HW0421

C10Q03 (a)

Subtract $E(x) = \gamma_1 + \theta_1 E(z)$ from $x = \gamma_1 + \theta_1 z + v$ to

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

Multiply both sides by $[z - E(z)]$ to obtain

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

Take the expected value of both sides to obtain

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

assuming $E(z-E(z))v = 0$. Solving for θ_1 we obtain

$$\theta_1 = \frac{E[(z-E(z))(x-E(x))]}{E[(z-E(z))^2]} = \frac{cov(z,x)}{Var(z)}$$

This is the OLS estimator of θ_1 in the regression $x = \gamma_1 + \theta_1 z + v$

(b) Subtract $E(y) = \pi_0 + \pi_1 E(z)$ from $y = \pi_0 + \pi_1 z + u$ to obtain

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

Multiply both sides by $(z - E(z))$ to obtain

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

Assuming $E(z-E(z))u = 0$, take the expected value of

both sides to obtain $E[(z-E(z))(y-E(y))] = \pi_1 E[(z-E(z))^2]$

Solving for π_1 we have : $\pi_1 = \dfrac{E[(z-E(z))(y-E(y))]}{E[(z-E(z))^2]} = \dfrac{cov(z,y)}{Var(z)}$

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

(c) The substitution leaves

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

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

(d) According to (c): $\pi_1 = \beta_2 \theta_1$

$\Rightarrow$ Solving $\pi_1 = \beta_2 \theta_1$ for β_2, we have $\beta_2 = \frac{\pi_1}{\theta_1}$

(e) From (a), $\hat{\theta}_1 = \frac{\widehat{cov(z,x)}}{\widehat{var(z)}} = \frac{\sum(z_i-\bar{z})(x_i-\bar{x})/N}{\sum(z_i-\bar{z})^2/N} = \frac{\sum(z_i-\bar{z})(x_i-\bar{x})}{\sum(z_i-\bar{z})^2}$

The estimator is consistent if z is uncorrelated with v. Similarly,

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

$\Rightarrow$ A consistent estimator if z is uncorrelated with u.

Then $\hat{\beta}_2 = \hat{\pi}_1 / \hat{\theta}_1 = \frac{\sum(z_i-\bar{z})(y_i-\bar{y})}{\sum(z_i-\bar{z})(x_i-\bar{x})} = \frac{\sum(z_i-\bar{z})(y_i-\bar{y})/N}{\sum(z_i-\bar{z})(x_i-\bar{x})/N}$

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

This is the IV estimator given in equation (10.17).

The consistency of this estimator is established using the fact that sample moments converge to population moments, so that $\widehat{cov(z,y)} \xrightarrow{P} cov(z,y)$ and $\widehat{cov(z,x)} \xrightarrow{P} cov(z,x)$.

It follows that $\hat{\beta}_2 = \hat{\pi}_1/\hat{\theta}_1 = \frac{\widehat{cov(z,y)}}{\widehat{cov(z,x)}} \xrightarrow{P} \frac{cov(z,y)}{cov(z,x)}$

$= \beta_2$

C10Q02

(a) β_2 (WAGE) : (+)

(higher wages are incentives for women to supply more labor hours)

β_3 (EDUC): (+) (more job opportunities → more labor supply)

β_4 (AGE): Uncertain (older women may have more experience, which increases their labor supply. However, they may also face greater health concerns, potentially limiting their labor supply.)

β_5 (KIDSL6) : (−) (taking care of young children : less time)

β_6 (NWIFE INC): (−) (other sources of income reduce the necessity for the wife to work, leading to a lower supply of labor.)

(b) Explanation: Endogeneity issues arising from

1. Simultaneity: WAGE may be influenced by the labor supply itself.
→ A simultaneous relationship that violates the OLS assumption of exogenous explanatory variables.

2. Omitted variable bias: Unobserved factors such as personal characteristics (motivation, family background) can affect both HOURS and WAGE.

3. Measurement error: Any error in measuring WAGE or other independent variables will bias the OLS estimates.

→ This supply equation cannot be consistently estimated by OLS regression.

(c) EXPER & $EXPER^2$ can be valid instruments:

1. Relevance: Both are expected to be positively correlated with WAGE. More experience → Higer wages, the square: capturing non-linear relationships.

2. Exogeneity: Neither EXPER nor $EXPER^2$ should have a direct effect on HOURS beyond their influence on WAGE.

3. Independence: Not correlated with the error term in the supply equation. If labor experience is not influenced by factors affecting both wage and hours, these variables can serve as effective instruments.

(d) Yes, the supply equation is identified because:

- Over-Identification: There is one endogenous variable (WAGE) and two instruments (EXPER and $EXPER^2$). Since we have more instruments than endogenous variables, the equation is over-identified.
- Validity of Instruments: Provided EXPER and $EXPER^2$ satisfy the relevance and exogeneity conditions, the supply equation can be consistently estimated using IV/2SLS methods.

(e) To obtain IV/2SLS estimates, follow these steps:

1. First-stage regression: $WAGE = \delta_0 + \delta_1 EXPER + \delta_2 EXPER^2 + \delta_1 EDUC + \ldots + u$

2. Obtain fitted values: Save the predicted values from the first stage ($\widehat{WAGE}$)

3. Second-stage regression: $HOURS = \beta_0 + \beta_1 \widehat{WAGE} + \beta_2 EDUC + \beta_3 AGE + \beta_4 KIDSL6 + \beta_5 NWIFEINC + e$

4. Assess model fit:
   - Check the first-stage regression results to ensure instruments are strong (check F-statistic).
   - Conduct over-identification tests (e.g., Sargan test) to validate the instruments.

5. Interprest results: Analyze the coefficients from the second stage to understand the impact of WAGE and other variables on labor supply.