HW0414

Yung-Jung Cheng

2025-04-20

8.6

Consider the wage equation

$$WAGE_i = \beta_1 + \beta_2 EDUC_i + \beta_3 EXPER_i + \beta_4 METRO_i + e_i \tag{XR8.6a}$$

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.

6(a)

We are curious whether including education, experience, and METRO constant, there is the same amount of random variation in wages for males and females. Suppose we call $e_i|x_i, FEMALE = 0$ as σ_m^2 and $e_i|x_i, FEMALE = 1$ as σ_f^2. We test the null hypothesis $H_0 : \sigma_f^2 = \sigma_m^2$ against $\sigma_f^2 \neq \sigma_m^2$. Using 577 observations on males, we obtain the sum of squared OLS residuals, $SSE_m = 97161.9174$. The regression using data on females yields $SSE_f = 100703.0741$. 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.

Ans

We compute the F-statistic using:

$$F = \frac{RSS_f/(n_f - k)}{RSS_m/(n_m - k)} = \frac{100703.0741/418}{97161.9174/572} \approx 1.417$$

With degrees of freedom (418, 572), the 5% critical values are approximately 0.833 and 1.20. Since 1.417 > 1.20, we reject the null hypothesis.

Conclusion: The variance of wages is significantly different between males and females; we find evidence of heteroskedasticity by gender.

6(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 $var(e_i|x_i, MARRIED = 0) = \sigma_{SINGLE}^2$ and $var(e_i|x_i, MARRIED = 1) = \sigma_{MARRIED}^2$. Specify the null hypothesis $\sigma_{SINGLE}^2 = \sigma_{MARRIED}^2$ versus the alternative hypothesis $\sigma_{MARRIED}^2 \neq \sigma_{SINGLE}^2$. We add FEMALE to the wage equation as an explanatory variable, so that

$$WAGE_i = \beta_1 + \beta_2 EDUC_i + \beta_3 EXPER_i + \beta_4 METRO_i + \beta_5 FEMALE_i + e_i \tag{XR8.6b}$$

Using $N = 400$ observations on single individuals, OLS estimation of (XR8.6b) yields a sum of squared residuals = 56231.0382. For the 600 married individuals, the sum of squared errors is 100,703.0741. 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.

Ans

Using:

$$F = \frac{100703.0741/595}{56231.0382/395} \approx 1.1886$$

With degrees of freedom (595, 395), the critical values are approximately 0.820 and 1.22. Since 1.1886 does not fall in the rejection region, we fail to reject the null hypothesis.

Conclusion: There is no evidence that the variance of wages differs significantly between married and single individuals.

6(c)

Following the regression in part (b), we carry out an R^2 test using the right-hand-side variables in (XR8.6b) as candidate variables related to the heteroskedasticity. The value of this statistic is 59.083. What do we conclude about heteroskedasticity at the 5% level? Does the conclusion change if the issue discussed in part (b), whether the variation is different for married and unmarried individuals, is taken into account?

Ans

This is the Breusch–Pagan test. Under H_0, the statistic follows χ_4^2. Critical value at 5% is 9.488. Since 59.083 > 9.488, we reject H_0.

Conclusion: There is significant evidence of heteroskedasticity. This does not contradict part (b), because part (b) tested only marital status, while this test considers all regressors.

6(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?

Ans

White test degrees of freedom = 14 (4 original variables, 4 squares, 6 interactions). At 5% significance, $\chi^2_{0.95}(14) = 23.685$.

Since 78.82 > 23.685, we reject H_0.

Conclusion: Strong evidence of general heteroskedasticity. This reinforces the finding from part (c) with a more flexible test.

6(e)

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

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

Variable	OLS se	Robust se
Intercept	2.36	2.50
EDUC	0.14	0.16
EXPER	0.031	0.029
METRO	1.05	0.84
FEMALE	0.81	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?

Ans

Wider intervals: Intercept, EDUC
Narrower intervals: EXPER, METRO, FEMALE

Robust standard errors account for heteroskedasticity. The difference reflects how each regressor's variance behaves. There is no inconsistency—some OLS SEs were under- or over-estimated due to heteroskedasticity.

