ECON 626: Empirical Microeconomics

How Much Should We Trust Difference-in-Differences Estimates?

Department of Economics
University of Maryland
Fall 2019

This exercise is based on Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan's paper "How Much Should We Trust Differences-in-Differences Estimates?" (*QJE*, 2004). The paper examines difference-in-difference estimation in a panel data framework when both outcomes and treatments are serially correlated.

1. The data set `wdidata.xlsx` contains information on PPP-adjusted GDP per capita for 148 countries and 26 years (from 1990 to 2015); data come from the World Bank's World Development Indicators Database. The Stata `.do` file `GDPsimulation.do` reads the file into Stata and reshapes into long form, so that the unit of observation is the country×year.

   (a) Extend the `.do` file by so that it does the following 100 times: (i) randomly assigns treatment at the country×year level (i.e. generate a treatment dummy that is randomly assigned across country×year observations); (ii) estimates a regression of log GDP per capita on the treatment dummy, country fixed effects, and year fixed effects; and (iii) stores the p-value associated with the treatment variable. When treatment is randomly assigned across country×year observations, does your test appear to be correctly sized?

   (b) Modify your program so that the estimation has a difference-in-difference structure: chose a year at random between 1995 and 2010 when the treatment starts, and select 74 of the 148 countries to be treated (in the sense that the treatment dummy is equal to one in those countries for every year after the treatment starts). Simulate this random treatment-assignment process a hundred times and record the p-value associated with a test that the coefficient on treatment is equal to zero when you control for country fixed effects and year fixed effects. Is your test correctly sized?

   (c) Implement one of the proposed fixes suggested in Bertrand, Duflo, and Mullainathan (2004): use the `egen` command with the `mean` option to calculate the average value of log GDP per capita in the pre-treatment and post-treatment period for every country in the data set. Use the `egen` command with the `tag` option option to tag one observation per country before treatment and and one observation per country after treatment. Simulate the random treatment assignment process one hundred times; in each iteration, regress mean log GDP per capita on the treatment dummy in your sample of $2N$ observations. How many times out of 100 do you reject the null hypothesis of no effect of a random treatment?

   (d) Implement the other fix proposed in Bertrand, Duflo, and Mullainathan (2004) by clustering your standard errors at the country level. Does this lead to a correctly-sized test?