

Summary of the work during Summer 2020

Jorge Valdebenito

October 10, 2020

This document presents the summary for the summer of 2020 work on the political uncertainty project.

1. Literature review: I updated the literature of the 2016 paper “Lobbying and Legislative Uncertainty” by Buzard and Saiegh (2016). This review contained the latest literature on ideal point models estimators and recent work on political uncertainty models using roll call data. An important filter for ideal point models was to focus on research that employs a multidimensional approach. Most of the papers that take a unidimensional approach were discarded, unless they specifically focus on uncertainty or efficiency of the unidimensional model. The Literature review was based on:

- Bafumi, Gelman, Park, Kaplan (2005)
- Clinton, Jackman, Rivers (2004)
- Crespín and Rohde (2010)

The document is called “Literature review” and it’s on the Github repository.

2. The code: There are 2 approaches mixed in the G drive and in the Github repository. The first one is from the early stages of the project and the developer is Yusuf, his work is very well detailed and documented in the “order of the files” word document. This word document explains the steps, data sets and order of how the program should be run. We discovered a problem with the code when we were running these models. This approach uses the command “destring” to merge the data sets, the problem was that it is ignoring the suffix on each law, therefore if 2 laws have the same number, but different suffix, the code will consider them as the same. For example, let’s take the laws H 2 and HR 2. These are different laws/actions, but the “destring” command removes the suffix and both laws are now called 2. Which is not correct. Another issue with this approach is that it calls datasets

and other pieces of code that were on Yusuf's computer, not on the G drive. It is important to consider also, that this work generates several output files that are called by different pieces of code later on, making them difficult to track and are located in different folders.

The second approach is an update of Yusuf's work and the developer is Yimin. His program is documented in the "Replication" file located in the Github repository. This approach consists of running 3 different pieces of code. We encountered a problem with the third of these files called `RStan_model.R`. This piece of code presents some errors when running the `rstan` command in R. This code avoids using the "destring" command and it should perform faster than Yusuf's code. It also makes the merging more efficient by doing a better job at cleaning the data set.

At the time of writing this document, I am still working on troubleshooting the errors in the `rstan` model.

KB 10/10/2020 : can you add some details about this now that your work is complete? You'll need to pull my updates to this document down to your repository to do this and then make a new pull request on the .tex and .pdf files for this summary

The other 2 programs `Generate_vote_data_for.R.do` and `Generate_group_dummy_by_action.do` run perfectly and their output is located in the Combined files folder in the G drive.

One detail that brings confusion is that Yusuf's and Yimin's work are located in the same folders. Yusuf's work is located inside folders `build` and `analysis`. It uses the subfolders `input`, `output` and `code` to call and export different files and datasets (following "order of the files" document must be used to work in these approach).

KB 10/10/2020 : I don't know what "following "order of the files" document must be used to work in these approach" means

Yimin's work is located in `build/code` (`Generate_vote_data_for-R.do` and `Generate_group_dummy_by_action.do`) and `analysis/code` `RStan_model.R`. His work is describe in "Replication", located in the political-uncertainty Github repository.

Therefore, future work will be to separate them and organize the files from these two different approaches, which will entail changing the output locations in all the individual R and Stata programs.

3. Data

Yusuf’s approach uses 3 different data sets, 2 from Maplight and one from Poole (Voteview.com). Bill Position Maplight, can be downloaded from their website using a key (need to request one for free) and Votes Maplight was emailed to professor Buzard.

KB 10/10/2020 Is “Bill Position Maplight” `build\input\maplight_bill_positions.dta`, or a raw version of this file that is not in the Github folder?

KB 10/10/2020 Similarly, is “Votes Maplight” `build\input\maplight_votes.dta`, or a raw version of this file that is not in the Github folder?

Poole’s data set can be downloaded from Voteview.com. Yimin’s programs only use the two previously-mentioned Maplight data sets.

I requested a key from Maplight to update the dataset, but, as of the writing of this summary, haven’t received a link to download the dataset. The API key is in `Road Map.txt` in the main folder of the Github repository. The `Road Map.txt` document contains more detailed notes about my work over the summer.

KB 10/10/2020 Is this the right way to characterize `Road Map.txt`?