Replication file for Glynn and Kashin, "Front-door Difference-in-Differences Estimators"

Last updated: October 1, 2016

Overview

This is the replication file for Glynn and Kashin (2016). The replication files are in R (run using version 3.2.2) and organized using ProjectTemplate, a project management package for R.

The entire analysis may be run simply by executing the `make.R` script in the main level of the directory. This will carry out data munging and analysis, as well as compile figures and tables seen in the paper.

Required packages

The versions that were used for this analysis are in parentheses.

- ProjectTemplate (0.7)
- knitr (1.13)
- xtable (1.8-2)
- ggplot2 (2.1.0)
- dplyr (0.5.0)
- reshape2 (1.4.1)
- scales (0.4.0)
- R.utils (2.4.0)
- stringr (1.0.0)

Additional detail about project structure

Raw data

We include raw data files for the JTPA analysis in the `raw/jtpa` subdirectory. Some of these files were obtained from the Upjohn Institute and others that were obtained in communication with Jeffrey Smith and Petra Todd.

For early voting, we do not include the raw data, but include a write-up of how to obtain the exact data we used as well as a codebook in the `raw/vote` subdirectory.

Pre-processing of data

The `src` directory contains files used for the pre-processing of raw data. We provide these files for full transparency of how we assembled the master dataset we work off of for our analyses that are included with this replication.

The cleaned up JTPA data (see `data/jtpa_earn.RData` and `data/jtpa.RData`) is constructed using the following scripts:

- `src/jtpa/construct_jtpa_earnings.R`: script to construct clean earnings data from the JTPA raw data.
- `src/jtpa/construct_jtpa.R`: script that merges treatment information, compliance information, relevant background characteristics (sex, age, race, site, marriage status) and merges them with income data.

The cleaned up voting data (see `data/transition_array.RData`) is constructed using the following scripts:

- `src/vote/clean_registration.py` parses raw registration and vote history records from the state of Florida and outputs tables in csv format.
- `src/vote/create-county-dt.R` takes voter registration files and voter history files (in csv form from the previous script), merges them, and outputs data tables that contain sequential voting behavior for each individual (one row per individual across the elections).
- `src/vote/create_transition_array.R` takes the data tables created in the previous step and formats the data into a transition array which contains counts of individuals who move from one "state" (e.g. voting early) to another "state' (e.g. voting on election day) for various election "transitions" (e.g. 2006-2008 or 2008-2010).

Datasets

The cleaned up datasets, along with codebooks, are located in the `data` subdirectory:

- `jtpa_earn.RData`: cleaned up earnings data.
- `jtpa.RData`: cleaned up JTPA data.
- `transition_array.RData`: transition array for early voting in FL.

Analysis

The analysis of the data is in the `munge` subdirectory and is executed sequentially by ProjectTemplate, which caches the outputs of the analysis in the `cache` subdirectory.

The files here are:

- `01-jtpa_gain_scores.R` : gain score analysis for JTPA.
- `02-jtpa_benchmarks.R` : calculate experimental benchmarks for JTPA.
- `03-jtpa_frontdoor.R` : front-door and front-door diff-in-diff estimates for JTPA.
- `04-vote_fddid.R` : front-door diff-in-diff estimates for early voting in Florida.
- `05-vote_fddid_race_robust.R` : robustness check for front-door diff-in-diff estimates for early voting in Florida.

These files utilize various helper files in the `lib` subdirectory to handle estimation.

Reports

Reports are automatically compiled from Rmd to pdf files using `knitr` when running the `make.R` file. The Rmd files and the compiled pdf files, one per application, are available in the `reports` subdirectory.