**Econometrics I**

**Fall 2006**

**Assignment 3**

*Today's Date: September 19*
*Due Date: September 26*

*Please show all of your work and clearly indicate your final response to each question. You are to write your own GAUSS (or MATLAB, Fortran, etc.) program in answering each of the computational questions.*

In questions 1-8, you are to consider estimation of models of the general form

$$y = X\beta + \varepsilon,$$

where $X$ is a $(N \times k)$ deterministic matrix, $\beta$ is an unknown $(k \times 1)$ vector and $\varepsilon$ is a $(n \times 1)$ vector of disturbance terms with the property $E(\varepsilon|X) = 0$ and $E(\varepsilon\varepsilon'|X) = \sigma_\varepsilon^2 I_N$, where $\sigma_\varepsilon^2$ is an unknown population parameter. Some questions will ask you to consider estimation issues when certain of these conditions are modified. When this is the case the modification will be clearly specified.

1. Using the data set described in the codebook below, fit a regression model in which the dependent variable is the natural logarithm of the wage rate, $\ln(w_i)$. Using regressors consisting of a column vector of 1's and the variables appearing in columns 2, 3, 4, and 5, compute $\hat{\beta}$, the covariance matrix of $\hat{\beta}$, and $R^2$. Provide a substantive interpretation of your estimates.

2. Assume that $\varepsilon$ is multivariate normal. Perform an appropriate test for the hypothesis that the coefficient associated with male is equal to 0. Interpret the results of the test.

3. Is it possible to estimate a specification (with $\ln(w)$ as the dependent variable) in which the $X$ matrix consists of a column vector of 1's and the variables in columns 4, 5, and 6? Why or why not?

4. Define a new variable

$$school = 10 * Less\_HS + 12 * HS\_dip + 14 * More\_HS.$$

Regress ln wage on a constant, age, and school. Assuming that $\varepsilon$ is multivariate normal, perform a test of the null hypothesis that the coefficient associated with age is equal to the coefficient associated with school. Interpret the results of your test.

5. If you knew that the value of $\sigma_\varepsilon^2 = .42$, would that modify your procedure for testing the null hypothesis in the previous question? If so, perform the appropriate test given this information.

6. Define the new indicator variable *hwage* which takes the value 1 when the wage is greater than 15. Regress hwage on the covariates listed in Question 1. Describe how you would test the null hypothesis that the coefficient associated with male is equal to 0. (Do not assume normally distributed disturbances).

7. 4.10

8. 4.11

**Codebook**
**Dataset: Ecn1_wag**

| *Column* | *Variable* | *Values* |
|---|---|---|
| 1 | Hourly Wage Rate | Dollars/Hour |
| 2 | Age | Years |
| 3 | Male | 1 if male<br>0 if female |
| 4 | Less_HS | 1 if less than HS graduate<br>0 if not |
| 5 | HS_Dip | 1 if HS graduate (exactly)<br>0 if not |
| 6 | More_HS | 1 if more than HS graduate<br>0 if not |