HW 0324

Question 10.2.

Part a.

In the assignment the labor supply of married women is given by the following formula:

$$\text{HOURS} = \beta_1 + \beta_2\text{WAGE} + \beta_3\text{EDUC} + \beta_4\text{AGE} + \beta_5\text{KIDSL6} + \beta_6\text{NWIFEINC} + e$$

I believe that the signs of the coefficients will be as follows:

Variable	Coefficient	Expected sign	Justification
WAGE	β_2	+	Higher wages increase the opportunity cost of not working, encouraging labor supply (substitution effect).
EDUC	β_3	+	More education often leads to higher wages and better job prospects, increasing labor force participation.
AGE	β_4	-	From my experience, I think that with age people tend to be: 1. More efficient and 2. Prioritize work-life balance both of which reduce work hours

Variable	Coefficient	Expected sign	Justification
KIDSL6	β_5	-	Having children under 6 increases home responsibilities and childcare needs, lowering available work hours.
NWIFEINC	β_6	-	More non-wife income (e.g., spouse's earnings) reduces the financial necessity for the wife to work.

Part b.

The equation presented earlier in the assignment cannot be consistently estimate using OLS due to endogeneity. In other words, the variable WAGE is likely correlated with the error term in the model. One reason for this could be simultaneity: a woman's hours worked may influence her observed wage (for example overtime pay or bigger bonus), especially in jobs where pay is linked to experience or part-time/full-time status. Additionally, there may be unobserved factors such as motivation, health, or household preferences that affect both wage and labor supply but are not included in the model, leading to omitted variable bias. Another issue is selection bias, as wages are only observed for women who choose to work, making the sample non-random and potentially unrepresentative.

Part c.

EXPER (experience) and EXPER2 (experience squared) are valid instruments for WAGE because they fulfill the two essential criteria for instrumental variables: relevance and exogeneity. They are relevant because experience is a key determinant of wages as usually, more experience leads to higher earnings, and EXPER2 accounts for diminishing returns to experience. This ensures a strong correlation between the instruments and the endogenous variable, WAGE.

They are also plausibly exogenous to the labor supply equation. After controlling for variables like age, education, and children, experience is unlikely to directly affect labor supply (HOURS) except through its impact on wages. That is, EXPER and EXPER2 do not capture unobserved factors (like preferences or constraints) that would directly influence hours worked.

Part d.

Yes, the supply equation is identified, because we have more instruments than endogenous regressors, satisfying the order condition for identification.

In this model, the endogenous variable is WAGE, and we are using two instruments: EXPER and EXPER2. Since we have 2 instruments for 1 endogenous variable, the equation is overidentified, meaning not only can we estimate the effect of WAGE on HOURS, but we can also test the validity of the instruments using overidentification tests (like the Sargan test).

Identification ensures that the model's structural parameters—especially the causal effect of WAGE—can be uniquely estimated using instrumental variable methods like two-stage least squares (2SLS). It is important to note, however, that while the model is identified in theory, this relies on the assumption that the instruments are both relevant (correlated with WAGE) and exogenous (uncorrelated with the error term in the HOURS equation).

If these assumptions hold, the supply equation is not only identified but also consistently estimable.

Part e.

To obtain IV/2SLS estimates in R, I would start by identifying the endogenous variable—in this case, WAGE—which is likely correlated with the error term in the labor supply equation. Next, I would select valid instruments that are correlated with WAGE but uncorrelated with the error term; here, EXPER and EXPER2 serve this purpose. In the first stage, I would regress WAGE on all exogenous variables and the instruments to obtain predicted values (WAGE_hat). In the second stage, I would regress HOURS on WAGE_hat and the remaining exogenous variables.

Question 10.3.

Part a.

Given:

$$\beta_2 = \frac{Cov(z, y)}{Cov(z, x)}$$

$$\beta_2 = \frac{Cov(z, y)/Var(z)}{Cov(z, x)/Var(z)} \Rightarrow \beta_2 = \frac{\theta_2}{\theta_1}$$

$$\theta_2 = \frac{Cov(z, y)}{Var(z)}$$

$$\theta_1 = \frac{Cov(z, x)}{Var(x)}$$

where θ_1 is the slope from the regression of x on z: $x = \gamma_1 + \theta_1 z + v$

This is the first-stage regression in 2SLS. Here, θ_1 measures how strongly z predicts x

2. θ_2 is the slope from the regression of y on z: $y = \delta_1 + \theta_2 z + u$

hence $\beta_2 = \frac{Cov(z, y)}{Cov(z, x)}$ can be expressed as:

$$\beta_2 = \frac{\text{slope of regression of } y \text{ on } z}{\text{slope of regression of } x \text{ on } z}$$

Part b.

Given:

$$\beta_2 = \frac{Cov(z, y)}{Cov(z, x)}$$

$$\beta_2 = \frac{Cov(z, y)/Var(z)}{Cov(z, x)/Var(z)}$$

$\frac{Cov(z, y)}{Var(z)} \Rightarrow$ definition of the slope coefficient frome:

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

where:

$\pi_1 = \frac{Cov(x, y)}{Var(z)}$ then the IV estimator can be written as:

$\beta_2 = \frac{\pi_1}{\theta_1}$ where π_1 is from the regression $y = \pi_0 + \pi_1 z + u$

θ_1 is from the regression $x = \gamma_1 + \theta_1 z + a$

Part c.

$y = \beta_1 + \beta_2 x + e$

Given:

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

$y = \beta_1 + \beta_2(\gamma_1 + \theta_1 z + v) + e = \beta_1 + \beta_2\gamma_1 + \beta_2\theta_1 z + \beta_2 v + e = (\beta_1 + \beta_2\gamma_1) + \beta_2\theta_1 z + (\beta_2 v + e) \Rightarrow \pi_0 + \pi_1 z + y$ where:

$\pi_0 = \beta_1 + \beta_2\gamma_1$

$\pi_1 = \beta_2\theta_1$

$u = \beta_2 v + e$

Part d.

$y = \beta_1 + \beta_2 x + e$

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

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

where:

$\pi_0 = \beta_1 + \beta_2\gamma_1$

$\pi_1 = \beta_2\theta_1$

$u = \beta_2 v + e$

Hence: $y = \pi_0 + \pi_1 z + u \Rightarrow \pi_1 = \beta_2\theta_1$

rearranging we get:

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

Part e.

Show that $\hat{\beta}_2 = \frac{\hat{\pi}_1}{\hat{\theta}_1}$ is a concistant estimator of $\beta_2 = \frac{\pi_1}{\theta_1}$

estimate β_2 using instrumental variable metme where z such that:

$cov(z, x) \neq 0$

$cov(z, u) = 0$

For $x = \theta_1 z + v$

For y: $y = \pi_1 z + u$

if $\hat{\pi}_1 \xrightarrow{p} \pi_1$ and $\hat{\theta}_1 \xrightarrow{p} \theta_1$ then $\frac{\hat{\pi}_1}{\hat{\theta}_1} \xrightarrow{p} \frac{\pi_1}{\theta_1} = \beta_2$ therefore $\hat{\beta}_2$ is consistent for β_2

hence:

$$\hat{\beta}_2 = \frac{cov(z, y)}{cov(z, x)} = \frac{\hat{\pi}_1}{\hat{\theta}_1}$$