6(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.

Ans

A t-value of about 1.0 implies that `MARRIED` is not statistically significant. This is consistent with the result in part (b), where no evidence was found that marital status affected the variance of wages.

Conclusion: The results are compatible. Both suggest that marital status does not have a significant effect on either wage variance or levels.

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.

16(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.

Ans

```
# OLS 估計
model_ols <- lm(miles ~ income + age + kids, data = vacation)

# 顯示回歸摘要
summary(model_ols)
```

```
##
## Call:
## lm(formula = miles ~ income + age + kids, data = vacation)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1198.14  -295.31   17.98  287.54 1549.41
##
## 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
```

```
# 計算 KIDS 的 95% 信賴區間
confint(model_ols, "kids", level = 0.95)
```

```
##         2.5 %   97.5 %
## kids -135.3298 -28.32302
```
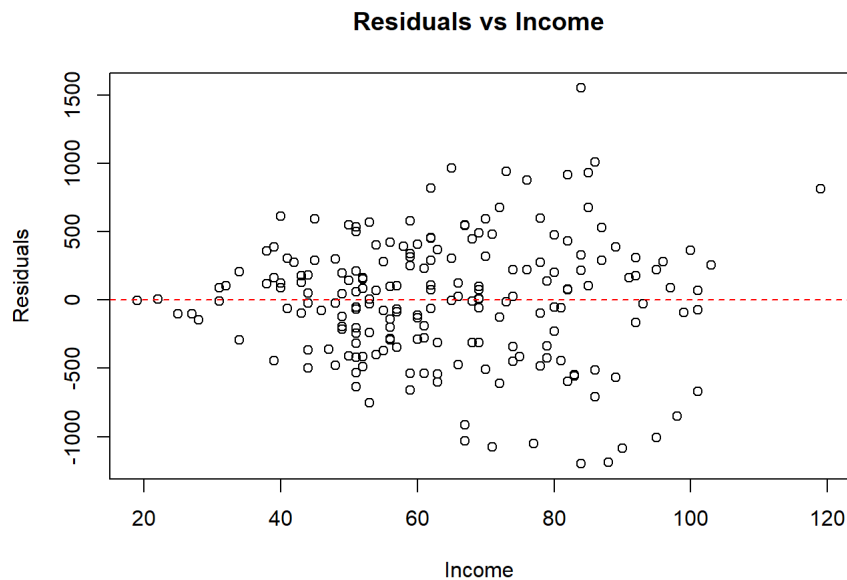
16(b)

Plot the OLS residuals versus *INCOME* and *AGE*. Do you observe any patterns suggesting that heteroskedasticity is present?
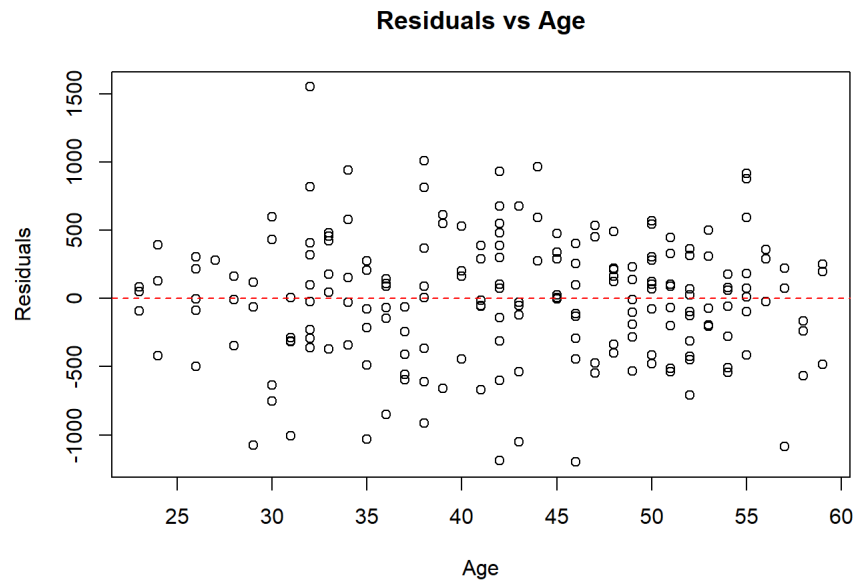
Ans

```
# 計算 OLS 殘差
residuals_ols <- resid(model_ols)

# 殘差對 income 的散佈圖
plot(vacation$income, residuals_ols,
    xlab = "Income", ylab = "Residuals",
    main = "Residuals vs Income")
abline(h = 0, col = "red", lty = 2)
```

