

# 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?

Ans.

母親受過大學教育 (*MOTHERCOLL* = 1) 的比例為 12.15%

父親受過大學教育 (*FATHERCOLL* = 1) 的比例為 11.68%

```
> mean(mroz_sub$mothercoll, na.rm = TRUE)
[1] 0.1214953
> mean(mroz_sub$fathercoll, na.rm = TRUE)
[1] 0.1168224
```

- 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*?

Ans.

*EDUC* 與 *MOTHERCOLL* 的相關係數為 0.3595

*EDUC* 與 *FATHERCOLL* 的相關係數為 0.3985

	educ	mothercoll	fathercoll
educ	1.0000000	0.3594705	0.3984962
mothercoll	0.3594705	1.0000000	0.3545709
fathercoll	0.3984962	0.3545709	1.0000000

相關係數的大小是否重要：父母是否上過大學與子女教育年數呈現中度正相關，顯示 *MOTHERCOLL* 與 *FATHERCOLL* 在統計上具備作為工具變數的「相關性條件 (Relevance)」。

為什麼 *MOTHERCOLL* 和 *FATHERCOLL* 可能更好：*MOTHERCOLL* 和 *FATHERCOLL* 是離散的二元變數，只揭示父母是否有大學經歷，資訊較粗略，較不容易直接影響工資，因此更有可能滿足工具變數的外生性條件 (Exogeneity)。

- 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*?

Ans.

*EDUC* 的 95% 信賴區間為 [-0.0012, 0.1533]

```
> confint(iv_model, "educ", level = 0.95)
          2.5 %    97.5 %
educ -0.001219763 0.1532557
```

- 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?

Ans.  $\hat{EDUC} = 12.0791 + 0.0562EXPER - 0.00196EXPER^2 + 2.5171MOTHERCOLL + \nu$

```
Call:
lm(formula = educ ~ exper + exper2 + mothercoll, data = mroz_sub)

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
exper2       -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 + exper2 + 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
```

$H_0 : MOTHERCOLL \text{係數} = 0 \quad H_1 : MOTHERCOLL \text{係數} \neq 0$

F統計量 = 63.563 > 10, 拒絕  $H_0$ , 表明  $MOTHERCOLL$  對  $EDUC$  之間存在高度顯著的關係。

$MOTHERCOLL$  是強工具變數, 可以可靠地用於 IV 估計。

- 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)?

Ans.

$EDUC$  的 95% 信賴區間為 [0.0275, 0.1482]

與 10.18c 的比較: 使用兩個工具變數的信賴區間更窄, 這表明額外的工具變數確實提高了估計效率。

```
> confint(iv_model_both, "educ", level = 0.95)
                2.5 %    97.5 %
educ 0.02751845 0.1481769
```

- 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?

Ans.  $\hat{EDUC} = 11.89026 + 0.04915EXPER - 0.00145EXPER^2$   
 $+ 1.74995MOTHERCOLL + 2.18661FATHERCOLL + \nu$

```
Call:
lm(formula = educ ~ exper + exper2 + mothercoll + fathercoll,
    data = mroz_sub)

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
exper2      -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 + exper2 + 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
```

$H_0 : MOTHERCOLL \text{係數} = 0, FATHERCOLL \text{係數} = 0$

$H_1 : MOTHERCOLL \text{係數} \neq 0, FATHERCOLL \text{係數} \neq 0$

F統計量 = 56.963 > 10, 拒絕  $H_0$ , 表明  $MOTHERCOLL$  和  $FATHERCOLL$  對  $EDUC$  之間存在高度顯著的關係。 $MOTHERCOLL$  和  $FATHERCOLL$  是強工具變數, 可以可靠地用於 IV 估計。

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

Ans.

Sargan 檢定 p值 = 0.626 > 0.05: 無法拒絕"所有工具變數都有效"的虛無假設。表明沒有證據顯示工具變數無效。支持工具變數的有效性。

```
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
```