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.

※ Endogeneity：
1. 遺漏變數
2. 互為因果
3. 測量誤差

a. $\beta_2 > 0$ ：時薪越高，工作意願上升

$\beta_3 > 0$ ：教育程度高，工作机会↑

β_4 not sure ：年齡↑累積經驗，但也可能：~~臨近退休~~而 hours↓

$\beta_5 < 0$ ：小孩多→照顧 time↑→工作時間↓

$\beta_6 < 0$ ：有其它額外收入降低工作意願

b. 遺漏變數：wage可能和誤差項有關，有未被觀察到的變數
ex：工作能力、健康狀況 → 產生內生性問題

c. 工具變數必須和 x_i 相關，但是不影响 y，即 e_i，
而 Experience 滿足和 Wage 高度正相關，且為非線性關係，所以用2次方捕捉關係
理論上 Experience 不会影响工作時數，故 Exper 與 e_i 無関

d. identified 定义：若工具變數(IV)數量 ≥ 內生性變數，則滿足 identify 條件
(在 IV or 2SLS 法中，必須滿足 identified 條件！
IV ($exper, exper^2$) 2個 > 內生變數(Wage)，故符合 identified 條件

e. $Wage = \gamma_0 + \gamma_1 EXPER + \gamma_2 EXPER^2 + \gamma_3 EDUC + \gamma_4 AGE + \gamma_5 KIDSL6 + \gamma_6 NWIFEINC$
可以得到預測值 $\widehat{WAGE}$ → 隔離掉 Wage 對誤差項的影响
再將預測值 $\widehat{WAGE}$ 代入原迴歸式 → $Hours = \beta_1 + \beta_2 \widehat{WAGE} + \beta_3 EDUC + \cdots\cdots$
得到的 β_i 估計量，即 2SLS 的一致估計量。

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. 一階段模型：$x = \gamma_1 + \theta_1 z + v$ —— ①

取期望值：$E(x) = \gamma_1 + \theta_1 E(z)$ —— ②

① − ② ⇒ $x - E(x) = \theta_1 (z - E(z)) + v$

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

$E\{[x - E(x)][z - E(z)]\} = \theta_1 E[(z - E(z))^2] + E[(z - E(z))v]$

∵ $z - E(z)$ 和 v 不相關 ∴ $E[(z - E(z))v] = 0$

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

$\Rightarrow \text{cov}(z, x) = \theta_1 \text{var}(z)$

$\Rightarrow \theta_1 = \dfrac{\text{cov}(z, x)}{\text{var}(z)}$

b.

$y = \pi_0 + \pi_1 z + u$ —— ①

$E(y) = \pi_0 + \pi_1 E(z)$ —— ②.

① − ② ⇒ $y - E(y) = \pi_1 (z - E(z)) + u$

$E([y - E(y)][z - E(z)]) = E(\pi_1 (z - E(z))^2) + E(u(z - E(z)))$ $= 0$.

$\Rightarrow E([y - E(y)][z - E(z)]) = E(\pi_1 (z - E(z))^2)$

$\Rightarrow \text{cov}(y, z) = \pi \,\text{var}(z)$

$\Rightarrow \pi = \dfrac{\text{cov}(y, z)}{\text{var}(z)}$

c. $x=\gamma_1+\theta_1 z+v$ 代入 $y=\beta_1+\beta_2 x+e$

$$\Rightarrow y=\beta_1+\beta_2(\gamma_1+\theta_1 z+v)+e$$

$$\Rightarrow y=\underbrace{(\beta_1+\beta_2\gamma_1)}_{\pi_0}+z(\underbrace{\beta_2\theta_1}_{\pi_1})+\underbrace{\beta_2 v+e}_{u}$$

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

e. $$\hat{\theta}_1=\frac{cov(z,x)}{var(z)}=\frac{\frac{1}{n}\Sigma(z_i-\bar{z})(x_i-\bar{x})}{\frac{1}{n}\Sigma(z_i-\bar{z})^2}=\frac{\Sigma(z_i-\bar{z})(x_i-\bar{x})}{\Sigma(z_i-\bar{z})^2}$$

$$\hat{\pi}_1=\frac{cov(z,y)}{var(z)}=\frac{\Sigma(z_i-\bar{z})(y_i-\bar{y})}{\Sigma(z_i-\bar{z})^2}$$

$$\hat{\beta}_2=\frac{\pi_1}{\theta_1}=\frac{\widehat{cov}(z,x)}{\widehat{cov}(z,y)}$$

由大數法則可知 $\hat{\beta}_2=\dfrac{\widehat{cov}(z,x)}{\widehat{cov}(z,y)}\xrightarrow{p}\dfrac{cov(z,x)}{cov(z,y)}=\beta_2$