Labor Economics: Problem Set 1.2

Kai Hillen

November 16, 2025

a)

The correlation between wages and education as well as hours worked is weaker compared to the PS1.1 data.

Table 1: Correlation Matrix of Key Variables

	education	**hours_ME**	**ln_wage_ME**
education	1.0000	0.5710	0.1451
hours_ME	0.5710	1.0000	0.1374
ln_wage_ME	0.1451	0.1374	1.0000

Table 2: Summary Statistics

Variable	**n**	**min**	**max**	**median**	**iqr**	**mean**	**sd**	**se**	**ci**
education	5000	10.00	24.00	12.00	4.00	12.800	2.420	0.034	0.067
hours_ME	5000	6.07	9.83	7.94	0.75	7.950	0.549	0.008	0.015
wage_premium	5000	0.00	1.00	0.00	1.00	0.342	0.474	0.007	0.013

b

regression1

R^2 is worse, the constant is larger and coefficient for log wage smaller.

Table 3: Bivariate Regression: Hours Worked on Log Wage

	Dependent variable:
	Hours Worked
Log Wage	0.088*** (0.071, 0.106)
Constant	7.651*** (7.590, 7.713)
Observations	5,000
R^2	0.019
Adjusted R^2	0.019
Residual Std. Error	0.544 (df = 4998)
F Statistic	96.197*** (df = 1; 4998)
Note:	*p<0.1; **p<0.05; ***p<0.01

PS1.1

Table 4: Bivariate Regression: Hours Worked on Log Wage

	Dependent variable:
	Hours Worked
Log Wage	1.285*** (1.238, 1.332)
Constant	3.636*** (3.478, 3.794)
Observations	5,000
R^2	0.365
Adjusted R^2	0.365
Residual Std. Error	0.390 (df = 4998)
F Statistic	2,874.511*** (df = 1; 4998)
Note:	*p<0.1; **p<0.05; ***p<0.01

c

Education and wage_premium have to be uncorrelated, which should be the case if treatment wage_premium is randomly assigned and there is no bias. The data passes this check, sample means are very close and given p-value 0.3334 we can't reject the null hypothesis.

d

The data passes the balance check:

Table 5: Balance Measures

Variable	Type	Corr.Un	Sample Sizes
wage_premium	Binary	0.0087	
		Total	5000

e

The instrument is highly relevant and the coefficient of 0.354 shows that wages of individuals who were assigned the wage premium increased by 42.49%.

Table 6: first stage

	Dependent variable:
	Log Wage
wage premium	0.354^{***} (0.305, 0.403)
Constant	3.235^{***} (3.206, 3.264)
Observations	5,000
R^2	0.039
Adjusted R^2	0.038
Residual Std. Error	0.839 (df = 4998)
F Statistic	200.562^{***} (df = 1; 4998)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: IV

	Dependent variable:
	hours worked
log wage	0.235^{***} (0.143, 0.327)
Constant	7.158^{***} (6.848, 7.468)
Observations	5,000
R^2	−0.033
Adjusted R^2	−0.034
Residual Std. Error	0.559 (df = 4998)
Note:	*p<0.1; **p<0.05; ***p<0.01

The substitution effect dominates and wage increases cause workers to supply more labor.

f

The IV model fixes both measurement errors, as in Regression b) in Table 8 and OVB in PS1.1.

g

Table 8: First Stage and IV Model Results

	Dependent variable:	
	Log Wage	Hours Worked
	OLS	*instrumental variable*
	First Stage (OLS)	IV Model (2SLS)
	(1)	(2)
Wage Premium (Instrument)	0.351*** (0.302, 0.399)	
Log Wage		0.212*** (0.133, 0.291)
Education	0.050*** (0.041, 0.060)	0.119*** (0.112, 0.125)
Constant	2.595*** (2.470, 2.719)	5.723*** (5.498, 5.948)
Observations	5,000	5,000
R^2	0.059	0.256
Adjusted R^2	0.058	0.255
Residual Std. Error (df = 4997)	0.830	0.474
F Statistic	156.284*** (df = 2; 4997)	
Note:		*p<0.1; **p<0.05; ***p<0.01

The treatment was randomly assigned, therefore a small change should be expected. The instrument is uncorrelated with education, adding education as a control variable won't change it's effect on wages.

h

Both models estimate the elasticity of labor supply at 0.21. PS1.1 used data without measurement error which is not realistic. The IV result matches the full regression from PS 1.1 despite measurement error and removes bias. Therefore I prefer the IV estimate.