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?

(sol)

a.

```
> cat("母親有大學教育者比例：", round(pct_mother, 2), "%\n")
母親有大學教育者比例： 12.15 %
> cat("父親有大學教育者比例：", round(pct_father, 2), "%\n")
父親有大學教育者比例： 11.68 %
```

b.

```
> print(round(corr_matrix, 4))
              educ mothercoll fathercoll
educ        1.0000     0.3595     0.3985
mothercoll  0.3595     1.0000     0.3546
fathercoll  0.3985     0.3546     1.0000
```

從結果可見，EDUC 與 MOTHERCOLL、FATHERCOLL 的相關係數分別約為 0.36 和 0.40，說明父母大學教育與子女受教程度之間具備適度的相關性。將父母教育簡化為是否具備大學門檻的二元變數，可減少連續年數的極端值影響，也更符合工具變數的外生性要求，因而比原始的教育年數更適合作為工具變數。

c.

```
> confint(iv_model, level = 0.95)
                   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.

```
Diagnostic tests:
                 df1 df2 statistic  p-value
Weak instruments   1 424    63.563 1.46e-14 ***
Wu-Hausman         1 423     0.742     0.39
Sargan             0  NA        NA       NA
---
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
```

$F = 63.563$ (p-value < 0.001)

MOTHERCOLL是一個強工具變數

e.

```
> confint(iv2_model, level = 0.95)
                 2.5 %        97.5 %
(Intercept) -1.04782153  4.896578e-01
educ         0.02751845  1.481769e-01
exper        0.01661839  6.873386e-02
I(exper^2)  -0.00162779 -6.940599e-05
```

f、

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

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
```

第一階段迴歸中對 MOTHERCOLL 與 FATHERCOLL 的聯合 F 統計量約為 56.96（$p < 2e\text{-}16$），明顯大於常用的 10 標準，顯示這兩個虛擬變數都是強工具。

g、Sargan 過度識別檢定的 p-value 約為 0.626，遠大於 0.05，不拒絕工具變數外生性假設，代表沒有過度識別問題，工具變數有效。

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.

(509) a.

```
Call:
lm(formula = ex_msft ~ ex_mkt, 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    
ex_mkt      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
```

微軟的 $\hat{\beta}\approx1.20$（t=9.84，p<0.001），顯著大於1，表示其報酬對市場波動反應更強，屬於比市場更高風險的股票。

b.

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

Residuals:
      Min        1Q    Median        3Q       Max 
-0.109075 -0.006481  0.000742  0.009279  0.027691 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -7.643e-02  2.187e-03  -34.95   <2e-16 ***
capm5$rank   9.040e-04  2.096e-05   43.14   <2e-16 ***
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01461 on 178 degrees of freedom
Multiple R-squared:  0.9127,	Adjusted R-squared:  0.9122 
F-statistic:  1861 on 1 and 178 DF,  p-value: < 2.2e-16
```

第一階段迴歸中，RANK 的係數極度顯著（t=43.14，p<2e-16），R^2=0.9127 且 F-statistic=1861（遠大於10），顯示 RANK 與市場超額報酬高度相關，是一個非常強的工具變數。

c.

```
Call:
lm(formula = ex_msft ~ ex_mkt + v_hat, data = capm5)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.26981 -0.04469 -0.00894  0.03434  0.35167 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.003041   0.005996   0.507   0.6127    
ex_mkt       1.270690   0.126925  10.011   <2e-16 ***
v_hat       -0.796194   0.430918  -1.848   0.0663 .  
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08028 on 177 degrees of freedom
Multiple R-squared:  0.3645,	Adjusted R-squared:  0.3573 
F-statistic: 50.77 on 2 and 177 DF,  p-value: < 2.2e-16
```

$\hat{v}$ 的係數約為 -0.796（$t = -1.848$，$p = 0.0663$），在 1% 水準下不顯著，因此無法拒絕市場超額報酬為外生的假設。

d.

```
Call:
ivreg(formula = ex_msft ~ ex_mkt | 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    
ex_mkt      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
```

IV/2SLS 的 $\hat{\beta}\approx1.278$（$t=9.99$，$p<0.001$），略高於 OLS 的 1.202，符合「市場報酬有測量誤差時，OLS 會向零偏誤」這一預期。

e.

