3137-7-049 俞懷蘊

10.2 The labor supply of married women has been a subject of a great deal of economic research. Consider the following supply equation specification

$$HOURS = \beta_1 + \beta_2 WAGE + \beta_3 EDUC + \beta_4 AGE + \beta_5 KIDSL6 + \beta_6 NWIFEINC + e$$

where *HOURS* is the supply of labor, *WAGE* is hourly wage, *EDUC* is years of education, *KIDSL6* is the number of children in the household who are less than 6 years old, and *NWIFEINC* is household income from sources other than the wife's employment.

a. Discuss the signs you expect for each of the coefficients.
b. Explain why this supply equation cannot be consistently estimated by OLS regression.
c. Suppose we consider the woman's labor market experience *EXPER* and its square, $EXPER^2$, to be instruments for *WAGE*. Explain how these variables satisfy the logic of instrumental variables.
d. Is the supply equation identified? Explain.
e. Describe the steps [not a computer command] you would take to obtain IV/2SLS estimates.

a. 、
1. beta 2（WAGE 的係數）>0
一般來說，工資越高，個人可能更願意工作（替代效應），因此預期 > 0。但如果收入效應主導（工資高到一定程度後，個人可能選擇減少工作以享受更多休閒），則可能為負。
2. beta 3（EDUC 的係數）>0
教育年數（EDUC）通常與更高的技能和生產力相關，這可能增加勞動供給（因為高教育者可能有更好的工作機會）。
3. beta 4（AGE 的係數）不確定
年齡（AGE）的影響可能因年齡階段而異。年輕時，勞動供給可能隨年齡增加而增加，但到一定年齡後（例如接近退休），勞動供給可能減少。對於已婚女性，考慮到家庭責任，年齡增加可能導致勞動供給減少（特別是在中年時），可能為負，但這取決於樣本年齡分佈。
4. beta 5（KIDS6 的係數）<0
6歲以下的子女數量（KIDS6）通常會增加照顧責任，特別是對已婚女性來說，這可能減少勞動供給（因為需要更多時間照顧孩子）。
5. beta 6（NWIFEINC 的係數）<0
除妻子收入外的家庭收入（NWIFEINC）越高，妻子可能減少勞動供給，因為家庭經濟壓力降低，她可以選擇更多休閒或家庭時間（收入效應）。
b.
1. 內生性問題（Endogeneity）
工資（WAGE）很可能是一個內生變數。勞動供給（HOURS）和工資（WAGE）之間存在雙向因果關係：工資會影響勞動供給，但勞動供給（例如工作小時數）也可能影響工資（例如，通過工作經驗或勞動市場的供需）。這種雙向因果導致 WAGE 與誤差項 e 相關，違反 OLS 的基本假設（自變數與誤差項不相關）。
2. 省略變數偏誤（Omitted Variable Bias）
方程中可能存在未觀察到的變數（例如個人能力、健康狀況或偏好），這些變數可能同時影響 HOURS 和 WAGE。如果這些變數未包含在模型中，它們會進入誤差項 e，導致 WAGE 與 e 相關。
由於上述原因，OLS 估計會有偏誤且不一致，需要使用工具變數（Instrumental Variables, IV）方法來解決內生性問題。

c.
工具變數（IV）需要滿足兩個條件：
1. 相關性（Relevance）：工具變數必須與內生變數（WAGE）顯著相關。
2. 外生性（Exogeneity）：工具變數不能與誤差項 e 相關（即不能直接影響 HOURS，只能通過 WAGE 影響 HOURS）。
分析 EXPER 和 $EXPER^2$ 是否滿足條件：
1. 相關性
勞動市場經驗（EXPER）通常與工資（WAGE）高度相關，因為更多的經驗往往意味著更高的技能和生產力，從而導致更高的工資。$EXPER^2$（經驗的平方）則捕捉了經驗與工資之間的非線性關係（例如，經驗對工資的影響可能隨著經驗增加而減弱）。
因此，EXPER 和 $EXPER^2$ 很可能是與 WAGE 相關的，滿足相關性條件。
2. 外生性
EXPER 和 $EXPER^2$ 需要與誤差項 e 不相關。勞動市場經驗通常是過去的累積（例如，過去的工作年數），因此它在某種程度上是外生的，不會直接受到當前勞動供給（HOURS）或未觀察到的當前因素（例如當前的健康狀況或偏好）影響。

d.

