

Econometrics II - Assignment 1

Floris Holstege, Stanislav Avdeev

Exercise 1

A

The only statistically significant variable (5% threshold) is the measure for schooling. On average, an increase of one year of schooling leads to an 0.216 in the logged value of wages.

Table 1:

	<i>Dependent variable:</i>
	Log of Wage
Schooling	0.216*** (0.032)
Age	-0.342 (0.521)
Age ²	-0.011 (0.008)
Intercept	26.409*** (8.057)
Observations	416
R ²	0.815
Adjusted R ²	0.813
Residual Std. Error	1.499 (df = 412)
F Statistic	604.261*** (df = 3; 412)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

B

The problem is that there is selection on unobservables and observables. We have a selected sample, excluding the unemployed. This means we have no info on their potential earnings, if they were employed. We cannot just extrapolate our findings in the selected sample to the non-employed, since the selection influences both unobservable variables (for example; the unemployed might have less motivation) and observable variables (for instance; the effect of years of schooling on earnings might be less strong for people who have not worked for some years).

We formally model this selection bias in three steps. First, we indicate if the dependent variable is observed as follows:

$$I_i^* = Z_i' \gamma + V_i$$

I_i takes value 1 if $I_i^* > 0$, and 0 if $I_i^* \leq 0$. Second, we have a model for our latent variable Y_i^* .

$$Y_i^* = X_i' \beta + U_i$$

Our observed dependent variable, Y_i , is then defined as follows:

$$Y_i = \begin{cases} Y_i^* & \text{for } I_i = 1 \\ \text{Missing} & \text{for } I_i = 0 \end{cases}$$

C

D & E

Exercise 2