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

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

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. $\begin{cases} H_0: \sigma_M^2 = \sigma_F^2 \\ H_1: \sigma_m^2 \neq \sigma_F^2 \end{cases}$

$df_m = n - K = 577 - 4 = 573$

$\hat{\sigma}_m^2 = \dfrac{SSE}{573} = \dfrac{97161.9174}{573} = 169.5670$

$F = \dfrac{\hat{\sigma}_m^2}{\hat{\sigma}_F^2} = \dfrac{169.5670}{12.024^2} = 1.1729$

$RR: \{F : F \geq 1.1968 \text{ or } F \leq 0.8377\}$

$F \notin RR \Rightarrow$ fail to reject H_0

$\Rightarrow$ We can't conclude the vandom error variantion is different for males and females.

b. $\begin{cases} H_0: \sigma_{single}^2 = \sigma_{married}^2 \\ H_1: \sigma_{single}^2 < \sigma_{married}^2 \end{cases}$ $\quad | \; < \dfrac{\sigma_m}{SS}$

$\hat{\sigma}_{single}^2 = \dfrac{56231.0382}{400-5} = 142.3571$

$\hat{\sigma}_{married}^2 = \dfrac{100703.0471}{600-5} = 169.2488$

$F = \dfrac{169.2488}{142.3571} = 1.1889$

$RR: \{F : F \geq 1.1647\}$

$F \in RR$

$\Rightarrow$ The random error variation is greater for married individuals than single individuals.

c. $\begin{cases} H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0 \\ H_1: H_0 \text{ isn't true.} \end{cases}$

$\chi^2_{0.95,4} = 9.4877$

$\chi^2 = 59.03$

$\Rightarrow \chi^2 > \chi^2_{0.94,4}$

$\Rightarrow$ Reject H_0

$\Rightarrow$ 只說明該差異與至少一變數相關，
但未指出是哪一個變數。

d. 原變數：EDUC、EXPER、METRO、FEMALE $\Rightarrow 4$

平方項：$EDUC^2$, $EXPER^2$, $METRO^2$ (Dummy), $FEMALE^2$ (Dummy) $\Rightarrow 4-2=2$

交乘項：$3 \times 2 = 6$

$df: 4+2+6=12$

$\begin{cases} H_0: \text{Homoskedasticity} \\ H_1: \text{Hetroskedasticity.} \end{cases}$

$\chi^2_{0.95,12} = 21.0267$

$\chi^2 = 78.82$

$\Rightarrow \chi^2_{0.95,12} > \chi^2$

$\Rightarrow$ Reject H_0

$\Rightarrow$ Hetroskedasticity.

e. gotten wider：EDUC、intercept.

gotten narrow：EXPER、METRO、FEMALE

$\Rightarrow$ 結果不一致

f. 加入 married 後，White 檢定下 t 值 = 1

$\Rightarrow$ married 的係數不顯著

8.b 和 8f 結果不衝突，∵ 8b 是檢驗是否為同質變異，8f 則是檢驗係數是否顯著異於零。

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?

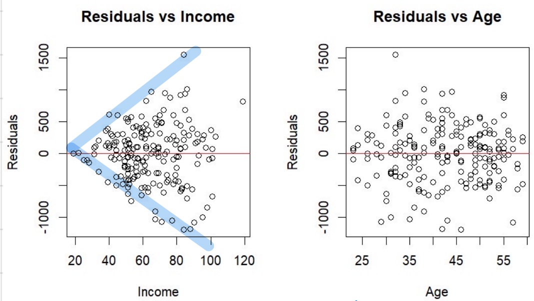
c. Sort the data according to increasing magnitude of income. Estimate the model using the first 90 observations and again using the last 90 observations. Carry out the Goldfeld–Quandt test for heteroskedastic errors at the 5% level. State the null and alternative hypotheses.

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.
```
> confint(model, "kids", level = 0.95)
         2.5 %    97.5 %
kids -135.3298 -28.32302
```

b.



Income ↑, Var(e) ↑

No patterns.

c.

$\begin{cases} H_0: \tilde{\sigma}_{first} = \tilde{\sigma}_{last} \\ H_1: \tilde{\sigma}_{first} > \tilde{\sigma}_{last} \end{cases}$

RR: $\{p: p < 0.05\}$

```
        Goldfeld-Quandt test

data:  miles ~ income + age + kids
GQ = 3.1041, df1 = 86, df2 = 86, p-value = 1.64e-07
alternative hypothesis: variance increases from segment 1 to 2
```

$p \in RR \Rightarrow$ reject H_0

$\Rightarrow$ Error variance depends on income.

d.
```
> coefci(model, vcov. = robust_se, level = 0.95)
                   2.5 %     97.5 %
(Intercept) -672.883378 -110.21263
income        10.377633   18.02503
age            7.919934   23.56191
kids        -139.322973  -24.32986
```

$\Rightarrow$ Wider than using the conventional OLS standard errors.

e.
```
> confint(model_gls, level = 0.95)
                 2.5 %     97.5 %
(Intercept) -664.50116 -185.49119
income        11.02744   16.86718
age           10.75260   22.68240
kids        -119.89450  -33.71808
> coefci(model_gls, vcov. = robust_se_gls)
                 2.5 %     97.5 %
(Intercept) -613.93428 -236.05807
income        11.29086   16.60376
age           11.20062   22.23438
kids        -121.41339  -32.19919
```

GLS and Robust GLS model are narrower than OLS and Robust OLS.

