10.18 Consider the data file *mroz* on working wives. Use the 428 observations on married women who participate in the labor force. In this exercise, we examine the effectiveness of a parent's college education as an instrumental variable.

a. Create two new variables. *MOTHERCOLL* is a dummy variable equaling one if *MOTHEREDUC* > 12, zero otherwise. Similarly, *FATHERCOLL* equals one if *FATHEREDUC* > 12 and zero otherwise. What percentage of parents have some college education in this sample?

b. Find the correlations between *EDUC*, *MOTHERCOLL*, and *FATHERCOLL*. Are the magnitudes of these correlations important? Can you make a logical argument why *MOTHERCOLL* and *FATHERCOLL* might be better instruments than *MOTHEREDUC* and *FATHEREDUC*?

c. Estimate the wage equation in Example 10.5 using *MOTHERCOLL* as the instrumental variable. What is the 95% interval estimate for the coefficient of *EDUC*?

d. For the problem in part (c), estimate the first-stage equation. What is the value of the *F*-test statistic for the hypothesis that *MOTHERCOLL* has no effect on *EDUC*? Is *MOTHERCOLL* a strong instrument?

e. Estimate the wage equation in Example 10.5 using *MOTHERCOLL* and *FATHERCOLL* as the instrumental variables. What is the 95% interval estimate for the coefficient of *EDUC*? Is it narrower or wider than the one in part (c)?

f. For the problem in part (e), estimate the first-stage equation. Test the joint significance of *MOTHERCOLL* and *FATHERCOLL*. Do these instruments seem adequately strong?

g. For the IV estimation in part (e), test the validity of the surplus instrument. What do you conclude?

a.

```
> mean(mroz$mothercoll)
[1] 0.1009296
> mean(mroz$fathercoll)
[1] 0.1075697
```

b.

```
                educ mothercoll fathercoll
educ       1.0000000  0.3370171  0.3193212
mothercoll 0.3370171  1.0000000  0.3674532
fathercoll 0.3193212  0.3674532  1.0000000
```

透過相關性來判斷工具變數的強度，相關性越強工具變數越有用

工具變數要跟educ 相關，但是跟wage不相關

因為OTHERCOLL/FATHERCOLL 是虛擬變數，代表是否上過大學，而不是實際的受教年數
用虛擬變數可以減少 measurement error 對估計的扭曲

c.

```
Call:
ivreg(formula = log(wage) ~ educ + exper + I(exper^2) | exper +
    I(exper^2) + mothercoll, data = mroz_clean)

Residuals:
     Min       1Q   Median       3Q      Max
-3.08719 -0.32444  0.04147  0.36634  2.35621

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.1327561  0.4965325  -0.267  0.78932
educ         0.0760180  0.0394077   1.929  0.05440 .
exper        0.0433444  0.0134135   3.231  0.00133 **
I(exper^2)  -0.0008711  0.0004017  -2.169  0.03066 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6703 on 424 degrees of freedom
Multiple R-Squared: 0.147,      Adjusted R-squared: 0.1409
Wald test:   8.2 on 3 and 424 DF,  p-value: 2.569e-05
```

```
                   2.5 %        97.5 %
(Intercept) -1.105942034  8.404298e-01
educ        -0.001219763  1.532557e-01
exper        0.017054428  6.963439e-02
I(exper^2)  -0.001658392 -8.385898e-05
```

d.

```
Call:
lm(formula = educ ~ exper + I(exper^2) + mothercoll, data = mroz_clean)

Residuals:
    Min      1Q  Median      3Q     Max
-7.4267 -0.4826 -0.3731  1.0000  4.9353

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.079094   0.303118  39.849  < 2e-16 ***
exper        0.056230   0.042101   1.336    0.182
I(exper^2)  -0.001956   0.001256  -1.557    0.120
mothercoll   2.517068   0.315713   7.973 1.46e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.133 on 424 degrees of freedom
Multiple R-squared:  0.1347,    Adjusted R-squared:  0.1285
F-statistic: 21.99 on 3 and 424 DF,  p-value: 2.965e-13
```

```
Linear hypothesis test:
mothercoll = 0

Model 1: restricted model
Model 2: educ ~ exper + I(exper^2) + mothercoll

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    425 2219.2
2    424 1929.9  1    289.32 63.563 1.455e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

可以發現F = 63.56 > 10，代表mothercoll是強工具變數

e.

