**ECO 5720**

**Problem Set 2**

1. In a study relating college grade point average to time spent in various activities, you distribute a survey to several students. The students are asked how many hours they spend each week in four activities: studying, sleeping, working, and leisure. Any activity is put into one of the four categories, so that for each student, the sum of hours in the four activities must be 168. In the model

does it make sense to hold *sleep*, *work*, and *leisure* fixed, while changing study?

2. Which of the following can cause the OLS estimator, i.e., , to be biased?

(i) Heteroskedasticity.

(ii) Omitting an important variable.

(iii) A sample correlation coefficient of .95 between two independent variables both included in the model.

3. The following equation describes the median housing price in a community in terms of amount of pollution (*nox* for nitrous oxide) and the average number of rooms in houses in the community (*rooms*):

(i) What is the interpretation of ? Explain in terms of percentage changes in *nox* and *price*.

(ii) Does the simple regression of log(*price*) on log(*nox*) produce an unbiased estimator of ? Explain in terms of omitted variables bias.

4. Use the data in HPRICE1 to estimate the model

where *price* is the price of a house in thousands of dollars; *sqrft* represents the size of a house in square feet; *bdrms* denotes the number of bedrooms.

(i) Write out the results in an equation form. While it is sufficient to report the estimates, you may also paste the Stata results.

(ii) What is the estimated increase in price for a house with one more bedroom, holding square footage constant?

(iii) What percentage of the variation in price is explained by square footage and number of bedrooms?

(iv) The first house in the sample has *sqrft* = 2,438 and *bdrms* = 4. Find the predicted selling price for this house from the OLS regression line.

(v) The actual selling price of the first house in the sample was $300,000 (so *price* = 300). Find the residual for this house.

5. Continue to use the data in NBASAL to estimate the model

Here, *wage* denotes annual salary in thousands of dollars; *points*, *rebounds*, and *assists* represent points, rebounds, and assists per game, respectively.

(i) What is the estimated value of ?

(ii) Next, estimate the model

and save the residuals, .

Finally, estimate the model

What is the estimated value of ? How does it compare to the value of estimated in (i)?

6. Simulate a data set from the following model

Estimate the model by OLS.

(i) For 1,000 repetitions and 900 observations in each repetition, graph the empirical distribution of . Please attach your Stata do file and graph.

(ii) Is the OLS estimator of unbiased? Explain briefly.