```
Call:
lm(formula = ex_mkt ~ pos + rank, 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 ***
pos         -0.0092762  0.0042156   -2.20   0.0291 *  
rank         0.0009819  0.0000400   24.55   <2e-16 ***
---
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
```

第一階段迴歸顯示 POS ($p=0.029$) 與 RANK ($p<2e-16$) 均顯著，聯合 F-statistic$\approx$951.3 ($p<2.2e-16$)，且 $R^2\approx0.915$，說明這兩個工具變數聯合起來十分強力。

f.

```
Call:
lm(formula = ex_msft ~ ex_mkt + v_hat2, 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    
ex_mkt       1.283118   0.126344  10.156   <2e-16 ***
v_hat2      -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
```

$\hat{v}_2$ 的係數約為 -0.955（$t = -2.205$，$p = 0.0287$），在 1% 水準下不顯著，因此我們無法拒絕「市場超額報酬為外生」的假設。

g.

```
Call:
ivreg(formula = ex_msft ~ ex_mkt | pos + rank, 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    
ex_mkt      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
```

此係數大於 OLS（1.202）且與單一工具的 IV 結果（1.278）相近，符合「測量誤差會使 OLS 向零偏誤」的預期。

h.

```
> cat("Sargan 統計量 =", round(S, 4), "\n")
Sargan 統計量 = 0.5585
> cat("p-value =", round(p_value, 4), "\n")
p-value = 0.4549
```

故在 5% 水準下不拒絕過度識別檢定的虛無假設，代表多餘的工具變數外生性成立，IV 有效。

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.

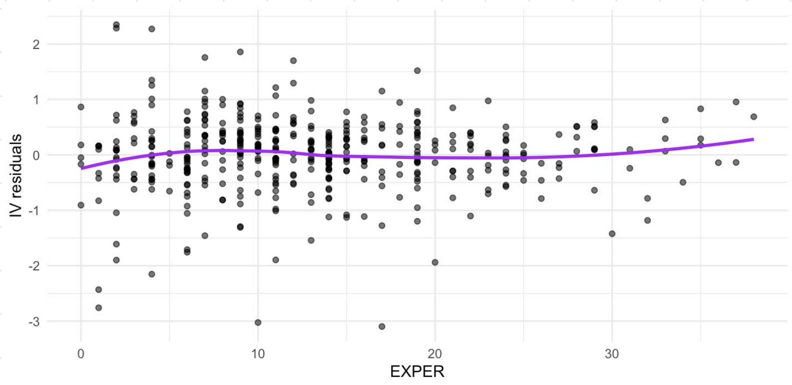
a. Calculate the IV/2SLS residuals, $\hat{e}_{IV}$. Plot them versus *EXPER*. Do the residuals exhibit a pattern consistent with homoskedasticity?

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.

(sol) a.



從殘差點雲圖及紫色趨勢線看，殘差在各個 EXPER 值附近散佈並未明顯呈現漏斗形（即「變異隨 EXPER 而系統上升或下降」），大致維持恆定，與同質變異性（homoskedasticity）相符。

b.

```
> cat("NR² =", round(NR2, 4), "\n")
NR² = 7.4386
> cat("p-value =", round(p_value, 4), "\n")
p-value = 0.0064
```

reject H₀、有 Hetroskedasticity

c.

```
95% confidence interval for EDUC: -0.0039 to 0.1267
```

d.

```
> cat("EDUC 係數 (bootstrap):", round(boot_est, 4), "\n")
EDUC 係數 (bootstrap): 0.0641
> cat("Bootstrap SE:", round(boot_se, 4), "\n")
Bootstrap SE: 0.0323
> cat("95% C.I. (normal method):",
+     round(boot_ci$normal[2], 4), "to", round(boot_ci$normal[3], 4), "\n")
95% C.I. (normal method): -0.0047 to 0.1221
```

Bootstrap 結果顯示 EDUC 係數約為 0.0641，對應的 bootstrap SE ≈ 0.0323，略大於基準模型的常規 SE 但略小於(c)中計算出的 robust SE；利用常態近似法得 95% 信賴區間為 [– 0.0047, 0.1221] 。