```
Call:
ivreg(formula = log(wage) ~ educ + exper + I(exper^2) | exper +
    I(exper^2) + mothercoll + fathercoll, data = mroz_clean)

Residuals:
     Min       1Q   Median       3Q      Max
-3.07797 -0.32128  0.03418  0.37648  2.36183

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.2790819  0.3922213  -0.712  0.47714
educ         0.0878477  0.0307808   2.854  0.00453 **
exper        0.0426761  0.0132950   3.210  0.00143 **
I(exper^2)  -0.0008486  0.0003976  -2.135  0.03337 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6679 on 424 degrees of freedom
Multiple R-Squared: 0.153,      Adjusted R-squared: 0.147
Wald test: 9.724 on 3 and 424 DF,  p-value: 3.224e-06
```

```
> confint(mroz.ivmocoll)
                   2.5 %        97.5 %
(Intercept) -1.105942034  8.404298e-01
educ        -0.001219763  1.532557e-01
exper        0.017054428  6.963439e-02
I(exper^2)  -0.001658392 -8.385898e-05
```

f.

```
Call:
lm(formula = educ ~ exper + I(exper^2) + mothercoll + fathercoll
    data = mroz_clean)

Residuals:
    Min      1Q  Median      3Q     Max
-7.2152 -0.3056 -0.2152  0.7627  5.0620

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.890259   0.290251  40.965  < 2e-16 ***
exper        0.049149   0.040133   1.225    0.221
I(exper^2)  -0.001449   0.001199  -1.209    0.227
mothercoll   1.749947   0.322347   5.429 9.58e-08 ***
fathercoll   2.186612   0.329917   6.628 1.04e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.033 on 423 degrees of freedom
Multiple R-squared:  0.2161,    Adjusted R-squared:  0.2086
F-statistic: 29.15 on 4 and 423 DF,  p-value: < 2.2e-16
```

```
Linear hypothesis test:
mothercoll = 0
fathercoll = 0

Model 1: restricted model
Model 2: educ ~ exper + I(exper^2) + mothercoll + fathercoll

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    425 2219.2
2    423 1748.3  2    470.88 56.963 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

F=56.96 > 10

所以mothercoll and fathercoll是好的工具變數

g.

P–value =0.626 代表兩個工具變數是有效的, fail to reject null hypothesis.

```
Call:
ivreg(formula = log(wage) ~ educ + exper + I(exper^2) | exper +
    I(exper^2) + mothercoll + fathercoll, data = mroz_clean)

Residuals:
     Min       1Q   Median       3Q      Max
-3.07797 -0.32128  0.03418  0.37648  2.36183

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.2790819  0.3922213  -0.712  0.47714
educ         0.0878477  0.0307808   2.854  0.00453 **
exper        0.0426761  0.0132950   3.210  0.00143 **
I(exper^2)  -0.0008486  0.0003976  -2.135  0.03337 *

Diagnostic tests:
                 df1 df2 statistic p-value
Weak instruments   2 423    56.963  <2e-16 ***
Wu-Hausman         1 423     0.519   0.472
Sargan             1  NA     0.238   0.626
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6679 on 424 degrees of freedom
Multiple R-Squared: 0.153,      Adjusted R-squared: 0.147
Wald test: 9.724 on 3 and 424 DF,  p-value: 3.224e-06
```

```
Wage equation: OLS, IV mothercoll, mroz.firstm ,IV mothercoll and fathercoll,mroz.firstmf,
=================================================================================================
                                          Dependent variable: wage
              -----------------------------------------------------------------------------------
                 OLS    IV mothercoll fist mothercoll IV mothercoll fathercoll first mothercoll fathercoll
-------------------------------------------------------------------------------------------------
Constant      -0.52204    -0.13276      12.07909          -0.27908                  11.89026
              (0.19863)   (0.49653)     (0.30312)         (0.39222)                 (0.29025)

educ           0.10749     0.07602                         0.08785
              (0.01415)   (0.03941)                       (0.03078)

exper          0.04157     0.04334       0.05623           0.04268                   0.04915
              (0.01318)   (0.01341)     (0.04210)         (0.01330)                 (0.04013)

I(exper2)     -0.00081    -0.00087      -0.00196          -0.00085                  -0.00145
              (0.00039)   (0.00040)     (0.00126)         (0.00040)                 (0.00120)

mothercoll                               2.51707                                     1.74995
                                        (0.31571)                                   (0.32235)

fathercoll                                                                           2.18661
                                                                                    (0.32992)

-------------------------------------------------------------------------------------------------
Observations     428         428           428               428                       428
=================================================================================================
```

