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

$$\widehat{WAGE} = -17.77 + 2.50EDUC + 0.23EXPER + 3.23METRO - 4.20FEMALE$$

(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. $H_0: \sigma_M^2 = \sigma_F^2$ $\quad \sigma_M^2 = \frac{SSE_M}{n_M} = \frac{97161.9174}{577} = 168.3915$ $\quad F = \frac{\sigma_M^2}{\sigma_F^2} = \frac{168.3915}{12.024^2} = 1.1647$

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

$df1 = 577-4 = 573$ $\quad df_2 = 1000-577-4 = 419$ $\quad F_{573, 419, \alpha=5\%} = 0.8377$ & 1.1968

$\because 0.8377 < 1.1647 < 1.1968$ $\therefore$ We fail to reject H_0.

There's no sufficient evidence to show that $\sigma_M^2 \neq \sigma_F^2$

b.

$H_0: \sigma_{MR}^2 = \sigma_S^2$ $\quad \sigma_{MR}^2 = \frac{100703.0471}{600} = 167.6384$ $\quad F = \frac{167.6384}{140.5776} = 1.1925$

$H_1: \sigma_{MR}^2 > \sigma_S^2$ $\quad \sigma_S^2 = \frac{56231.0382}{400} = 140.5776$ $\quad F_{595, 395, \alpha=5\%} = 0.8366$ & 1.1994

$\because 0.8366 < 1.1925 < 1.1994$ $\therefore$ We fail to reject H_0.

There's no sufficient evidence to show that $\sigma_{MR}^2 \neq \sigma_S^2$

c.

$NR^2 = 59.03$ $\quad \chi^2_{(4, 0.05)} = 9.4877$ $\quad \because 59.03 > 9.4877$ $\therefore$ We reject H_0,

there's sufficient evidence to show that heteroskedasticity exists, and this is consistent with (b)

d. $df = 12$ $\quad \chi^2_{12, 0.05} = 21.0267$ $\quad \because 78.82 > 21.0267$ $\therefore$ We reject H_0.

there's sufficient evidence to show that heteroskedasticity exists, and this is consistent with (b)

e. Wider: Intercept, EDUC $\quad$ Narrower: EXPER. METRO. FEMALE

No contradiction. but it illustrates heteroskedasticity since it affects OLS SE unevenly.

f.

It's compatible with (b). $\because$ (b) test the heteroskedasticity but (f) tests the influence of MARRIED to the model, there's no conflict.

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)?

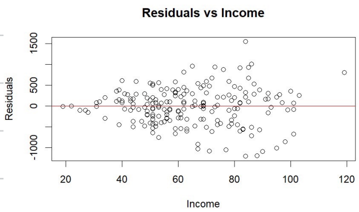
a.

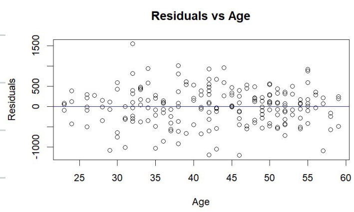
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -391.548    169.775  -2.306   0.0221 *
income        14.201      1.800   7.889 2.10e-13 ***
age           15.741      3.757   4.189 4.23e-05 ***
kids         -81.826     27.130  -3.016   0.0029 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 452.3 on 196 degrees of freedom
Multiple R-squared:  0.3406,	Adjusted R-squared:  0.3305
F-statistic: 33.75 on 3 and 196 DF,  p-value: < 2.2e-16
```

```
> confint(mod1, "kids", level = 0.95)
          2.5 %    97.5 %
kids -135.3298 -28.32302
```

b.





→ spreads is getting larger as income increases, the heteroskedasticity exists.

→ spread is even with age, there's no pattern here.

c. $H_0: \sigma_L^2 = \sigma_H^2$

$H_1: \sigma_L^2 < \sigma_H^2$

```
	Goldfeld-Quandt test

data:  mod1
GQ = 3.1041, df1 = 86, df2 = 86, p-value = 1.64e-07
alternative hypothesis: variance increases from segment 1 to 2
```

→ $F = 3.1041 > F_{86, 86, 0.95} = 1.4286$ $\therefore$ We reject H_0. there's no sufficient evidence to say that heteroskedasticity exists.

d.

```
t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept) -391.5480   142.6548 -2.7447 0.0066190 **
income        14.2013     1.9389  7.3246 6.083e-12 ***
age           15.7409     3.9657  3.9692 0.0001011 ***
kids         -81.8264    29.1544 -2.8067 0.0055112 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
      kids       kids
-138.96900  -24.68383
```

$\because$ SE of robust is larger. which considers heteroskedasticity. $\therefore$ the CI of robust is larger.

e.

```
Generalized least squares fit by REML
  Model: miles ~ income + age + kids
  Data: vacation
       AIC      BIC    logLik
  2960.626 2977.017 -1475.313

Variance function:
 Structure: fixed weights
 Formula: ~income^2

Coefficients:
                Value Std.Error   t-value p-value
(Intercept) -424.9962 121.44414 -3.499520   6e-04
income        13.9473   1.48056  9.420293   0e+00
age           16.7175   3.02458  5.527212   0e+00
kids         -76.8063  21.84844 -3.515413   5e-04

 Correlation:
       (Intr) income age
income -0.339
age    -0.738 -0.294
kids    0.031  0.040 -0.334

