Econ 640 - Problem Set 2

Instructions: *Please submit all answers as a physical copy in a separate document. For Stata output, ensure that your **do-files and results** are printed without any error messages. In essence, I should be able to clearly understand what you did (Please do not handwrite your code!). Full credit will only be awarded to answers that demonstrate clear reasoning and a solid understanding of the problem. This problem set is due at the **beginning of class on October 9th**.*

1. For a linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

   (a) Derive the OLS estimators for β_0 and β_1.

   (b) What assumption(s) is(are) necessary to prove that the OLS estimator $\hat{\beta}_1^{OLS}$ is unbiased?

   (c) Derive $\hat{\beta}_1^{OLS}$ again two alternative methods:

      1. Using the projection matrix approach.
      2. Using the method of moments approach.

   (d) Derive the variance of $\hat{\beta}_1^{OLS}$

   (e) What assumption(s) is(are) necessary to prove that the OLS estimator $\hat{\beta}_1^{OLS}$ is consistent?

   (f) What assumption(s) is(are) necessary to prove that the OLS estimator $\hat{\beta}_1^{OLS}$ is BLUE (best linear unbiased estimator)?

   (g) Having derived the expression of the variance of $\hat{\beta}_1^{OLS}$, how can you estimate it using residual? Propose an estimator for $Var(\hat{\beta}_1^{OLS} \mid X)$

   (h) Suppose someone proposed the following simplified model:

$$y_i = \beta x_i + u_i$$

   Does this model require more or fewer assumptions?

   (i) Derive a least squares estimator for $\hat{\beta}$

   (j) Is this estimator $\hat{\beta}$ unbiased? What assumptions do you need?

2. Empirical Task: Linear Regression. Use the following dataset:
   `use http://fmwww.bc.edu/ec-p/data/wooldridge/wage1, clear`

   (a) run a naive OLS regression with outcome = wage; and independent variable = education. Interpret the coefficient's magnitude and significance level. Then, explain the information provided in Stata regarding standard error and F-statistic.

   (b) Next, calculate the natural logarithm of wage and re-estimate the regression using log(wage) as the dependent variable. Interpret the results and compare them to the previous model. If the interpretations differ substantially, what concerns might arise?

   (c) Next, expand your control variable set by sequentially including female, experience, tenure, non-white, and married. This will result in five additional models alongside the main model. How does the coefficient on education change in each model? Interpret the results. Use `esttab` command to present all models in a **single** table (so you can compare coefficients across models), reporting standard errors with significance levels of 0.1, 0.05, and 0.01. Based on the table, write a paragraph discussing the type of bias each control variable reduces and whether the results make sense.

   (d) Finally, re-estimate all models using robust standard errors (hint: `reg y x, robust`), create a new table, and compare the standard errors to those in the previous table. Identify which set of standard errors is larger and explain the potential reasons for the difference.