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.

1.WAGE: 預期：$\beta_2 > 0$，表示工資越高，機會成本越高，替代效果讓女性願意投入更多工作時速

EDUC: 預期：$\beta_3 > 0$，表示教育提升→預期工資或職業選擇更好→刺激勞動供給

AGE: 預期：$\begin{cases}\beta_4 > 0\text{，年輕}\\ \beta_4 < 0\text{，老年}\end{cases}$，表示年輕時經驗累積供給增加，但老時供給可能下降（呈倒 U 形）

KIDSL6: 預期：$\beta_5 < 0$，表示需要照顧小孩時間，照顧多，上班時間減少

NWIFEINC: 預期：$\beta_6 < 0$，表示家庭收入越高，女性越不用出來工作

2.無法以 OLS 一致估計的原因有：

1.同時性：WAGE 同時由勞動需求跟供給共同決定，誤差項包含了影響女性工作的偏好因素，這些因素會影響工資的談判結果，因此$Cov(WAGE, e) \neq 0$

2.遺漏變數：如果漏掉影響工資跟工作意願的第三因素，而誤差項中如果有混入這類影響，就會造成內生性

3.測量誤差：工資如果有測量誤差，也會讓 OLS 估計量偏向 0

因此 OLS 估計會產生偏誤而且不一致的情況下，就需要用到工具變數

3.把 EXPER 跟 $EXPER^2$ 做為 WAGE 的工具變數下

EXPER 跟 $EXPER^2$ 跟個人技能、市場談判能力呈現高度正相關，因此與 WAGE 也具有高度相關

可以透過回歸式加入 EXPER 跟 $EXPER^2$ 去做檢驗，要求 F-static 要顯著大於 10，這樣就可以證明工具變數不弱

而經驗年數理論上只透過提高工資而影響勞動供給，並不會直接改變工作意願，因此會假設$Cov(\{\text{EXPER}, \text{EXPER2}\}, e) = 0$

4.內生變數：WAGE、工具變數：EXPER 跟 $EXPER^2$，Over-identified：當工具變數>內生變數時，可以識別，而且還可以進行過度識別檢定

5. 第一階段：將內生變數 WAGE 用所有外生變數與工具變數回歸：$\widehat{WAGE_\iota} = \widehat{\pi_1} + \widehat{\pi_2} EXPER + \widehat{\pi_3} XPER^2 \ldots\ldots + \widehat{u_\iota}$

檢查工具變數的整體顯著性，確認為不弱工具

第二階段：以第一階段之預測值 $\widehat{WAGE_\iota}$替代原本的 WAGE，並與其他外生變數一起做 OLS：

$HOURS_i = \beta_1 + \beta_2 \widehat{WAGE_\iota} + \beta_3 EDUC_i \ldots\ldots + u_i$，而所得的$\hat{\beta}$為一致的 2SLS 估計值

第三階段：過度識別檢定：Sargan/Hansen's J test，檢驗工具變數是否真與誤差項無關。

內生性檢定：使用 Wu–Hausman 檢定或比較 OLS 與 2SLS 估計結果差異。

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.

1. 因從$x = r_1 + \theta_1 z + v$而知$\text{E}(x) = r_1 + \theta_1 E(z)$，兩個相減而得$x - \text{E}(x) = \theta_1\big(z - E(z)\big) + v$

兩邊同乘$\big(z - E(z)\big)$而得$((x - \text{E}(x))\big(z - E(z)\big) = \theta_1\big(z - E(z)\big)^2 + \big(z - E(z)\big)v$

兩邊同取期望值而得$\text{E}[(\big(x - \text{E}(x)\big)\big(z - E(z)\big)] = \text{E}[\theta_1\big(z - E(z)\big)^2 + \big(z - E(z)\big)v] = \theta_1\text{E}[\big(z - E(z)\big)^2]$

而$\theta_1 = \frac{\text{E}[(\big(x-\text{E}(x)\big)\big(z-E(z)\big)]}{\text{E}[\big(z-E(z)\big)^2]} = \frac{cov(z,x)}{var(z)}$，而這就是$x = r_1 + \theta_1 z + v$中的斜率$\theta_1$

2. $y_i = \pi_0 + \pi_1 z_i + u_i$，而可知$E(y) = \pi_0 + \pi_1 E(z_i) + u_i$，兩個相減而得$y_i - \text{E}(y) = \pi_1\big(z_i - E(z_i)\big) + u_i$

兩邊同乘$\big(z_i - E(z_i)\big)$而得$(y_i - \text{E}(y))\big(z_i - E(z_i)\big) = \pi_1\big(z_i - E(z_i)\big)^2 + \big(z_i - E(z_i)\big)u_i$，兩邊同取期望值而得

$\text{E}[(y_i - \text{E}(y))\big(z_i - E(z_i)\big)] = \pi_1\text{E}[\big(z_i - E(z_i)\big)^2] + E[\big(z_i - E(z_i)\big)u_i]$，假設$E[u_i|z_i] = 0$

而得$\pi_1 = \frac{\text{E}[(y_i-\text{E}(y))\big(z_i-E(z_i)\big)]}{\text{E}[\big(z_i-E(z_i)\big)^2]} = \frac{cov(z,y)}{var(z)}$，而這個就是$y = \pi_0 + \pi_1 z + u$的 OLS 斜率

3. $y_i = \beta_1 + \beta_2 x_i + e_i$，而$x_i = r_1 + \theta_1 z + v_i$，把 x 帶入 y 裡面而得

$y_i = \beta_1 + \beta_2(r_1 + \theta_1 z_i + v_i) + e_i = (\beta_1 + \beta_2 r_1) + (\beta_2\theta_1)z_i + (\beta_2 v_i + e_i)$

對照 y 的縮減形式$y = \pi_0 + \pi_1 z + u$可知$\pi_0 = \beta_1 + \beta_2 r_1$、$\pi_1 = \beta_2\theta_1$、$u_i = \beta_2 v_i + e_i$

4.由 C 可以知道$\pi_1 = \beta_2\theta_1$，兩邊同除θ_1而得$\beta_2 = \frac{\pi_1}{\theta_1}$

5.令$\widehat{\theta_1}$和$\widehat{\pi_1}$分別是 a 跟 b 中的 OLS 斜率估計量，依據工具變數條件和 CLS 定理可知$\widehat{\theta_1} \xrightarrow{p} \theta_1$以及$\widehat{\pi_1} \xrightarrow{p} \pi_1$

由連續映射定理而得$\widehat{\beta_2} = \frac{\widehat{\pi_1}}{\widehat{\theta_1}} \xrightarrow{p} \frac{\pi_1}{\theta_1} = \beta_2$

因此$\widehat{\beta_2} = \frac{\widehat{\pi_1}}{\widehat{\theta_1}}$在大樣本下是一致估計量，稱間接最小平方法