Erasmus School of Economics

MOOC Econometrics

Training Exercise 4.3

Questions

Consider the linear model $y = X\beta + \varepsilon$, where some variable in the $n \times k$ matrix X may be correlated with ε. As a result X may be endogenous. Denote by Z an $m \times n$ matrix of instruments. In general the 2SLS estimator is given by

$$b_{2SLS} = (X'H_Z X)^{-1} X' H_Z y,$$

with $H_Z = Z(Z'Z)^{-1}Z'$.

(a) Show that if $m = k$ we can rewrite the 2SLS estimator to $b_{2SLS} = (Z'X)^{-1}Z'y$. Clearly give the steps that you take and the assumptions that you use.

(b) Suppose that there is only a single explanatory variable, that is, the model equals $y = \beta x + \varepsilon$ and that there is only a single instrument z ($m = k = 1$). Furthermore suppose that the means of x, y and z over the sample are equal to 0. Show that we can write the 2SLS estimator of β as

$$b_{2SLS} = \frac{\mathrm{Cov}(y, z)}{\mathrm{Cov}(z, x)}.$$

Cov(u, v) denotes the (sample) covariance between u and v, which is defined as

$$\mathrm{Cov}(u, v) = \frac{1}{n-2} \sum_{i=1}^{n} (u_i - \bar{u})(v_i - \bar{v}),$$

where $\bar{u}$ and $\bar{v}$ denote the sample mean of u and v, respectively.

(c) Use the formula in (b) to explain what happens to the 2SLS estimator when the correlation between instruments and the endogenous variable is very small.