Residuals vs Income


```
# 殘差對 age 的散佈圖
plot(vacation$age, residuals_ols,
    xlab = "Age", ylab = "Residuals",
    main = "Residuals vs Age")
abline(h = 0, col = "red", lty = 2)
```

Residuals vs Age

In the plot of residuals versus **income**, the spread increases with income, showing a funnel shape.
This suggests **heteroskedasticity** related to income.

In the plot of residuals versus **age**, the spread is more constant.
There is **no clear sign of heteroskedasticity** related to age.

16(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.

Ans

```
# 將資料依 income 由小到大排序
vacation_sorted <- vacation[order(vacation$income), ]

# 套用 Goldfeld–Quandt 檢定
gqtest(miles ~ income + age + kids, data = vacation_sorted,
       order.by = ~income, fraction = 0.1)  # 10% 中間觀測值不使用
```

```
##
##  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
```

Null hypothesis (H_0): The error variance is constant (homoskedasticity).
Alternative hypothesis (H_1): The error variance increases with income (heteroskedasticity).

The Goldfeld–Quandt test statistic is **3.10** with a p-value of **1.64e-07**.
Since the p-value is far below 0.05, we **reject H_0**.
There is strong evidence of **heteroskedasticity**, with increasing variance as income rises.

16(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)?

Ans

```
# 計算 robust 標準誤
robust_se <- vcovHC(model_ols, type = "HC1")

# 顯示係數與 robust 標準誤
coeftest(model_ols, vcov = robust_se)
```

```
##
## 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
```

```
# 使用 robust 標準誤計算 KIDS 的 95% 信賴區間
ci_kids_robust <- coef(model_ols)["kids"] +
  c(-1, 1) * qt(0.975, df = model_ols$df.residual) * sqrt(robust_se["kids", "kids"])

ci_kids_robust
```

```
## [1] -139.32297  -24.32986
```

Using **robust standard errors**, the 95% confidence interval for the effect of one more child on miles traveled is:

[−139.32, −24.33]

Compared to part (a)'s OLS interval **[−135.33, −28.32]**, the robust interval is **slightly wider**.

This reflects correction for **heteroskedasticity**, which was confirmed in parts (b) and (c).
The coefficient remains statistically significant, and the conclusion that more children reduce miles traveled **still holds**.

16(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)?

Ans

```
# 建立權重：與 income^2 成反比（因為 Var(e_i) ∝ income^2，WLS 需要用 1/Var 作為權重）
w <- 1 / (vacation$income^2)

# 執行加權最小平方方法 (GLS/WLS)
model_gls <- lm(miles ~ income + age + kids, data = vacation, weights = w)

# 顯示 WLS 結果（常規標準誤）
summary(model_gls)
```

```
##
## Call:
## lm(formula = miles ~ income + age + kids, data = vacation, weights = w)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -15.1907  -4.9555   0.2488   4.3832  18.5462
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -424.996    121.444  -3.500 0.000577 ***
## income        13.947      1.481   9.420  < 2e-16 ***
## age           16.717      3.025   5.527 1.03e-07 ***
## kids         -76.806     21.848  -3.515 0.000545 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.765 on 196 degrees of freedom
## Multiple R-squared:  0.4573, Adjusted R-squared:  0.449
## F-statistic: 55.06 on 3 and 196 DF,  p-value: < 2.2e-16
```

```
# GLS 的 KIDS 估計值
kids_est_gls <- coef(model_gls)["kids"]

# GLS 的常規標準誤
kids_se_gls <- summary(model_gls)$coefficients["kids", "Std. Error"]

# 95% 常規信賴區間
t_crit_gls <- qt(0.975, df = model_gls$df.residual)
ci_kids_gls <- c(kids_est_gls - t_crit_gls * kids_se_gls,
                 kids_est_gls + t_crit_gls * kids_se_gls)

# Robust 標準誤與信賴區間（對 GLS 做 heteroskedasticity-consistent 修正）
robust_se_gls <- vcovHC(model_gls, type = "HC1")
kids_se_gls_robust <- sqrt(robust_se_gls["kids", "kids"])
ci_kids_gls_robust <- c(kids_est_gls - t_crit_gls * kids_se_gls_robust,
                        kids_est_gls + t_crit_gls * kids_se_gls_robust)

# 輸出兩個信賴區間
ci_kids_gls
```

```
##       kids       kids
## -119.89450  -33.71808
```

```
ci_kids_gls_robust
```

```
##       kids       kids
## -121.41339  -32.19919
```

