Description of Replication Material

Voter Expectations of Government Formation in Coalition Systems: The Importance of the Information Context

European Journal of Political Research

Shaun Bowler Gail McElroy Stefan Müller

Introduction

This document describes the datasets and R scripts required to reproduce all plots and tables included in the paper and the Supporting Information.

The header of each R script lists the package versions that were used to run the code.

If you are using RStudio, I recommend to create a new R Project[1] in the folder that contains the scripts and data. Working with an RProject will not require that you manually change the working directory in the scripts to load the files. Alternatively, you need to specify the path to the folder that contains the data at the beginning of each script using `setwd()`.

[1] https://support.rstudio.com/hc/en-us/articles/200526207-Using-Projects

Summary and Descriptions of all R Scripts and Required Data

- `01a_calculate_majorities_land.R` contains the code to import the opinion polls with standing of parties in each German state and calculate the coalition probabilities using the *coalitions* package by Bender et al. The script is used to create and save the file **data_coalition_predictions_land.rds** which contains the coalition probabilities for each potential coalition for each election. To run this code script, you require the following files:
    - **data_polls_land_raw.csv**: csv file that contains all raw polls scraped from https://wahlrecht.de (in June 2018)
    - **data_survey_dates_land.csv**: csv file that contains the start and end dates of the survey period of each pre-election study. These dates are required to calculate the time frame of polls to be included in the estimation of coalition probabilities
- `01b_calculate_majorities_federal.R` contains the code to import the opinion polls with standing of parties on the federal level and calculate the coalition probabilities using the *coalitions* package by Bender et al. The script is used to create and save the file **data_coalition_predictions_federal.rds** which contains the coalition probabilities for each potential coalition for each election. To run this code script, you require the following files:
    - **data_polls_federal_raw.csv**: csv file that contains all raw polls scraped from https://wahlrecht.de (in June 2018)
    - **data_survey_dates_federal.csv**: csv file that contains the start and end dates of the survey period of each pre-election study. These dates are required to calculate the time frame of polls to be included in the estimation of coalition probabilities.
    - **data_election_results_federal.csv**: csv file with the result of the federal elections
- `01c_get_coalitions_since_1990.R` extracts the number of coalitions formed on the Länder level between 1990 and 2017. This variable is used as a covariate in the regression models. The file returns a plot for the appendix and the file **data_coalitions_laender.csv** which is merged to the respondent-level data in `04_merge_metadata_total.R`.
- `02a_import_recode_land_elections.R` imports the Stata (.dta) files for each state election survey, selects the relevant variables, recodes and harmonises the variables across the elections, and estimates the ideological range of coalition options and the distance between the average left-right position of each government and a respondent. To run this code script, you require the following files:
    - The election surveys from the GESIS website. This requires (1) creating a free account at GESIS, downloading each survey (the name of the files is indicated after foreign::read.dta()), and including the files into a new folder called "data_gesis" in your working directory. Afterwards, you will be able to reproduce the data wrangling and recoding. For copyright reasons, the files cannot be included in this Dataverse repository, but the files (with the exact same file names) can be retrieved from the GESIS website. SI Section J lists all election surveys used in this paper.
- `02b_import_recode_federal_elections.R` imports the Stata (.dta) files for each federal election survey, selects the relevant variables, recodes and harmonises the variables across the elections, and estimates the ideological range of coalition options and the

distance between the average left-right position of each government and a respondent. To run this code script, you require the following files:

  - The election surveys from the GESIS website. This requires (1) creating a free account at GESIS, downloading each survey (the name of the files is indicated after foreign::read.dta()), and including the files into a new folder called "data_gesis" in your working directory. Afterwards, you will be able to reproduce the data wrangling and recoding. For copyright reasons, the files cannot be included in this Dataverse repository, but the files (with the exact same file names) can be retrieved from the GESIS website. SI Section J lists all election surveys used in this paper.

