README

Replication Archive for "Federal Tax Deductions and the Demand for Local Public Goods"

Brent W. Ambrose & Maxence Valentin

December 22, 2023

This document describes the structure of the replication files that reproduce all the figures and tables in “Federal Tax Deductions and the Demand for Local Public Goods”. It details the different datasets, subcodes, and computational requirements. From *"0.Main_script.R"*, all results of the paper can be replicated.

1 Data files and dictionary

The final dataset that reproduces all exhibits in the paper is located in *"Data/data_for_analysis.R"*. All the variables of the main dataset ("df" when loaded in **R**) are described in Table 1 below.

To reproduce this dataset from the raw data, the source code *"Rscripts/Create_dataset_for_analysis.R"* loads and merges these different datasets:

- "CA_schools_final.csv": contains the merged and cleaned referendum results in California school districts (source: (California Secretary of State, 2022)).
- "soi_school_districts.csv": contains merged and cleaned panel data of SOI variables at the school district levels (source: (SOI, 2018)).
- "ACS_SchoolDistricts_2010-2021.csv": contains merged and cleaned American Community Survey 5-years data at the school district levels (source: ACS).
- *"mean_bond_yields_CA_schools.csv"*: contains the mean bond yields in California school districts. (source: Mergent)
- "City_referendums.csv": contains the merged and cleaned referendum results in California cities (source: (California Secretary of State, 2022)).
- "ACS_City_2010-2021.csv": contains merged and cleaned American Community Survey data at the city level – i.e. places (source: ACS).

In addition, to perform results in the Online Appendix, additional datasets (all located in "Data/Raw Data") are needed. Here is a description of the data and sources:

- *"soi_zipcode_panel_with_all_noagi.csv"*: contains merged and cleaned panel data of SOI variables at the zip code levels (source: (SOI, 2018)) – used to produce Figure A1.
- *"ppic.csv"*: contains merged and cleaned survey answers from the April Public Policy Institute of California Statewide Survey (source: PPIC) – used to produce Figure A3.
- *"SLGF_2017.csv"*: contains the data from the Survey of Local Government Finances for the year 2017 (source SLGF) – used to produce Figure A7.
- *"Instructional_modes.csv"*: contains the share of instructions performed online or in-person during the school year 2020-2021 (source: (COVID-19 School Data Hub, 2023)) – used to produce Table A4.
- *"land_use_school_districts_all.csv"*: contains land cover data aggregated at the school district level (source: National Land Cover Database) – used to produce Table B2.

2 Directory organization and scripts description

The folder *"Rscripts"* contains all the sub-R files to produce all the exhibits in the papers. By default, figures and tables will be saved in the folder *"Output"* with corresponding Table and Figure names. The different files are:

- *"1.Create_dataset_for_analysis.R"*: Loads and merges main data. It relies on:
  - *"2.Load_clean_referendums.R"*: Load school district referendums and clean datasets
  - *"3.merging_soi.R"*: Merges with SOI data
  - *"4.merging_acs.R"*: Merges with ACS data
  - *"5.Assemble_city_referendums.R"*: Load, merge, and clean city referendums
- *"6.sumstat.R"*: Creates Table 1 – Summary statistics
- *"7.main_result.R"*: Creates Table 2 – Average Causal Response Estimates
- *"8.school_vs_city_raw.R"*: Creates Figure 1 – Share of California Local Tax Referendums Approved
- *"9.main.parallel_trend.R"*: Creates Figure 2 – Dynamic Average Causal Responses and least wiggly path of confound. Relies on:
  - *"9bis.LWP.R"*: Search for the least wiggly path of confound
- *"10.main.elasticity.R"*: Shows the coefficient results discussed in the text of Section 4.
- *"11.figures_in_appendix.R"*: Reproduces all figures of the online appendix. Relies on:
  - *"11bis.construct_matrix_of_refdums.R"*: Construct a balanced panel of school districts with entries regarding whether they have had referendums.
- *"12.tables_in_appendix.R"*: Reproduces all tables of the online appendix. Relies on:
  - *"11bis.construct_matrix_of_refdums.R"*: Construct a panel dataset of school districts with entries regarding whether they have had referendums.

3 Version and packages

The current code runs on **R** version 4.3.1. To produce all exhibits, following packages are required:

- ggplot2
- stargazer
- DescTools
- lfe
- doParallel
- scales
- tigris
- ggthemes

References

California Secretary of State (2022). County, city, school district ballot measure election results. Available at https://www.sos.ca.gov/elections/county-city-school-district-ballot-measure-election-results.

COVID-19 School Data Hub (2023). Percentage of school year spent in-person, hybrid, or virtual, district overall shares. Accessed at https://www.covidschooldatahub.com/data-resources on June 21st, 2023.

SOI (2018). Statistics of income of the internal revenue service. Accessed at https://www.irs.gov/statistics/soi-tax-stats-statistics-of-income.

Table 1: Variables in main datasets

Variable name	Description	Source
LEAID	Unique school district identifier	CA SOS
year	Year of the referendum	CA SOS
date	Date of the referendum	CA SOS
type	Type of referendums (bond, levy, property tax)	CA SOS
threshold	Percentage of Yes votes for referendum to be approved	CA SOS
bond_amount	proposed bond amount	CA SOS
levy_parcel	Proposed levy parcel increase	CA SOS
vote_for	indicator variable for passed referendums	CA SOS
vote_against	indicator variable for passed referendums	CA SOS
vote_total	indicator variable for passed referendums	CA SOS
yield	Mean bond yield in CA districts in the year of the referendum	Mergent
cost_bond	Annual interest payment: bond_amount x yield	Mergent
perc_yes	Share of Yes votes on the referendums	CA SOS
Pass	indicator variable for passed referendums	CA SOS
election_FE	Election fixed effects (see footnote 6 in paper)	CA SOS
failed_recently	indicator if referendum in same school district failed in the last 4 years	CA SOS
presidential_el	indicator for referendums occurring during a presidential election	CA SOS
midterm_el	indicator for referendums occurring during a midterm election	CA SOS
odd_year	indicator for referendums occurring during in odd years	CA SOS
Bond	Indicator for bond referendums	CA SOS
post	indicator for referendums after 2019	CA SOS
tight_election	Indicator for election results within 40-90%	CA SOS
one_election	Indicator for school district with one referendum only	CA SOS
itm_prop_perc_17	Ratio of property tax deducters in 2017	SOI
itm_prop_change	Change in the ratio of property tax deducters between 2017 and 2018	SOI
salt_amt_change_perc	Percentage change in SALT deductions	SOI
wasted_salt	Dollar amount of SALT not deducted because of the cap	SOI
wasted_salt_perc	Wasted SALT over the total possible SALT deductions	SOI
salt_amt_change	Dollar change in SALT deductions	SOI
highpropchange	Chg.Ded greater than the median school district chg.Ded (0.139)	SOI
avg_costLPG_weighted_AGI	Weighted average cost of local public goods (see Equation 3)	SOI
housing_units	number of housing units	ACS
population_votingage	Population over 19 years old	ACS
income_mean	Average income in school district	ACS
realestate_taxes_median	Median property tax bills in the school district	ACS
ownership_2017	rate of home-ownership in 2017	ACS
housing_units18	housing units in 2018	ACS
bond_ph	Proposed bond amount per housing unit	
turnout	Total votes over the voting population	
salt_change_ph	salt_amt_change per housing unit	
wasted_salt_ph	wasted_salt per housing unit	
cost_bond_ph	cost_bond per housing unit	
levy_increase	Levy parcel proposed over current real estate taxes	