

ECON 523: Program Evaluation for International Development

In-Class Activity 5

In this exercise, we'll be using a data set on primary school enrollment and completion in 35 African countries, 15 of which eliminated primary school fees between 1990 and 2015. Raw data on country-level enrollment and completion rates comes from the World Bank's World Development Indicators Database. We will use this data set to generate two-way fixed effects (TWFE) estimates of the impact of eliminating school fees on enrollment and completion. Since this policy was phased in by different countries at different times, it is a useful setting for exploring the strengths and weaknesses of two-way fixed effects.

Before you begin, create a do file (with all the standard stuff at the top) that downloads the data set `E5-fpdata.dta` from the course website. In this part of the activity, we'll be using `enroll` as our outcome variable. Drop country-years for which `enroll` is missing.

Start by familiarizing yourself with the data set. What years does it cover? What variable indicates the year in which a country implemented free primary education (FPE)? How many countries implemented it? The gross primary enrollment ratio is 100 times the number of students enrolled in primary school divided by the number of primary-school-aged children. This number can be greater than 100 when over-age children are enrolled in primary school - which often happens when school fees are eliminated. What was the average level of primary school enrollment in 1981 (at the beginning of the data set)? What was the average year of primary school enrollment in the last year for which data is available? In how many country-years does the gross primary enrollment ratio above 100?

Extend your do file as you answer the following questions:

1. Generate a treatment dummy `fpe` that is equal to one for years where primary school is free (i.e. all years starting from the year where FPE was implemented). Label this variable "Free primary education". What is the mean of this variable across all country-years in the data set?
2. Regress gross enrollment on `fpe` controlling for country and year fixed effects. Restrict the sample to countries that eventually implemented FPE (an easy way to do this is to go up a few lines and make sure `fpe` is missing for observations in countries that never implemented free primary). Cluster your standard errors at the country level. What is the estimated impact of eliminating school fees on enrollment?
3. Next we are going to construct the TWFE estimate of the impact of FPE "by hand" using the residuals.
 - (a) Regress `fpe` on country and year fixed effects, and generate a variable `tresid` (short for treatment residual) equal to the residuals from the regression (make sure that you dropped the country-years with missing values of the outcome variable).
 - (b) Regress `enroll` on country and year fixed effects, and generate a variable `yresid` that contains the residuals from that regression.

- (c) Regress `yresid` on `tresid`. Confirm that you recover your TWFE estimate from Question 2.
 - (d) What fraction of the treated country-years received negative weight in our TWFE regression?
 - (e) The TWFE coefficient is a linear combination of the observed values of the outcome variable, with each value of Y is weighted by the associated residualized value of treatment (`tresid`) divided by the sum of all the squared values of `tresid`. Confirm that this is correct by: (1) calculating a variable `tr2` equal to the square of `tresid`, (2) using `egen`'s `sum` option to calculate a variable `tvar` equal to the sum of `tr2` across all observations, (3) generate a `weight` variable equal to `tresid` divided by `tvar`, (4) generate a variable `yxweight` that is equal to the observed value of `enroll` (the outcome variable) times the regression weight, and (5) calculate the TWFE coefficient as the sum of `yxweight` across all observations.
 - (f) Now that you know you can do this by hand, we will never do it again. Drop `yresid`, `tresid`, `tr2`, `weight`, `yxweight`, and `betahat`.
4. Now rerun your TWFE regression including the never-treated countries. How many treated country-years are negatively weighted now?