HW0414

6.

8.6 Consider the wage equation

$$WAGE_i = \beta_1 + \beta_2 EDUC_i + \beta_3 EXPER_i + \beta_4 METRO_i + e_i \qquad \text{(XR8.6a)}$$

where wage is measured in dollars per hour, education and experience are in years, and $METRO = 1$ if the person lives in a metropolitan area. We have $N = 1000$ observations from 2013.

a. We are curious whether holding education, experience, and *METRO* constant, there is the same amount of random variation in wages for males and females. Suppose $\text{var}(e_i|\mathbf{x}_i, FEMALE = 0) = \sigma_M^2$ and $\text{var}(e_i|\mathbf{x}_i, FEMALE = 1) = \sigma_F^2$. We specifically wish to test the null hypothesis $\sigma_M^2 = \sigma_F^2$ against $\sigma_M^2 \neq \sigma_F^2$. Using 577 observations on males, we obtain the sum of squared OLS residuals, $SSE_M = 97161.9174$. The regression using data on females yields $\hat{\sigma}_F = 12.024$. Test the null hypothesis at the 5% level of significance. Clearly state the value of the test statistic and the rejection region, along with your conclusion.

b. We hypothesize that married individuals, relying on spousal support, can seek wider employment types and hence holding all else equal should have more variable wages. Suppose $\text{var}(e_i|\mathbf{x}_i, MARRIED = 0) = \sigma_{SINGLE}^2$ and $\text{var}(e_i|\mathbf{x}_i, MARRIED = 1) = \sigma_{MARRIED}^2$. Specify the null hypothesis $\sigma_{SINGLE}^2 = \sigma_{MARRIED}^2$ versus the alternative hypothesis $\sigma_{MARRIED}^2 > \sigma_{SINGLE}^2$. We add *FEMALE* to the wage equation as an explanatory variable, so that

$$WAGE_i = \beta_1 + \beta_2 EDUC_i + \beta_3 EXPER_i + \beta_4 METRO_i + \beta_5 FEMALE + e_i \qquad \text{(XR8.6b)}$$

Using $N = 400$ observations on single individuals, OLS estimation of (XR8.6b) yields a sum of squared residuals is 56231.0382. For the 600 married individuals, the sum of squared errors is 100,703.0471. Test the null hypothesis at the 5% level of significance. Clearly state the value of the test statistic and the rejection region, along with your conclusion.

c. Following the regression in part (b), we carry out the NR^2 test using the right-hand-side variables in (XR8.6b) as candidates related to the heteroskedasticity. The value of this statistic is 59.03. What do we conclude about heteroskedasticity, at the 5% level? Does this provide evidence about the issue discussed in part (b), whether the error variation is different for married and unmarried individuals? Explain.

d. Following the regression in part (b) we carry out the White test for heteroskedasticity. The value of the test statistic is 78.82. What are the degrees of freedom of the test statistic? What is the 5% critical value for the test? What do you conclude?

---

e. The OLS fitted model from part (b), with usual and robust standard errors, is

$$\begin{array}{lllllll}
\widehat{WAGE} = & -17.77 & + 2.50EDUC & + 0.23EXPER & + 3.23METRO & - 4.20FEMALE \\
\text{(se)} & (2.36) & (0.14) & (0.031) & (1.05) & (0.81) \\
\text{(robse)} & (2.50) & (0.16) & (0.029) & (0.84) & (0.80)
\end{array}$$

For which coefficients have interval estimates gotten narrower? For which coefficients have interval estimates gotten wider? Is there an inconsistency in the results?

f. If we add *MARRIED* to the model in part (b), we find that its t-value using a White heteroskedasticity robust standard error is about 1.0. Does this conflict with, or is it compatible with, the result in (b) concerning heteroskedasticity? Explain.

