Codebook

Irish Polling Indicator Datasets

Tom Louwerse and Stefan Müller

Version: 18 June 2023

Contents

1 Introduction

The Irish Polling Indicator (https://pollingindicator.com) provides daily aggregated estimates of support for Irish political parties. The project started in 2014 and is maintained by Tom Louwerse and Stefan Müller. Louwerse (2016) describes the underlying method and background in detail. In 2022, we set up repositories of all raw polling results and aggregated estimates. In this codebook, we describe the variables and structure of each dataset and provide an overview of the available data.

We provide two datasets:

- `data_polls` contains the raw polling results, information on the field period (the days when the survey was conducted), the publication date, and the polling company. As of 18 June 2023, the dataset reports results from 687 polls released between 1982 and 2023.
- `data_pollingindicator` provides the daily aggregated estimates for parties, along with 95 per cent credible intervals. As of 18 June 2023, the dataset contains estimates for 13265 days from 1987 to 2023.

2 Dataset Versions

2.1 Development Version

Both of the datasets have been published in two data repositories. Everyone is welcome and free to use the estimates and polling results as long as the corresponding datasets are cited.

The GitHub repository IrishPollingIndicator/ipi-data contains the "development version" of the datasets. These datasets are updated after the release of new polls and thus change over time. Moreover, the estimates for previous dates in the current term may change after the releases of new polls (for details see see Louwerse 2016).

If you use data from the development version, please cite:

- Tom Louwerse and Stefan Müller. 2023. *Irish Polling Indicator Datatsets: Development Version.* URL: https://github.com/Irish-Polling-Indicator/ipi-data.

2.2 Stable Version

Users who would like to access the most recent data should download and analyse the development version. Users who do not need access to the most recent files but only need data until the end of a calendar year should use and cite the stable version. The "stable version" is stored at Harvard's Dataverse. New releases are published after an election cycle.

The stable version has a unique identifier (DOI: 10.7910/DVN/BY5GXC), and the daily estimates will not change since all polls in an election cycle are considered.

If you use data from the stable version, please cite:

- Tom Louwerse and Stefan Müller. 2022. *Irish Polling Indicator Datasets: Stable Version*. Harvard Dataverse, V1. DOI: 10.7910/DVN/F0Y7RQ

3 File Formats

We provide `data_pollingindicator` and `data_polls` in four file formats.

- `csv`: The comma-separated values file ensures inter-operability as it can be opened in R, Python, Stata, SPSS, and Excel.
- `xlsx`: The Excel spreadsheets allow for an even more straightforward import of the data into Microsoft Excel.
- `dta`: This file can be used to import the datasets with correct variable encodings into Stata.
- `rds`: The RDS file is optimised for the R statistical programming language and stores the variables in the correct data type.

4 Party Codings

We use consistent codings of party names in all datasets. The following political parties are included for at least one legislative period.

- `FF`: Fianna Fáil
- `FG`: Fine Gael
- `SF`: Sinn Féin
- `LAB`: Labour Party
- `GP`: Green Party
- `SD`: Social Democrats
- `SPBP`: Solidarity-People Before Profit
- `PD`: Progressive Democrats
- `DL`: Democratic Left
- `AU`: Aontú
- `RENUA`: Renua (only included in `data_polls`)
- `OTH_IND`: Other parties and independent candidates (calculated as 100 - sum of support for all remaining parties)

5 Variables and Data Structure

Using the R statistical programming language, we explain the structure of both datasets.

5.1 Polling Results (`data_polls`)

The dataset contains the support for the parties listed above. In addition, the dataset includes the following variables:

- `date`: date when poll was published (usually in an Irish newspaper)
- `date_start`: start of survey period
- `date_end`: end of survey period
- `date_middle`: middle date of survey period
- `pollster`: name of polling company
- `sample_size`: number of respondents in poll

The code below shows how to import the dataset, describes the structure, and visualises the number of available polls per year.

```
# load required packages
library(dplyr)
library(ggplot2)
library(scales)

# load dataset of raw polls
data_polls <- read.csv("data_polls.csv")

# inspect variable names
names(data_polls)
```

