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

$$GPA = \beta_0 + \beta_1 study + \beta_2 sleep + \beta_3 work + \beta_4 leisure + u,$$

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

Answer: No, since *study*, *sleep*, *work*, and *leisure* are perfectly collinear. Moreover, *study* + *sleep* + *work* + *leisure* = 168. We cannot change *study*, without changing at least one of the other categories.

2. Which of the following can cause the OLS estimator, i.e., $\hat{\beta}$, 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.

Answer: An omitted variable that affects the dependent variable and is correlated with the included explanatory variables can cause bias. The homoskedasticity assumption, plays no role in unbiasedness of the OLS estimators. Further, high (but not perfect) collinearity between the explanatory variables in the sample does not affect the assumptions needed for unbiasedness. However, it does inflate the corresponding standard errors. Thus, our answer is only (ii).

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*):

$$\log(price) = \beta_0 + \beta_1 \log(nox) + \beta_2 rooms + u.$$

(i) What is the interpretation of β_1? Explain in terms of percentage changes in *nox* and *price*.

Answer: β_1 is the elasticity of *price* with respect to *nox*. Thus, a 1% increase in *nox* is associated with a β_1% change in *price*.

(ii) Does the simple regression of log(*price*) on log(*nox*) produce an unbiased estimator of β_1? Explain in terms of omitted variables bias.

Answer: This is unlikely due to omitted variables bias. For example, quality of houses could be such an unobserved characteristic. Better quality houses may have more rooms and be located in neighborhoods with less pollution.

4. Use the data in HPRICE1 to estimate the model

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u,$$

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 $\hat{\beta}$ estimates, you may also paste the Stata results.

Answer: Estimated equation:

$$\widehat{price} = -19.32 + 0.128sqrft + 15.198bdrms.$$

```
. reg price sqrft bdrms

      Source |       SS           df       MS      Number of obs   =        88
-------------+----------------------------------   F(2, 85)        =     72.96
       Model |  580009.152         2  290004.576   Prob > F        =    0.0000
    Residual |  337845.354        85  3974.65122   R-squared       =    0.6319
-------------+----------------------------------   Adj R-squared   =    0.6233
       Total |  917854.506        87  10550.0518   Root MSE        =    63.045

------------------------------------------------------------------------------
       price | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       sqrft |   .1284362   .0138245     9.29   0.000     .1009495    .1559229
       bdrms |   15.19819   9.483517     1.60   0.113    -3.657582    34.05396
       _cons |    -19.315   31.04662    -0.62   0.536    -81.04399      42.414
------------------------------------------------------------------------------
```

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

Answer: Holding square footage constant, price increases by 15.198, i.e., $15,198.

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

Answer: About 63.2%.

(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.

Answer: The predicted price is –19.315 + 0.128(2,438) + 15.198(4) = 353.541, or $353,541.

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

Answer: From part (iv), the estimated value of the home based only on square footage and number of bedrooms is $353,541. The actual selling price was $300,000 and the residual is -53541.

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

$$wage = \beta_0 + \beta_1 points + \beta_2 rebounds + \beta_3 assists + u.$$

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 β_3?

Answer: : $\hat{\beta}_3 = 24.347.$

```
. reg wage points rebounds assists

      Source |       SS           df       MS      Number of obs   =       269
-------------+----------------------------------   F(3, 265)       =     80.07
       Model |   127366839         3  42455612.8   Prob > F        =    0.0000
    Residual |   140512078       265  530234.258   R-squared       =    0.4755
-------------+----------------------------------   Adj R-squared   =    0.4695
       Total |   267878917       268  999548.197   Root MSE        =    728.17

------------------------------------------------------------------------------
        wage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      points |   81.19369   11.56929     7.02   0.000     58.41426    103.9731
    rebounds |   92.23602     19.911     4.63   0.000     53.03213    131.4399
     assists |   24.34695   26.98747     0.90   0.368    -28.79021     77.4841
       _cons |   130.2154   96.50168     1.35   0.178    -59.79217     320.223
------------------------------------------------------------------------------
```

(ii) Next, estimate the model

$$assists = \delta_0 + \delta_1 points + \delta_2 rebounds + v,$$

and save the residuals, $\hat{v}$.

