Introductory Econometrics

Tutorial 9

<u>PART A:</u> To be done before you attend the tutorial. The solutions will be made available at the end of the week.

1. *A bit of practice with dummy variables and heteroskedasticity*: Consider the following model for the logarithm of wage given years of university education and gender of person i :

$$\log(wage_i) = \beta_0 + \delta_0 female_i + \beta_1 totuni_i + \delta_1 female_i \times totuni_i + u_i \tag{1}$$

where log is the natural logarithm, $female_i$ is a dummy variable that is equal to 1 if person i is female is 0 otherwise, and $totuni_i$ is the number of years of university education that person i has. Based on a random sample of 6763 individuals, we have estimated this model using OLS and obtained the following estimated equation:

$$\begin{aligned} \widehat{\log(wage_i)} &= \underset{(0.011)}{3.289} - \underset{(0.015)}{0.360} female_i + \underset{(0.003)}{0.050} totuni_i + \underset{(0.005)}{0.030} female_i \times totuni_i, \\ i &= 1, 2, \ldots, 6763, \; R^2 = 0.202. \end{aligned}$$

   (a) Explain how you would test the hypothesis that the conditional expectation of $\log(wage)$ conditional on years of university education is exactly the same for men and women. You need to specify the null and the alternative, the test statistic and its distribution under the null, the regression that you should run so that you can compute the test statistic, and the rule for rejection or non-rejection of the null hypothesis.

   (b) Explain what insights the estimated model reported above provide about the conditional expectation of $\log(wage)$ given years of university education for men and women.

   (c) Using the estimated equation, find the value of $totuni$ (total number of years in university) such that the predicted values of $\log(wage)$ are the same for men and women. Can women realistically get enough years of university education so that their earnings catch up to those of men? Explain.

   (d) Suppose we suspect that the variation of $\log(wage)$ around its conditional mean is larger for women than it is for men. That means that in the population regression model, we believe that $Var(u \mid totuni, female = 1) > Var(u \mid totuni, female = 0)$. Explain the consequences of this for the OLS estimator and for the t-tests based on OLS estimates and their OLS standard errors.

   (e) Suppose we are told that the variation of $\log(wage)$ around its conditional mean is twice as large as that for men, i.e. if $Var(u \mid totuni, female = 0) = \sigma^2$, then $Var(u \mid totuni, female = 1) = 2\sigma^2$. Verify that multiplying equation (1) by $\frac{1}{\sqrt{female_i + 1}}$ will lead to an equation with the same unknown parameters as equation (1) but with homoskedastic errors. What is the significance of this?

2. The workfile profits.wf1 includes data on profits and assets of 88 firms (some firms having missing data). The variables in the data set are $profits$ and $assets$, which are each firm's profit and assets in million dollars, and mno, which is a dummy variable which is equal to 1 if the CEO of the firm is not the owner of the firm, and is zero otherwise.

   (a) Get the scatterplot of $profits$ against $assets$. What does the scatterplot suggest about the sign of the slope coefficient in the regression of $profits$ on a constant and $assets$?

   (b) What does the scatterplot suggest about the properties of the OLS estimators of the intercept and slope parameters in the regression of $profits$ on a constant and $assets$?

Do not forget to bring your answers to PART A and a copy of the tutorial questions to your tutorial.

<u>Part B:</u> **This part will be covered in the tutorial. It is still a good idea to attempt these questions before the tutorial.**

1. (*This is a continuation of question 2 in Part A*) The workfile profits.wf1 includes data on profits and assets of 88 firms (some firms having missing data). The variables in the data set are $profits$ and $assets$, which are each firm's profit and assets in million dollars, and mno, which is a dummy variable which is equal to 1 if the CEO of the firm is not the owner of the firm, and is zero otherwise. Our objective is to test the hypothesis that the relationship between profits and assets is the same for owner-managed and non-owner-managed firms. The general model is:

$$profits_i = \beta_0 + \delta_0 mno_i + \beta_1 assets_i + \delta_1 \left(mno_i \times assets_i\right) + u_i \tag{2}$$

   (a) Formulate the null hypothesis that the nature of ownership of the firm (i.e. whether the firm is managed by its owner or not) does not effect the relationship between profits and assets in a firm and the alternative that it does.

   (b) Estimate the model using OLS and test whether the errors are homoskedastic using the following tests (in each case, answer the question as if this was a question on the final exam, i.e. write down the null and the alternative, the test statistic and its distribution under the null, the auxiliary regression you should estimate to compute the test statistic, then compute the test statistic and compare it with the critical value you get from statistical tables):

      i. Breusch-Pagan test when the alternative hypothesis is $Var\left(u_i \mid mno_i, assets_i\right) = \alpha_0 + \alpha_1 mno_i + \alpha_2 assets_i$

      ii. White test

      iii. The special form of the White test that uses the predicted value of $profits$ and its square as predictors of variance.

   (c) Is it likely that a log-log formulation that uses $\log\left(profits\right)$ and $\log\left(assets\right)$ would solve the heteroskedasticity problem in this application? Explain.

   (d) In each of the following scenarios, determine the appropriate weight that solves the problem of heteroskedasticity when it multiplies both sides of equation (2):

      i. $Var\left(u_i \mid mno_i, assets_i\right) = \sigma^2 \times assets_i$

      ii. $Var\left(u_i \mid mno_i, assets_i\right) = \sigma^2 \times assets_i^2$

      iii. $Var\left(u_i \mid mno_i, assets_i\right) = \sigma^2 \log(assets_i)$

   (e) Suppose we know $Var\left(u_i \mid mno_i, assets_i\right) = \sigma^2 \times assets_i$. Test the hypothesis that you formulated in part (a), i.e. that the nature of the ownership of a firm does not effect the relationship between its profits and its assets against the alternative that it does.

2. *This question requires conceptual thinking. If there is no time to cover it in the tutorial, that is fine. You can do it on your own as an exercise in conceptual thinking. The purpose of it is to show how weighted least squares works in a simple situation where you know what the best estimator is. Some may get a light bulb moment when answering this question, others may not.* In order to estimate the population mean of the salary of BCom graduates in their first job after graduation denoted by μ, we have selected a random sample of 10 BCom graduates. They were interviewed by 3 administrators. The first administrator interviewed the first two graduates and reported their wages denoted by w_1 and w_2. The other two administrators interviewed four BCom graduates each, but only reported the average wage of each group of 4, denoted by $wbar_1$ and $wbar_2$. So, we ended up with 4 observations $\{w_1, w_2, wbar_1, wbar_2\}$.

(a) Intuitively, how would you use $\{w_1, w_2, wbar_1, wbar_2\}$ to produce what you would consider to be the best estimate for μ? *(Tutors will manage the discussion and lead you to the desired answer!)*

(b) We arrange $\{w_1, w_2, wbar_1, wbar_2\}$ into a 4×1 vector. Consider the model:

$$\begin{pmatrix} w_1 \\ w_2 \\ wbar_1 \\ wbar_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \beta_0 + \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix}$$

Show that the OLS estimator of β_0 is an unbiased estimator of μ. However, argue that because the variance of the errors of this model are heteroskedastic, the OLS estimator is not BLUE (in particular, its variance is larger than the variance of the estimator you suggested in (a)).

(c) Suggest multiplying this model by appropriate weights to transform it to a model that is homoskedastic. Then compute the OLS estimator of β_0 in this weighted model. Compare it to the estimator you suggested in (a). Hopefully this will show you the wisdom of weighted least squares.