HW0421

Yung-Jung Cheng

2025-04-26

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.

2(a)

Discuss the signs you expect for each of the coefficients.

Ans

- β_2 (WAGE): Expected to be positive. Higher wages increase the opportunity cost of not working, thus encouraging greater labor supply (substitution effect).
- β_3 (EDUC): Expected to be positive. More education typically increases labor market opportunities and potential earnings, leading to greater labor force participation.
- β_4 (AGE): Ambiguous. Younger women may supply more hours, but labor supply might decrease later in life due to family responsibilities or retirement.
- β_5 (KIDSL6): Expected to be negative. More young children increase household responsibilities, reducing labor supply.
- β_6 (NWIFEINC): Expected to be negative. Higher non-wife income reduces the economic need for the wife to work (income effect).

2(b)

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

Ans

Because $WAGE$ is likely endogenous. Unobserved factors such as ability, motivation, or health may influence both $WAGE$ and $HOURS$, violating the OLS exogeneity assumption. As a result, OLS estimates would be biased and inconsistent.

2(c)

Suppose we consider the woman's labor market experience `EXPER` and its square, `EXPER²`, to be instruments for `WAGE`.
Explain how these variables satisfy the logic of instrumental variables.

Ans

- **Relevance**: $EXPER$ and $EXPER^2$ are expected to be strongly correlated with $WAGE$ because labor market experience is an important determinant of earnings.
- **Exogeneity**: Assuming that $EXPER$ and $EXPER^2$ affect $HOURS$ only through $WAGE$, and not directly through the error term e, they satisfy the exogeneity requirement for valid instruments.

2(d)

Is the supply equation identified? Explain.

Ans

Yes, the equation is identified.
There is one endogenous regressor, $WAGE$, and two instruments, $EXPER$ and $EXPER^2$. Since the number of instruments is greater than the number of endogenous regressors, the model is overidentified and thus can be consistently estimated.

2(e)

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

Ans

1. **First Stage**:
   Regress $WAGE$ on $EXPER$ and $EXPER^2$, and obtain the fitted values $\widehat{WAGE}$.
2. **Second Stage**:
   Replace $WAGE$ with $\widehat{WAGE}$ in the original supply equation and estimate:

   $$HOURS = \beta_1 + \beta_2\widehat{WAGE} + \beta_3EDUC + \beta_4AGE + \beta_5KIDSL6 + \beta_6NWIFEINC + u$$
3. **Instrument Validity Checks**:
   Check the relevance of instruments in the first stage (e.g., via F-test) and, if overidentified, test the exogeneity of the instruments (e.g., using Hansen's J-test).

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 $\beta_2 = \operatorname{cov}(z, y)/\operatorname{cov}(z, x)$.

3(a)

Divide the denominator of $\beta_2 = \operatorname{cov}(z, y)/\operatorname{cov}(z, x)$ by $\operatorname{var}(z)$.
Show that $\operatorname{cov}(z, x)/\operatorname{var}(z)$ is the coefficient of the simple regression of x on z.

Ans

Consider the simple linear regression model:

$$x = \gamma_1 + \theta_1 z + \nu$$

The ordinary least squares (OLS) estimator for θ_1 minimizes the residual sum of squares:

$$\sum (x_i - \gamma_1 - \theta_1 z_i)^2$$

Taking the first-order condition with respect to θ_1, and assuming the data are mean-centered (i.e., $\mathrm{E}[x] = \mathrm{E}[z] = 0$ so that $\gamma_1 = 0$), the normal equation simplifies to:

$$\sum x_i z_i - \theta_1 \sum z_i^2 = 0$$

Dividing by n and recognizing sample covariances and variances:

$$\mathrm{cov}(z, x) - \theta_1 \mathrm{var}(z) = 0$$

Thus:

$$\theta_1 = \frac{\mathrm{cov}(z, x)}{\mathrm{var}(z)}$$

Therefore, dividing $\mathrm{cov}(z, x)$ by $\mathrm{var}(z)$ yields the OLS coefficient θ_1 from the regression of x on z. This is the **first-stage regression** in two-stage least squares (2SLS).

3(b)

Divide the numerator of $\beta_2 = \mathrm{cov}(z, y)/\mathrm{cov}(z, x)$ by $\mathrm{var}(z)$.
Show that $\mathrm{cov}(z, y)/\mathrm{var}(z)$ is the coefficient of the simple regression of y on z.

Ans

Consider the simple linear regression model:

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

The OLS estimator for π_1 minimizes the residual sum of squares:

$$\sum (y_i - \pi_0 - \pi_1 z_i)^2$$

Assuming mean-centered data ($\mathrm{E}[y] = \mathrm{E}[z] = 0$, so that $\pi_0 = 0$), the normal equation simplifies to:

$$\sum y_i z_i - \pi_1 \sum z_i^2 = 0$$

Dividing by n:

$$\mathrm{cov}(z, y) - \pi_1 \mathrm{var}(z) = 0$$

Thus:

$$\pi_1 = \frac{\mathrm{cov}(z, y)}{\mathrm{var}(z)}$$

Therefore, dividing $\mathrm{cov}(z, y)$ by $\mathrm{var}(z)$ yields the OLS coefficient π_1 from the regression of y on z.

3(c)

In the model $y = \beta_1 + \beta_2 x + e$, substitute for x using $x = \gamma_1 + \theta_1 z + \nu$ and simplify to obtain $y = \pi_0 + \pi_1 z + u$.
Identify π_0, π_1, and u.

Ans

Substituting $x = \gamma_1 + \theta_1 z + \nu$ into $y = \beta_1 + \beta_2 x + e$:

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

Expanding:

$$y = \beta_1 + \beta_2\gamma_1 + \beta_2\theta_1 z + \beta_2\nu + e$$

Grouping terms:

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

Thus, comparing with the reduced-form equation $y = \pi_0 + \pi_1 z + u$, we identify:

- $\pi_0 = \beta_1 + \beta_2\gamma_1$
- $\pi_1 = \beta_2\theta_1$
- $u = \beta_2\nu + e$

This is the **reduced-form regression**, expressing y directly in terms of the instrument z.

3(d)

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

Ans

From the result of 3(c), we have:

$$\pi_1 = \beta_2\theta_1$$

Solving for β_2:

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

Thus, β_2 can be obtained by dividing the reduced-form coefficient π_1 by the first-stage coefficient θ_1.
This reflects the core idea behind **two-stage least squares (2SLS)** and **indirect least squares (ILS)**.

3(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 $\beta_2 = \pi_1/\theta_1$.

Ans

First, by the consistency of OLS under the classical assumptions (exogeneity and no perfect multicollinearity), we have:

$$\hat{\pi}_1 \xrightarrow{p} \pi_1, \quad \hat{\theta}_1 \xrightarrow{p} \theta_1$$

Since $\theta_1 \neq 0$ (the instrument z must be relevant), the function $g(\hat{\pi}_1, \hat{\theta}_1) = \hat{\pi}_1 / \hat{\theta}_1$ is continuous at (π_1, θ_1). By the **Continuous Mapping Theorem**, it follows that:

$$\hat{\beta}_2 = \frac{\hat{\pi}_1}{\hat{\theta}_1} \xrightarrow{p} \frac{\pi_1}{\theta_1} = \beta_2$$

Thus, $\hat{\beta}_2$ is a consistent estimator of β_2.
This method of estimation is known as **indirect least squares (ILS)**.