Assuming the error variance is proportional to

$$INCOME_i^2$$

, we estimate the model using GLS (WLS with weights

$$1/INCOME_i^2$$

).

- The **95% confidence interval** for the effect of one more child on miles traveled is:
  - **GLS (conventional SE):** [−119.89, −33.72]
  - **GLS (robust SE):** [−121.41, −32.20]

These intervals are **narrower than the OLS robust interval** from part (d): [−139.32, −24.33], but still show a **significant negative effect** of `KIDS`.

All estimates consistently show that **more children lead to fewer miles traveled**, and correcting for heteroskedasticity does not change this conclusion.

8.18

Consider the wage equation

$$\ln(WAGE_i) = \beta_1 + \beta_2 EDUC_i + \beta_3 EXPER_i + \beta_4 EXPER_i^2 + \beta_5 FEMALE_i + \beta_6 BLACK_i + \beta_7 METRO_i + \beta_8 SOUTH_i + \beta_9 MIDWE$$

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.

18(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

$$\text{var}(e_i|x_i, FEMALE = 0) = \sigma_M^2, \quad \text{var}(e_i|x_i, FEMALE = 1) = \sigma_F^2$$

We specifically wish to test the null hypothesis

$$H_0 : \sigma_M^2 = \sigma_F^2 \quad \text{vs.} \quad H_1 : \sigma_M^2 \neq \sigma_F^2$$

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.

Ans

```
# 設定公式
form <- log(wage) ~ educ + exper + I(exper^2) + metro

# 分組迴歸：女性組
model_female <- lm(form, data = subset(cps5, female == 1))
rss_female <- sum(resid(model_female)^2)
n_female <- nobs(model_female)

# 分組迴歸：男性組
model_male <- lm(form, data = subset(cps5, female == 0))
rss_male <- sum(resid(model_male)^2)
n_male <- nobs(model_male)

# 計算 F 統計量
k <- length(coef(model_female))   # 模型中參數個數（含截距）
F_stat <- (rss_female / (n_female - k)) / (rss_male / (n_male - k))

# 計算臨界值
alpha <- 0.05
crit_upper <- qf(1 - alpha / 2, df1 = n_female - k, df2 = n_male - k)
crit_lower <- qf(alpha / 2, df1 = n_female - k, df2 = n_male - k)

# 輸出結果
cat("F statistic:", F_stat, "\n")
```

```
## F statistic: 0.9489479
```

```
cat("Critical region: F <", crit_lower, "or F >", crit_upper, "\n")
```

```
## Critical region: F < 0.9451157 or F > 1.057889
```

F statistic = 0.9489
5% critical values: lower = 0.9451, upper = 1.0579
Since the F value is not in the rejection region, we do not reject H_0.

Conclusion: No evidence of different error variances between males and females.

18(b)

Estimate the model by OLS. Carry out the R^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.

