Data Availability Guidelines and Code Base

for

"Diversity Effects or Dissent Aversion? Identification and Estimation in Judicial Panel Voting"

Charles M. Cameron
Center for the Study of Democratic Politics, Princeton University

Lealand Morin
College of Business, University of Central Florida

Harry J. Paarsch
College of Business, University of Central Florida

June 2, 2021

Reflection_Appeals

This is the code base to accompany the manuscript *Diversity Effects or Dissent Aversion? Identification and Estimation in Judicial Panel Voting* by Cameron, Morin, and Paarsch in the [Journal Name], 2021.

Any updates will be available on the GitHub code repository called `Reflection_Appeals` available at the following link:

    https://github.com/LeeMorinUCF/Reflection_Appeals

Data Availability

We extract data of three main types: information about cases in the U.S. Courts of Appeals, information about the corresponding trials, and information about the judges involved in both courts.

Some of the information is drawn from the Westlaw database, which is a proprietary data source and requires a subscription to Westlaw, a subsidiary of Thompson-Reuters. Some of the information is publicly available on the Website of the Department of Justice.

Westlaw Data

The primary data source is the Westlaw legal database.

Department of Justice

Some information is publicly available on the Website of the Department of Justice.
[To be completed.]

Instructions

The workflow proceeds in three stages: one set of instructions outlines the operations to draw the raw data from the Westlaw database. The data are then joined with other publicly-available datasets, to produce the final datasets that are the inputs for the statistical analysis in the final stage.

Data Manipulation

0.1   Text Mining

These procedures were performed on the `ms.economics` computing cluster at the Collee of Business at the University of Central Florida to generate the primary datasets. These scripts are stored in the `Code/Data_Prep` folder.

1. First run this script.

2. This script calls this function.

3. This other script calls this other function.

4. And everything goes from there.

Datasets

The above operations will produce the following datasets in `csv` format.

Main datasets

`appeals_cases_wl.csv`
    This dataset contains observations of cases in the U.S. Courts of Appeals that were heard from 2000-2019. The data were drawn from the Westlaw database. This dataset contains the following variables:

- `var_name`

- `var_name`

- `var_name`

- `var_name`

- `var_name`

`trial_cases.csv`
    This dataset contains observations of trials held U.S. Federal Courts for which the verdicts were appealed and the data were recorded in the `appeals_cases.csv` dataset. It contains the following variables:

- `var_name`

- `var_name`

- `var_name`

- `var_name`

- `var_name`

Auxiliary datasets

`appeal_cases_doj.csv`
    This dataset contains observations of cases in the U.S. Courts of Appeals that were heard from 2000-2019. The data were drawn from the Website of the Department of Justice. This dataset contains the following variables:

- `var_name`

- `var_name`

- `var_name`

- `var_name`

- `var_name`

These data sources are used in the statistical analysis that follows.

Statistical Analysis

These procedures were performed on a microcomputer to generate the tables and figures in the paper. These scripts are stored in the `Code/Stats` folder.

All Files in One Script:

1. Place all datasets in the `Data` folder, including the main datasets, including `appeals_cases_wl.csv`, `appeals_cases_doj.csv`, and `trial_cases.csv`.

2. Run `Reflection_Appeals.sh` in a terminal window from the `Reflection_Appeals` folder.

This shell script calls the main `python` programs, in the `Code/Westlaw` folder, which generate the `appeals_cases_wl.csv` dataset. Then it calls the R programs `TVN_Probit_Estn.R` and `TVN_Probit_Post_Estn.R`, which are found in the `Code/Stats` folder, which analyze the datasets stored in the `Data` folder. These scripts create the tables and figures for the entire manuscript, by writing `tex` files to the `Tables` folder and `eps` files to the `Figures` folder.

Generating Tables and Figures Separately

Tables

Table 1: Title

This table contains information from...

The numbers are combined into the file `Table_1.tex`.

Figures

Figure 1: Title

Computing Requirements

Data Manipulation

The `csv` files in the Data folder were generated on the `ms.economics` computing cluster at the Collee of Business at the University of Central Florida.