10.20 The CAPM [see Exercises 10.14 and 2.16] says that the risk premium on security j is related to the risk premium on the market portfolio. That is

$$r_j - r_f = \alpha_j + \beta_j(r_m - r_f)$$

where r_j and r_f are the returns to security j and the risk-free rate, respectively, r_m is the return on the market portfolio, and β_j is the jth security's "*beta*" value. We measure the market portfolio using the Standard & Poor's value weighted index, and the risk-free rate by the 30-day LIBOR monthly rate of return. As noted in Exercise 10.14, if the market return is measured with error, then we face an errors-in-variables, or measurement error, problem.

a. Use the observations on Microsoft in the data file *capm5* to estimate the CAPM model using OLS. How would you classify the Microsoft stock over this period? Risky or relatively safe, relative to the market portfolio?

b. It has been suggested that it is possible to construct an IV by ranking the values of the explanatory variable and using the rank as the IV, that is, we sort $(r_m - r_f)$ from smallest to largest, and assign the values $RANK = 1, 2, \ldots, 180$. Does this variable potentially satisfy the conditions IV1–IV3? Create *RANK* and obtain the first-stage regression results. Is the coefficient of *RANK* very significant? What is the R^2 of the first-stage regression? Can *RANK* be regarded as a strong IV?

c. Compute the first-stage residuals, $\hat{v}$, and add them to the CAPM model. Estimate the resulting augmented equation by OLS and test the significance of $\hat{v}$ at the 1% level of significance. Can we conclude that the market return is exogenous?

d. Use *RANK* as an IV and estimate the CAPM model by IV/2SLS. Compare this IV estimate to the OLS estimate in part (a). Does the IV estimate agree with your expectations?

e. Create a new variable $POS = 1$ if the market return $(r_m - r_f)$ is positive, and zero otherwise. Obtain the first-stage regression results using both *RANK* and *POS* as instrumental variables. Test the joint significance of the IV. Can we conclude that we have adequately strong IV? What is the R^2 of the first-stage regression?

f. Carry out the Hausman test for endogeneity using the residuals from the first-stage equation in (e). Can we conclude that the market return is exogenous at the 1% level of significance?

g. Obtain the IV/2SLS estimates of the CAPM model using *RANK* and *POS* as instrumental variables. Compare this IV estimate to the OLS estimate in part (a). Does the IV estimate agree with your expectations?

h. Obtain the IV/2SLS residuals from part (g) and use them (not an automatic command) to carry out a Sargan test for the validity of the surplus IV at the 5% level of significance.

a.

```
Call:
lm(formula = excess_msft ~ excess_rm, data = capm5)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.27424 -0.04744 -0.00820  0.03869  0.35801 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.003250   0.006036   0.538    0.591    
excess_rm   1.201840   0.122152   9.839   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08083 on 178 degrees of freedom
Multiple R-squared:  0.3523,	Adjusted R-squared:  0.3486 
F-statistic:  96.8 on 1 and 178 DF,  p-value: < 2.2e-16
```

Microsoft's beta is 1.2018, indicating that this stock is relatively correlated to market portfolio.

b.

```
Call:
lm(formula = excess_rm ~ rank, data = capm5)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.110497 -0.006308  0.001497  0.009433  0.029513 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -7.903e-02  2.195e-03   -36.0   <2e-16 ***
rank         9.067e-04  2.104e-05    43.1   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01467 on 178 degrees of freedom
Multiple R-squared:  0.9126,	Adjusted R-squared:  0.9121 
F-statistic:  1858 on 1 and 178 DF,  p-value: < 2.2e-16
```

```
Linear hypothesis test:
rank = 0

Model 1: restricted model
Model 2: excess_rm ~ rank

  Res.Df     RSS Df Sum of Sq      F    Pr(>F)    
1    179 0.43784                                  
2    178 0.03829  1   0.39955 1857.6 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The variable rank has no relation with Microsoft.
R^2: 0.9121 thus is relatively related to excess_rm.
F:1857.6 > 10, which is an extremely strong IV.

c.