工具變數的數量必須至少等於內生變數的數量（這是必要條件，稱為秩序條件，order condition）。這裡我們有 1 個內生變數（WAGE）和 2 個工具變數（EXPER 和 EXPER²），因此滿足秩序條件。
此外，工具變數還需滿足前面提到的相關性和外生性條件（即秩條件，rank condition）。假設 EXPER 和 EXPER² 與 WAGE 充分相關，且與誤差項不相關，則方程可以識別。

e.

1. **第一階段（First Stage）**
   對內生變數 $WAGE$ 進行回歸，使用工具變數 $EXPER$ 和 $EXPER$　以及其他外生變數（$EDUC$、AGE、$KIDS6$、$NWIFEINC$）作為解釋變數：
   $$WAGE = \pi_0 + \pi_1 \cdot EXPER + \pi_2 \cdot EXPER \;\; + \pi_3 \cdot EDUC + \pi_4 \cdot AGE + \pi_5 \cdot KIDS6 + \pi_6 \cdot NWIFEINC + v$$
   從中得到 $WAGE$ 的預測值 $W\hat{A}GE$。
2. **第二階段（Second Stage）**
   將第一階段得到的 $W\hat{A}GE$ 代入原始的勞動供給方程，然後用 OLS 估計：
   $$HOURS = \beta_1 + \beta_2 \cdot W\hat{A}GE + \beta_3 \cdot EDUC + \beta_4 \cdot AGE + \beta_5 \cdot KIDS6 + \beta_6 \cdot NWIFEINC + e$$
   這樣得到的 β_2（以及其他係數）是工具變數法下的估計值，能夠解決 $WAGE$ 的內生性問題。

10.3 In the regression model $y = \beta_1 + \beta_2 x + e$, assume x is endogenous and that z is a valid instrument. In Section 10.3.5, we saw that $\beta_2 = \text{cov}(z, y)/\text{cov}(z, x)$.

a. Divide the denominator of $\beta_2 = \text{cov}(z, y)/\text{cov}(z, x)$ by $\text{var}(z)$. Show that $\text{cov}(z, x)/\text{var}(z)$ is the coefficient of the simple regression with dependent variable x and explanatory variable z, $x = \gamma_1 + \theta_1 z + v$. [*Hint*: See Section 10.2.1.] Note that this is the first-stage equation in two-stage least squares.

b. Divide the numerator of $\beta_2 = \text{cov}(z, y)/\text{cov}(z, x)$ by $\text{var}(z)$. Show that $\text{cov}(z, y)/\text{var}(z)$ is the coefficient of a simple regression with dependent variable y and explanatory variable z, $y = \pi_0 + \pi_1 z + u$. [*Hint*: See Section 10.2.1.]

c. In the model $y = \beta_1 + \beta_2 x + e$, substitute for x using $x = \gamma_1 + \theta_1 z + v$ and simplify to obtain $y = \pi_0 + \pi_1 z + u$. What are π_0, π_1, and u in terms of the regression model parameters and error and the first-stage parameters and error? The regression you have obtained is a **reduced-form** equation.

d. Show that $\beta_2 = \pi_1/\theta_1$.

e. If $\hat{\pi}_1$ and $\hat{\theta}_1$ are the OLS estimators of π_1 and θ_1, show that $\hat{\beta}_2 = \hat{\pi}_1/\hat{\theta}_1$ is a consistent estimator of $\beta_2 = \pi_1/\theta_1$. The estimator $\hat{\beta}_2 = \hat{\pi}_1/\hat{\theta}_1$ is an **indirect least squares** estimator.

a.

