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.

a. $H_0: \sigma_M^2 = \sigma_F^2$ $\quad N_m = 577$

$H_a: \sigma_M^2 \neq \sigma_F^2$

$$\hat{\sigma}_M^2 = \frac{97161.9174}{577-4} = 169.5670$$

$$\hat{\sigma}_F^2 = 12.024^2 = 144.5766$$

$$GQ: \frac{\hat{\sigma}_M^2}{\hat{\sigma}_F^2} = \frac{169.5670}{144.5766} \approx 1.173$$

$F_{0.025}(573, 1000-577-4)$ $F_{0.025}(573, 419) \approx (0.838, 1.197)$

$0.838 < 1.173 < 1.197 \Rightarrow$ do not reject Ho, we don't have evidence to show that $\hat{\sigma}_M^2 \neq \hat{\sigma}_F^2$ at 95% level of significance.

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.

$H_0: \sigma_{single}^2 = \sigma_{married}^2$

$H_a: \sigma_{married}^2 > \sigma_{single}^2$.

$$\hat{\sigma}_{single}^2 : \frac{56231.0382}{400-5} \approx 142.36.$$

$$\hat{\sigma}_{married}^2 : \frac{100703.0471}{600-5} \approx 169.25$$

$$\hat{GQ} = \frac{169.25}{142.36} \approx 1.189.$$

$F_{0.95}(595, 395) \approx 1.165$

$1.189 > 1.165$

$\Rightarrow$ reject Ho, we have evidence show that $\sigma_{married}^2 > \sigma_{single}^2$.

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?

c. $NR^2 = 59.03$, $K = 5$.

$59.03 \sim \chi^2(5-1) = \chi^2(4)$

$\chi^2(0.95, 4) \approx 9.488$,

$NR^2 > \chi^2(4) \Rightarrow$ reject Ho and we have evidence to show the existing of heteroskedasticity at 95% significant level.

d. $\chi^2(4) \approx 9.488$,

$78.82 > 9.458$, reject Ho,

EDUC, EXPER, METRO, FEMALE

$EDUC^2$, $EXPER^2$

EDUC × EXPER, EDUC × METRO, EDUC × FEMALE,
EXPER × METRO, EXPER × FEMALE, METRO × FEMALE

df $= 4 + 2 + 6 = 12$.

$\alpha = 0.05$, $\chi^2_{0.95, 12} \approx 21.026$
$NR^2 = 78.82$

reject Ho. at 5% critical value.

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

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

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.

e. EXPER, METRO, FEMALE have narrower interval estimates because their robse smaller than se.

intercept and EDUC have wider interval estimates.

有不一致現象，因原本預期 robse 會較大，因其有考慮異質變異數，較保守，但結果不一致，此結果雖不一致但不矛盾，因 robse 是在異質變異時較準確 可能變小 & 變大。

f. 結論是 compatible，因(b)做了異質性檢定，且有存在異質性，而(f)用了 robust se 來估計且不顯著，因此是合理的，因是使用 robse。

8.16

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?

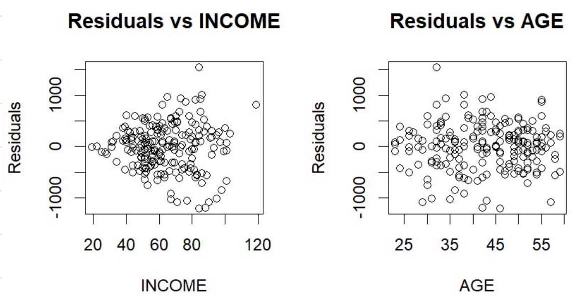
c. Sort the data according to increasing magnitude of income. Estimate the model using the first 90 observations and again using the last 90 observations. Carry out the Goldfeld–Quandt test for heteroskedastic errors at the 5% level. State the null and alternative hypotheses.

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. [−135.3298 , −28.32302]

b.



income ↑, residuals ↑

可能存在異質變異數

c. GQ - test at 5% significance level

GQ statistic = 3.104

critical value : 1.429

⇒ reject H0

d. [−138.969, −24.684]

e. GLS: [−119.8945, −33.7181]

robust GLS: [−121.1388, −32.4738]

(a) [−135.3298 , −28.32302]

(d) [−138.969, −24.684]

相較 a.d 都有更窄的 C.I.

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) F-statistic: 1.0538.

Critical value: 1.0581.

⇒ do not reject Ho.

(b) NR^2 = 23.557.

reject Ho at 1% significance level.

即 METRO. FEMALE. BLACK 與 error variance 有相關

(c) NR^2 = 194.445.

⇒ reject Ho. #

(d)

OLS:

(Intercept)	1.1384	1.2643
educ	0.0978	0.1047
exper	0.0271	0.0322
I(exper^2)	-0.0005	-0.0004
female	-0.1842	-0.1468
black	-0.1447	-0.0783
metro	0.0949	0.1431
south	-0.0723	-0.0192
midwest	-0.0916	-0.0363
west	-0.0348	0.0216

robust:

ROB_Lower	ROB_Upper	
1.1371	1.2657	Narrow
0.0975	0.1050	wider
0.0270	0.0322	wider
-0.0005	-0.0004	narrow
-0.1841	-0.1469	Narrow
-0.1431	-0.0800	Narrow
0.0963	0.1417	Narrow
-0.0730	-0.0185	wider
-0.0908	-0.0370	narrow
-0.0351	0.0219	wider

⇒ the result inconsistently

e.

(Intercept)	1.1267	1.2506
educ	0.0984	0.1053
exper	0.0276	0.0327
I(exper^2)	-0.0005	-0.0004
female	-0.1843	-0.1472
black	-0.1441	-0.0776
metro	0.0948	0.1401
south	-0.0712	-0.0181
midwest	-0.0906	-0.0358
west	-0.0337	0.0226

: C.I. 變窄,
其餘的反而變寬,

f.

	Estimate	Lower	Upper
(Intercept)	1.1886	1.1253	1.2520
educ	0.1019	0.0982	0.1056
exper	0.0302	0.0276	0.0327
I(exper^2)	-0.0005	-0.0005	-0.0004
female	-0.1658	-0.1843	-0.1473
black	-0.1108	-0.1419	-0.0798
metro	0.1174	0.0948	0.1401
south	-0.0446	-0.0718	-0.0175
midwest	-0.0632	-0.0901	-0.0363
west	-0.0055	-0.0340	0.0229

```
>
```

用 robust FGLS 後,
C.I. 變窄, 估計更精確

g. use FGLS with robust se.

因其 C.I. 較窄, 且能提供較精確的結果