ReadMe

Article details:

- Title: “Mass political information on social media: Facebook ads, electorate saturation, and electoral accountability in Mexico”
- Authors: José Ramón Enríquez, Horacio Larreguy, John Marshall, and Alberto Simpser
- Journal: Journal of the European Economic Association

Description of datasets used for final estimation:

- **“ad_reach_characteristics.dta”**:
    - Description: Facebook campaign engagement data aggregated at the municipality level. This dataset is sourced directly from Facebook analytics reports for each Facebook page that delivered ad campaigns. As in the rest of the analysis, we pool the common knowledge and non-common knowledge treatment variants of treatment within high saturation municipalities.
    - Purpose: This data is used to estimate differences engagement with Facebooks ads across campaigns in high and low saturation municipalities in Appendix Table D.1.
- “**DIST_S_ALL.dta**”:
    - Description: This dataset calculates the distance from each electoral precinct within the experimental sample to all other segments in the same municipality within 10km. This was computed using Mexico’s electoral precinct shapefile using ARC GIS; this entailed identifying all precincts within 10km of a given precinct and restricting attention to those in different segments within the same municipality.
    - Purpose: This implements Appendix Table H.3, one of several tests designed to assess the extent of ad mistargeting. The test is described in detail in Appendix Section H.2.
- **“Facebook_analytics_FS_Data.dta”**:
    - Description: Facebook campaign engagement data between June 15th and July 9th, 2018 aggregated at the municipality-day level. This dataset draws from Facebook analytics reports for each Facebook page that delivered ad campaigns. As in the rest of the analysis, we pool the common knowledge and non-common knowledge treatment variants of treatment within high saturation municipalities.
    - Purpose: This dataset is used to plot engagement dynamics across ad campaigns in Appendix Figure D.1.
- **“facebook_ads.dta”**:
    - Description: This is a Facebook ad campaign-level dataset derived from Facebook engagement metrics summed across days during the intervention period.
    - Purpose: This dataset is used to characterize levels of engagement with the intervention on Facebook in Table 1, Appendix Figure D.2, and Appendix Table I.1.
- **“Main_Precinct_Dataset.dta”**:

    - Description: This is the core precinct-level dataset used throughout the analysis. It includes the following municipality-level variables: state and municipality identifiers, audit results assembled from ASF audit reports, 2015 and 2018 municipal incumbents and election winners, and 2010 census data from INEGI. It further includes the following precinct- and segment-level variables: 2015 and 2018 municipal and federal election returns (and related variables) from state and federal electoral institutes, Facebook data from the Social Connectedness Index, and 2010 census data from INEGI. Finally, it includes experimental information including municipal and segment treatment assignments, randomization blocks, and the number of municipalities within a block. The randomization code is available upon request.
    - Purpose: This dataset is the building block for the majority of analyses in the paper.
- **"Municipal_Level_Census_2015_Subset.dta"**:
    - Description: This dataset includes municipal-level variables from the 2015 intercensus, originally source from INEGI.
    - Purpose: This dataset is used for producing the sample summary statistics in Appendix Table B.1, municipal-level covariate balance tests in Appendix Table F.1, and municipal summary statistics by reported irregularities in Appendix Table F.5.
- **"Newspapers.xls"**:
    - Description: This xls files includes a list of articles scraped from 263 local newspaper that directly mention the ASF reports, indirectly mention the ASF reports, or mention corruption more generally. Appendix Section I.4 provides details on the construction of this dataset.
    - Purpose: This dataset is used in Appendix Table I.7 to examine the effect of the intervention on newspaper reporting on corruption in general.
- **"Permuted datasets"**:
    - Description: Each of the "Main_Precinct_Dataset_X.dta" datasets contained within this folder is a subset of variables from "Main_Precinct_Dataset.dta" for one of 5,000 alternative random assignments. The alternative random assignments followed identical protocols to the original two-level randomization.
    - Purpose: This set of datasets is used to conduct the randomization inference robustness checks in Figure 6.
- **"Randomized_Municipal_Saturations.dta"**:
    - Description: This is a municipal-level dataset for all municipalities, including the two without electoral data. The dataset includes municipality identifiers, audit results assembled from ASF audit reports, and experimental blocks and treatment assignments. This file was generated by our municipal-level randomization.
    - Purpose: This dataset is used to report the distribution of irregularities across municipalities in our sample in Figure 2.
- **"Spillover_Dataset.dta"**:
    - Description: A similar dataset to "Main_Precinct_Dataset.dta", except where observations are precincts in non-experimental municipalities that are within 5kms of a single segment in an experimental municipality. These distances were first computed using ARC GIS, with the sample being narrowed down based on the set of nearby municipalities.

- Purpose: This dataset is used to estimate the effect of the intervention on precinct-level outcomes in nearby municipalities in Appendix Table G.2.

All variables are labelled within these datasets.

Description of program:

- **"replication_code.do"**: Stata do file that implements all of figures and tables in the main paper and online appendix, as well as a Stata log file entitled "Replication_Results_Log.smcl". This was implemented in Stata 17, using the following installed packages: "estout" and "reghdfe".