(a) 男女的誤差變異是否相同

- **檢定方法**：使用 F 檢定 (Goldfeld–Quandt Test)，檢定男性與女性的誤差變異是否相同。
- **虛無假設** H0: σ f2=σ m2H_0: \sigma_f^2 = \sigma_m^2H0: σ f2=σ m2，對立假設為 H1: σ f2≠ σ m2H_1: \sigma_f^2 \neq \sigma_m^2H1: σ f2 =σ m2

- **結論依據**：使用給定的 SSE 和樣本數計算 F 值，再與臨界值比較，若 F 值落入拒絕域，拒絕 H0H_0H0，表示有異質變異。

(b) 已婚與未婚者的誤差變異是否相同

- **檢定方法**：同樣為 F 檢定
- **虛無假設** H0: σ MARRIED2= σ SINGLE2H_0: \sigma_{MARRIED}^2 = \sigma_{SINGLE}^2H0: σ MARRIED2= σ SINGLE2
- **統計值與結論**：根據 sum of squared residuals 計算 F 統計量，若大於臨界值，拒絕 H0H_0H0，表示已婚者的變異顯著不同。

(c) 進行 NR2NR^2NR2 檢定

- **檢定方法**：使用對異質變異可能相關的變數進行輔助回歸，計算 NR2NR^2NR2 值
- **統計量**：59.03，自由度為變數數目，查卡方臨界值後比較。
- **結論**：若 NR2NR^2NR2 超過臨界值，表示存在異質變異，並可能與性別、婚姻狀態等變數有關。

(d) White Test 檢定

- **統計量**：78.82，自由度為輔助回歸中的變數數量。
- **結論**：若超過 5% 的卡方臨界值，表示存在異質變異，且無特定模式。

(e) 比較 OLS 與 Robust 標準誤

- **觀察重點**：比較原始與 robust 的標準誤，若差距大表示存在異質變異。
- **結論**：若區間變寬，代表原本低估標準誤，robust 更可信。

(f) 若加入婚姻狀態變數，White 檢定 p 值約為 1，說明什麼？

- **解釋**：表示婚姻狀態的加入有助於解釋變異，殘差誤差變異變得較一致，異質變異現象減少。

16.

8.16 A sample of 200 Chicago households was taken to investigate how far American households tend to travel when they take a vacation. Consider the model

$$MILES = \beta_1 + \beta_2 INCOME + \beta_3 AGE + \beta_4 KIDS + e$$

MILES is miles driven per year, *INCOME* is measured in \$1000 units, *AGE* is the average age of the adult members of the household, and *KIDS* is the number of children.

a. Use the data file *vacation* to estimate the model by OLS. Construct a 95% interval estimate for the effect of one more child on miles traveled, holding the two other variables constant.

b. Plot the OLS residuals versus *INCOME* and *AGE*. Do you observe any patterns suggesting that heteroskedasticity is present?

c. Sort the data according to increasing magnitude of income. Estimate the model using the first 90 observations and again using the last 90 observations. Carry out the Goldfeld–Quandt test for heteroskedastic errors at the 5% level. State the null and alternative hypotheses.

d. Estimate the model by OLS using heteroskedasticity robust standard errors. Construct a 95% interval estimate for the effect of one more child on miles traveled, holding the two other variables constant. How does this interval estimate compare to the one in (a)?

e. Obtain GLS estimates assuming $\sigma_i^2 = \sigma^2 INCOME_i^2$. Using both conventional GLS and robust GLS standard errors, construct a 95% interval estimate for the effect of one more child on miles traveled, holding the two other variables constant. How do these interval estimates compare to the ones in (a) and (d)?

(a) 使用 OLS 估計模型

- **重點**：估計子女對旅遊距離的影響，並計算 95% 信賴區間。

(b) 畫出殘差與收入、年齡的關係圖

- **觀察重點**：殘差是否隨收入或年齡變動增加，有沒有「漏斗型」或「扇形」模式，判斷是否有異質變異。

(c) Goldfeld – Quandt Test

- **方法**：對收入排序，前 90 與後 90 分別估計回歸，計算 SSE 比例，進行 F 檢定。
- **假設**：
  - H0H_0H0: 同質變異
  - H1H_1H1: 異質變異
- **結論**：若拒絕 H0H_0H0，代表高收入群體的誤差變異不同。

(d) 使用 heteroskedasticity-robust 標準誤重新估計

- **比較**：robust 結果的信賴區間與 (a) 相比，若變寬表示原本低估標準誤。

(e) GLS 與 robust GLS

- GLS 假設：σ i2=σ 2·INCOMEi2\sigma_i^2 = \sigma^2 \cdot INCOME_i^2σ i2=σ 2·INCOMEi2
- 目的：校正異質變異，提高估計效率。
- 比較：(e) 中 GLS 的信賴區間應該比 (a)(d) 更窄，代表效率提升。

18.