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 $\operatorname{var}(e_i|\mathbf{x}_i, FEMALE = 0) = \sigma_M^2$ and $\operatorname{var}(e_i|\mathbf{x}_i, FEMALE = 1) = \sigma_F^2$. 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.

```
> cat(F_stat)
1.05076
> cat(F_critical_lower,F_critical_upper)
0.9452566 1.058097
```

F ∈ RR ⇒ fail to reject Ho

b.

Only METRO. FEMALE. BLACK

```
        Breusch-Pagan test

data:  model
BP = 34.427, df = 3, p-value = 1.61e-07
```

< 0.01

All Variables.

```
        Breusch-Pagan test

data:  model
BP = 159.92, df = 9, p-value < 2.2e-16
```

< 0.01

⇒ 兩模型 p < 0.01 ⇒ reject Ho

c.

```
        studentized Breusch-Pagan test

data:  model
BP = 184.81, df = 36, p-value < 2.2e-16
```

< 0.05

⇒ Reject Ho

d.

```
> print(confint_ols)
                    2.5 %        97.5 %
(Intercept)  1.1384302204  1.2643338265
educ         0.0977830603  0.1046761665
exper        0.0270727569  0.0321706349
I(exper^2)  -0.0004974407 -0.0003941203
female      -0.1841810529 -0.1468229075
black       -0.1447358548 -0.0783146449
metro        0.0948966363  0.1431441846
south       -0.0723384657 -0.0191724010
midwest     -0.0915893895 -0.0362971859
west        -0.0348207138  0.0216425095
```

```
> print(confint_robust)
            lower_robust  upper_robust
(Intercept)  1.137098683  1.2656653641
educ         0.097493811  0.1049654160
exper        0.027044205  0.0321991870
I(exper^2)  -0.000499876 -0.0003916849
female      -0.184100928 -0.1469030324
black       -0.143072211 -0.0799782888
metro        0.096316998  0.1417238226
south       -0.073005508 -0.0185053588
midwest     -0.090845662 -0.0370409129
west        -0.035123519  0.0219453146
```

The highlighted Robust OLS intervals are wider than OLS's.
And the resultes aren't consistency.

e.

> print(result_table)

	Variable	OLS_CI_Lower	OLS_CI_Upper	OLS_Width	FGLS_CI_Lower	FGLS_CI_Upper	FGLS_Width	CI_Change
(Intercept)	(Intercept)	1.137098683	1.2656653641	0.1285666813	1.1302695254	1.2541279001	0.1238583747	narrower
educ	educ	0.097493811	0.1049654160	0.0074716051	0.0982024458	0.1051204663	0.0069180205	narrower
exper	exper	0.027044205	0.0321991870	0.0051549824	0.0275467064	0.0326335081	0.0050868017	narrower
I(exper^2)	I(exper^2)	-0.000499876	-0.0003916849	0.0001081911	-0.0005086498	-0.0004036251	0.0001050246	narrower
female	female	-0.184100928	-0.1469030324	0.0371978956	-0.1847977976	-0.1476290326	0.0371687649	narrower
black	black	-0.143072211	-0.0799782888	0.0630939222	-0.1441623553	-0.0775448504	0.0666175049	wider
metro	metro	0.096316998	0.1417238226	0.0454068242	0.0953066354	0.1402324253	0.0449257899	narrower
south	south	-0.073005508	-0.0185053588	0.0545001491	-0.0713493481	-0.0183363498	0.0530129983	narrower
midwest	midwest	-0.090845662	-0.0370409129	0.0538047496	-0.0906033967	-0.0357807800	0.0548226168	wider
west	west	-0.035123519	0.0219453146	0.0570688335	-0.0336747637	0.0226870709	0.0563618347	narrower

f.

> print(result_table)

	Variable	OLS_CI_Lower	OLS_CI_Upper	FGLS_CI_Lower	FGLS_CI_Upper	CI_Change
(Intercept)	(Intercept)	1.137098683	1.2656653641	1.1287671947	1.2556302309	narrower
educ	educ	0.097493811	0.1049654160	0.0979512563	0.1053716558	narrower
exper	exper	0.027044205	0.0321991870	0.0275327897	0.0326474248	narrower
I(exper^2)	I(exper^2)	-0.000499876	-0.0003916849	-0.0005098633	-0.0004024116	narrower
female	female	-0.184100928	-0.1469030324	-0.1847139905	-0.1477128397	narrower
black	black	-0.143072211	-0.0799782888	-0.1419596068	-0.0797475989	narrower
metro	metro	0.096316998	0.1417238226	0.0951039024	0.1404351583	narrower
south	south	-0.073005508	-0.0185053588	-0.0719612018	-0.0177244961	narrower
midwest	midwest	-0.090845662	-0.0370409129	-0.0900718148	-0.0363123619	narrower
west	west	-0.035123519	0.0219453146	-0.0339339956	0.0229463028	narrower

g. I would report Robust OLS results. ∵ OLS estimates with robust standard errors are all significant at 0.001 level, except the WEST, which is not significant using FGLS either.