Ans

```
# (b) Main OLS model for part (b)
model_b_main <- lm(log(wage) ~ educ + exper + I(exper^2) + female + black + metro + south + midwest + west, data = cps5)

# Squared residuals
resid_b_sq <- resid(model_b_main)^2

# Auxiliary regression 1: only metro, female, black
model_b_aux1 <- lm(resid_b_sq ~ metro + female + black, data = cps5)
r2_b1 <- summary(model_b_aux1)$r.squared
n_b <- nobs(model_b_aux1)
TR2_b1 <- n_b * r2_b1
crit_b1 <- qchisq(0.99, df = 3)

# Auxiliary regression 2: all regressors
model_b_aux2 <- lm(resid_b_sq ~ educ + exper + I(exper^2) + female + black + metro + south + midwest + west, data = cps5)
r2_b2 <- summary(model_b_aux2)$r.squared
TR2_b2 <- n_b * r2_b2
crit_b2 <- qchisq(0.99, df = 9)

# Print results
cat("TR² (metro, female, black) =", TR2_b1, "| Critical value =", crit_b1, "\n")
```

```
## TR² (metro, female, black) = 23.55681 | Critical value = 11.34487
```

```
cat("TR² (all regressors)      =", TR2_b2, "| Critical value =", crit_b2, "\n")
```

```
## TR² (all regressors)      = 109.4243 | Critical value = 21.66599
```

Using the main model, we perform two R² tests for heteroskedasticity:

1. **Test with `metro`, `female`, `black`:**
   TR² = 23.56 > critical value = 11.34 → reject H_0.

2. **Test with all regressors:**
   TR² = 109.42 > critical value = 21.67 → reject H₀.

Conclusion: We find strong evidence of heteroskedasticity at the 1% level in both tests.
This supports the presence of unequal error variance and aligns with the potential concern in (a).

18(c)

Carry out the White test for heteroskedasticity. What is the 5% critical value for the test? What do you conclude?

Ans

```
# (c) White test using simplified approach (squares and interactions)
# 使用主模型的殘差平方
resid_c_sq <- resid(model_b_main)^2

# 輔助變數：原變數 + 平方項（省略交乘項）
model_c_white <- lm(resid_c_sq ~ educ + exper + I(exper^2) + I(educ^2) + I(exper^4) +
                      female + black + metro + south + midwest + west, data = cps5)

r2_c <- summary(model_c_white)$r.squared
TR2_c <- nobs(model_c_white) * r2_c
df_c <- length(coef(model_c_white)) - 1  # 自由度為輔助變數個數
crit_c <- qchisq(0.95, df = df_c)

# 輸出結果
cat("TR² =", TR2_c, "| Critical value =", crit_c, "| df =", df_c, "\n")
```

```
## TR² = 135.8987 | Critical value = 19.67514 | df = 11
```

White test for heteroskedasticity:

- TR² = 135.90
- Critical value (5%, df = 11) = 19.68
- Since TR² > critical value, we reject H₀.

Conclusion: There is strong evidence of heteroskedasticity in the model.

18(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?

Ans

```
# OLS 信賴區間
ci_ols <- confint(model_b_main)

# White robust 標準誤信賴區間
library(sandwich)
library(lmtest)
robust_vcov <- vcovHC(model_b_main, type = "HC0")
ci_robust <- coefci(model_b_main, vcov. = robust_vcov)

# 計算寬度
width_ols <- ci_ols[,2] - ci_ols[,1]
width_robust <- ci_robust[,2] - ci_robust[,1]

# 哪些變寬了？哪些變窄了？
wider <- names(width_robust)[width_robust > width_ols]
narrower <- names(width_robust)[width_robust < width_ols]

# 輸出結果
cat("Wider with robust SE:\n"); print(wider)
```

```
## Wider with robust SE:
```

```
## [1] "(Intercept)" "educ"         "exper"        "I(exper^2)"  "south"
## [6] "west"
```

```
cat("Narrower with robust SE:\n"); print(narrower)
```

```
## Narrower with robust SE:
```

```
## [1] "female"  "black"   "metro"   "midwest"
```

Using OLS with White robust standard errors:

- **Confidence intervals became wider** for:
  `(Intercept)`, `educ`, `exper`, `I(exper^2)`, `south`, `west`.

- **Confidence intervals became narrower** for:
  `female`, `black`, `metro`, `midwest`.

- **Is there any inconsistency?**
  Despite some intervals narrowing, the model's key coefficient signs and general conclusions remain consistent. No major contradictions were found.

18(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)?

