CH 10 Q18

(a)

```
> mean(mroz_lfp$parentcoll)
[1] 0.1869159
```

父母有18.69%的比例有部分大學教育

(b)

```
> cor_matrix
                 educ mothercoll fathercoll
educ        1.0000000  0.3594705  0.3984962
mothercoll  0.3594705  1.0000000  0.3545709
fathercoll  0.3984962  0.3545709  1.0000000
```

mothercoll 和 fathercoll 保留父母最高學歷的資訊，

去除可能與誤差項相關的部分. 因此可能比

mothereduc 和 fathereduc 更適合做 IV

(c)

```
> confint(iv_model, level = 0.95)["educ", ]
       2.5 %        97.5 %
-0.001219763  0.153255678
```

95% C.I. = [−0.0012 , 0.1533]

(d)

```
> anova(first_stage)
Analysis of Variance Table

Response: educ
             Df  Sum Sq Mean Sq F value    Pr(>F)
exper         1    0.52   0.516  0.1133    0.7366
I(exper^2)    1   10.46  10.464  2.2990    0.1302
mothercoll    1  289.32 289.317 63.5631 1.455e-14 ***
Residuals   424 1929.90   4.552
```

F-test statistic : F值大於10為強工具變數

mothercoll 為強工具變數

(e)

```
> confint(iv_model_2, level = 0.95)["educ", ]
     2.5 %     97.5 %
0.02751845 0.14817686
```

用 mothercoll 和 fathercoll 做工具变數，CI 变窄

(f)

```
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 ***
---
```

```
> linearHypothesis(first_stage_2, c("mothercoll = 0", "fathercoll

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
```

对 mothercoll 和 fathercoll 做联合 F 检定，為強 IV

(g)

```
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
```

Sargan test 的 H_0 : IV 皆有效

p-value = 0.626 ⇒ do not reject H_0

兩者皆為有效的工具變數

CH 10 Q 20

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)

```
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 ***
---
```

$\beta = 1.2$. Microsoft Stock 比市場投組更 Risky

(b)

```
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 和 market excess return 有顯著相關. 且 F 統計量 > 10

是強 IV

(c)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.003018   0.005984   0.504   0.6146
mkt_excess   1.278318   0.126749  10.085   <2e-16 ***
v           -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
```

$\hat{v}$ 在 99% 顯著水準下不顯著, 無法拒絕 market excess 是外生

但在 95% 水準下, 可以拒絕 market excess 為外生變數

(d)

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.003018   0.006044   0.499    0.618
mkt_excess  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
```

t值變得更顯著，可能消除了內生性

(e)

```
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: mkt_excess ~ 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 ‘
```

第一階迴歸的 $R^2 = 0.91$

聯合檢驗結果顯著，皆為強IV

(f)

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.003004   0.005972   0.503   0.6157
mkt_excess  1.283118   0.126344  10.156   <2e-16 ***
v_2        -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
```

$\hat{v}_2$ 在99%水準下不顯著，不拒絕 H_0：market excess 是外生

但在95%水準下，可以拒絕 market excess 是外生變數

(g)

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.003004   0.006044   0.497     0.62
mkt_excess  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
```

係數與t值都增加，IV解決內生性問題

(b)

```
> cat("NR-square:", J_stat, "\n")
NR-square: 0.5584634
> cat("P-value:", p_value, "\n")
P-value: 0.45488
>
```

p-value > 0.05 , do not reject H_0

∴ IV 都有效

CH 10 Q 24

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.

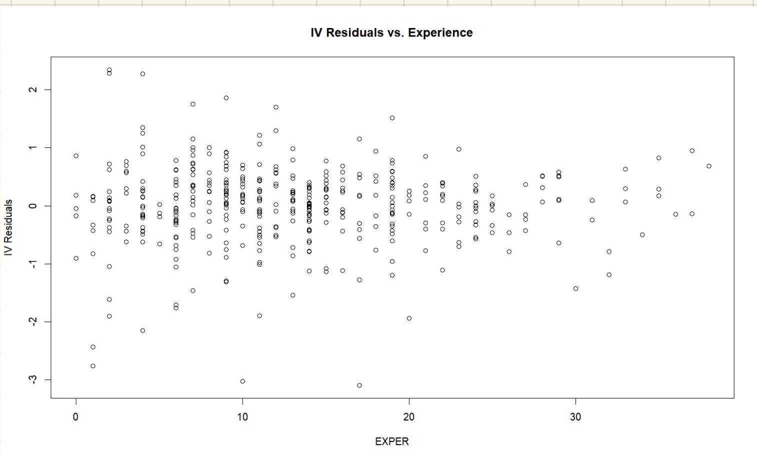
a. Calculate the IV/2SLS residuals, $\hat{e}_{IV}$. Plot them versus *EXPER*. Do the residuals exhibit a pattern consistent with homoskedasticity?

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)



Residual 隨著 EXPER 增加，分散程度降低

(b)

```
> pchisq(NR2_stat, df = 1, lower.tail = FALSE)
[1] 0.006384122
> p_value <- pchisq(NR2_stat, df = 1, lower.tail = FALSE)
> cat("NR2 test stastic =", NR2_stat, "\n")
NR2 test stastic = 7.438552
> cat("p-value =", p_value, "\n")
p-value = 0.006384122
```

p-value is very small，拒絕 H_0，

存在 heteroskedasticity

(c)

```
> print(se_comparison)
            Estimate Baseline_SE Robust_SE Increased_SE
(Intercept)  0.04810     0.40033   0.42980          Yes
exper        0.04417     0.01343   0.01555          Yes
exper2      -0.00090     0.00040   0.00043          Yes
educ         0.06140     0.03144   0.03334          Yes
> cat("95% Robust CI for EDUC: [", round(ci_lower, 4), ",", round(ci_upper, 4), "]\n")
95% Robust CI for EDUC: [ -0.0039 , 0.1267 ]
```

Robust SE is larger than the baseline model

(d)

```
> print(se_compare)
                   Coef Baseline_SE Robust_SE Bootstrap_SE Larger_than_Baseline_SE
(Intercept) (Intercept)     0.40033   0.42980      0.42199                     Yes
exper             exper     0.01343   0.01555      0.01640                     Yes
exper2           exper2     0.00040   0.00043      0.00047                     Yes
educ               educ     0.03144   0.03334      0.03291                     Yes
            Larger_than_Robust_SE
(Intercept)                    No
exper                         Yes
exper2                        Yes
educ                           No
```

```
> (boot_ci_educ <- boot.ci(boot_result, type = "norm", index = 4))
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 200 bootstrap replicates

CALL :
boot.ci(boot.out = boot_result, type = "norm", index = 4)

Intervals :
Level      Normal
95%   (-0.0055,  0.1235 )
Calculations and Intervals on Original Scale
```

Bootstrap SE is larger than baseline SE

exper , exper 2 bs SE is larger than robust SE