Homework 11

Heteroskedasticity (25 points)

(20 points)

Due Date: Monday April 26 at 11:59 pm

Instruction:

- This HW must be done in Rmarkdown!
- Please submit both the .rmd and the Microsoft word files. (Do not submit a PDF or any other image files as the TAs are going to give you feedback in your word document)
- Name your files as: HW11_groupnumber_name
- All the HW assignments are individual work. However, I highly encourage you to discuss it with your group members.
- The answer key will be uploaded on Canvas a couple of days after the due date.

Question 1 Multiple choice:
Which of the following are consequences of heteroskedasticity?

1. The OLS estimators, $\hat{\beta}_j$ are inconsistent.
2. The usual F statistic no longer has an F distribution.
3. The OLS estimators are no longer BLUE.
4. All of the above.

Question 2 Given the following:

$$beer = \beta_0 + \beta_1 inc + \beta_2 price + \beta_3 educ + \beta_4 female + u$$
$$E(u|inc, price, educ, female) = 0$$
$$Var(u|inc, price, educ, female) = \sigma^2 inc^2$$

Write the transformed equation that has a homoskedastic error term. Hint: You need to use WLS.

Question 3 True or False: WLS is preferred to OLS when an important variable has been omitted from the model. Hint: think of unbiasedness of OLS vs WLS estimators!

Question 4 There are different ways to combine features of the Breusch-Pagan and White tests for heteroskedasticity. One possibility not covered in the text is to run the regression

$$\hat{u}_i^2 \text{ on } x_{i1},\ x_{i2}, ..., x_{ik},\ \hat{y}_i^2,\ i = 1, ...,\ n,$$

where the $\hat{u}_i$ are the OLS residuals and the $\hat{y}_i$ are the OLS fitted values. Then, we would test joint significance of $x_{i1},\ x_{i2}, ..., x_{ik}$, and $\hat{y}_i^2$. (Of course, we always include an intercept in this regression.)

1. What are the *df* associated with proposed F test for heteroskedasticity?
2. Explain why the R-squared from the regression above will always be at least as large as the R-squareds for the BP regression and the special case of the White test.
3. Does part 2 imply that the new test always delivers a smaller p-value than either the BP or special case of the White statistic? Explain.

4. Suppose someone suggests also adding $\hat{y}_i$ to the newly proposed test. What do you think of this idea?

Computer Exercises

Question 5 Use the data in SLEEP75 for this exercise.

Given the following:

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + \beta_4 age^2 + \beta_5 yngkid + \beta_6 male + u$$

1. Write down a model that allows the variance of u to differ between men and women. The variance should not depend on other factors.
2. Estimate the parameters of the model for heteroskedasticity. (You have to estimate the *sleep* equation by OLS, first, to obtain the OLS residuals.) Is the estimated variance of u higher for men and for women?
3. Is the variance of u statistically different for men and for women?

Question 6 Use the data in HPRICE1 for this exercise.

Given the following:

$$\widehat{price} = \underset{29.48}{-27.77} + \underset{0.00064}{0.00207} lotsize + \underset{0.013}{0.123} sqrft + \underset{9.01}{13.85} bdrms$$
$$n = 88, \quad R^2 = 0.672$$

and

$$\widehat{log(price)} = \underset{0.65}{-1.30} + \underset{0.038}{0.168} log(lotsize) + \underset{0.093}{0.700} log(sqrft) + \underset{0.028}{0.37} bdrms$$
$$n = 88, \quad R^2 = 0.643$$

1. Obtain the heteroskedasticity-robust standard errors for the first equation. Discuss any important differences with the usual standard errors.
2. Repeat part 1 for the second equation.
3. What does this example suggest about heteroskedasticity and the transformation used for the dependent variable?

Question 7 Use the data in VOTE1 for this exercise.

1. Estimate a model with *voteA* as the dependent variable and *prtystrA, democA,* $\log(expendA)$ and $\log(expendB)$ as independent variables. Compute the Breusch-Pagan test for heteroskedasticity. Use both F and LM statistics and report the p-values.
2. Compute the special case of the White test for heteroskedasticity, again use both F and LM statistics and report the p-values. How strong is the evidence for heteroskedasticity now?

Question 8 Use the data in 401KSUBS for this exercise.

1. Using OSL, estimate a linear probability model for *e401k*, using as explanatory variables *inc,* inc^2*, age,* age^2*,* and *male*. Obtain both the usual OLS standard errors and the heteroskedasticity-robust versions. Are there any important differences?
2. After verifying that the fitted values from part 1 are all between zero and one, obtain the weighted least squares estimates of the linear probability model. Do they differ in important ways from the OLS estimates?