HW0421 – Pinyo 312712017

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.

bata_1 – positive

beta_2 – positive

beta_3 – positive

beta_4 – Negative

beta_5 – Negative

beta_6 – Negative

b.

The labor supply equation for married women cannot be consistently estimated by OLS regression because the hourly wage (WAGE) is likely to be **endogenous** in the model. In labor supply studies, a woman's observed wage is determined in part by her decision to participate in the labor market and by factors that also influence her labor supply (such as ability, motivation, or unobserved preferences for work). These unobserved factors are part of the error term e*e*, and if they are correlated with WAGE, the OLS assumption that the regressors are uncorrelated with the error Cov(WAGE,e)=0 is violated

c.

EXPER and **EXPER²** are **excluded from the labor supply equation** (they do not appear as regressors), ensuring they only affect **HOURS** indirectly via their impact on **WAGE**.

They are unlikely to be correlated with unobserved factors in e*e* (e.g., personal preferences for work-life balance) that might directly influence labor supply.

Potential Weakness:

If **EXPER** is correlated with omitted variables affecting labor supply (e.g., unobserved ability or career commitment), the exogeneity condition could fail. However, this is often assumed negligible in IV applications.

In summary, **EXPER** and **EXPER²** are valid instruments because they (1) strongly predict **WAGE** and (2) plausibly satisfy the exclusion restriction, making them suitable for addressing endogeneity in the wage variable.

d.

The supply equation is **identified** if the instruments (EXPER and $EXPER^2$) satisfy both the **order condition** and **rank condition** for identification. Here's the analysis:

1. Order Condition:

- The equation has **one endogenous regressor**: WAGE.
- We have **two excluded instruments**: EXPER and $EXPER^2$.
- Since the number of instruments (2) ≥ number of endogenous regressors (1), the order condition is satisfied.

2. Rank Condition:

- The instruments (EXPER, $EXPER^2$) must be **correlated with WAGE** (relevance) and **uncorrelated with the error term e** (exogeneity).
- **Relevance**: As discussed in part (c), EXPER and $EXPER^2$ are likely strong predictors of WAGE because labor market experience affects wages.
- **Exogeneity**: Assuming EXPER and $EXPER^2$ do not directly affect HOURS (except through WAGE) and are uncorrelated with omitted variables (e.g., unobserved ability), they satisfy the exclusion restriction.

Conclusion:

The supply equation is **identified** because:

- The order condition is met (2 instruments ≥ 1 endogenous regressor).
- The rank condition holds if EXPER and $EXPER^2$ are valid instruments (relevant and exogenous)

e.

Step 1: First Stage Regression

- Regress the endogenous variable (WAGE) on the instruments (EXPER and $EXPER^2$) and all other exogenous variables in the original equation.

$$WAGE = \alpha 1 + \alpha 2 EXPER + \alpha 3 EXPER^2 + \alpha 4 EDUC + \alpha 5 AGE + \alpha 6 KIDSL6 + \alpha 7 NWIFEINC + v$$

where 'v' is the error term for this regression.

- Obtain the predicted values of WAGE from this first-stage regression. We'll call these predicted values WAGE_hat.

Step 2: Second Stage Regression

- Replace the actual endogenous variable (WAGE) in the original equation with its predicted values (WAGE_hat) from the first stage.
- Estimate the original labor supply equation using OLS, but with WAGE_hat instead of WAGE.

$$HOURS = \beta 1 + \beta 2 WAGE\_hat + \beta 3 EDUC + \beta 4 AGE + \beta 5 KIDSL6 + \beta 6 NWIFEINC + e$$

The coefficients obtained from this second-stage regression (especially the coefficient on WAGE_hat) are the IV/2SLS estimates.

Important Considerations:

- **Standard Errors**: The standard errors obtained directly from the second-stage regression are incorrect. This is because they don't account for the fact that WAGE_hat is a predicted value from the first stage. You need to adjust the standard errors. Most statistical software packages have built-in commands (e.g., in R, using the ivreg function from the AER package) that perform 2SLS and automatically calculate the correct standard errors.
- **Validity of Instruments**: The IV/2SLS estimates are only valid if the instruments (EXPER and $EXPER^2$) are valid. They must be relevant (correlated with WAGE) and exogenous (uncorrelated with the error term in the original equation). If the instruments are weak or invalid, the IV/2SLS estimates can be even more biased than OLS.

