Tutorial 9

Hong Xiang Yue

02/05/2019

Assignment Feedback

Question 1

Part A

- Everybody got the correct answer, everybody lost a few marks though because of incorrect/incomplete working
- Most common mistake is jumping from $Var(h_1 + h_2)$ immediately to $Var(h_1) + Var(h_2)$
- This is only correct if h_1 and h_2 are independent
- The full formula is $Var(h_1 + h_2) = Var(h_1) + Var(h_2) + 2Cov(h_1, h_2)$

Question 1

Part C

- Reasoning was a bit flawed for some of the assumptions
- Model can be thought of as $h_i = \mu + u_i$, where $u_i \sim N(0, \sigma^2)$
- Linear regression with just a constant
- Linear in parameters E.1 satisfied
- Columns of X are linearly independent (there's only one column!) E.2 satisfied

Question 1

Part C

- $E(\mathbf{u}|\mathbf{X}) = E(\mathbf{u})$ b/c no explanatory variables and $E(\mathbf{u}) = 0$, thus $E(\mathbf{u}|\mathbf{X}) = 0$ and E.3 satisfied
- Homoskedasticity because the variance of errors is the same, diagonals of $Var(\mathbf{u}|\mathbf{X})$ are the same
- Random sampling means independence and $Cov(u_i, u_j) = 0, i \neq j$ off diagonals of $Var(\mathbf{u}|\mathbf{X})$ are zero
- Hence $Var(\mathbf{u}|\mathbf{X}) = \sigma^2 \mathbf{I}_n$

Question 2

2.2

- Most marks lost because of lack of exploratory data analysis
- Histograms, summary statistics, scatterplots
- Important to get a feel for the data before doing analysis

Question 2

2.3

- A few groups lost a lot of marks because of misreading the question and estimated the wrong model
- Many groups did not really explain how this model helped us test the existence of the Environmental Kuznets Curve
- Many lost marks because they didn't convert GDPPC to PPP dollars as hinted
- Some groups did not derive the turning point at all
- Many groups did not interpet the turning point sufficiently thoroughly

Question 2

2.4

- A few groups lost a lot of marks because of misreading the question and estimated the wrong model
- Some thought this was supposed to be a time series regression for their particualr country
- A few groups did not explain at all why government prescriptions to change CO2 levels needs to be implemented carefully
- A few groups made mistakes in calculating the change in predicted CO2PC when changing fosspct to the sample median

Part A

Question 1a

$$\ln(wage_i) - \beta_0 + \delta_0 female_i + \beta_1 totuni_i + \delta_1 female_i \times totuni_i + u_i$$

$$\widehat{\ln(wage_i)} = \underset{(0.011)}{3.289} - \underset{(0.015)}{0.360} female_i + \underset{(0.003)}{0.050} totuni_i + \underset{(0.005)}{0.030} female_i \times totuni_i$$

$$n = 6763, R^2 = 0.202$$

Testing the null hypothesis that the conditional expectation of log(wage) for men and women is the same, given education.

$$H_0 : \delta_0 = \delta_1 = 0$$

Question 1b,c

Insights on gender, education and wage

- Women are predicted to have an inital lower wage with no education, but will eventually catch up to men with each additional year of education
- This point is after twelve years of university
- Quite unrealistic

Question 1d

$$Var(u|totuni, female = 1) = 2\sigma^2 > Var(u|totuni, female = 0) = \sigma^2$$

- Assumption E.4 $Var(\mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{I}_n$ is violated
- OLS is no longer BLUE
- Provided E.3 is still met, OLS is still unbiased, but no longer efficient
- In addition, t and F tests no longer valid

Question 1e

Multiply both sides by $w_i = \frac{1}{\sqrt{female_i+1}}$

$$w_i \times \ln(wage_i) = \beta_0 w_i + \delta_0 w_i \times female_i$$

$$+\beta_1 w_i \times totuni_i + \delta_1 w_i \times female_i \times totuni_i + w_i u_i$$

- Transformed equation will still estimate the same coefficients that we want
- But now the variance of the error term is constant for males $Var(u_i) = \sigma^2$ and females $Var(w_i u_i) = \sigma^2$

Part B

Question 1a

$$profits_i = \beta_0 + \delta_0 mno_i + \beta_1 assets_i + \delta_1 mno_i \times assets_i + u_i$$

Suppose we wanted to test the null hypothesis that the nature of ownership does not affect the relationship between profits and assets in a firm, against the alternative it does. What would the null hypothesis be?

https://flux.qa/LDPPHD

Question 1b: Testing for Heteroskedasticity

Estimate the model using OLS and and test for heteroskdasticity in the errors by using:

1. A Breusch-Pagan test where $H_1 : Var(u_i|mno_i, assets_i) = \alpha_0 + \alpha_1 mno_i + \alpha_2 assets_i$. What distribution does $n \times R^2_{\hat{u}}$ have under the null hypothesis?
2. A White test. What is the alternative hypothesis?
3. The special form of the White test which uses the predicted values of *profits* and $profits^2$

https://flux.qa/LDPPHD

Question 1c: log transformation

Would a log transformation of both profits and assets fix our heteroskedasticity problem?

Question 1d: Weighted Least Squares

What should w_i be if the conditional variance of the error takes the form:

1. $Var(u_i|mno_i, assets_i) = \sigma^2 \times assets_i$
2. $Var(u_i|mno_i, assets_i) = \sigma^2 \times assets_i^2$
3. $Var(u_i|mno_i, assets_i) = \sigma^2 \times \ln(assets_i)$

Question 1e

Suppose we know that $Var(u_i|mno_i, assets_i) = \sigma^2 \times assets_i$, correct for the heteroskedasticity using Weighted Least Squares and test the hypothesis from part a, that the nature of ownership a firm has does not affect the relationship between its profits and assets.

Question 2

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 $\{w_1, w_2, wbar_1, wbar_2\}$

Question 2a

In your groups of four to five, discuss what would be the best estimator for μ? Assume that the salaries for each BCom graduate are independently and identically distributed (i.i.d.)

Think about

- Unbiasedness
- Efficiency

Question 2b

Consider the model

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

What is the mean and variance of $\hat{\beta}_0$?

Question 2c

How can we transform the model so that the errors are homoskedastic?