Erasmus School of Economics

MOOC Econometrics

Training Exercise 4.1

Questions

We want to explain the income y_i of an individual $i = 1, \ldots, n$ using the individual's intelligence x_i^*. Suppose that the true relationship between these two variables is

$$y_i = \alpha + \beta x_i^* + u_i,$$

where β gives the impact of intelligence on income. Furthermore, suppose that this model satisfies all the standard assumptions of the linear model. However, the intelligence (x_i^*) cannot be observed directly. We can only observe a test score that equals the true intelligence plus a measurement error, that is, $x_i = x_i^* + w_i$. The measurement error process satisfies the following conditions:

- Mean zero: $\mathrm{E}[w_i] = 0$
- Constant variance: $\mathrm{Var}[w_i] = \sigma_w^2$
- Zero correlation across individuals: $\mathrm{Cov}[w_i, w_j] = 0$ for all $i \neq j$
- Uncorrelated with unexplained income and true intelligene: $\mathrm{Cov}[w_i, u_i] = 0$ and $\mathrm{Cov}[w_i, x_i^*] = 0$

We have data on (y_i, x_i), for $i = 1, \ldots, n$. Suppose we ignore measurement error and simply apply OLS to

$$y_i = \alpha + \beta x_i + \varepsilon_i. \tag{1}$$

(a) Show that by definition $\varepsilon_i = -\beta w_i + u_i$.

(b) Derive the covariance between x_i and ε_i. Note: under the assumptions above, the measurement error w_i is uncorrelated with x_i^*.

(c) Use the above results to give the formal conditions under which x_i is endogenous in (1).