

ECON 523: Program Evaluation for International Development

Empirical Exercise 2

This exercise makes use of the data set E2-BanerjeeEtAl-data.dta, a subset of the data used in the paper “A multifaceted program causes lasting progress for the very poor: Evidence from six countries” by Abhijit Banerjee, Esther Duflo, Nathanael Goldberg, Dean Karlan, William Pariente, Jeremy Shapiro, Bram Thuysbaert, and Chris Udry, published in *Science* in 2015.

The authors examine the impacts of a “graduation” program first designed by the Bangladeshi NGO BRAC. The program offers extremely poor households an asset transfer, temporary consumption support, skills training, home visits, and access to savings technologies. The program was evaluated through a randomized trial in six countries.

In this exercise, we use data on the program’s impacts on food security to explore the mechanics of fixed effects. For this exercise, we’re going to drop all the observations in the treatment group, and then simulate alternative scenarios to better understand how fixed effects work.

Before you begin, save a new do file containing the Stata code below.

```
** preliminaries
clear all
set more off
set seed 12345

** load the data from the course website
webuse set https://pjakiela.github.io/ECON523/exercises
webuse E2-BanerjeeEtAl-data.dta

** drop observations in the treatment group
drop if treatment==1
drop treatment

** randomly assign observations to four equally-sized groups
gen randnum = runiform()
sort country randnum
by country: gen within_id = _n
gen group = mod(within_id,4)
replace group = 4 if group==0
sort country within_id
```

Extend your do file as you answer the following questions:

1. **Fixed effects when p is constant across countries.**

- (a) Create a treatment variable `t1` and assign observations in groups 1 and 2 to treatment. Then, create a variable `impact1` that is equal to 2 for observations in the treatment group and 0 otherwise. This is the treatment effect for the purposes of this (first) simulation. Generate an outcome variable `y1` that is endline foodsecurity (`e_foodsec`) plus `impact1`. Now regress `y1` on `t1` with and without country fixed effects. How do the estimated treatments effect and the levels of statistical significant compare across the two specifications?
- (b) When treatment does not vary across countries, including country fixed effects is not necessary – but it may increase statistical power. In the example above, fixed effects did not improve statistical power much because the mean does not vary across countries (it is normalized to zero in the control group in every country). Change this by increasing `y1` by 10 in two countries and decreasing `y1` by 20 in two other countries. Now rerun your two regressions (with and without fixed effects). You should see that including fixed effects now changes the standard error on your estimated treatment effect substantially (though it still should not impact your estimated coefficient much).
- (c) When p is fixed, the estimated coefficient from a regression with fixed effects is a weighted average of the estimated country-specific treatment effects (i.e. the within-country differences in means). The weights are the share of the total sample size within each country. Given this, if you increased the treatment effect in Peru from 2 to 11, what you expect the treatment effect to be? See whether this is true in practice (by changing the treatment effect in Peru and then re-running your fixed effects regression).

2. When are fixed effects necessary?

Fixed effects are needed when treatment probabilities vary across countries **and** the mean of the outcome variable also varies across countries (because then treatment is correlated with the outcome, even in the absence of a treatment effect). To see this, generate a variable `t2` that is equal to 1 for all observations in group 1 plus the observations in group 2 in Ethiopia, Ghana, and Honduras (countries 1, 2, and 3). In this case, we are not going to add any treatment effect. Generate an outcome variable `y2` that is equal to food security, and then add 5 to it in Ethiopia, Ghana, and Honduras (for observations in the treatment and control groups in those countries). How do the results of regressions with and without country fixed effects compare?

3. How observations are weighted with fixed effects.

For the last question, we need to have the same number of observations in each country. The code below does this. You can see that we now have equal numbers of observations from groups 1, 2, 3, and 4 in each country as well.

```
keep if within<=360 // 360 obs per country
tab country group
```

Now generate a treatment variable `t3`. `t3` should be equal to one for observations in group 1 in Ethiopia and Ghana. `t3` should be equal to one for observations in groups 1 and 2 in Honduras and India. `t3` should be equal to 1 for observations in groups 1, 2, and 3 in Pakistan and Peru. Given this, what is the proportion treated in each country?

- (a) First, consider what happens when we **only** have a treatment effect in the countries with the lowest proportion treated. Create a variable `impact3` that is equal to 10 for treated observations in Ethiopia and Ghana, and equal to zero for everybody else. Then, create an outcome variable `y3a` that is the sum of `e_foodsec` and `impact3a`. You can see the average treatment effect across all the treated observations in the sample summarizing `impact3a` among all treated individuals. How does that compare to the results of regressions with and without fixed effects, or to the results from a regression that only includes data from Ethiopia and Ghana?
- (b) Now replicate the exercise above, but have the treatment effect occur in Honduras and India (where the proportion treated is one half) rather than in Ethiopia and Ghana (where the proportion treated is one quarter). Generate new variables `impact3b` and `y3b` and repeat your analysis.
- (c) Now replicate the exercise again, but have the treatment effect occur in Pakistan and peru (where the proportion treated is three quarters) rather than in Honduras and India (where the proportion treated is one half). Generate new variables `impact3c` and `y3c` and repeat your analysis.

4. Takeaways.

Based on the above, which countries received relatively low weight in the analysis of Banerjee et al. because the proportion treated was relatively low? How do you think that might have impacted their results?