PROBLEM SET 2

PROBLEMS DUE FRIDAY 2/4

Do problems 8 and 9 from Greene's chapter 4, as well as the following problems:

Problem 1:

The CNLR model applies to

$$E\left(Y|X_1, X_2\right) = X_1\beta_1 + X_2\beta_2$$

A sample of size $n = 102$ gives

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}; (X'X) = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}; \hat{\varepsilon}'\hat{\varepsilon} = 80$$

Let $\theta = \beta_1 - \beta_2$. Test at the 5% significance level the null hypothesis that $\theta = 1$.

Problem 2:

The CNLR model applies to

$$E\left(Y|X_1, X_2\right) = X_1\beta_1 + X_2\beta_2$$

with $\sigma^2 = V\left(Y_i|X_1, X_2\right) = 2$ and

$$X'X = \begin{pmatrix} 5 & 2 \\ 2 & 4 \end{pmatrix}$$

The sample produces

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

(a) Construct a 95% confidence interval for $\theta = \beta_1 + \beta_2$.
(b) Construct a 90% confidence region for the pair $\left(\beta_1, \beta_2\right)$.

Problem 3:
A multiple regression of y on a constant, x_1 and x_2 produces the following results:

$$\hat{y} = 4 + 0.4x_1 + 0.9x_2, \quad R^2 = 8/60, \quad \hat{\varepsilon}'\hat{\varepsilon} = 520, \quad n = 29, \quad X'X = \begin{bmatrix} 29 & 0 & 0 \\ 0 & 50 & 10 \\ 0 & 10 & 80 \end{bmatrix}$$

(a) Test the hypothesis that the two slopes sum to 1 at the 5% significance level under normality of y.

(b) Test the hypothesis at the 5% significance level that the slope on x_1 is 0 by running the restricted regression and comparing the two sums of squared deviations. As in part (a), carry the test assuming normality of y.

Problem 4:
Production data for 22 firms in a certain industry produce the following, where $y = ln(\text{output})$ and $x = ln(\text{labor hours input})$:

$$\overline{y} = 20,\ \overline{x} = 10,\ \sum_{i=1}^{n} (y_i - \overline{y})^2 = 100,\ \sum_{i=1}^{n} (x_i - \overline{x})^2 = 60,\ \sum_{i=1}^{n} (y_i - \overline{y})(x_i - \overline{x}) = 30$$

(a) Write down the model in matrix notation and state the assumptions that justify running OLS to compute the unknown coefficients. Form the $(X'X)$, $(X'X)^{-1}$ and $(X'Y)$ matrices and compute the least squares estimator of $\beta = (\beta_1, \beta_2)'$.
(b) Test the statistical significance of your estimates at the 5% significance level assuming that the y_i's are jointly normally distributed with variance-covariance matrix the identity matrix.
(c) Test the hypothesis that there exist constant returns to labor at the 5% significance level under the same assumption as in part (b).

Problem 5:
This problem uses the data set CPS85 that may be downloaded from the class web site, along with the file ReadCPS that describes the data. The data set is a random sample from the May 1985 Current Population Survey conducted by the U.S. Census Bureau. It contains observations on 12 variables for 534 individuals. The first variable in the data set is years of schooling (EDU), and the next six entries are 0-1 dummy variables taking on the value 1 if the individual resides in the south ($SOUTH$), is non-white and non-Hispanic ($NONWH$), is Hispanic ($HISP$), is female (FE), is married (MAR) and is female and married ($MARRFE$). The next two variables measure potential years of experience (EX), computed as age minus years of schooling minus 6, and this potential experience measure squared ($EXSQ$). The next entry is a dummy variable taking on the value 1 if the individual works at a union job ($UNIO$). The next column is the natural logarithm of the individuals average hourly in dollars earnings ($LNWAGE$). The next variable is the individual's age in years (AGE). The rest of the variables in the data set will not be used in this problem. Whenever necessary in the questions below, assume that the assumptions of the Classical Normal Regression model hold.
(a) Compute the average hourly wage for the entire sample. There are two ways of doing this. Either compute the arithmetic mean of $LNWAGE$ and exponentiate it (which gives the geometric mean of average hourly wage, $WAGE = e^{LNWAGE}$), or exponentiate $LNWAGE$ for each individual and then compute the arithmetic mean. Are these two measures identical?
(b) Compute the sample means of the following dummy variables: $SOUTH$, FE, $UNIO$, $NONWH$, $HISP$. How many males, females, south residents, non-south residents, union workers, non-union workers, whites, Hispanics, non-whites & non-Hispanics are in the sample?
(c) Compute the means and standard deviations of $LNWAGE$, EDU, and EX for the entire sample, and then by gender (male/female), by race (white/nonwhite/Hispanic), and by union status (union/non-union). Within each of the three groups sorted by gender, race, and union status, find which subgroup has the highest average $LNWAGE$ and the highest dispersion as measured by the standard deviation. Do the same for EDU.

