Tutorial 6 Answers

Rohit Armstrong

April 2025

1 Question 1

log model

$$\log(\widehat{\text{wage}}_i) = \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{educ}_i + \hat{\beta}_2 \cdot \text{exper}_i + \hat{\beta}_3 \cdot \text{expersq}_i + u_i \tag{1}$$

a) Estimate equation (1) by OLS and report the estimated regression in equation form. Provide robust standard errors in parentheses.

Answer for part (a)

Estimated logit regression

$$\begin{aligned} \log(\widehat{\text{wage}}_i) = \quad & -0.2981 + 0.0984\text{educ}_i + 0.0426\text{exper}_i - 0.0008\text{expersq}_i + u_i \\ & (0.3174) \quad (0.0230) \quad (0.0152) \quad (0.0004) \end{aligned} \tag{2}$$

b) Suppose we use *huseduc* as an instrument for *educ*. What conditions does this variable have to satisfy to be considered a valid instrument?

Answer for part (b)

To be considered a valid instrument, the variable *huseduc* must satisfy the following two conditions:

1. **Relevance:** *huseduc* must be correlated with the endogenous regressor *educ*. Formally,

   $$\text{Cov}(\text{huseduc}, \text{educ}) \neq 0$$

   This tells us that *huseduc* does indeed have useful variation that can explain *educ* in the first-stage regression.

2. **Exogeneity:** *huseduc* must be uncorrelated with the error term u_i in the structural equation. That is,

   $$\text{Cov}(\text{huseduc}, u_i) = 0$$

   This tells us that *huseduc* does indeed affect the dependent var *infl* only through its effect on *educ* and not from any omitted variable bias.

c) How will you test for the relevance of *huseduc* for *educ*? Write down the reduced form for *educ* and perform a hypothesis test. Be sure to mention the null and alternative hypotheses, test-statistic, critical value, asymptotic distribution under the null, conclusion, and interpretation.

Answer for part (c)

1) Hypotheses:

$$H_0 : \alpha_1 = 0 \quad \text{(huseduc is not a relevant instrument)}$$

$$H_A : \alpha_1 \neq 0 \quad \text{(huseduc is a relevant instrument)}$$

2) Reduced Form Regression:

$$\text{educ}_i = \alpha_0 + \alpha_1 \cdot \text{huseduc}_i + \alpha_2 \cdot \text{exper}_i + \alpha_3 \cdot \text{expersq}_i + v_i$$

3) Test-Statistic and Distribution:
From the first-stage regression output:

$$t\text{-statistic on } \alpha_1 = 18.46, \quad p\text{-value} < 0.0001$$

t-stat follows a standard normal distribution asymptotically under the null:

$$t \sim N(0, 1)$$

4) Critical Value and Decision Rule:
At a 5% significance level for a two-sided test:

$$|t| > 1.96 \Rightarrow \text{Reject } H_0$$

$$18.46 > 1.96 \Rightarrow \text{Reject } H_0$$

5) Conclusion and Interpretation:
We reject the null hypothesis in favour of the alternative and conclude that *huseduc* is valid instrument and is statistically significant for explaining *educ*.

d) Now estimate the regression given in (1) by IV using *huseduc* as an instrument for *educ*. Compare the estimate on *educ* with the one you obtain in part b). Produce a table containing OLS and IV estimates together with robust standard errors in parentheses.

Answer for part (d)

We estimate the wage equation using Two-Stage Least Squares (2SLS), instrumenting *educ* with *huseduc*.

	OLS		IV (2SLS)	
educ	0.1075	(0.0132)	0.0894	(0.0230)
exper	0.0416	(0.0153)	0.0426	(0.0151)
expersq	-0.0008	(0.0004)	-0.0008	(0.0004)
intercept	-0.5220	(0.2017)	-0.2981	(0.3174)

Table 1: Comparison of OLS and IV Estimates (Robust Std errors in parentheses)

The IV estimate of educ is marginally lower than the OLS estimate. It is 0.089 vs. 0.107 and this tells us that the OLS coefficient may be upward biased due to omitted variables such as ability, which were positively correlated with both education and wage. Though, the std error for educ in the IV estimate is marginally larger than OLS. This is a byproduct of the IV estimation, where additional variation is brought from it only relying on the variation in educ that is explained by the instrument huseduc.

Overall, we can confirm that *huseduc* is a strong and relevant instrument. Furthermore, the IV estimates are likely more reliable than the OLS estimates(with exogeneity as an assumption).