

EXERCISE SHEET 8: LA FERRARA ET AL. (2012) - PART 2

Main learning goals: 1. Regional fixed effects in individual data vs. regional panel

2. Clustering of standard errors
3. Dealing with specification/estimation errors
4. Estimating and plotting yearly coefficients in panel data
5. Accessing coefficient estimates

Hand-in: As usual: do-file and the corresponding log-file (saved in .log-format) containing the following steps (numbered in the do-file).

Moreover, upload a replicated version of the figure from exercise 6, saved under the name `familyname_graph8.pdf`.

1. Use the dataset `LCD_new.dta` and `collapse` the dataset to a regional level panel, containing the average values of the dependent and explanatory variables and population size (remember the sampling weights); save it under a different name.

Use `xtset` to define the resulting dataset as panel data and `xtdes` to inspect it.

2. Rerun columns (3) and (6) of table 2, now using district level averages instead of individual data and applying two different techniques. First estimate the models unweighted, then use population to weight each individual estimate. Compare the four sets of results from the regional panel to the individual level estimates from table 2 of the paper and briefly comment on the differences in your do-file.
3. Think about the i.i.d.-assumptions of the error terms, in particular about the independence assumption. To do so, estimate the previous regression for column (6) of table 2 first as pooled OLS (with and without clustered standard errors) and then with AMC fixed effects, but without cluster robust standard errors. Then, compare the results to the previous results with clustered standard errors. Use your predicted errors to compare the distribution across regions and briefly discuss.

4. Return to the dataset `LCD_new.dta`.

The variables `yrsexp*` are supposed to measure for how many years a given female was exposed to Globo programs while being of a specific age group (10 to 19, 20 to 29, 30 to 39, and 40 to 49). Sadly, these variables suffer from data generation errors. Look at the code in the file `corrected_exposure.do`, which generates these age exposure variables from scratch. Include this do-file in your code.

Comment on the data generation of `_yrsexp*` and the difference between `yrsexp*` and `_yrsexp*` in your do-file.

5. Re-estimate columns (4) and (5) of table (4), first using the `yrsexp*` and then your own `_yrsexp*` variables. Compare the two sets of results and comment shortly on the difference in the results in your do-file.
6. Re-estimate the specification in equation (4), collect coefficient estimates and generate the graph in figure 4 using `twoway line`. Display 90% confidence intervals. This graph is to be handed in.

Further comments: Submit 3 files (.do, .log, .pdf(with a nice graph)) to Ilias. Usual procedures apply.