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.

β_2: + A higher wage increases the opportunity cost of not working, so women are more likely to supply more labor hours.

β_3: + More educated women often have better job opportunities and may have a stronger attachment to the labor force, leading to more hours worked.

β_4: ∩ Labor supply might initially increase with age due to experience, but at older ages, it might decrease due to retirement or health considerations.

β_5: − More young children usually reduce the labor supply due to childcare responsibilities.

β_6: − Higher non-wife household income may reduce the incentive for the wife to work, lowering labor supply.

b. Explain why this supply equation cannot be consistently estimated by OLS regression.

Because of endogeneity. Wage and labor supply may be determined simultaneously.

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.

1. They likely influence WAGE.
2. They may not directly influence HOURS (exogeneity), though this depends on assumptions about behavior and preferences.

d. Is the supply equation identified? Explain.

Yes, the equation is identified. In fact, it is overidentified, because there are two instruments EXPER and EXPER^2 for a single endogenous regressor WAGE. This allows estimation the model using IV or 2SLS, and also test the validity of the instruments.

e. Describe the steps [not a computer command] you would take to obtain IV/2SLS estimates.

1. Regress WAGE on the instruments (EXPER and EXPER^2) and all exogenous variables (EDUC,AGE,KIDSL6,NWIFEINC). Save the predicted values of WAGE.
2. Regress HOURS on the predicted values of WAGE from stage one, along with all other exogenous variables.
3. The coefficients from the second stage are the IV/2SLS estimates.

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.

$$E(x) = \gamma_1 + \theta_1 E(z)$$

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

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

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

$$= \theta_1 \text{Var}(z)$$

$$\Rightarrow \theta_1 = \frac{E(z - E(z))(x - E(x))}{\text{Var}(z)} = \frac{\text{cov}(x, z)}{\text{Var}(z)}$$

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.]

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

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

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

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

$$\text{cov}(y, z) = \pi_1 \text{Var}(z) + 0$$

$$\pi_1 = \frac{\text{cov}(y, z)}{\text{Var}(z)}$$

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.

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

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

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

$$\pi_0 = \beta_1 + \beta_2 \gamma_1, \quad \pi_1 = \beta_2 \theta_1, \quad u = \beta_2 v + e$$

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

$$\pi_1 = \beta_2 \theta_1$$

$$\Rightarrow \beta_2 = \frac{\pi_1}{\theta_1}$$

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.

$$\hat{\theta}_1 = \frac{\widehat{cov}(x,z)}{\widehat{var}(z)} \quad . \quad \hat{\pi}_1 = \frac{\widehat{cov}(y,z)}{var(z)}$$

$$\hat{\beta}_2 = \frac{\dfrac{\sum(y-\bar{y})(z-\bar{z})/N}{\sum(z-\bar{z})^2/N}}{\dfrac{\sum(x-\bar{x})(z-\bar{z})/N}{\sum(z-\bar{z})^2/N}} = \frac{\sum(y-\bar{y})(z-\bar{z})/N}{\sum(x-\bar{x})(z-\bar{z})/N}$$

$$= \frac{\widehat{cov}(y,z) \xrightarrow{p} cov(y,z)}{\widehat{cov}(x,z) \xrightarrow{p} cov(x,z)}$$

$$\hat{\beta}_2 \xrightarrow{p} \beta_2 \quad ※$$