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.

0.1009 = 10.09% 的人母親有大學學歷

0.1076 = 10.76% 的人父親有大學學歷

b.

cor(EDUC, MOTHERCOLL) = 0.3370

cor(EDUC, MOTHERCOLL) = 0.3193

cor(MOTHERCOLL, FATHERCOLL) = 0.3675

c.

```
> summary(iv1)

Call:
ivreg(formula = log(wage) ~ exper + I(exper^2) + educ | 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
exper        0.0433444  0.0134135   3.231  0.00133 **
I(exper^2)  -0.0008711  0.0004017  -2.169  0.03066 *
educ         0.0760180  0.0394077   1.929  0.05440 .
---
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
> # 95% 信賴區間
> confint(iv1)
                    2.5 %        97.5 %
(Intercept) -1.105942034  8.404298e-01
exper        0.017054428  6.963439e-02
I(exper^2)  -0.001658392 -8.385898e-05
educ        -0.001219763  1.532557e-01
```

d.

$EDUC = \beta_0 + \beta_1 MOTHERCOLL$

$H_0: \beta_0 = \beta_1 = 0$

H_1: 至少一個不為0

$F = 63.22$

p-value < 0.05

∴ reject H_0, MOTHERCOLL在此模型具有解釋力

```
> summary(educ.ols)

Call:
lm(formula = educ ~ MOTHERCOLL, data = mroz_clean)

Residuals:
    Min      1Q  Median      3Q     Max
-7.3537 -0.3537 -0.3537  0.7684  4.6463

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.3537     0.1101 112.194  < 2e-16 ***
MOTHERCOLL    2.5117     0.3159   7.951 1.68e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.135 on 426 degrees of freedom
Multiple R-squared:  0.1292,	Adjusted R-squared:  0.1272
F-statistic: 63.22 on 1 and 426 DF,  p-value: 1.68e-14
```

e.

```
Call:
ivreg(formula = log(wage) ~ exper + I(exper^2) + educ | 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
exper        0.0426761  0.0132950   3.210  0.00143 **
I(exper^2)  -0.0008486  0.0003976  -2.135  0.03337 *
educ         0.0878477  0.0307808   2.854  0.00453 **
---
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
```

```
                   2.5 %        97.5 %
(Intercept) -1.04782153  4.896578e-01
exper        0.01661839  6.873386e-02
I(exper^2)  -0.00162779 -6.940599e-05
educ         0.02751845  1.481769e-01
```

f.

```
Call:
lm(formula = educ ~ MOTHERCOLL + FATHERCOLL, data = mroz_clean)

Residuals:
    Min      1Q  Median      3Q     Max
-7.1897 -0.1897 -0.1897  0.8103  4.8103

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.1897     0.1076 113.310  < 2e-16 ***
MOTHERCOLL    1.7436     0.3215   5.423 9.84e-08 ***
FATHERCOLL    2.2031     0.3270   6.737 5.27e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.032 on 425 degrees of freedom
Multiple R-squared:  0.2132,	Adjusted R-squared:  0.2095
F-statistic:  57.6 on 2 and 425 DF,  p-value: < 2.2e-16
```

9

```
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
```

Weak instruments p-value < 0.05 顯著
表示 mothercoll 和 fathercoll 並非弱工具變數
即工具變數與內生變數之間有足夠相關性

根據 Wu-Hausman Test p-value = 0.472 > 0.05, don't reject H_0,
即回歸中的變數可以被認為是外生變數

根據 Sargan Test p-value = 0.626, don't reject H_0, 即工具變數是有效的

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 = risk_premium_msft ~ risk_premium_market, 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    
risk_premium_market 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
```

b.

```
Call:
lm(formula = risk_premium_market ~ 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
```

RANK p-value < 0.05 顯著

$R^2 = 0.9126$

RANK can be regarded as a strong IV.

c.

```
Call:
lm(formula = risk_premium_msft ~ risk_premium_market + residuals_first_stage, 
    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    
risk_premium_market   1.278318   0.126749  10.085   <2e-16 ***
residuals_first_stage -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
```

d

```
Call:
ivreg(formula = risk_premium_msft ~ risk_premium_market | 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    
risk_premium_market 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
```

e

```
Call:
lm(formula = risk_premium_market ~ 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
```

$R^2 = 0.9149$

f.

