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?

母親有大學 (>12 年)：12.15% 父親有大學 (>12 年)：11.68%

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*?

	educ	MOTHERCOLL	FATHERCOLL
educ	1.000	0.359	0.398
MOTHERCOLL	0.359	1.000	0.355
FATHERCOLL	0.398	0.355	1.000

這兩個變數跟 educ 有一定關聯，所以可以當作工具變數來用。父母有沒有大學學歷，通常代表家庭的資源、地位或對教育的重視，會影響子女的受教育程度，但不會直接影響子女的薪水

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*?

(c) educ 係數 95% CI: [-0.0012, 0.1533]

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?

	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	***

因為 F 統計量大於 10，符合強工具的標準，因此認為 MOTHERCOLL 是一個強工具變數

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)?

(e) educ 係數 95% CI: [0.0275, 0.1482] 比 c 還要窄

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?

	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	***

F 為 56.963>10，p 值 < 2.2e-16，代表這兩個變數對 educ 有強影響，是夠強的工具變數

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

df1	df2	statistic	p-value
1.0000000	NA	0.2375850	0.6259558

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?

```
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    
x           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
```

beta 估計值約為 1.20，且高度顯著（$p < 0.001$）。因為 beta > 1，代表微軟在這段期間相對市場投資組合是偏高風險的股票，報酬波動大於市場，屬於風險較高的資產

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?

```
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 在第一階段回歸中非常顯著，R^2=0.91，顯示它是非常強的工具變數，適合作為 IV 使用。

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?

```
Call:
lm(formula = y ~ x + v_hat, 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    
x            1.278318   0.126749  10.085   <2e-16 ***
v_hat       -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 值約 0.043，大於 1% 顯著水準，因此我們無法在 1% 顯著水準下拒絕市場報酬 exogenous 的假設，也就是市場報酬可以被視為外生變數

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?

```
Call:
ivreg(formula = y ~ x | 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    
x           1.278318   0.128011   9.986   <2e-16 ***

Diagnostic tests:
                 df1 df2 statistic p-value    
Weak instruments   1 178  1857.587  <2e-16 ***
Wu-Hausman         1 177     4.164  0.0428 *  
Sargan             0  NA        NA      NA    
---
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
```

用 RANK 做 IV 估計後，x 的係數是 1.28，與 (a) 題 OLS 估計的 1.20 接近，符合預期，說明 IV 模型的估計結果穩健。

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?

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

結果顯示 RANK 和 POS 聯合顯著，R^2=0.9149，代表這兩個工具變數非常強，可以放心用作 IV。

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?

```
Call:
lm(formula = y ~ x + 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    
x            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
```

v_hat2 的 p 值約 0.029，在 1% 顯著水準下，我們無法拒絕市場報酬是外生的假設

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?

```
Call:
ivreg(formula = y ~ x | 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    
x           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
```

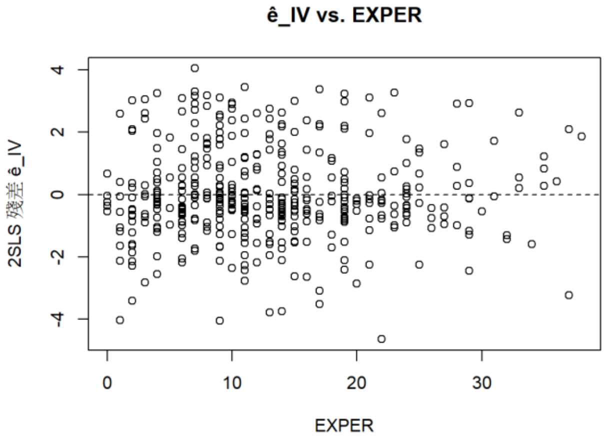
用 RANK 和 POS 當工具跑出來的結果跟 OLS 差不多，表示市場超額報酬的影響很穩定，符合預期。

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.

```
      df1       df2 statistic   p-value 
1.0000000        NA 0.5584634 0.4548800 
```

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

```
n = 428
NR2 = 0.2409118
p-value = 0.6235484
R2 = 0.0005628779
```

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.

```
t test of coefficients:

             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.5941566 20.1943995  0.3265   0.7442
educ        -0.4678190  1.6316602 -0.2867   0.7745
exper        0.0740668  0.0968656  0.7646   0.4449
I(exper^2)  -0.0019068  0.0032607 -0.5848   0.5590

> # 計算 EDUC 95% 信賴區間 ( robust )
> b_educ <- coef(iv_base)["educ"]
> se_educ_rob <- sqrt(vcovHC(iv_base, type = "HC1")
> CI_robust <- b_educ + qnorm(c(0.025, 0.975)) * se
> cat(sprintf("EDUC 95%% CI (robust): [%.4f, %.4f]\
EDUC 95% CI (robust): [-3.6658, 2.7302]
```

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.

```
Bootstrap SE (educ) = 0.7327437
> cat("95% CI (educ, bootstrap) =", boot_CI[1],
boot_CI[2], "\n")
95% CI (educ, bootstrap) = -1.013175 - 1.617065
```