

Econometrics - Week 4

William Schill

Each of the answers for the exam question are posted below their respective question reference letter. Additionally, there is a bug in the mathjax html that leaves a trailing vertical bar after equations. I could not find a way to fix this.

=====

(a):

The equation has γy_{i0} on the right hand side which corresponds to a correlation effect with the left hand side. Additionally, a possible reason for the d_i variable (for having followed the diet in the past year) might be endogenous is the length at which the diet was actually followed. If the individual had only followed the diet for a short time but had great weight loss, then it may have been due to factors found within the error or vice verse.

Additionally, if there were significant variation of the length of time to which the diet was followed over the past year with significant variation of weight loss among these times, then the model may underestimate if individuals dropped off of the diet early due to lack of weight loss results or stayed on for the entire year and still had insignificant results.

The weight loss variable may increase with time spent on diet.

=====

(b):

(i): The equation, $1/n * Z' \epsilon \rightarrow 0$ refers to the instrument variable (in this case the door-to-door advertising influence) and it is essentially a sum over observations and including $1/n$ term is like an average and it has a value that converges to zero as the number of individuals increases. In other words, as the number of individuals (the sample size) increases, and if the advertising is not correlated to the error then their dot product will not increase with increasing n and the full equation will get smaller and smaller, converging to zero. Mathematically, it says that Z is orthogonal to ϵ and has 0 correlation.

(ii): The equation, $1/n * Z' X \rightarrow Q \neq 0$, is saying that as the number of individuals increase, then the dot product of the instrument variable if correlated with X will converge to some value Q as the dot product will continually change due to the correlation and the full value of the equation will not decrease with increasing n . Door-to-door advertising is correlated with the diet's efficacy and an improvement (for example) gives a direct improvement to the other.

=====

(c):

For part **b(i)**, it is difficult to test whether or not this is true in this case. Remember as stated above, that this is approximately equal to :

$$\text{cov}(Z, \epsilon) \rightarrow 0$$

Since there is no residual estimator to take into account here, $\text{cov}(Z, \epsilon)/\text{var}(\epsilon)$, this has to be confirmed using some sort of intuition.

For part **b(ii)**, and following the same information above, we would have to test if:

$$\text{cov}(Z, X) \rightarrow Q \neq 0$$

This can be confirmed by using the left hand side of the inequality and assuming that:

$$Q = \text{cov}(Z, X)/\text{var}(X)$$

Then we can test for $H_0 : Q = 0$ against $H_1 : Q \neq 0$ so for our regression here we can reject H_0 as that would imply we have a unique choice of instrument variable.

=====

(d):

For two stage least squares, given $y_{i1} - y_{i0} = \alpha + \beta d_i + \mu_i$,

Assume z_i satisfies the conditions in (b), is an instrument for d_i , and z_i is uncorrelated with y_{i0} and x_i .

We want to show that $\hat{\beta} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{d}_1 - \bar{d}_0}$.

We know that,

$$\hat{\beta} \approx (z_i' d_i)^{-1} (z_i' y) \approx \frac{\text{cov}(z_i, y_i)}{\text{cov}(z_i, d_i)} = \frac{\sum_{i=1}^n (y_i - \bar{y})(z_i - \bar{z})}{\sum_{i=1}^n (d_i - \bar{d})(z_i - \bar{z})}$$

Since z_i is binary as described in the problem, let n_1 be the number of 1's in the sample and n_0 be the number of 0's in the sample, then we could say that $\bar{z} = \frac{n_1}{n_1 + n_0}$.

By expanding the terms in the summation, we simplify to the following:

$$\hat{\beta} = \frac{\sum_{z_i=1} y_i - \bar{z} \sum_{i=1}^n y_i - \bar{y} \sum_{i=1}^n z_i + n \bar{y} \bar{z}}{\sum_{z_i=1} d_i - \bar{z} \sum_{i=1}^n d_i - \bar{d} \sum_{i=1}^n z_i + n \bar{d} \bar{z}} = \dots$$

Then rearranging, we have:

$$\dots = \frac{n_1 \bar{y}_1 - \bar{z} (n_1 \bar{y}_1 + n_0 \bar{y}_0)}{n_1 \bar{d}_1 - \bar{z} (n_1 \bar{d}_1 + n_0 \bar{d}_0)} = \dots = \frac{\frac{n_1 n_0}{n_1 + n_0} \bar{y}_1 - \frac{n_1 n_0}{n_1 + n_0} \bar{y}_0}{\frac{n_1 n_0}{n_1 + n_0} \bar{d}_1 - \frac{n_1 n_0}{n_1 + n_0} \bar{d}_0} = \dots$$

Then further simplification and cancelation of terms yields:

$$\hat{\beta} = \frac{\frac{n_1 n_0}{n_1 + n_0} (\bar{y}_1 - \bar{y}_0)}{\frac{n_1 n_0}{n_1 + n_0} (\bar{d}_1 - \bar{d}_0)} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{d}_1 - \bar{d}_0}$$

In []: