Confounding, spillovers, and interactions influence estimates of social distancing policy effects

README

1 System requirements:

We use three pieces of software:

- Python 3.7.6 via Jupyter Notebook: scraping of government social media accounts and newspapers to generate part of the policy database.
- R 4.0.2 via RStudio 1.3.1073 (all packages used on R as specified in the code): production of figures.
- Stata/MP 16.0 and Stata/MP 17.0 (Stata packages pdslasso (v1.1), lassopack (v1.3.1), and reghdfe (5.8.0)): statistical analysis.

2 Installation guide:

To open Jupyter Notebook:

- Open Terminal
- type `cd [NEW WORKING DIRECTORY]`, for example, `cd Desktop`
- type `jupyter notebook`—a new window of default browser will pop up

To open/install R Studio:

- Download and follow instructions for RStudio Desktop in this link. This is an open source software.

To open/install Stata/MP 17.0:

- Download and follow instructions for Stata/MP 17.0 in this link. This is not a free software.
- Once installed, start a new session in Stata.
- Install the following packages `pdslasso`, `lassopack`, and `reghdfe` by typing: `ssc intall [NAME OF THE PACKAGE]` followed by `ENTER`.

3 Instructions for use:

Running the statistical analysis

To conduct the statistical analyses, run *DoFile_master.do* in Folder "Analysis" to run analyses (highlight code and type `Command + Shift + D` or `Ctrl + Shift + D`). Output is the final database (`panel_COVIDTreatMR`) and five csv files containing the estimates of all models: `m1.csv`, `m2.csv`, `m3.csv`, `m4.csv`, and `m5.csv`. First three models take about 3 minutes to run each; fourth model takes about 10 minutes; and fifth model takes about 3 hours. Further below in this README file we describe how to produce the figures.

Further details on the structure of the statistical analysis:

There is no need to run the following components separately; they are all run by the master do file (*DoFile_master.do*), as explained above.

Construction of databases

Folder "Dataset".

COVID

Run *DoFile_CleanPanelCOVID.do* on Stata. The code produces a clean database at the municipality-day level, `panel_COVID_fixdates`, based on `08232020COVID19MEXICO.csv`, which is a database directly downloaded from the Mexican Ministry of Health website on August 2020 (the present analysis is restricted to the first semester of 2020).

Lockdowns

Run *DoFile_CleanLockdowns.do* on Stata. This code cleans and harmonizes multiple datasets of social distancing policies at the municipal-day level. The database produced is `panel_COVID_Lockdowns`. In particular, it appends data from newspapers, Twitter, Facebook, as well as local governments' websites and social media accounts.

The first part of this database is generated by running all files with the extension ipynb: `0-scrape`, `search_accounts`, `local-accounts`, `local-gob_mapping`, `tweets_gobiernos-locales`, `tweets_gobiernos_estatales`, `facebook-local_gob`, and `policy_actions` in that order on Jupyter Notebook. The output is curated by hand in the file `policy_action.csv`.

The second part of the database is coded by hand searching terms online using a dictionary. These data is contained in `Entidad.xlsx`.

Facebook Mobility data

Run *DoFile_CleanFBMovementRange.do* on Stata. This code both cleans and consolidates a database of Facebook mobility data at the municipal-day level and also merges the previous two databases (`panel_COVID_fixdates` and `panel_COVID_Lockdowns`). The output is a municipal-day level database, `panel_COVIDTreatMR`, that contains epidemiological data, social distancing policies data, and Facebook mobility data including mobility weights.

Analysis

Folder "Analysis".

Main analysis

Main analysis is performed by running *DoFile_analysis.do* on Stata. The code yields, with the help of *DoFile_append_m**, five .csv files: `m1.csv`, `m2.csv`, `m3.csv`, `m4.csv`, and `m5.csv` named after the model number. Each row in such files contains a point estimate, 95% confidence intervals, number of observations, as well as the indicator of the model.

Regression estimates

Regression table in Supplementary materials (Table S1) is produced by running *DoFile_regressiontable.do*, which estimates four models and calculates differences between point estimates. The code produces `pvalues_models1-4.csv`.

Creation of figures:

Folder "Figures".

Main (Figures 2, 3, 4, S5, S6, S7)

Based on the five .csv files produced in the main analysis (`m1.csv`, `m2.csv`, `m3.csv`, `m4.csv`, and `m5.csv`), *Rscript_graphs_threepanels.R* produces the figueres included in the main text (Figures 2 and 4) and the Supplementary materials (Figures S5, S6, and S7). Finally, *Rscript_heatmap_noquadratic_signs.R* produces the heatmaps that show relevant pairwise policy interactions (Figure 3) based on results from model 5 from the main analysis, which are formatted using *DoFile_csv_nodiagonal.do*.

Population (Figures 1 and S1)

Based on `panel_COVID_Lockdowns` and Mexican census data (`population_2020`), we produce Figure 1 that shows the total population and the share of the total population subject to each of the social distancing policies in sample. *population_date_graph.R* produces Figure 1 (to run code highlight lines of code and click `Command + Enter` or `Ctrl + Enter`). Likewise, and using data directly downloaded from the CoronaNet project (Database Version 1.0 (core)), we wrangle the data to have the same format as Figure 1 using *DoFile_Coronanet.do*. Figure S1, which shows the proportion of the world population subject to specific social distancing policies is produced by *policies_nature_graph.R*.

Histogram (Figures S2 and S3)

Figures S2 and S3 in Supplementary materials show the distribution of local and inflow weights. These figures are produced running *histograms.R* based on `weights_dow` that includes weights at the municipality-day level.

Maps (Figure S4)

Finally, we include maps showing the distribution of our sample—municipalities for which we have post-pandemic Facebook mobility data and for those that we are able to calculate weighted poli-

cies based on pre-pandemic Facebook mobility data. We produce such maps using *RScript_maps.R* based on our sample shown in `maps_sample.csv` and Mexican municipalities shapefiles (in folder "muni_2018gw").