PLSC 597
Causal Inference
Homework 2
Due: April 4th 11:59pm

You are going to replicate some of the results in the following paper,

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti. 2004. "Economic Shocks and Civil Conflict: An Instrumental Variable Approach." Journal of Political Economy. 112(4):725-753.

The paper and replication material are here.

Carry out the following:

1. Discuss the specific contribution of each of the following aspects of their empirical approach in addressing concerns related to non-random assignment of year-to-year GDP growth:

(a) Using rainfall growth as an instrument for GDP growth.
(b) Country fixed effects.
(c) Country time trends.

With these elements included, can you think of sources of confounding that may still be a reason for concern?

2. Replicate the estimates in Tables 2, 3, and 4. Briefly discuss the main results. (This is not meant to be punishing, but rather get you used to downloading and replicating published papers. Don't worry about getting all the formatting right.)

3. Take advantage of the country-specific heterogeneity to construct a placebo test to provide circumstantial evidence on the exclusion restriction:

• Run the simple first-stage regression (that is, the one with no covariates, country FEs, or time trend/FEs) going country-by-country. Look at the point estimates (ignore pvalues or confidence intervals) for the coefficients on the rainfall growth variable and its lag. For which countries do these coefficients take values that are contrary to the one hypothesized in the paper — that is, both coefficients are either negative or zero through the first and second decimal places, implying no substantial relationship?
• Re-run the first stage regression just using the countries identified in the previous step. You should find that the first stage, when restricted for these countries, is not valid for an IV analysis of the sort proposed in the paper.

• Now, estimate the reduced form regression just using the countries identified as having no valid first stage. Are your findings in favor or against Miguel et al.'s identification strategy? Why?

Now I'd like you to complete the following exercise for two papers published within the past five years using the "close election RDD." For each paper, make sure there is replication data available (should be the case if published in reputable journals…)

For each paper, answer the following questions:

- What is the "running variable" that the author(s) use for the regression discontinuity and how is this used in the regression specification?
- What standard error estimator does the author(s) use and what types of dependencies does this estimator address?
- What are the bandwidths that the author(s) use to estimate the effect?

Now, for each paper, download the replication data and generate two plots of the RDD for each paper. It's likely that the authors have done some kind of plot themselves, either to demonstrate the outcome or to justify their identification assumptions; you can either adapt theirs or generate your own.

One plot should demonstrate the plausibility of the RDD identification assumption – either the absence of "heaps" or "bunching," or smoothness in the covariates around the discontinuity. Explain how the plot you generate is important for the RDD identification argument.

The other plot should demonstrate the LATE itself. You can decide how to visualize the data, including the bandwidths or the inclusion of lines on top of the data points. Explain what is being shown, with specific reference to the L(ocal) in LATE.

Finally, compare the theoretical arguments about the relevance of the LATE and the RDD identification strategy across the two papers. When read in combination, do any contradictions or complications emerge, or are the results possibly consistent?