- `03_text_analysis_newspaper_articles.R` loads the newspaper articles retrieved from NexisLexis which are stored in a text corpus, applies the Newsmap (Watanabe 2018) classifier, creates the dictionaries, applies the dictionaries to a sentence-level corpus, returns the most frequent terms and multiword expressions, and aggregates the mentions of each government option within a window of 2 years prior to the start of each election survey. The R script creates the file **data_news_coalitions.rds** which contains one observation per election-government-observations and lists the absolute number and relative proportion of mentions of the respective government option. To run this code, you require the following files:
  - **corpus_ger_elections.Rdata**: text corpus containing 15,980 news articles retrieved from NexisLexis that potentially deal with coalitions on the federal and state level. See the Supporting Information of the paper for more details on how the articles were retrieved. For copyright reasons established by NexisLexis, this file cannot be shared in this Dataverse repository. Please get in touch with the authors if you have any questions.
  - **metadata_elections_land.rds**: file with metadata on all state elections in the sample
  - **metadata_elections_federal.rds**: file with metadata on all federal elections in the sample
- `04_merge_metadata_total.R` merges the files constructed and stored in `01a_calculate_majorities_land.R`, `01a_calculate_majorities_federal.R`, `02a_import_recode_land_elections.R`, `02a_import_recode_federal_elections.R`, and `03_text_analysis_newspaper_articles.R`. The file also loads the manifesto-level Wordscores estimates from Bräuininger et al. (2020) which are stored in **data_parties_all_wordscores.csv**. The script is also used for constructing new variables and selecting the variables required for the analysis. The resulting file is called **data_reg_full.rds** and is used in all subsequent analyses. Note that respondents who did not make any prediction of coalitions are still included in this dataset in order to check the proportions of respondents who made a prediction (e.g., Table 1 in the paper).
- `05_descriptive_plots.R` creates the descriptive plots for the paper and the Supporting Information. To run this code script, you require the following files:
  - **data_reg_full.rds**: file with merged survey results and information on each coalition
  - **data_coalition_predictions_land.rds** and **data_coalition_predictions_federal.rds**: files that contain the coalition probabilities for each government option, based on polling results

    - **data_tstat_freq_translated.csv**: most frequent terms and multi-word expressions (German and English translation) in sentences classified as coalition signals `03_text_analysis_newspaper_articles.R`
- `06_validate_text_analysis.R` compares the automated detection of coalition signals with the manual classification of coalition signals by Best (2015). To run this code script, you require the following file:
    - **data_reg_full.rds**: file with merged survey results and information on each coalition
    - **data_coalition_options_coded.csv**: coding of pre-electoral coalition signals by Best (2015)
- `07a_reg_logit.R` runs the logistic regression models that predict whether a respondent predicted the government option that was formed after the election. To run this code script, you require the following file:
    - **data_reg_full.rds**: file with merged survey results and information on each coalition
- `07b_reg_clogit.R` runs the conditional logit models and produces plots and tables with the (standardised) coefficients. To run this code, you require the following files:
    - **data_reg_full.rds**: file with merged survey results and information on each coalition
- `07c_random_forest.R` runs the random forest analysis that tests for the relative importance of variables for the correct prediction of the government formed after the election. To run this code script, you require the following file:
    - **data_reg_full.rds**: file with merged survey results and information on each coalition
- `07d_reg_dontknow.R` predicts which respondents are more likely not to predict any coalition. The results from this analysis are reported in the Supporting Information. To run this code script, you require the following file:
    - **data_reg_full.rds**: file with merged survey results and information on each coalition
- `function_theme_base.R` contains the ggplot2 theme used for all plots for the paper and the Supporting Information

```
> sessionInfo()
R version 4.0.2 (2020-06-22)
Platform: x86_64-apple-darwin17.0 (64-bit)
Running under: macOS  10.16

Matrix products: default
LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib

locale:
[1] en_IE.UTF-8/en_IE.UTF-8/en_IE.UTF-8/C/en_IE.UTF-8/en_IE.UTF-8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

loaded via a namespace (and not attached):
[1] compiler_4.0.2 tools_4.0.2
```