$x = r_1 + \theta_1 z + v$

$E(x) = r_1 + \theta_1 E(z)$

$x - E(x) = \theta_1 (z - E(z)) + v$

$(z - E(z))(x - E(x)) = \theta_1 (z - E(z))^2 + v(z - E(z))$

$E[(z - E(z))(x - E(x))] = \theta_1 E[(z - E(z))^2] + vE(z - E(z))$

Assuming $vE(z - E(z)) = 0 \Rightarrow E[(z - E(z))(x - E(x))] = \theta_1 E[(z - E(z))^2]$

Solve $\theta_1 = \dfrac{E[(z - E(z))(x - E(x))]}{E[(z - E(z))^2]} = \dfrac{\text{Cov}(z, x)}{\text{Var}(z)}$ #

This is the OLS estimator of θ_1 in the 1st regression $x = r_1 + \theta_1 z + v$

b.

$y = \pi_0 + \pi_1 z + u$

$E(y) = \pi_0 + \pi_1 E(z)$

$y - E(y) = \pi_1 (z - E(z)) + u$

$(z - E(z))(y - E(y)) = \pi_1 (z - E(z))^2 + u(z - E(z))$

$E(z - E(z))(y - E(y)) = \pi_1 E\left[(z - E(z))^2\right] + u E(z - E(z))$

Assuming $E(z - E(z)) = 0 \Rightarrow E(z - E(z))(y - E(y)) = \pi_1 E\left[(z - E(z))^2\right]$

Solve $\pi_1 = \dfrac{E(z - E(z))(y - E(y))}{E\left[(z - E(z))^2\right]} = \dfrac{Cov(z, y)}{Var(z)}$

This is the OLS estimation of π_1 in the regression $y = \pi_0 + \pi_1 z + u$

c. $y = \beta_1 + \beta_2 x + e$

x用1st regression : $x = r_1 + \theta_1 z + v$ 代入原 model

$y = \beta_1 + \beta_2 (r_1 + \theta_1 z + v) + e = (\beta_1 + \beta_2 r_1) + \beta_2 \theta_1 z + (\beta_2 v + e)$

compare with $y = \pi_0 + \pi_1 z + u$, we got $\begin{cases} \pi_0 = \beta_1 + \beta_2 r_1 \\ \pi_1 = \beta_2 \theta_1 \\ u = \beta_2 v + e \end{cases}$

d. $\pi_1 = \beta_2 \theta_1 \Rightarrow \beta_2 = \dfrac{\pi_1}{\theta_1}$

e.

$\theta_1 = \dfrac{E\left[(z - E(z))(x - E(x))\right]}{E\left[(z - E(z))^2\right]} = \dfrac{Cov(z, x)}{Var(z)}$

$\hat{\theta}_1 = \dfrac{\widehat{Cov}(z, x)}{\widehat{Var}(z)} = \dfrac{\Sigma(z_i - \bar{z})(x_i - \bar{x}) \cancel{/N}}{\Sigma(z_i - \bar{z})^2 \cancel{/N}}$, This estimator is consistent if z is uncorrelated with v

$\pi_1 = \dfrac{E(z - E(z))(y - E(y))}{E\left[(z - E(z))^2\right]} = \dfrac{Cov(z, y)}{Var(z)}$

$\hat{\pi}_1 = \dfrac{\widehat{Cov}(z, y)}{\widehat{Var}(z)} = \dfrac{\Sigma(z_i - \bar{z})(y_i - \bar{y})}{\Sigma(z_i - \bar{z})^2}$, This estimator is consistent if z is uncorrelated with u

$\beta_2 = \dfrac{\pi_1}{\theta_1}$

$\hat{\beta}_2 = \dfrac{\hat{\pi}_1}{\hat{\theta}_1} = \dfrac{\Sigma(z_i - \bar{z})(y_i - \bar{y}) \div N}{\Sigma(z_i - \bar{z})(x_i - \bar{x}) \div N} = \dfrac{\widehat{Cov}(z, y)}{\widehat{Cov}(z, x)}$

if $N \to \infty$, $\widehat{Cov}(z, y) \xrightarrow{P} Cov(z, y)$; $\widehat{Cov}(z, x) \xrightarrow{P} Cov(z, x)$

$\therefore \hat{\beta}_2 = \dfrac{\widehat{Cov}(z, y)}{\widehat{Cov}(z, x)} \xrightarrow{P} \dfrac{Cov(z, y)}{Cov(z, x)}$