Heteroskedasticity has to do with the estimates of error variance, which has nothing to do with the expected value of the parameter estimates. Further, high correlations only pose a problem for obtaining a precise estimate of variance.