Causal Inference - Problem Set 4

Gustavo Baroni

9/29/2020

Question 1

Consider the model

$$Y_i t = \gamma_t + \lambda_i + \beta_1 D_i t + \epsilon_i t$$

[1] where, $E[\epsilon_i t | D_i t, t] = 0$. $D_i t$ is equal to 1 if the individual i in period t was treated and 0, otherwise. β gauges the causal effect of treatment on $Y_i t$.

(a) Assuming that $E[\epsilon_i t Dit] = 0$. Show that β is equal to the two differences in the mean of Y.

By difference-in-differences[2],

$$[E[Y_i t | D_i t = 1, t = B] - E[Y_i t | D_i t = 1, t = A]] - [E[Y_i t | D_i t = 0, t = B] - E[Y_i t | D_i t = 0, t = A]] = \beta$$

Where $t = B$ and $t = A$ are different periods of time.

(b) How does this result relates to the assumption necessary in a standard difference-in-difference model?

This result is given by the difference-in-differences model, so it relates perfectly.

(c) Consider the example: The outcome is the incidence with the dengue diseases. The treatment consists of volunteers who assist in the cleansing of potential foci of the disease transmitter. What types of unobserved variables do the difference-in-differences regression equation control? What types of unobserved variables do not control?

The difference-in-differences regression equation control

Question 2

Using pset4.csv, restrict your sample to years 2002 and 2007. Consider the following model:

$$larreacd_i t = c_i + \beta_0 + \beta_1 lpib_i t + \beta_2 lprop_s erv_i t + \beta_3 lpop_i t + \delta_0 a2007_t + u_i t$$

Where i is the unit of cross section, t is the period and a200 is 1 if $t = 2007$ and 0 if $t = 2002$.

(a) Assume c_i is non-stochastic and represents the fixed effect. What is the interpretation of ci? What is it different to $u_i t$?

(b) Estimate the Pooled OLS model?

(c) Estimate the first difference equation of this model. Describe the hypotheses needed for this estimator.

(d) Estimate the fixed effect model. Describe the hypotheses needed for this estimator.

(e) Compare the estimates on (c) and (d).

(f) Compare the estimates on (b) and (c). Explain the pros e cons of each estimator.

[1] My $Y_i t$ isn't in the way that I want: I tryed to put the "t" small next to the "i", but I couldn't neither by doing $Y_(it)$ nor $Y_[it]$. I would appreciate any help whith that for the next HWs!

[2] Angrist, Joshua D. and Jorn-Steffen Pischke. 2008. Mostly Harmless Econometrics: An Empiricist's Companion. Mimeo. (Available *HERE*)

Question 3

Using pset4.csv, restrict your sample to years 2002, 2004 and 2007. Consider the following model:

$$larreacd_it = c_i + \beta_0 + \beta_1 lpib_it + \beta_2 lprop_serv_it + \beta_3 lpop_it + \delta_0 a2004_t + \delta_1 a2007_t + u_it$$

Where i is the unit of cross section, t is the period, a2007 is 1 if $t = 2007$ and 0 if otherwise and a2004 is 1 if $t = 2004$ and 0 otherwise.

(a) Estimate the first difference equation of this model.

(b) Estimate the fixed effect model.

(c) Compare the estimates on (a) and (b).