8.18 Consider the wage equation,

$$\ln(WAGE_i) = \beta_1 + \beta_2 EDUC_i + \beta_3 EXPER_i + \beta_4 EXPER_i^2 + \beta_5 FEMALE_i + \beta_6 BLACK + \beta_7 METRO_i + \beta_8 SOUTH_i + \beta_9 MIDWEST_i + \beta_{10} WEST + e_i$$

where *WAGE* is measured in dollars per hour, education and experience are in years, and $METRO = 1$ if the person lives in a metropolitan area. Use the data file *cps5* for the exercise.

a. We are curious whether holding education, experience, and *METRO* equal, there is the same amount of random variation in wages for males and females. Suppose $\text{var}(e_i|\mathbf{x}_i, FEMALE = 0) = \sigma_M^2$ and $\text{var}(e_i|\mathbf{x}_i, FEMALE = 1) = \sigma_F^2$. We specifically wish to test the null hypothesis $\sigma_M^2 = \sigma_F^2$ against $\sigma_M^2 \neq \sigma_F^2$. Carry out a Goldfeld–Quandt test of the null hypothesis at the 5% level of significance. Clearly state the value of the test statistic and the rejection region, along with your conclusion.

b. Estimate the model by OLS. Carry out the NR^2 test using the right-hand-side variables *METRO*, *FEMALE*, *BLACK* as candidates related to the heteroskedasticity. What do we conclude about heteroskedasticity, at the 1% level? Do these results support your conclusions in (a)? Repeat the test using all model explanatory variables as candidates related to the heteroskedasticity.

c. Carry out the White test for heteroskedasticity. What is the 5% critical value for the test? What do you conclude?

d. Estimate the model by OLS with White heteroskedasticity robust standard errors. Compared to OLS with conventional standard errors, for which coefficients have interval estimates gotten narrower? For which coefficients have interval estimates gotten wider? Is there an inconsistency in the results?

e. Obtain FGLS estimates using candidate variables *METRO* and *EXPER*. How do the interval estimates compare to OLS with robust standard errors, from part (d)?

f. Obtain FGLS estimates with robust standard errors using candidate variables *METRO* and *EXPER*. How do the interval estimates compare to those in part (e) and OLS with robust standard errors, from part (d)?

g. If reporting the results of this model in a research paper which one set of estimates would you present? Explain your choice.

(a) 男女誤差變異是否相同

- 方法：Goldfeld–Quandt test
- 結論：若檢定結果顯示有異質變異，表示模型殘差受性別影響。

(b) 使用指定變數進行 NR2NR^2NR2 檢定（如 FEMALE, BLACK）

- 結果判讀：是否超過卡方臨界值，並與 (a) 結論做比較。

(c) White test 統計量與檢定

- 自由度：與輔助回歸變數數量有關。

- **結論**：若統計量顯著，則表示誤差具有異質變異。

(d) 使用 White robust 標準誤重新估計

- 分析：比較 OLS 與 robust 標準誤，確認是否低估或高估標準誤。

(e) 使用 FGLS（如 METRO 與 EXPER）

- 目的：修正誤差異質性，增加估計效率。

(f) 比較不同估計方法的信賴區間

- **比較**：OLS、White robust、FGLS，判斷何者較窄、估計效率較好。

(g) 若為研究論文應報告哪個估計結果？

- **建議**：若異質變異存在，應使用 robust 標準誤或 FGLS 估計，因為 OLS 在異質變異下不具一致性與效率。