```
Call:
lm(formula = excess_msft ~ excess_rm + rankhat, data = capm5)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.27140 -0.04213 -0.00911  0.03423  0.34887 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.003018   0.005984   0.504   0.6146    
excess_rm    1.278318   0.126749  10.085   <2e-16 ***
rankhat     -0.874599   0.428626  -2.040   0.0428 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08012 on 177 degrees of freedom
Multiple R-squared:  0.3672,	Adjusted R-squared:   0.36 
F-statistic: 51.34 on 2 and 177 DF,  p-value: < 2.2e-16
```

P–value: 0.0428, it is not significant at 1% level , but significant at 5% level. Thus, we can not reject null hypothesis test at 1% level.

因此我們不拒絕excess_rm是外生變數的虛無假設

d.

```
Call:
ivreg(formula = excess_msft ~ excess_rm | rank, data = capm5)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.271625 -0.049675 -0.009693  0.037683  0.355579 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.003018   0.006044   0.499    0.618    
excess_rm   1.278318   0.128011   9.986   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08092 on 178 degrees of freedom
Multiple R-Squared: 0.3508,	Adjusted R-squared: 0.3472 
Wald test: 99.72 on 1 and 178 DF,  p-value: < 2.2e-16
```

相比於a小題的結果，可以發現使用工具變數之後的回歸係數與a小題相似

e.

```
Call:
lm(formula = excess_rm ~ rank + POS, data = capm5)

Residuals:
      Min        1Q    Median        3Q       Max
-0.109182 -0.006732  0.002858  0.008936  0.026652

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0804216  0.0022622  -35.55   <2e-16 ***
rank         0.0009819  0.0000400   24.55   <2e-16 ***
POS         -0.0092762  0.0042156   -2.20   0.0291 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01451 on 177 degrees of freedom
Multiple R-squared:  0.9149,    Adjusted R-squared:  0.9139
F-statistic: 951.3 on 2 and 177 DF,  p-value: < 2.2e-16
```

```
Linear hypothesis test:
rank = 0
POS = 0

Model 1: restricted model
Model 2: excess_rm ~ rank + POS

  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1    179 0.43784
2    177 0.03727  2   0.40057 951.26 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

f.

```
Call:
lm(formula = excess_msft ~ excess_rm + vhat2, data = capm5)

Residuals:
     Min       1Q   Median       3Q      Max
-0.27132 -0.04261 -0.00812  0.03343  0.34867

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.003004   0.005972   0.503   0.6157
excess_rm    1.283118   0.126344  10.156   <2e-16 ***
vhat2       -0.954918   0.433062  -2.205   0.0287 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07996 on 177 degrees of freedom
Multiple R-squared:  0.3696,    Adjusted R-squared:  0.3625
F-statistic: 51.88 on 2 and 177 DF,  p-value: < 2.2e-16
```

H_0 : excess_rm 是外生的
H_1 : excess_rm 是內生的
用1%去檢驗

P–value 是0.0287, 不拒絕虛無假設

g.

```
Call:
ivreg(formula = excess_msft ~ excess_rm | rank + POS, data = capm5)

Residuals:
     Min       1Q   Median       3Q      Max
-0.27168 -0.04960 -0.00983  0.03762  0.35543

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.003004   0.006044   0.497     0.62
excess_rm   1.283118   0.127866  10.035   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08093 on 178 degrees of freedom
Multiple R-Squared: 0.3507,     Adjusted R-squared: 0.347
Wald test: 100.7 on 1 and 178 DF,  p-value: < 2.2e-16
```

可以發現 IV 的值比 OLS 大，代表OLS的值可能有偏誤

h.

10.24 Consider the data file *mroz* on working wives. Use the 428 observations on married women who participate in the labor force. In this exercise, we examine the effectiveness of alternative standard errors for the IV estimator. Estimate the model in Example 10.5 using IV/2SLS using both *MOTHEREDUC* and *FATHEREDUC* as IV. These will serve as our baseline results.

