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 $\text{var}(e_i|\mathbf{x}_i, MARRIED = 0) = \sigma^2_{SINGLE}$ and $\text{var}(e_i|\mathbf{x}_i, MARRIED = 1) = \sigma^2_{MARRIED}$. Specify the null hypothesis $\sigma^2_{SINGLE} = \sigma^2_{MARRIED}$ versus the alternative hypothesis $\sigma^2_{MARRIED} > \sigma^2_{SINGLE}$. 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

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

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.

a. $H_0: \sigma_M^2 = \sigma_F^2$

$H_1: \sigma_M^2 \neq \sigma_F^2$

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

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

$$F^* = \frac{\hat{\sigma}_M^2}{\hat{\sigma}_F^2} = \frac{169.5670}{144.5766} = 1.173$$

$$F_{0.05}(573, 419) = (0.838, 1.197)$$

$0.838 < 1.173 < 1.197$, don't reject H_0

We don't have evidence to show that $\hat{\sigma}_M^2 \neq \hat{\sigma}_F^2$ at 5% level of significance.

b.

$H_0: \sigma^2_{single} = \sigma^2_{married}$

$H_1: \sigma^2_{single} \neq \sigma^2_{married}$

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

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

$$F^* = \frac{169.25}{142.36} = 1.189 > 1.165 = F_{0.95}(595, 395)$$

reject H_0. we have evidence show that $\sigma^2_{married} > \sigma^2_{single}$

c.

test statistic $\chi^2 = NR^2 \sim \chi^2(k-1)$

$NR^2 = 59.03 > 9.488 = \chi^2_{0.95}(4)$

reject H_0 and we have evidence to show the existing of heteroskedasticity at the 5% level.

d.

$NR^2 = 78.82 > 9.4877 = \chi^2_{0.95}(4)$

we reject H_0 and conclude that the error term (or at least one of them) is not zero.

e.

Narrower: EXPER, METRO, FEMALE

Wider: EDUC, intercept

The result are inconsistent

f.

Compatible. (b) 做了 GQ-test 顯示存在 heterskedasticity 的問題，而(f) 用了 robust se 來估計且不顯著，因此是合理的

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.

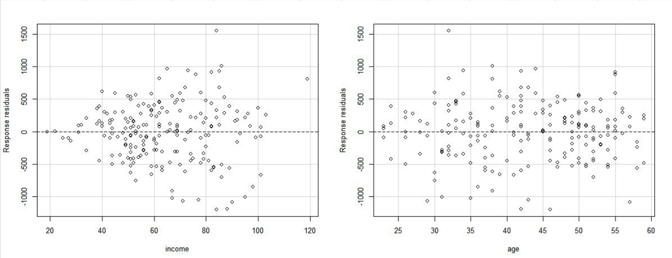
d. Estimate the model by OLS using heteroskedasticity robust standard errors. Construct a 95% interval estimate for the effect of one more child on miles traveled, holding the two other variables constant. How does this interval estimate compare to the one in (a)?

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. 95% interval

[-135.3298, -28.32302]

(b)



當 Icome 上升 risidual 也上升

表示可能有 heterskedasticity 的問題

(C)

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

```
        Goldfeld-Quandt test

data:  model
GQ = 3.4142, df1 = 76, df2 = 76, p-value = 1.106e-07
alternative hypothesis: variance increases from segment 1 to 2
```

p-value < 0.05 ∴ reject H_0

σ_1^2 and σ_2^2 aren't equal.

d.

95% interval

[-138.969, -24.68383]

e.

95% interval

[-119.8945, -33.71808]

相較 (a) (d) C.I. 更窄

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)

```
        Goldfeld-Quandt test

data:  model_a
GQ = 0.97394, df1 = 3910, df2 = 3909, p-value = 0.4092
alternative hypothesis: variance changes from segment 1 to 2
```

p-value > 0.05

don't reject H_0

b.

test statistic: $\chi^2 = NR^2 \sim \chi^2(S-1)$

$NR^2 = 23.55681$

reject H_0 at 1% significance level.

METRO, FEMALE, BLACK as a candidates related to the heteroskedasticity

c

```
        studentized Breusch-Pagan test

data:  model_a
BP = 109.42, df = 9, p-value < 2.2e-16
```

p-value < 0.05 reject H_0

d.

```
> coeftest(model_a, vcov = vcovHC(model_a, type = "HC1"))

t test of coefficients:

               Estimate  Std. Error  t value  Pr(>|t|)    
(Intercept)  1.2014e+00  3.2794e-02  36.6340 < 2.2e-16 ***
educ         1.0123e-01  1.9058e-03  53.1160 < 2.2e-16 ***
exper        2.9622e-02  1.3149e-03  22.5276 < 2.2e-16 ***
I(exper^2)  -4.4578e-04  2.7597e-05 -16.1533 < 2.2e-16 ***
female      -1.6550e-01  9.4883e-03 -17.4428 < 2.2e-16 ***
black       -1.1153e-01  1.6094e-02  -6.9297 4.482e-12 ***
metro        1.1902e-01  1.1582e-02  10.2762 < 2.2e-16 ***
south       -4.5755e-02  1.3902e-02  -3.2914  0.001001 ** 
midwest     -6.3943e-02  1.3724e-02  -4.6591 3.217e-06 ***
west        -6.5891e-03  1.4557e-02  -0.4526  0.650813    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(e)

```
> confint(fgls_model)
                    2.5 %        97.5 %
(Intercept)   1.127694057  1.2515350381
educ          0.098351366  0.1052682659
exper         0.027590905  0.0326693606
I(exper^2)   -0.000509177 -0.0004041652
female       -0.184317568 -0.1471399412
black        -0.144166923 -0.0776164205
metro         0.094808099  0.1401225846
south        -0.071252312 -0.0182311336
midwest      -0.090708494 -0.0358393299
west         -0.033747215  0.0226111169
```

(f)

```
> #f
> coeftest(fgls_model, vcov = vcovHC(fgls_model, type = "HC1"))

t test of coefficients:

               Estimate  Std. Error  t value  Pr(>|t|)    
(Intercept)  1.1896e+00  3.2326e-02  36.8008 < 2.2e-16 ***
educ         1.0181e-01  1.8906e-03  53.8505 < 2.2e-16 ***
exper        3.0130e-02  1.3042e-03  23.1022 < 2.2e-16 ***
I(exper^2)  -4.5667e-04  2.7403e-05 -16.6649 < 2.2e-16 ***
female      -1.6573e-01  9.4379e-03 -17.5599 < 2.2e-16 ***
black       -1.1089e-01  1.5862e-02  -6.9911 2.906e-12 ***
metro        1.1747e-01  1.1557e-02  10.1636 < 2.2e-16 ***
south       -4.4742e-02  1.3833e-02  -3.2344  0.001223 ** 
midwest     -6.3274e-02  1.3708e-02  -4.6158 3.965e-06 ***
west        -5.5680e-03  1.4504e-02  -0.3839  0.701060    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```