Coursera MOOC Econometrics - Test Exercise 3

Thijs Meijerink
2022-12-18

a) First of all the researcher uses OLS to estimate the parameters of the model

$$y_{i1} - y_{i0} = \alpha + \beta d_i + \gamma y_{i0} + x_i' \delta + \varepsilon_i$$

The OLS estimator for *β* is possibly not consistent as the variable d_i may be endogenous. Clearly explain why this may be the case. Indicate whether your reason would lead OLS to overestimate or underestimate the true effect of the diet.

Answer:
Being on a diet is not random, it is choice.

b) In general there are two important conditions for variables *Z* to be useful as instruments. In formal terms these conditions are $\frac{1}{n} Z' \varepsilon \to 0$ and $\frac{1}{n} Z' X \to Q \neq 0$ as the sample size *n* grows large. Rephrase these two conditions in words in the context of this application for the above mentioned advertising variable (no formulas!).

Answer:
1. Advertisement should be independent of the other variables. So the advertisement should not be selected only in high weight areas.
2. The advertisement should result in higher probability of being on a diet

c) For both assumptions in (b), indicate whether it can be tested statistically given the available variables. If yes, indicate how. If no, why not?

Answer:
1. No, one should know what area a person belongs to.
2. Yes, one can see if the is significant higher amount of people on diet when advertised than not advertised.

d) Suppose that z_i satisfies the conditions in (b) and suppose that z_i is uncorrelated with y_{i0} and x_i. In this case the 2SLS-estimator for *β* in the model $y_{i1} - y_{i0} = \alpha + \beta d_i + \eta_i$ is consistent when a constant and z_i are used as instruments.
Show that we can write this 2SLS estimator for *β* in terms of simple sample averages. You can use the following averages:

- Average weight change over all individuals: Δ
- Average weight change over individuals with $Z_i = 1$: Δ^1
- Average weight change over individuals with $Z_i = 0$: Δ^0
- Proportion of people taking the diet: $\bar{d}$
- Proportion of people with $z_i = 1$ taking the diet: $\bar{d}^1$
- Proportion of people with $z_i = 0$ taking the diet: $\bar{d}^0$

To further explain the notation, for example:

$$\bar{d}^1 = \frac{1}{\sum_{i=1}^{n} z_i} \sum_{i=1}^{n} z_i d_i$$

Answer:

$$Z = \begin{pmatrix} 1 & \mathrm{z}_1 \\ 1 & \mathrm{z}_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & \mathrm{z}_\mathrm{n} \end{pmatrix} \qquad \text{and} \qquad \mathrm{X} = \begin{pmatrix} 1 & \mathrm{d}_1 \\ 1 & \mathrm{d}_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & \mathrm{d}_\mathrm{n} \end{pmatrix}$$

$$(\mathrm{Z'X})^{-1}\mathrm{Z'y} = \begin{pmatrix} \mathrm{n} & \sum \mathrm{d_i} \\ \mathrm{Zz_i} & \sum \mathrm{d_i z_i} \end{pmatrix}^{-1} \begin{pmatrix} \sum \mathrm{y_i} \\ \sum \mathrm{z_i y_i} \end{pmatrix}$$

$$\hat{\beta} = \frac{\mathrm{n} \sum \mathrm{z_i y_i} - \sum \mathrm{y_i} \sum \mathrm{z_i}}{\mathrm{n} \sum \mathrm{d_i z_i} - \sum \mathrm{z_i} \sum \mathrm{d_i}} = \frac{\frac{\mathrm{i}}{\sum \mathrm{z_i}} \sum \mathrm{z_i y_i} - \frac{1}{\mathrm{n}} \sum \mathrm{y_i}}{\frac{\mathrm{i}}{\sum \mathrm{z_i}} \sum \mathrm{d_i z_i} - \frac{1}{\mathrm{n}} \sum \mathrm{d_i}} = \frac{\Delta' - \Delta}{\overline{\mathrm{d}}^{\mathrm{i}} - \overline{\mathrm{d}}}$$