

ECON 626: Empirical Microeconomics

Difference-in-Differences Estimation: the Basics

Department of Economics
University of Maryland
Fall 2019

The data set `DD-data1.dta` contains observations of student test scores (the `score` variable) for students in Grade 7 and 8, normalized to be out of 100 percent. The data set includes observations from two years, 2008 and 2009, for two schools. Students in the sample were in Grade 7 in 2008 and in Grade 8 in 2009. In 2009, School 1 was selected to receive a package of educational inputs (textbooks, flipcharts, and study guides) from a local non-profit. School 2 was not selected to participate in the program.

1. Generate a `treatment` variable that is equal to 1 for the program school in the year after the program (2009) and equal to zero otherwise.
2. Calculate the naive treatment vs. comparison estimator by estimating a regression of `score` on the `treatment` variable. Restrict the sample so that you only use data from 2009 (i.e. from the post-program period).
3. Calculate the naive pre-treatment vs. post-treatment estimator by estimating a regression of `score` on a dummy for the post-treatment period (2009). How do you need to restrict the sample when you run this regression?
4. Estimate the impact of the treatment in a difference-in-difference framework, controlling for selection bias and time effects. Discuss the relationship between the estimated coefficients and your answers to (1b) and (1c).
5. Reshape the data using the `reshape` command with the `wide` option so that you have one observation per student (i.e. per ID number). Show that a first-difference specification generates the same results difference-in-difference estimation.