

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 4, 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 Thomson-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 College 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_w1.csv`, `appeals_cases_doj.csv`, and `trial_cases.csv`.
2. Run `Appeals_Voting.sh` in a terminal window from the `Appeals_Voting` folder.

This shell script calls the main `python` program `Get_Westlaw_Cases.py`, in the `Code/Westlaw` folder, which generate the `appeals_cases_w1.csv` dataset. Then it calls the `R` programs `Appeals_Voting_Estn.R`, `Appeals_Voting_Post_Estn.R`, and the function library `TVN_Probit_Lib.R`, which are found in the `Code/Stats` folder. These scripts 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...

Run the script `Appeals_Voting_Estn.R` up to line X. Lines Y to Z generate the file `Table_1_output.tex`.

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

Figures

Figure 1: Title

This figure shows...

Run the script `Appeals_Voting_Estn.R` up to line X. Lines Y to Z generate the file `Figure_1.tex`.

Computing Requirements

Data Manipulation

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

To be updated: It is a cluster of [48 Nvidia Tesla K99 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 \times 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, on either a 64-bit version of Windows 10 or a Linux platform running Red Hat Enterprise Linux 7.

The following python modules were imported:

- `os`, version 9.9.9, for interacting with the operating system when handling files.
- `glob`, version 9.9.9, to organize lists of directories and file names. It was used to facilitate iteration over the set of case files drawn from the Westlaw database.
- `win32com`, version 9.9.9, to perform operations with Microsoft Windows applications, which was used to translate the Westlaw case files from `doc` format into `txt` format.
- `pandas`, version 1.1.3, to store and manipulate the dataset of information surrounding cases in the U.S. Courts of Appeals.

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:

- `mvtnorm`, version 1.1-1, to calculate the probability mass under rectangular regions under the multivariate normal density. This was used to calculate the probabilities of voting combinations in the evaluation of the likelihood function.

- `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 voting frequency, which created the plots in Figure 3.

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

- `package` version 9.9.9
- `package` version 9.9.9
- `package` version 9.9.9
- `package` version 9.9.9

Acknowledgements

We thank the customer service team at Westlaw for their excellent assistance with the Westlaw database. We are grateful to Clint Cummins for his advice relating to the multivariate probit model. We also thank Han Hong just for being such a nice guy.

References

Westlaw [Name of Database with Cases], Thomson-Reuters, accessed June 2021.

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

[Name of Dataset from U.S. Courts of Appeals], Table: 99-999-99-999, Department of Justice, accessed June 2021.