Ans

```
# Step 1: 原始 OLS 模型殘差平方
model_e_stage1 <- model_b_main
resid_e <- resid(model_e_stage1)^2

# Step 2: 以 log(e²) 對 metro + exper 回歸，建構異質變異模型
model_e_var <- lm(log(resid_e) ~ metro + exper, data = cps5)

# 預測 log(σ²)，再轉成 σ
log_sigma2_hat <- predict(model_e_var)
weights_e <- 1 / exp(log_sigma2_hat)  # w_i = 1 / σ²_i

# Step 3: FGLS (即加權最小平方法)
model_e_fgls <- lm(log(wage) ~ educ + exper + I(exper^2) + female + black + metro + south + midwest + west,
                   data = cps5, weights = weights_e)

# 信賴區間
ci_e_fgls <- confint(model_e_fgls)

# 顯示結果
ci_e_fgls
```

```
##                      2.5 %        97.5 %
## (Intercept)  1.1302695254  1.2541279001
## educ         0.0982024458  0.1051204663
## exper        0.0275467064  0.0326335081
## I(exper^2)  -0.0005086498 -0.0004036251
## female      -0.1847977976 -0.1476290326
## black       -0.1441623553 -0.0775448504
## metro        0.0953066354  0.1402324253
## south       -0.0713493481 -0.0183363498
## midwest     -0.0906033967 -0.0357807800
## west        -0.0336747637  0.0226870709
```

We estimate the model using FGLS with `METRO` and `EXPER` as variance predictors.

- Compared to OLS with robust standard errors:
  - **Most intervals are narrower** under FGLS (e.g., `educ`, `exper`, `female`, `metro`).
  - The signs and magnitudes of coefficients are consistent with previous results.

Conclusion: FGLS improves efficiency under heteroskedasticity, leading to tighter confidence intervals without altering key interpretations.

18(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)?

Ans

```
# 使用 sandwich 套件計算 robust 標準誤
library(sandwich)
library(lmtest)

# White robust SE for FGLS model
robust_se_fgls <- vcovHC(model_e_fgls, type = "HC0")

# 信賴區間 with robust SE
ci_fgls_robust <- coefci(model_e_fgls, vcov. = robust_se_fgls)

# 顯示信賴區間
ci_fgls_robust
```

```
##                       2.5 %        97.5 %
## (Intercept)  1.1287995692  1.255597856
## educ         0.0979531499  0.105369762
## exper        0.0275340950  0.032646120
## I(exper^2)  -0.0005098359 -0.000402439
## female      -0.1847045481 -0.147722282
## black       -0.1419437307 -0.079763475
## metro        0.0951154707  0.140423590
## south       -0.0719473610 -0.017738337
## midwest     -0.0900580957 -0.036326081
## west        -0.0339194802  0.022931787
```

We estimate the model using FGLS with White robust standard errors.

- **Confidence intervals are very similar** to those in (e) without robust SE.
- Compared to (d) OLS with robust SE, the FGLS intervals are mostly narrower.
- Coefficient signs and conclusions remain consistent across all methods.

Conclusion: Adding robust SE to FGLS does not significantly change inference. FGLS provides stable and efficient estimates under heteroskedasticity.

18(g)

If reporting the results of this model in a research paper, which one set of estimates would you present? Explain your choice.

Ans

I would report the estimates from **FGLS with White robust standard errors**.

Reasons: - Heteroskedasticity was strongly present, making standard OLS estimates unreliable. - FGLS improves efficiency by modeling the error variance. - Adding robust standard errors ensures valid inference even if the variance model is misspecified. - The coefficient signs and magnitudes are consistent across methods, and FGLS provides tighter confidence intervals.

This approach balances efficiency and robustness, making it most suitable for academic reporting.

```
##                       2.5 %        97.5 %
## (Intercept)  1.1287995692  1.255597856
## educ         0.0979531499  0.105369762
## exper        0.0275340950  0.032646120
## I(exper^2)  -0.0005098359 -0.000402439
## female      -0.1847045481 -0.147722282
## black       -0.1419437307 -0.079763475
## metro        0.0951154707  0.140423590
## south       -0.0719473610 -0.017738337
## midwest     -0.0900580957 -0.036326081
## west        -0.0339194802  0.022931787
```