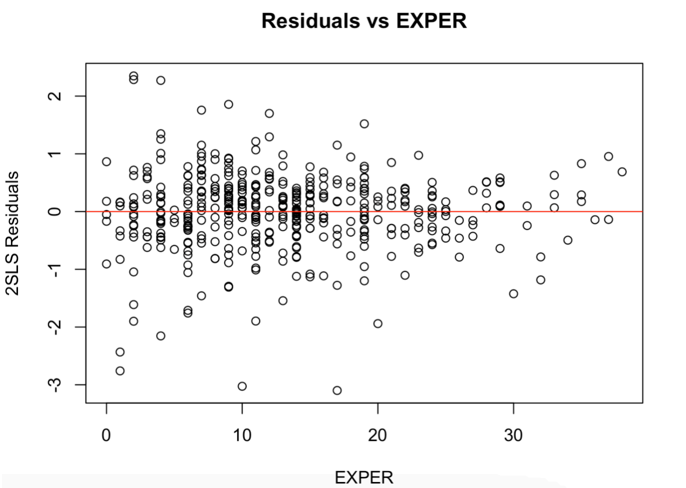
a. Calculate the IV/2SLS residuals, $\hat{e}_{IV}$. Plot them versus *EXPER*. Do the residuals exhibit a pattern consistent with homoskedasticity?

```
Call:
ivreg(formula = log(wage) ~ educ + exper + I(exper^2) | exper +
    I(exper^2) + mothereduc + fathereduc, data = mroz_clean)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0986 -0.3196  0.0551  0.3689  2.3493

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0481003  0.4003281   0.120  0.90442
educ         0.0613966  0.0314367   1.953  0.05147 .
exper        0.0441704  0.0134325   3.288  0.00109 **
I(exper^2)  -0.0008990  0.0004017  -2.238  0.02574 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6747 on 424 degrees of freedom
Multiple R-Squared: 0.1357,     Adjusted R-squared: 0.1296
Wald test: 8.141 on 3 and 424 DF,  p-value: 2.787e-05
```



殘差是否有同質性變異？
在EXPER<5的時候，殘差變化比較大，因此有異質性變異。

b. Regress $\hat{e}_{IV}^2$ against a constant and *EXPER*. Apply the NR^2 test from Chapter 8 to test for the presence of heteroskedasticity.

c. Obtain the IV/2SLS estimates with the software option for Heteroskedasticity Robust Standard Errors. Are the robust standard errors larger or smaller than those for the baseline model? Compute the 95% interval estimate for the coefficient of *EDUC* using the robust standard error.

d. Obtain the IV/2SLS estimates with the software option for Bootstrap standard errors, using $B = 200$ bootstrap replications. Are the bootstrap standard errors larger or smaller than those for the baseline model? How do they compare to the heteroskedasticity robust standard errors in (c)? Compute the 95% interval estimate for the coefficient of *EDUC* using the bootstrap standard error.

b.

```
Call:
lm(formula = iv_resquare ~ exper, data = mroz_clean)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.6740 -0.4341 -0.2685 -0.0168  9.2188 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.676563   0.096573   7.006 9.65e-12 ***
exper       -0.017303   0.006303  -2.745  0.00631 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.049 on 426 degrees of freedom
Multiple R-squared:  0.01738,	Adjusted R-squared:  0.01507 
F-statistic: 7.535 on 1 and 426 DF,  p-value: 0.006308
```

```
NR^2 statistic: 7.438552
P-value: 0.006384122
```

異質性變異檢定：

H_0：殘差是同值性變異

H_1：殘差是異質性變異

P–value：0.006 < 0.05，所以我們拒絕虛無假設，也就是有異質性變異

c.

```
> cat("95% CI for EDUCI (robust SE): [", ci_lower, ",", ci_upper, "]\n")
95% CI for EDUCI (robust SE): [ -0.003947005 , 0.1267403 ]
```

d.

```
Bootstrap SE: 0.03307752
> cat("95% CI for EDUCI (bootstrap): [", ci_boot[1], ",", ci_boot[2], "]\n")
95% CI for EDUCI (bootstrap): [ -0.005394204 , 0.1289739 ]
```