(d) Using Least Squares, estimate the parameters in a simple model where $LNWAGE$ is regressed on a constant, years of schooling (EDU), and experience (EX), and report these along with their estimated standard errors. What do the slope coefficients on EDU and EX measure? Intuitively, age may affect wages as well. What would be the problem of including AGE as an additional regressor?
(e) Compute and interpret the R^2 coefficient for this model.
(f) According to the human capital theory, experience affects wages. Test this hypothesis at the 5% significance level.
(g) Because human capital depreciates with age (and hence with the measure of experience we are using), we expect decreasing returns to experience. To see whether this assumption is correct, estimate a linear model with an additional quadratic in experience variable:

$$LNWAGE = \beta_0 + \beta_1 EDU + \beta_2 EX + \beta_3 EXSQ + \varepsilon$$

Is the sign of β_3 consistent with what you would expect? Is β_3 statistically significant? At what level of experience is $LNWAGE$ maximized? What happens to the estimated coefficients of EDU and EX and their standard errors?
(h) The specification in part (g) assumes that the intercept and slope coefficient on EDU are the same for all individuals. We may think that the effect of schooling on wages differs by a *constant factor of proportionality* for males and females, i.e. that for males,

$$WAGE = \alpha_M e^{\beta_1 EDU + \beta_2 EX + \beta_3 EXSQ} e^{\varepsilon}$$

while for females,

$$WAGE = \alpha_F e^{\beta_1 ED + \beta_2 EX + \beta_3 EXSQ} e^{\varepsilon}$$

where α_M and α_F are differing factors of proportionality, β is the common return to schooling and ε is a random error term. Show that this implies that when the dependent variable is $LNWAGE$ rather than $WAGE$, males and females have different intercept terms but common slope coefficients $\beta_1, .\beta_2 \beta_3$ To estimate these different intercepts, run the following regression

$$LNWAGE = \alpha_1 + \alpha_2 FE + \beta_1 EDU + \beta_2 EX + \beta_3 EXSQ + \varepsilon$$

Interpret the estimates of α_1 and α_2 relating them to α_M and α_F above. Formulate and test the hypothesis (at a 5% significance level) that there is no gender discrimination using your estimates of α_1 and α_2.
(i) An alternative procedure for formulating the regression relationship in the previous question is as follows: First create a dummy variable called MA, defined as $MA = 1 - FE$. (What does this variable denote?) Then estimate the following model

$$LNWAGE = \gamma_1 MA + \gamma_2 FE + \beta_1 EDU + \beta_2 EX + \beta_3 EXSQ + \varepsilon$$

Interpret γ_1 and γ_2 and relate them to α_M and α_F above. According to econometric theory, what should be the relationship among the estimates α_1 and γ_1 and α_2 and γ_2? Are your estimates consistent with this relationship? Formulate a test for the null hypothesis that there is no gender discrimination using the estimates of γ_1 and γ_2. What would be the problem of including an intercept term in the model?

(j) The specifications in parts (h) and (i) above take the returns to education to be constant for males and females. How would you test whether this assumption is correct? Perform the test.
(k) Consider the specification

$$LNWAGE = \beta_0 + \beta_1 EDU + \beta_2 EX + \beta_3 EXSQ + \varepsilon$$

Does the return to experience differ by years of schooling? How would you allow that the return to experience differs by education level? Test whether the return to experience differs by education level.
(l) Describe how you would test whether there is a wage premium for union jobs.