Introductory Econometrics
Tutorial 10

<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) Which Classical Linear Regression Model assumption is violated under the presence of heteroskedasticity? Briefly explain.

i. No perfect multicollinearity

ii. The error term has equal variance

iii. The error term has zero conditional mean

iv. The OLS estimators have equal variance

b) What is the effect of the presence of heteroskedasticity on the properties of OLS estimators? Briefly explain.

i. The OLS estimators are inconsistent.

ii. The usual F statistic no longer has an F distribution.

iii. The OLS estimators are no longer efficient.

2. We wish to study the hypothesis that the effect on trade tax revenue from higher trade volume is positive, while an increase in income would result in the government collecting more direct taxes (e.g. income tax) than rely on trade taxes. For this purpose, we run the following regression of the ratio of trade (import and export) taxes to total government revenue, on the ratio of the sum of exports and imports to GNP (Gross National Product) and GNP per capita, based on cross-sectional data on 41 countries (all in log form):

$$\ln Taxes_i = \beta_0 + \beta_1 \ln Trade_i + \beta_2 \ln GNP_i + u_i. \tag{1}$$

a) Intuitively, given the information provided would you expect heteroskedasticity in the error term?

b) You are asked to apply White's heteroskedasticity test in order to formally verify the existence of heteroskedasticity in this setting. What regression would you run?

c) Assuming that the R^2 of the regression in b) is equal to 0.11148, compute White's heteroskedasticity test statistic. What conclusion do you draw?

d) Given the number of restrictions imposed, what issue might you detect when applying White's heteroskedasticity test?

3. We wish to study the relationship between compensation and employment size using the data in wages.wf1. Compensation is measured as average compensation per employee in \$. Employment size comprises of 9 categories such that 1 (1-4 employees), 2 (5-9 employees), ..., 9 (1000-2499 employees). We are also given, σ_i, the standard deviations of wages.

a) Run the OLS regression of compensation on employment size. Comment on your results.

b) Rerun the regression in a) using weighted least squares. How are your variables of a) transformed? Compare your regression output with that in a).

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

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

This question is based on the work of James Tobin (winner of the Nobel prize in economics in 1981) on family food consumption, which was published in the Journal of the Royal Statistical Society in 1950. It is based on the US 1941 Family Budget Survey data set, which contains data on food consumption expenditure, income and number of people in the family, for some randomly selected families. However, data for each individual family was not reported. Instead, families were grouped according to the number of people in the family (1,2,3,4,5+) and their income range (0-500, 500-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-4000, 4000+), and only average number of people in the family (nf), average income (inc) and average expenditure on food ($food$) for families in each group were reported (a total of 37 observations). The number of families in each group (ng) was also known. In a conference in his honour in 1997, James Tobin said that it took him two to three days in 1949 to run a regression with 3 explanatory variables! The data is in Tobin.wf1.

a. Suppose that food consumption is related to family income and number of people in each family according to the following model.

$$food_i = \beta_0 + \beta_1\, inc_i + \beta_2\, nf_i + u_i \tag{2}$$

Argue that even if the above model is a fair description of each family's demand for food and all Gauss-Markov assumptions are satisfied for this model, we may expect heteroskedastic errors when we estimate the model using group averaged data. Show that it is logical to expect that variance of errors are inversely proportional to the number of families in each group, i.e., $\text{Var}(u_i) = \frac{\sigma^2}{ng_i}$.

b. If errors are heteroskedastic, would OLS be unbiased? Explain.

c. Estimate equation (2) using OLS. Then explore visually if the error variance is likely to be inversely proportional to the number of families in each group (ng), and then formally test the hypothesis of no heteroskedasticity against the alternative that $\text{Var}(u_i) = \frac{\sigma^2}{ng_i}$. (Hint: Breusch-Pagan auxiliary regression would be a regression of $\hat{u}^2$ on a constant and $\frac{1}{ng}$)

d. Suppose we reject homoskedasticity in favor of $\text{Var}(u_i) = \frac{\sigma^2}{ng_i}$. Describe how we should transform the variables and which regression we should run to obtain best linear unbiased estimator for β_0, β_1 and β_2.

e. Based on the weighted least squares estimates test the hypothesis that, other things staying the same, a 1 dollar increase in family income will result in 20 cents increase in family food consumption, against the alternative that it will result in a less than 20 cents increase in food consumption. Perform this test at the 5% level of significance.

Note that you can perform the Breusch-Pagan test by either running the auxiliary regression yourself and calculating the test statistic using the output of the auxiliary regression, or you can use the eviews built in commands. If you need help with this, please refer to the eviews screen shots in the lecture slides. Similarly, you can run the weighted least squares by multiplying each variable by the appropriate weight and running an OLS regression with these weighted variables, or you can use eviews' weighted least squares option as shown in the lecture.