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.

In a **simple linear regression** of x on z, the OLS coefficient θ_1 is:

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

This is a standard result in regression.

We have $\beta_2 = \text{cov}(z,y) / \text{cov}(z,x)$

Now divide both numerator and denominator by var(z)

$\beta_2 = (\text{cov}(z,y) / \text{var}(z)) / (\text{cov}(z,x) / \text{var}(z))$

Since both the numerator and denominator are divided by the same value, this doesn't change the value:

$\beta_2 = (\text{cov}(z,y) / \text{var}(z)) / (\text{cov}(z,x) / \text{var}(z))$ = (slope of regression of y on z) / (slope of regression of x on z)

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

So the **denominator** of the IV estimator can be seen as the **coefficient from the first-stage regression**:

$X = \gamma_1 + \theta_1 * z + v$

This is important in **2SLS**, where:

- First stage: regress x on z → get x^
- Second stage: regress y on x^ to estimate β_2

b.

In simple linear regression where you regress y on z, the OLS slope coefficient $\pi 1$ is:

$\pi 1 = cov(z,y) / var(z)$

The regression model is:

$y = \pi 0 + \pi 1 * z + u$

where:

- $\pi 0$ is the intercept,
- $\pi 1$ is the slope,
- u is the error term.

Thus, **cov(z, y) divided by var(z) gives the slope** $\pi 1$ from regressing y on z.

So,

$Cov (z,y) / var(z) = \pi 1$

where $\pi 1$ is the slope coefficient in the regression:

$y = \pi 0 + \pi 1 * z + u$

c.

From equation $y = \beta 1 + \beta 2 * x + e$

Given:

$x = \gamma 1 + \theta 1 * z + v$

Substitute into the y equation:

$y = \beta 1 + \beta 2 * (\gamma 1 + \theta 1 * z + v) + e$

Distribute $\beta 2$:

$y = \beta 1 + \beta 2 * \gamma 1 + \beta 2 * \theta 1 * z + \beta 2 * v + e$

Group terms:

$y = (\beta1 + \beta2*\gamma1) + (\beta2*\theta1)*z + (\beta2*v + e)$

Thus, by matching terms:

- $\pi0 = \beta1 + \beta2*\gamma1$
- $\pi1 = \beta2*\theta1$
- $u = \beta2*v + e$

π0\pi_0π0 is the new intercept, made from the original intercept β1 and the relationship between x and z (through γ1).

π1 is the **effect of z on y** in the reduced form — it's the product of the structural effect β2 and the strength of instrument θ1 .

u is a **new error term**, made up of the first-stage error v and the original error e

d.

From (c)

$y = (\beta1 + \beta2*\gamma1) + (\beta2*\theta1)*z + (\beta2*v + e)$

Thus, by matching terms:

- $\pi0 = \beta1 + \beta2*\gamma1$
- $\pi1 = \beta2*\theta1$
- $u = \beta2*v + e$

So,

$\beta2 = \pi1 / \theta1$

e.

An estimator $\hat{\beta}_2$ is **consistent** if:

$\hat{\beta}_2 \to \beta_2$ as sample size $n \to \infty$

This means that **as the sample gets large**, the estimator **converges in probability** to the true value.

From part (c), we know:

- $\pi_1 = \beta_2 * \theta_1$

That is:

$\pi_1 = \beta_2 * \theta_1$ or $\beta_2 = \pi_1 / \theta_1$

Since OLS is **consistent** under the usual assumptions (including valid instrument z):

- $\hat{\pi}_1 \to \pi_1$
- $\hat{\theta}_1 \to \theta_1$

Meaning: both OLS estimators **converge** to the true values as the sample gets larger.

The continuous mapping theorem says:

- If $\hat{a} \to a$ and $\hat{b} \to b$
- Then any continuous function of $\hat{a}$ and $\hat{b}$ will converge to the same function of a and b.

Here, $\hat{\pi}_1 / \hat{\theta}_1$ is a continuous function (division, assuming $\theta_1 \neq 0$).

Thus:

$\hat{\pi}_1 / \hat{\theta}_1 \to \pi_1 / \theta_1$

That is:

$\hat{\beta}_2 \to \beta_2$

Therefore, $\hat{\beta}_2$ is a **consistent estimator** of β_2.