HW0428

18.

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.

建立兩個虛擬變數：

- MOTHERCOLL = 1 若母親教育年數（MOTHEREDUC）> 12，否則為 0。
- FATHERCOLL = 1 若父親教育年數（FATHEREDUC）> 12，否則為 0。

母親或父親教育年數大於 12 表示他們有某種形式的大學教育。計算這兩個變數的平均數，即可得知有多少比例的父母有大學教育背景（例如：mean(MOTHERCOLL) 表示母親有大學教育的比例）。

B.

計算三個變數之間的皮爾森相關係數（cor()）：EDUC 與 MOTHERCOLL、FATHERCOLL。

若 MOTHERCOLL 與 EDUC 的相關係數高，代表 MOTHERCOLL 是一個有力的工具變數（IV）。不過，太高的相關性也可能意味著潛在的內生性風險。

邏輯上，MOTHERCOLL 和 FATHERCOLL 可能是比 MOTHEREDUC 與 FATHEREDUC 更佳的工具變數，因為將連續變

數轉為虛擬變數可能會減少 measurement error，並更清楚地傳達是否有高等教育的效果。

C.
進行兩階段最小平方法（2SLS）估計：

- 第一階段：以 MOTHERCOLL 解釋 EDUC。
- 第二階段：以預測的 EDUC 解釋工資（WAGE）。

取回 EDUC 的係數與其 95% 信賴區間，這將告訴我們母親是否受過大學教育對孩子教育的影響進而影響工資。

D.
檢驗 MOTHERCOLL 是否為強工具變數：在第一階段迴歸中使用 F 檢定檢驗 MOTHERCOLL 是否顯著影響 EDUC。F 值若大於 10，通常被視為強工具變數。若 F 值很小，表示 MOTHERCOLL 可能是弱工具。

E.
改用 MOTHERCOLL 與 FATHERCOLL 兩個工具變數做 2SLS 估計，並取 EDUC 的係數與其 95% 信賴區間。

比較這組信賴區間與 (c) 小題，通常兩個工具變數會讓估計更精確，但若工具變數品質不佳，也可能導致區間更寬。

F.
針對第一階段迴歸，檢驗 MOTHERCOLL 與 FATHERCOLL 的聯合顯著性（joint significance），檢驗它們是否共同強烈地解釋 EDUC。

若 F 統計量顯著，表示這兩個工具變數合適。

G.
進行超額識別檢定（overidentification test），例如 Hansen J 檢定或 Sargan 檢定，用來驗證是否所有工具變數都有效（工具變數是否外生）。

若檢定結果顯示不顯著，表示沒有證據反對工具變數的有效性。

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.

以 OLS 對 CAPM 模型作迴歸：

rj−rf=αj+βj(rm−rf)+εjr_j - r_f = \alpha_j + \beta_j(r_m - r_f) + \varepsilon_jrj−rf=αj+βj(rm−rf)+εj

估計 β 值來衡量 Microsoft 相對於市場的風險。

- 若 β > 1：風險高於市場，屬於高風險股票。
- 若 β < 1：風險低於市場，屬於防禦型股票。

B.

構建 IV：將 (r_m - r_f) 按大小排序，分配排序變數 RANK = 1, ..., 180 作為工具變數。此排序變數符合工具變數三大條件的潛力（相關性、外生性、排除限制）。

進行第一階段迴歸檢查 RANK 對 (r_m - r_f) 的解釋力與顯著性，若係數顯著、R^2高，表示 RANK 是強工具。

C.
計算第一階段殘差 v^\hat{v}v^，將其加入 CAPM 模型檢驗是否顯著。若 v^\hat{v}v^ 在第二階段迴歸中顯著，則市場報酬可能是內生變數（說明 OLS 估計有偏誤）。

D.
使用 2SLS 估計 CAPM 模型，以 RANK 為工具變數，比較此 β 值與 OLS 所得 β 是否相近。若相近，表示內生性問題不大。

E.
新增虛擬變數 POS = 1 若 (r_m - r_f) > 0，否則為 0，與 RANK 一起作為工具變數。再進行第一階段回歸，檢驗是否共同顯著，是否具有解釋力（看 R^2）。

F.
執行 Hausman 檢定，檢驗內生性：若顯著，表示 (r_m - r_f) 是內生變數，應該使用 IV/2SLS 而非 OLS。

G.
以 RANK 和 POS 為工具變數，取得 2SLS β 值及其信賴區間。若 β 與 OLS 結果差異不大，表示內生性影響不大。

H.
使用 Sargan 檢定檢驗超額識別限制。若不顯著，表示兩個工具變數都是有效的（外生性沒被拒絕）。

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.

計算 2SLS 的殘差 ε^IV\hat{\varepsilon}_{IV}ε^IV，並繪製與 EXPER 的散佈圖。若殘差與 EXPER 呈現明顯圖案（如扇形），表示可能存在異質變異（heteroskedasticity）。

B.

對 ε^IV2\hat{\varepsilon}^2_{IV}ε^IV2 對常數與 EXPER 進行迴歸，檢驗 EXPER 的係數是否顯著。計算 NR^2 統計量（等於樣本數乘以該迴歸的 R^2），對照卡方分配檢定異質變異。

C.

使用異質變異-穩健（robust）標準誤重新估計 2SLS，檢查標準誤是否變大，並計算 EDUC 的 95% 信賴區間。

D.

使用 Bootstrap（重抽樣，B=200 次）方法計算 2SLS 的標準誤與信賴區間。觀察 bootstrap 標準誤是否比 robust 或 baseline 小或大，並得出 EDUC 係數的 95% 信賴區間。