

This replication package contains files required to reproduce results, tables, and figures using Matlab and Stata. We divide the project into instructions to replicate the simulation, the result from Huang et al. (2006), and the application.

Simulation

For reproducing the simulation results

Included files in `*\Simulation` with short descriptions:

- `SSML_simfunc`: function that produces individual simulation runs
- `SSML_simulation`: script that loops over the `SSML_simfunc` for different DGP and multiple simulation runs
- `SSML_figures`: script that generates all figures for the paper
- `SSML_compilefunc`: function that compiles the results from `SSML_simulation` for the `SSML_figures` script

Steps for replicating simulation:

1. Save `SSML_simfunc`, `SSML_simulation`, `SSML_figures`, `SSML_compilefunc` to the same folder. This location will be referred to as the `FILEPATH`.
2. Create `OUTPUT` folder inside the `FILEPATH` location.
3. Change the `FILEPATH` location inside `SSML_simulation` and `SSML_figures`.
4. Run `SSML_simulation` to produce simulation data and results.
5. Run `SSML_figures` to produce figures.

Huang et al replication

For reproducing the Huang et. al. (2006) replication results.

Included files in `*\HuangetalReplication` with short descriptions:

- `SSML_huangrep`: script that replicates the results from Huang et. al. (2006)

Obtaining the dataset:

Go to <https://archive.ics.uci.edu/dataset/14/breast+cancer> and save file as "breast-cancer-wisconsin.data"

Steps for replicating results:

1. Save `SSML_huangrep` and the breast cancer data to the same folder. This location will be referred to as the `FILEPATH`.
2. Change the `FILEPATH` location inside `SSML_huangrep`
3. Run `SSML_huangrep` to produce results and figures.

Application

For reproducing the application section results.

Included program files in `*\Application` with short descriptions:

- `G0_main_202308.do`: Stata wrapper code that will run all application replication files
- `G1_cqclean_202308.do`: Cleans election outcomes data
- `G2_cqopen_202308.do`: Cleans open elections data
- `G3_demographics_cainc30_202308.do`: Cleans demographics data
- `G4_fips_202308.do`: Cleans FIPS code data
- `G5_klarnerclean_202308.do`: Cleans Klarner gubernatorial data
- `G6_merge_202308.do`: Merges cleaned datasets together
- `G7_summary_202308.do`: Generates summary statistics tables and figures
- `G8_firststage_202308.do`: Runs L1 penalized probit for the first stage
- `G9_prediction_202308.m`: Trains learners and makes predictions
- `G10_figures_202308.m`: Generates figures of prediction patterns
- `G11_final_202308.do`: Generates final figures and tables of results
- `r1_lasso_alwayskeepCF_202308.do`: Examines the effect of requiring the control function is not dropped from LASSO
- `latexTable.m`: Code by Eli Duenisch to write LaTeX tables from Matlab (<https://www.mathworks.com/matlabcentral/fileexchange/44274-latextable>)

Included non-confidential data in subdirectory `*\Application\Data\`:

- `\CAINC30`: County level income and demographics data from the BEA
- `\CPI`: CPI data from the BLS
- `\KlarnerGovernors`: Carl Klarner's Governors Dataset available at <https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/20408>

Confidential data suppressed in subdirectory `*\Application\CD\`:

These data cannot be transferred as part of the data use agreement with the CQ Press. Thus, the files are not included.

- `\CQ_county`: County level election outcomes available from <http://library.cqpress.com/elections/login.php?requested=%2Felections%2Fdownload-data.php>
- `\CQ_open`: Open elections available from http://library.cqpress.com/elections/advsearch/elections-with-open-seats-results.php?open_year1=1968&open_year2=2019&open_office=4

There is no batch download--downloads for each year must be done by hand. For each year, download as many state outcomes as possible and name the files `YYYYa.csv`, `YYYYb.csv`, etc. (Example: `1970a.csv`, `1970b.csv`, `1970c.csv`, `1970d.csv`). See line 18 of `G1_cqclean_202308.do` for file structure information.

Steps for replicating application:

1. Download confidential data from the CQ Press.
2. Change the working directory in `G0_main_202308.do` on line 18 to the application folder.
3. Change local `matlabpath` in `G0_main_202308.do` on line 18 to the appropriate location.
4. Set directory and file path in `G9_prediction_202308.m` and `G10_figures_202308.m` as necessary.
5. Run `G0_main_202308.do` in Stata to run all programs.
6. All output (figures and tables) will be saved to subdirectory `*\Application\Output`.

Contact

Contact Dylan Brewer (brewer@gatech.edu) or Alyssa Carlson (carlsonah@missouri.edu) for help with replication.