Answer:

```
. reg assists points rebounds

      Source |       SS           df       MS      Number of obs   =       269
-------------+----------------------------------   F(2, 266)       =     81.47
       Model |  445.978613         2  222.989306   Prob > F        =    0.0000
    Residual |  728.019988       266  2.73691725   R-squared       =    0.3799
-------------+----------------------------------   Adj R-squared   =    0.3752
       Total |   1173.9986       268  4.38059179   Root MSE        =    1.6544

------------------------------------------------------------------------------
     assists | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      points |   .2635213   .0207322    12.71   0.000     .2227013    .3043413
    rebounds |  -.2618239   .0422923    -6.19   0.000    -.3450942   -.1785537
       _cons |    .870579   .2126489     4.09   0.000     .4518898    1.289268
------------------------------------------------------------------------------

. predict vhat, res
```

Finally, estimate the model

$$wage = \alpha_0 + \alpha_1 \hat{v} + \varepsilon.$$

What is the estimated value of α_1? How does it compare to the value of β_3 estimated in (i)?

Answer:

```
. reg wage vhat

      Source |       SS           df       MS      Number of obs   =       269
-------------+----------------------------------   F(1, 267)       =      0.43
       Model |  431551.213         1  431551.213   Prob > F        =    0.5121
    Residual |   267447366       267  1001675.53   R-squared       =    0.0016
-------------+----------------------------------   Adj R-squared   =   -0.0021
       Total |   267878917       268  999548.197   Root MSE        =    1000.8

------------------------------------------------------------------------------
        wage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        vhat |   24.34695   37.09299     0.66   0.512    -48.68502    97.37892
       _cons |   1423.828   61.02213    23.33   0.000     1303.682    1543.973
------------------------------------------------------------------------------
```

Here, $\hat{\alpha}_1 = \hat{\beta}_3 = 24.347$.

6. Simulate a data set from the following model

$$y = 1 + 0.7x_1 + 0.5x_2 + u$$
$$x_1 \sim N(0,1)$$
$$x_2 \sim N(0,1)$$
$$u \sim N(0,1)$$
$$\text{Corr}(x_1, u) = 0.4$$
$$\text{Corr}(x_2, u) = 0.2$$
$$\text{Corr}(x_1, x_2) = -0.3.$$

