8.6

(a) $\hat{\sigma}_M^2 = \frac{SSE_M}{577-4} = 169.57$

$\hat{\sigma}_F^2 = 12.024^2 = 144.58$ $\quad df = 1000-577-4 = 419$

Test-statistic $\frac{\hat{\sigma}_M^2}{\hat{\sigma}_F^2} = \frac{169.57}{144.58} = 1.1728 \sim F_{573,419}$

$(F_{0.025,573,419}, F_{0.975,573,419}) = (0.8377, 1.197)$

rejection region : $F < 0.8377$ or $F > 1.197$

Non-reject Ho
No enough evidence to say $\sigma_M^2 \neq \sigma_F^2$

(b) $\hat{\sigma}_S^2 = \frac{56231.0382}{400-5} = 142.36$

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

test-statistic $\frac{\hat{\sigma}_M^2}{\hat{\sigma}_S^2} = 1.189$

$F_{0.95,595,395} = 1.1647$

rejection region : $F > 1.1647$

reject Ho
there's statistical evidence showing that $\sigma_M^2 > \sigma_S^2$

(c) $NR^2 = 59.03 \sim \chi^2_{(4)}$

rejection-region : $\chi^2 > 9.4877$

reject Ho , it's heteroskedasticity , accordance to (b)

(d) 78.82 , $df = \cancel{1} + 4 + \cancel{4}^2 + C_2^4 - \cancel{1} - \underset{\substack{metro^2 \\ female^2}}{\cancel{2}} = 12$

$\chi^2_{0.95,12} = 21.026$

reject Ho , it's heteroskedasticity

(e) narrower using robust : exper , metro , female
wider using robust : intercept , educ
Not consistent

(f) It’s compatible. B check whether it’s heteroskedasticity or not. F check the importance of the explanatory variable.

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.

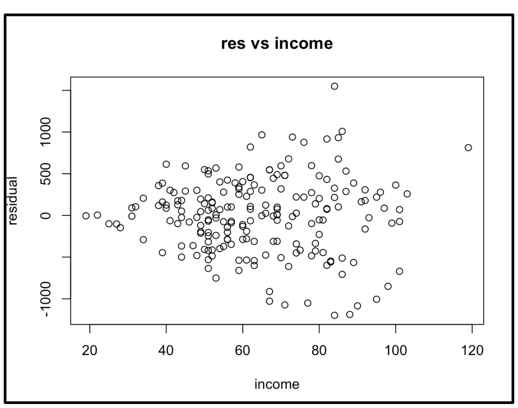
d. Estimate the model by OLS using heteroskedasticity robust standard errors. Construct a 95% interval estimate for the effect of one more child on miles traveled, holding the two other variables constant. How does this interval estimate compare to the one in (a)?

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. β4 95% CI

```
> lb
[1] -135.3298
> ub
[1] -28.32302
```

b. residual vs income 隨著 income 增加有上升的 pattern



c.

```
#c
vacation_sorted = vacation[order(vacation$income), ]
low_income = vacation_sorted[1:90, ]
high_income = vacation_sorted[111:200, ]
low_model = lm(miles ~ income + age + kids, data = low_income)
high_model = lm(miles ~ income + age + kids, data = high_income)
low_df = low_model$df.residual
high_df = high_model$df.residual
f = summary(low_model)$sigma^2 / summary(high_model)$sigma^2
(f > qf(0.025, low_df, high_df))*(qf(0.975, low_df, high_df) < f)
#reject H0
```

d. 估計值一樣，但是 robust 方法的 standard error 比較小

```
> lb
[1] -139.323
> ub
[1] -24.32986
```

e. 與 a 和 d 的估計值不同，區間估計 conventional GLS 最窄

Conventenal GLS

```
> gls_est
[1] -76.80629
> lb
[1] -119.8945
> ub
[1] -33.71808
```

Robust GLS

```
> gls_robust_est
[1] -76.80629
> lb
[1] -121.4134
> ub
[1] -32.19919
```

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.

```
#a
modela = lm(log(wage)~educ+exper+I(exper^2)+metro, data = cps5)
male = cps5[cps5$female==0, ]
female = cps5[cps5$female==1, ]
male_lm = lm(log(wage)~educ+exper+I(exper^2)+metro, data = male)
female_lm = lm(log(wage)~educ+exper+I(exper^2)+metro, data = female)

male_df = male_lm$df.residual
female_df = female_lm$df.residual
f = summary(male_lm)$sigma^2 / summary(female_lm)$sigma^2
f_low = qf(0.025, male_df, female_df)
f_up = qf(0.975, male_df, female_df)
(f > f_low)*(f < f_up) #1 non-reject H0
#non-reject H0
```

b.

```
#b
model1 = lm(log(wage)~educ+exper+I(exper^2)+female+black+metro+south+midwest+west, data = cps5)
ressq = resid(model1)^2
modres2 = lm(ressq~educ+exper+I(exper^2)+female+black+metro+south+midwest+west, data = cps5)
n = nobs(modres2) #9799
S = nobs(modres2) - df.residual(modres2) #10
rsqres = summary(modres2)$r.squared
chisq = n * rsqres
pval = 1-pchisq(chisq, S - 1) #0 reject H0
```

Under 1% level, we conclude that it's heteroskedasticity. Different from part a.

c. p-value < 0.05 reject H0

```
#c
modres3 = lm(ressq~educ+I(educ^2)+exper+I(exper^2)+I(exper^4)+female+I(female^2)+black+I(b
             educ*exper+educ*I(exper^2)+I(exper^3), data = cps5)
n = nobs(modres3) #9799
S = nobs(modres3) - df.residual(modres3) #15
rsqres = summary(modres3)$r.squared
chisq = n * rsqres
pval = 1-pchisq(chisq, S - 1) #0 reject H0
```

d.

```
"intercept has wider interval after robust"
"educ has wider interval after robust"
"exper has wider interval after robust"
"I(exper^2) has wider interval after robust"
"female has narrower interval after robust"
"\black has narrower interval after robust"
"metro has narrower interval after robust"
"south has wider interval after robust"
"midwest has narrower interval after robust"
"west has wider interval after robust"
```

	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	

e. wider interval: black, midwest
narrower interval: Intercept, educ, exper, exper^2, female, metro, south, west

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.190e+00	3.159e-02	37.659	< 2e-16	***
educ	1.018e-01	1.764e-03	57.705	< 2e-16	***
exper	3.013e-02	1.295e-03	23.260	< 2e-16	***
I(exper^2)	-4.567e-04	2.679e-05	-17.049	< 2e-16	***
female	-1.657e-01	9.483e-03	-17.476	< 2e-16	***
black	-1.109e-01	1.698e-02	-6.532	6.79e-11	***
metro	1.175e-01	1.156e-02	10.163	< 2e-16	***
south	-4.474e-02	1.352e-02	-3.308	0.000942	***
midwest	-6.327e-02	1.400e-02	-4.521	6.23e-06	***
west	-5.568e-03	1.438e-02	-0.387	0.698523	

f. similar SE with d.

	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	

g. I will choose FGLS+Robust, it takes the heteroskedasticity into consider, but we should check out more model of the residual to take the most use of FGLS.