Readme file for "Tax Breaks for Swing States? Political Bargaining, Targeted Policies, and Firm Outcomes" by Sahil Raina and Sheng-Jun Xu

Overview

This replication package contains two Stata do-files that (i) build the analysis datasets from raw inputs, (ii) produce the main tables and figures, and (iii) an R script file that produces additional figures.

- **1_prepdata.do**: constructs and cleans the analysis datasets used in the paper.
    - The **first section is commented out** and contains code used to create **masked/synthetic versions of Compustat and CRSP-derived inputs** for sharing/replication.
    - The active portion of the script reads the (synthetic) firm-level and state-level raw files, merges in political and economic covariates, and writes cleaned **intermediate** datasets used by the analysis script.
- **2_analysis.do**: runs the empirical analysis and produces the main tables and figures.
    - Tables are generated using estpost/esttab and are printed to the Stata output/log.
    - Figures are exported to output/figures/.
- **3_figures.R**: produces additional figures.
    - Figures are exported to output/figures/.

Recommended directory structure

The do-files assume a relative folder structure like:

- raw/ — raw inputs (Compustat/CRSP-derived files, state/political datasets, etc.)
- intermediate/ — cleaned datasets created by 1_prepdata.do
- output/figures/ — figures exported by 2_analysis.do and 3_figures.R
- Stata working directory set so relative paths resolve (e.g., run from the project root)

How to run

1. **Run data construction**
    - do 1_prepdata.do

- Outputs: cleaned intermediate .dta files in intermediate/

2. **Run analysis**
    - do 2_analysis.do
    - Outputs:
        - Tables printed in the Stata Results window and captured in the log created by the script (log using 2_analysis, replace)
        - Figures exported to output/figures/
3. **Produce additional figures**
    - Rscript 3_figures.R
    - Outputs:
        - Figures exported to output/figures/

Major data sources used throughout (by theme)

Below is a high-level inventory of the raw inputs referenced in 1_prepdata.do, grouped by category. File names reflect the package's internal conventions in raw/.

a) Firm fundamentals and identifiers

- Compustat fundamentals (firm-year): raw/firmpanel/compustat_funda*.dta (ultimate source: Compustat database)
- Compustat state identifiers / location mapping: raw/firmpanel/compustat_states.dta (ultimate source: Compustat database)
- Compustat tax variables: raw/firmpanel/compustat_taxvars*.dta (ultimate source: Compustat database)

b) Government spending, contracts, and subsidies

- U.S. government subsidies/loans (firm-year): raw/firmpanel/firm_subsidydata.dta (ultimate source: Subsidy Tracker databse)
- State government spending panel: raw/statepanel/govspend_statepanel.dta (ultimate source: usaspending.gov)

c) Politics and electoral variables

- State partisan control / political variables (state-year): raw/statepartisan.dta (ultimate source: MIT Election Lab)
- Senate "tightness" measures (year-level): raw/senatetightness.dta (ultimate source: MIT Election Lab)
- Pre-election closeness / preselection measures: raw/preselectcloseness.dta (ultimate source: MIT Election Lab)
- Committee chair indicators: raw/committeechairs.dta (ultimate source: Professor Charles Stewart III's website)
- Legislative voting, bill-level dataset: raw/legvoting/legvoting_billlvl.dta (ultimate source: Voteview Database)

d) Trade exposure and industry characteristics

- Industry-year import exposure (NAICS-based): raw/usimports_indyr.dta (ultimate source: Professor Peter Schott's trade database)
- Tradable-sector classification (Mian–Sufi style): raw/tradable_miansufi.dta (ultimate source: Mian and Sufi (2014))

e) Regulation exposure (industry-year)

- RegData / regulation intensity by NAICS-year: raw/regdata_naics.dta (ultimate source: QuantGov RegData database)

f) Returns and asset pricing controls

- CRSP-based returns linked to firm location: raw/returns/crsp_locationlinked*.dta (ultimate source: CRSP databse)
- Fama–French daily factors (FF3): raw/returns/ff3_daily.dta (ultimate source: Professor Ken French's website)

g) State-level macro and labor market controls

- QCEW state employment data: raw/statepanel/qcew_state.dta (ultimate source: BLS QCEW database)
- QWI state firm-size / employment composition: raw/statepanel/qwi_state_firmsize.dta (ultimate source: Census QWI database)
- State population: raw/statepanel/statepop.dta (ultimate source: U.S. Census)
- State GDP: raw/statepanel/gdp_state.dta (ultimate source: BEA)

- State code crosswalks: raw/state_codes.dta (ultimate source: U.S. Census)

h) National economic measures

- Annual unemployment rate (year): raw/unrate.dta (ultimate source: BLS)
- Annual GDP (year): raw/realgdp.dta (ultimate source: BEA)

Notes for replicators

- **Confidential data**: If you are using the shared replication package (rather than original licensed data), the scripts are configured to use the *_syn.dta synthetic inputs for Compustat/CRSP-derived files.