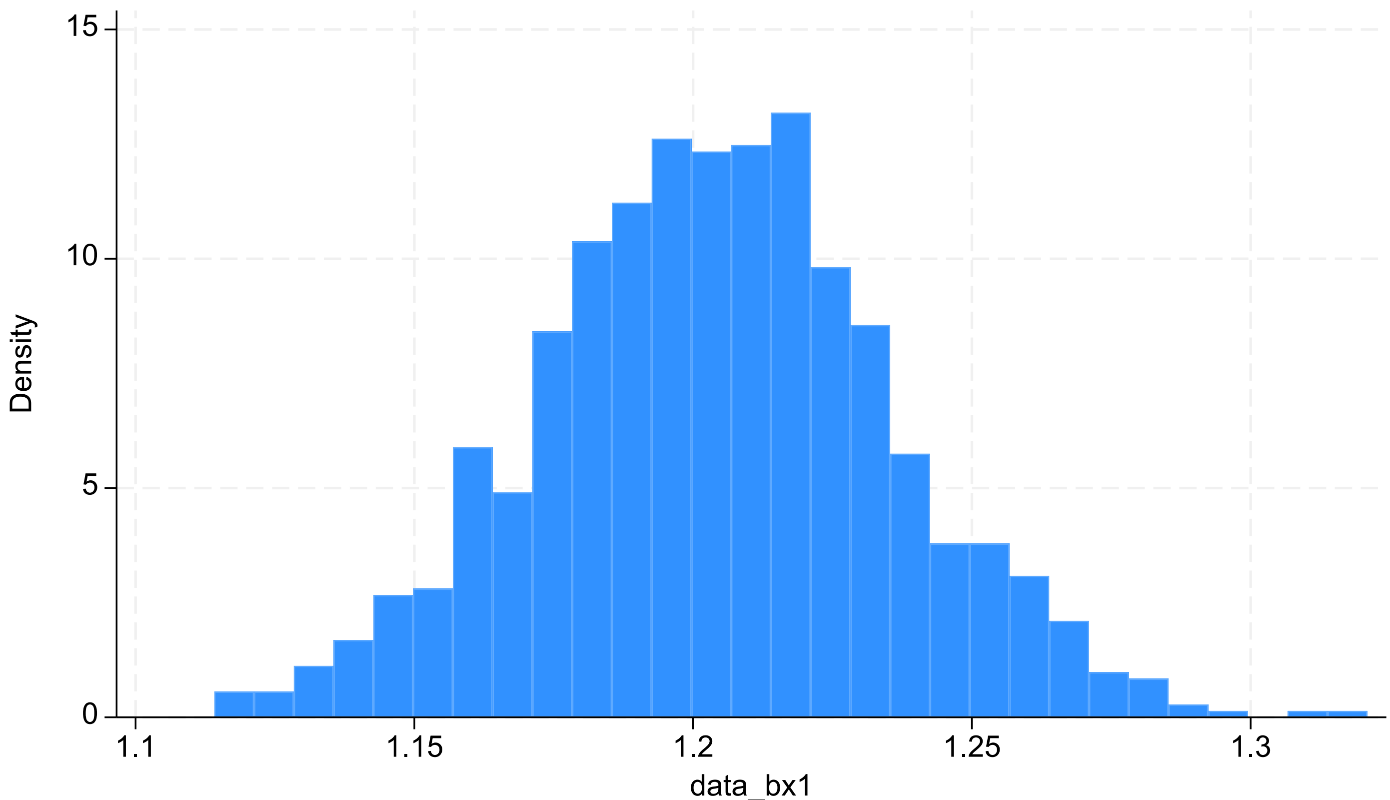
Estimate the model by OLS.

(i) For 1,000 repetitions and 900 observations in each repetition, graph the empirical distribution of $\widehat{\beta_1}$. Please attach your Stata do file and graph.

Answer:

```
**********************************************************************************************
*** Example: Simulation - 1000 reps; n=900; corr(x1,u) = 0.4; corr(x2,u) = 0.2; corr (x1,x2) = -0.3; bias ***
**********************************************************************************************

clear

* Generating 2 variables with missing values to store estimated values of beta hat from the simulation *
set obs 1000
g data_bx1=.
g data_bx2=.

* Simulating data based on correlation values and storing estimates of beta hat *
forval i=1/1000 {

	preserve
	clear
	set obs 900
	matrix C = (1, -0.3, 0.4 \ -0.3, 1, 0.2 \ 0.4, 0.2, 1)
	drawnorm x1 x2 u, corr(C)
```

```
        g y = 1 + 0.7*x1 + 0.5*x2 + u
        reg y x1 x2
        local x1coef =_b[x1]
        local x2coef =_b[x2]
        restore

        replace data_bx1=`x1coef' in `i'
        replace data_bx2=`x2coef' in `i'

}

* Summary and histogram of beta hat values *
su data_bx1 data_bx2
hist data_bx1
hist data_bx2
```



(ii) Is the OLS estimator of $\widehat{\beta_1}$ unbiased? Explain briefly.

Answer: No, since the explanatory variables and the unobserved factors are correlated.