ECON 523: Program Evaluation for International Development
In-Class Activity 6

In this exercise, we'll be using data from the paper "The Miracle of Microfinance? Evidence from a Randomized Evaluation" by Abhijit Banerjee, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. The paper reports the results of one of the first randomized evaluations of a microcredit intervention. The authors worked with an Indian MFI (microfinance institution) called Spandana that was expanding into the city of Hyderabad, India. Spandana identified 104 neighborhoods where it would be willing to open branches. They couldn't open branches in all the neighborhoods simultaneously, so they worked with the researchers to assign half of them to a treatment group where branches would be opened immediately. Spandana held off on opening branches in the control neighborhoods until after the study.

We'll be using a small slice of the data to explore the use of instrumental variables techniques to estimate impacts of treatment on the treated - and to think about when such methods are appropriate. To get started, open `ECON523-E6-inclass.do`, change the file path as needed in the do file, and download the data set `E6-BanerjeeEtAl-data.dta` from the course website.

Start by familiarizing yourself with the data set. We are going to make use of the variables treatment and `spandana_1`. Before you begin, add a line to your do file that drops any observations with one of these variables missing.

Extend your do file as you answer the following questions:

1. Estimate the impact of treatment on the likelihood of taking a loan from Spandana (the variable `spandana_1`). What is the estimated coefficient on treatment? Because treatment is randomly assigned at the neighborhood level, we need to cluster our standard errors by neighborhood. Do this. This is the **first stage** regression. Store the coefficient on treatment as the local `beta_fs`.

2. Now extend your do file so that you also run the **reduced form** regression of microenterprise profits (the variable `bizprofi_1`) on treatment. What is the estimated impact of being randomly assigned to a treatment (at the neighborhood level) on business profits? Store the coefficient on treatment as the local `beta_rf`.

3. Based on your answers to Questions 1 and 2, what is the estimated impact of **treatment-on-the-treated**? Write this down.

4. Now we want to output our results to Excel.

   (a) We're going to use the `putexcel` command to write our results into an Excel file. `putexcel` is a simple command that allows you to write Stata output to a particular cell or set of cells in an Excel file. Before getting started with `putexcel`, use the `pwd` ("print working directory") command in the Stata command window to make sure that you are writing your results to an appropriate file. Use the `cd` command to change your file path if necessary. Then set up the Excel file that will receive your results using the command `putexcel set`. Use the code below to setup a blank Excel table where you can store your results:

```
putexcel set E6-TOT-in-class.xlsx, replace
putexcel A1=" ", hcenter border(top)
putexcel A2=" ", hcenter border(bottom)
putexcel B1="Borrowed", hcenter bold border(top)
putexcel B2="(1)", hcenter bold border(bottom)
putexcel C1="Profits", hcenter bold border(top)
putexcel C2="(2)", hcenter bold border(bottom)
putexcel A3="Treatment", bold
putexcel A6="Observations", bold border(bottom)
```

(b) At this point, it is worth opening your Excel file to make sure that you are writing to it successfully. Be sure to close the file after you look at it; Stata won't write over an open Excel file. The column and row labels should all appear in bold font (the `bold` option), and the column headings in cells B1 and C1 should be centered (the `hcenter` option) and have a border above them (the `border()` option).

The next step is to write your regression results to Excel. We are going to do this by writing a program. We've already seen that we can store our regression results in a Stata matrix using the command

```
mat V = r(table)
```

after running a regression. This allows us to extract both the standard error and the p-value associated with each regression coefficient (something that is difficult to do using esttab). The program below adds a column to our Excel file containing the results of an additional regression. Review the code below carefully to make sure that you understand each line. Then add two lines to the program to write the p-value associated with the regression coefficient in Row 5 of the spreadsheet. Put the p-value in square brackets rather than parentheses.

```
cap program drop tabcolumn
program define tabcolumn // tabrow var columnletter
reg `1' treatment, cluster(areaid)
mat V = r(table)
local beta = string(V[1,1],"%04.3f")
local se = string(V[2,1],"%04.3f")
putexcel `2'3="`beta'", hcenter
putexcel `2'4="(`se')", hcenter
putexcel `2'6=`e(N)', hcenter border(bottom)
end

tabcolumn spandana_1 B
tabcolumn bizprofit_1 C
```

(c) **Optional.** You may want to set the widths of the columns in your Excel file. Unfortunately, there is no way to do this using `putexcel`. The code below invokes Stata's `mata` programming language to adjust the column widths. You can also just adjust them as needed by hand before you print your table to a pdf.

```
mata
b = xl()
b.load_book("E6-TOT-in-class.xlsx")
b.set_sheet("Sheet1")
b.set_column_width(1,1,20) // make variable name column widest
b.set_column_width(2,3,16) // width for subsequent columns
b.set_row_height(7,7,32)
b.close_book()
end
```

(d) Add a note at the bottom of your table that explains the contents of the table.