DS102 - Discussion 7

Wednesday, 23rd October, 2019

In this section we will further develop the connection between two-stage least squares (2SLS) and instrumental variables that we touched upon in lecture.

Assume that we are interested in determining whether reading more books causes a set of n student's SAT test scores to improve. Let's also assume that the i^{th} student's SAT score is generated through the following linear model that we don't observe[1]

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i,$$

where y_i is the SAT score, $\beta = (\beta_1, \beta_2)^\top$ are fixed coefficients, x_{i1} is the number of books the student has read over the last month, x_{i2} is the student's family's net worth, and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ is a noise term.

Now we run a survey asking students how many books they have read in the last month and also record their SAT scores. However we do not know x_{i2}, the net worth of each family. This can either be because we were not able to record that data or because the people that designed the survey did not think that was relevant information.

1. We first decide to run a linear regression model on the observed variables only. Define the matrix

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix}$$

and the vector $y = (y_1, y_2, \ldots, y_n)^\top$. we observe the first column of the matrix (x_1) but not the second (x_2). We then compute the least squares estimate of β_1

$$\hat{\beta}_1 = (x_1^\top x_1)^{-1} x_1^\top y.$$

Assuming that $n = 1$ for simplicity, can you think of a plausible situation where $\hat{\beta}_1$ is a biased estimator?

Solution: A simple situation would be a case where x_1 depends on x_2, for example we could have

$$x_1 = \frac{x_2}{50000} + \epsilon'$$
$$\implies x_2 = 50000 x_1 + \epsilon'',$$

[1]Obviously these are not the only factors that influence SAT scores in the real world, but let's assume that they are for the sake of this exercise.

where ϵ' and ϵ'' are both zero mean noise variables. This corresponds to the situation where a student is more likely to be encouraged to read more books if their family is wealthy.

$$\begin{aligned}\mathbb{E}\left[\hat{\beta}_1\right] &= \mathbb{E}\left[\frac{y}{x_1}\right] \\ &= \mathbb{E}\left[\frac{\beta_1 x_1 + \beta_2 x_2 + \epsilon}{x_1}\right] \\ &= \beta_1 + 50000\beta_2.\end{aligned}$$

Depending in β_2, this can be an extremely biased estimator.

2. Now assume you decide to incentivize a randomly chosen subset of students to read more books by organizing a "readathon" at their school. Let $z_i = 1$ if student i was assigned to participate in the readathon and $z_i = -1$ otherwise. Furthemore let $z = (z_1, z_2, \ldots, z_n)^\top$.

   Since we can informally think of β_1 as the rate of change of y_i with respect to x_{i1} and since it follows from the chain rule that

   $$\frac{dy_i}{dx_{i1}} = \frac{dy_i/dz_i}{dx_{i1}/dz_i},$$

   we can derive a natural estimator of β_1 by estimating the denominator and numerator of the above fraction:

   $$\hat{\beta}_{IV} = \frac{(z^\top z)^{-1} z^\top y}{(z^\top z)^{-1} z^\top x_1}.$$

   Show that

   $$\hat{\beta}_{IV} = (z^\top x_1)^{-1} z^\top y.$$

Solution:

$$\begin{aligned}\hat{\beta}_{IV} &= \frac{(z^\top z)^{-1} z^\top y}{(z^\top z)^{-1} z^\top x_1} \\ &= ((z^\top z)^{-1} z^\top x_1)^{-1} (z^\top z)^{-1} z^\top y \\ &= (z^\top x_1)^{-1} (z^\top z)(z^\top z)^{-1} z^\top y \\ &= (z^\top x_1)^{-1} z^\top y.\end{aligned}$$

3. Now consider the two-stage least squares estimator.

   1. Find the least squares estimate with x_1 as the output and z as the input

      $$\hat{\alpha} = (z^\top z)^{-1} z^\top x_1.$$

2. Find the least squares estimate with y as the output and $\hat{x}_1 = z\hat{\alpha}$ as the input

$$\hat{\beta}_{2SLS} = (\hat{x}_1^\top \hat{x}_1)^{-1}\hat{x}_1^\top y.$$

Show that $\hat{\beta}_{2SLS} = \hat{\beta}_{IV}$.

Solution:

$$\begin{aligned}
\hat{\beta}_{2SLS} &= (z^\top \hat{\alpha} z \hat{\alpha})^{-1} z^\top \hat{\alpha} y \\
&= \hat{\alpha}^{-1}(z^\top z)^{-1} z^\top y \\
&= (z^\top x_1)^{-1}(z^\top z)(z^\top z)^{-1} z^\top y \\
&= \hat{\beta}_{IV}
\end{aligned}$$