It is a cluster of [Nvidia Tesla K80 GPU Accelerators], each with [12 GB of GDDR5] on-board memory, running [2496] processor cores, with base core clock speed of [560 MHz] boost clocks from [562 MHz to 875 MHz], and with a memory clock speed of [2.5 GHz] on [48 pieces of 256M × 16 GDDR5 SDRAM], producing a memory bandwidth of [240GB/s per CPU].

For the queries that generated the datasets, [36] CPUs with [240 GB] of memory were sufficient to create the datasets within at most [24] hours each.

Statistical Analysis

Once the datasets have been saved in the `Data` folder, the remaining analysis, including the generation of all the tables and figures in the paper can be performed on a single microcomputer, such as a laptop computer. The particular model of computer on which the statistical analysis was run is a Dell Precision 3520, running a 64-bit Windows 10 operating system, with a 4-core x64-based processor, model Intel(R) Core(TM) i7-7820HQ CPU, running at 2.90GHz, with 16 GB of RAM.

Software

Data Manipulation

The data manipulation was conducted using Python 3.8.

The following python modules were imported:

- `os`, version 9.9.9

- `glob`, version 9.9.9

- `win32com`, version 9.9.9

- `pandas`, version 9.9.9

Statistical Analysis

The statistical analysis was conducted in R, version 4.0.2, which was released on June 22, 2020, on a 64-bit Windows platform x86_64-w64-mingw32/x64.

The attached packages include the following:

- `data.table`, version 1.13.0 (using 4 threads), to handle the main data table for analysis in the _prelim.R and _estim.R scripts.

- `xtable`, version 1.8-4, to generate LaTeX tables for Tables 1, 2, and 3.

- `plot3D`, version 1.3, to produce a 3-D bar chart of transition frequency, which created the plots in Figure 3.

- `MASS`, version 7.3-51.6, was also used to estimate the smoothed surface of the transition density as an alternative to that in Figure 3 but was not included in the paper.

Upon attachment of the above packages, the following packages were loaded via a namespace, but not attached, with the following versions:

- `Rcpp` version 1.0.5

- `lattice` version 0.20-41

- `grid` version 4.0.2

- `DTMCPack` version 0.1-2

- `stats4` version 4.0.2

- `magrittr` version 1.5

- `RcppParallel` version 5.0.2

- `misc3d` version 0.8-4

- `markovchain` version 0.8.5-3

- `Matrix` version 1.2-18

- `tools` version 4.0.2

- `igraph` version 1.2.6

- `parallel` version 4.0.2

- `compiler` version 4.0.2

- `pkgconfig` version 2.0.3

- `matlab` version 1.0.2

- `nnet` version 7.3-14

- `expm` version 0.999-5

- `zip` version 2.1.1

- `cellranger` version 1.1.0

- `pillar` version 1.6.0

- `forcats` version 0.5.1

- `lifecycle` version 1.0.0

- `tibble` version 3.1.0

- `gtable` version 0.3.0

- `rlang` version 0.4.10

- `curl` version 4.3

- `haven` version 2.3.1

- `rio` version 0.5.26

- `stringr` version 1.4.0

- `withr` version 2.4.2

- `hms` version 1.0.0

- `generics` version 0.1.0

- `vctrs` version 0.3.7

- `grid` version 4.0.5

- `tidyselect` version 1.1.0

- `glue` version 1.4.2

- `R6` version 2.5.0

- `rstatix` version 0.7.0

- `fansi` version 0.4.2

- `readxl` version 1.3.1

- `foreign` version 0.8-81

- `carData` version 3.0-4

- `purrr` version 0.3.4

- `tidyr` version 1.1.3

- `car` version 3.0-10

- `scales` version 1.1.1

- `backports` version 1.2.1

- `ellipsis` version 0.3.1

- `abind` version 1.4-5

- `colorspace` version 2.0-0

- `ggsignif` version 0.6.1

- `utf8` version 1.2.1

- `stringi` version 1.5.3

- `munsell` version 0.5.0

- `broom` version 0.7.6

- `crayon` version 1.4.1

Acknowledgements

References

Westlaw, Cases, Thompson-Reuters, accessed June 2021.

Westlaw, Litigation Analytics, Thompson-Reuters, accessed June 2021.

Information about U.S. Courts of Appeals, Table: 99-999-99-999, Department of Justice, accessed June 2021.