Standardized residuals:
        Min          Q1         Med          Q3         Max
-2.24561294 -0.73256459  0.03677301  0.64796222  2.74164318
```

	CI	
	2.5 %	97.5 %
OLS (a)	-135.3298	-28.32302
Robust OLS (d)	-138.96900	-24.68383
GLS	-119.89450	-33.71808
Robust GLS	-121.41339	-32.19919

GLS is narrower than OLS, it's more accuracy.

Robust GLS is narrower & better than OLS & 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 $\text{var}(e_i|\mathbf{x}_i, FEMALE = 0) = \sigma_M^2$ 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. $H_0: \sigma_M^2 = \sigma_F^2$

$H_1: \sigma_M^2 \neq \sigma_F^2$ $\quad F = \dfrac{\sigma_M^2}{\sigma_F^2} = 1.05076$ $\quad F_{5\%} = 0.9453$ & 1.0581

$\because 0.9453 < 1.05076 < 1.0581$ $\therefore$ Fail to reject H_0, there's sufficient evidence to show that $\sigma_F^2 = \sigma_M^2$

b. $H_0: \alpha_5 = \alpha_6 = \alpha_7 = 0$ (Homoskedasticity) $\quad \because NR^2 = 23.55681 > 11.34487 = \chi^2_{0.99,3}$

H_1: not all α_i in H_0 is 0 (Hetroskedasticity) $\quad \therefore$ We reject H_0, there's sufficient evidence to show that the model exists heteroskedasticity.

$H_0: \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = \alpha_7 = \alpha_8 = \alpha_9 = \alpha_{10} = 0$ (Homoskedasticity)

H_1: not all α_i in H_0 is 0 (Hetroskedasticity) $\quad \because NR^2 = 109.4243 > 21.66599 = \chi^2_{0.99,9}$

$\therefore$ We reject H_0, there's sufficient evidence to show that the model exists heteroskedasticity.

c. H_0: Homoskedasticity $\quad \because NR^2 = 194.4447 > 60.48089 = \chi^2$

H_1: Hetroskedasticity $\quad \therefore$ We reject H_0, there's sufficient evidence to show that the model exists heteroskedasticity.

d.

	OLS_Width	OLS_Robust_Width	CI_Change
(Intercept)	0.1259036061	0.1285010626	wider
educ	0.0068931063	0.0074677917	wider
exper	0.0050978780	0.0051523513	wider
I(exper^2)	0.0001033205	0.0001081359	wider
female	0.0373581454	0.0371789103	narrower
black	0.0664212100	0.0630617199	narrower
metro	0.0482475482	0.0453836492	narrower
south	0.0531660647	0.0544723330	wider
midwest	0.0552922035	0.0537772884	narrower
west	0.0564632233	0.0570397063	wider

EDUC, EXPER, EXPER², SOUTH, WEST are wider, means that SE of White robust is more conservative.

FEMALE, BLACK, METRO, MIDWEST are narrower, means that SE of White robust is more accuracy.

	OLS_p	OLS_Significant	OLS_Robust_p	OLS_Robust_Significant
(Intercept)	2.122456e-286	TRUE	1.002569e-275	TRUE
educ	0.000000e+00	TRUE	0.000000e+00	TRUE
exper	5.712676e-112	TRUE	1.030809e-109	TRUE
I(exper^2)	2.760763e-63	TRUE	5.285645e-58	TRUE
female	1.427018e-66	TRUE	3.454261e-67	TRUE
black	4.859552e-11	TRUE	4.371440e-12	TRUE
metro	5.016366e-22	TRUE	1.140396e-24	TRUE
south	7.438266e-04	TRUE	9.945792e-04	TRUE
midwest	5.861847e-06	TRUE	3.180265e-06	TRUE
west	6.473209e-01	FALSE	6.506470e-01	FALSE

There's no inconsistency.

e.

	OLS_Robust_Width	FGLS_Width	CI_Change
(Intercept)	0.1285010626	0.124843079	narrower
educ	0.0074677917	0.006905656	narrower
exper	0.0051523513	0.005092118	narrower
I(exper^2)	0.0001081359	0.000104173	narrower
female	0.0371789103	0.037265303	wider
black	0.0630617199	0.066513034	wider
metro	0.0453836492	0.046510222	wider
south	0.0544723330	0.053091297	narrower
midwest	0.0537772884	0.055064111	wider
west	0.0570397063	0.056413445	narrower

EDUC, EXPER, EXPER², SOUTH, WEST are narrower, means that FGLS is more accuracy.

FEMALE, BLACK, METRO, MIDWEST are wider, means that the uncertainty of the estimate increases.

f.

	OLS_Robust_Width	FGLS_Robust_Width	CI_Change
(Intercept)	0.1285010626	0.1274490902	narrower
educ	0.0074677917	0.0074304492	narrower
exper	0.0051523513	0.0051241957	narrower
I(exper^2)	0.0001081359	0.0001075916	narrower
female	0.0371789103	0.0370289927	narrower
black	0.0630617199	0.0625640260	narrower
metro	0.0453836492	0.0453173516	narrower
south	0.0544723330	0.0542654167	narrower
midwest	0.0537772884	0.0536708228	narrower
west	0.0570397063	0.0568719873	narrower

	FGLS_Width	FGLS_Robust_Width	CI_Change
(Intercept)	0.124843079	0.1274490902	wider
educ	0.006905656	0.0074304492	wider
exper	0.005092118	0.0051241957	wider
I(exper^2)	0.000104173	0.0001075916	wider
female	0.037265303	0.0370289927	narrower
black	0.066513034	0.0625640260	narrower
metro1	0.046510222	0.0453173516	narrower
south	0.053091297	0.0542654167	wider
midwest	0.055064111	0.0536708228	narrower
west	0.056413445	0.0568719873	wider

All CI are narrower, means that FGLS robust is more accuracy.

EDUC, EXPER, EXPER², SOUTH, WEST are wider, means that SE of White robust is more conservative.

FEMALE, BLACK, METRO, MIDWEST are narrower, means that SE of White robust is more accuracy.

g.

I would present the test FGLS with robust, these can effectively correct the error brought with heteroskedasticity.