ECON8010, Week 8 Exercise.

1. We want to investigate returns to schooling, but we only have group averages such as mean education and mean salary (let's say at a county level). We somehow could justify the homoskedasticity assumption for the wage equation at the individual level. However, we still face heteroskedastic error for a model at a county level.

(1) Show why we have to worry about the heteroskedasticity for our county level analysis.
(2) We decide to use the weighted least squares estimation to deal with the heteroskedastic error. What additional data do you need to use as the best weights? Explain briefly the rationale how your weighted least squares estimation works.

2. Let's consider the model:
$price = \beta_0 + \beta_1 rooms + \beta_2 sqfeet + u$

(1) You estimate the model of *price* using 88 observations and get the residual, $\hat{u}$. Set up a model of the squared residual (i.e. u(hat)-squared) for a Breusch-Pagan test, and state the null hypothesis for the test.
(2) You estimate the model of $\hat{u}^2$ for Breusch and Pagan's heteroskedasticity test (that you set up in Q2(1)) and have $R^2_{\hat{u}^2} = 0.120$. What is the LM statistic in the case? The critical value of χ^2_2 is 5.99 at the 5% level. In the case, is the error u in the model of *price* heteroskedastic or homoskedastic at the 5% level?
(3) Specify a model of $\hat{u}^2$ for the White's test, and state the null hypothesis for the test.

3. Use the R data ECON8010F10gpa1.RData for this exercise.

(1) Use the OLS to estimate a model relating colGPA to hsGPA, ACT, skipped, and PC. Report the estimation result.
(2) Test for heteroskedasticity based on the BP test and White test. Is there any difference regarding the test results? Discuss.
(3) Obtain the robust standard errors for the same model in Q3(1), and discuss any importance difference regarding the significance of the variables.
(4) Use a feasible GLS estimator to estimate the same model in Q3(1), and discuss any importance difference regarding the significance of the variables.

ECON8010, Week 9 Exercise.

1. Let's suppose that the parameter of interest is the probability that an individual is male for the UK population (let's say, p). However, we have only a sample with n observations from the UK population, and here we estimate $\hat{p}$ using the Maximum Likelihood Estimation (MLE). First, we define X_i as a dummy variable that is 1 if the observation i is male and 0 if the observation i is female.

(1) Define the probability density function that determines the probability of an individual i being male or female given the unknown probability p.

(2) Find the likelihood function L as the joint distribution of $\{X_i\}$ and the log likelihood function for our MLE.

(3) Briefly describe how you would estimate $\hat{p}$ using the MLE once you have the log likelihood function. No need to derive the estimate but give the basic idea of the MLE.

2. You will use the R to practice the following exercises. Find the R data file, ECON8010F12beauty, from Canvas → Modules → R-related. Copy-and-paste both your R commands and outputs from the R Console to MS Word (or other editor then make a pdf file). Add your annotation.

(1) Find the fraction that are classified as having above average looks (based on the variable, *abvavg*=1 if having above average looks) out of all observations. Find the fractions for men and women separately as well.

(2) Test the null hypothesis that the population fractions of above-average-looking women and men are the same at the 5% level.

(3) See 'desc' for the variable definitions, and estimate the following model:

$$married = \beta_0 + \beta_1 belavg + \beta_2 abvavg + \beta_3 bigcity + \beta_4 educ + u. \tag{A}$$

Report the R regression summary output, and interpret $\hat{\beta}_1$ and $\hat{\beta}_2$.

(4) Calculate the Percent Correctly Predicted.

(5) Predict the probability of being married based on your LPM estimation. Check if there are observations who have the predicted probabilities outside [0,1]. Discuss if the OLS estimation for the model (A) can be valid for the sample that we have.

(6) Use the Chow test to test whether we have to specify the model (A) differently for men and women at the 5% level.