```
Call:
ivreg(formula = risk_premium_msft ~ risk_premium_market | 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    
risk_premium_market 1.283118   0.127866  10.035   <2e-16 ***

Diagnostic tests:
                 df1 df2 statistic p-value    
Weak instruments   2 177   951.262  <2e-16 ***
Wu-Hausman         1 177     4.862  0.0287 *  
Sargan             1  NA     0.558  0.4549    
---
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
```

Weak instruments p-value < 0.01 ∴ Reject H_0, meaning that at least one instrument is strong.

g.

加入 POS 後，$\beta = 1.283$，與 OLS 不同

原始 OLS 可能有 measure error.

h.

Sargan test p-value = 0.4549 > 0.05

$\therefore$ don't reject H_0 表示工具變數是有效的

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.

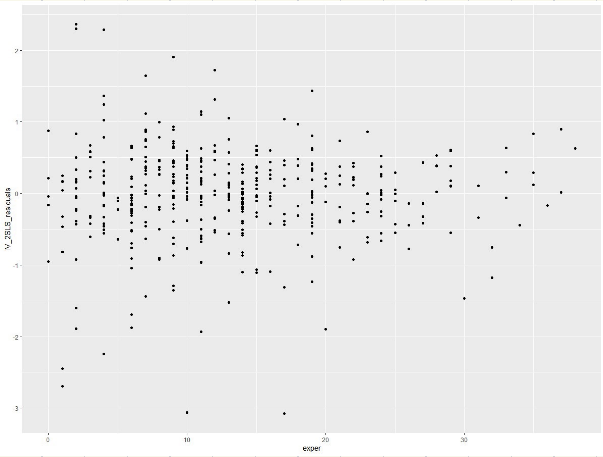
a. Calculate the IV/2SLS residuals, $\hat{e}_{IV}$. Plot them versus *EXPER*. Do the residuals exhibit a pattern consistent with homoskedasticity?

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.

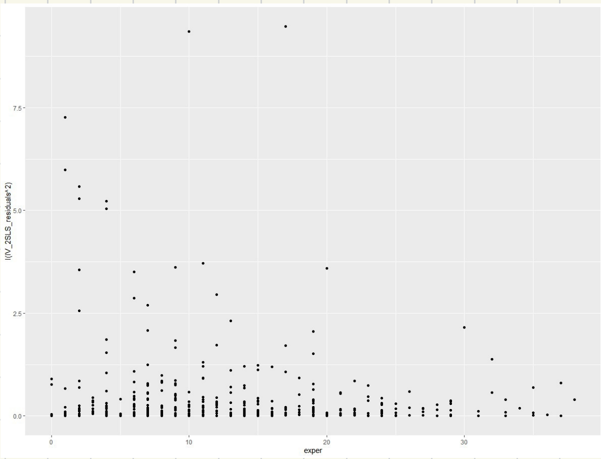
d. Obtain the IV/2SLS estimates with the software option for Bootstrap standard errors, using $B = 200$ bootstrap replications. Are the bootstrap standard errors larger or smaller than those for the baseline model? How do they compare to the heteroskedasticity robust standard errors in (c)? Compute the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

a.



前期殘差大，後期殘差小
可能有 heteroskedasticity

b.




```
> bptest(modres)

	studentized Breusch-Pagan test

data:  modres
BP = 3.4306, df = 1, p-value = 0.064
```

reject H_0. 有 heteroskedasticity

c.

```
Call:
ivreg(formula = log(wage) ~ exper + I(exper^2) + educ | 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
exper        0.0426761  0.0132950   3.210  0.00143 **
I(exper^2)  -0.0008486  0.0003976  -2.135  0.03337 *
educ         0.0878477  0.0307808   2.854  0.00453 **
---
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
```

```
> coeftest(iv2, vcov = vcovHC(iv2, type = "HC1"))

t test of coefficients:

               Estimate  Std. Error t value Pr(>|t|)
(Intercept) -0.27908185  0.43224043 -0.6457 0.518847
exper        0.04267612  0.01548203  2.7565 0.006095 **
I(exper^2)  -0.00084860  0.00042745 -1.9852 0.047762 *
educ         0.08784765  0.03395005  2.5876 0.009998 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> confint(iv2, level = 0.95)
                  2.5 %        97.5 %
(Intercept) -1.04782153  4.896578e-01
exper        0.01661839  6.873386e-02
I(exper^2)  -0.00162779 -6.940599e-05
educ         0.02751845  1.481769e-01
```

d.

bootstarp se > IV/2SLS

C.I. [-0.0047, 0.1221] 包含0

比 robust 大