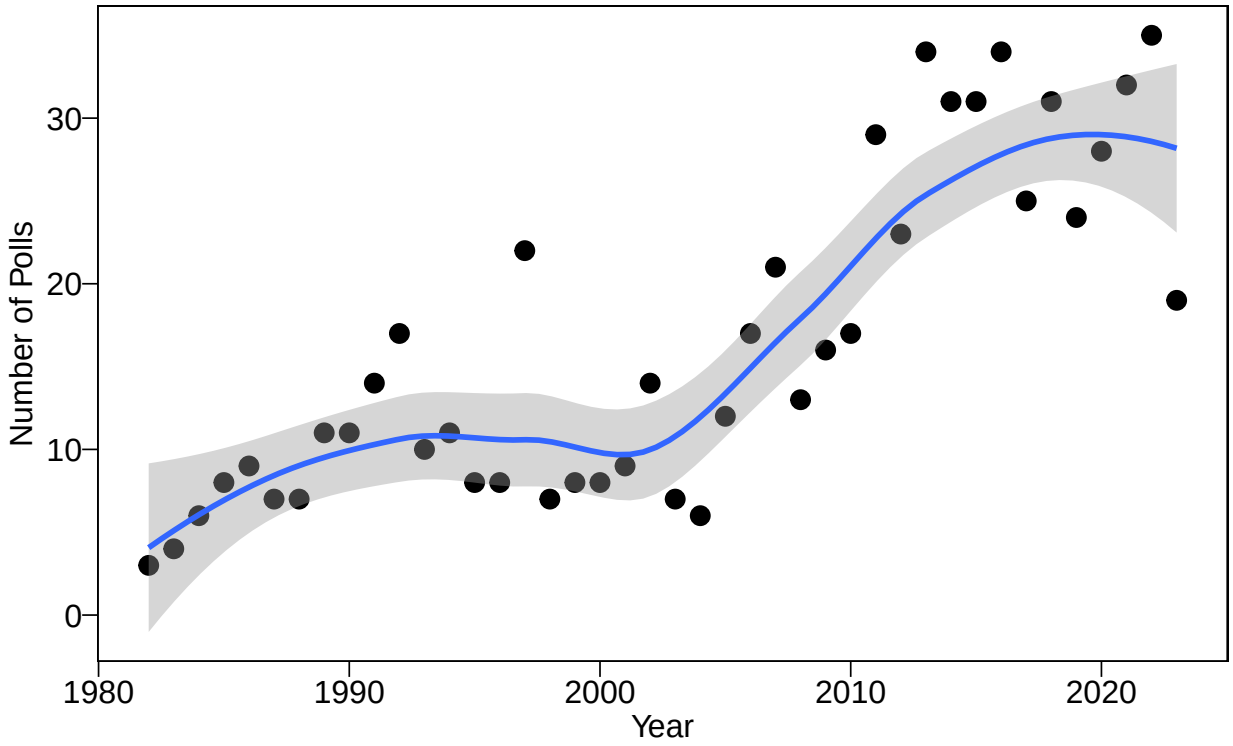
```
##  [1] "date"        "date_start"  "date_end"    "date_middle" "pollster"
##  [6] "sample_size" "FF"          "FG"          "SF"          "LAB"
## [11] "GP"          "PD"          "WP"          "DL"          "SPBP"
## [16] "RENUA"       "SD"          "AU"          "OTH_IND"
```

```
# create "year" variable based on date
data_polls <- data_polls |>
    mutate(year = substr(date, 1, 4))

# group data by year and count observations (=polls) per year
data_polls_years <- data_polls |>
    group_by(year) |>
```

```
    count()

# plot the number of polls per year
ggplot(data_polls_years, aes(x = as.integer(year), y = n)) +
    geom_point(size = 3) +
    geom_smooth() +
    labs(x = "Year", y = "Number of Polls")
```



5.2 Irish Polling Indicator Estimates (`data_pollingindicator`)

The dataset `data_pollingindicator` contains the daily aggregated estimates for parties listed above.[1] We limit the estimates to parties whose support was reported consistently throughout an electoral cycle (Louwerse 2016). If consistent polling results are unavailable, we add this party to `OTH_IND` (Other Parties/Independent). The dataset also contains the following variables.

- `date`: support on a given day
- `cycle`: electoral cycle
- `..._lo`: lower 95 per cent credible interval of estimated support for party
- `..._hi`: upper 95 per cent credible interval of estimated support for party

[1] For a summary of the methodological approach, see https://pollingindicator.com/method

Note that we do not include election results but only polls. This means that measures of support in each cycle only range up until the mid-date of the most recent poll before the election date.

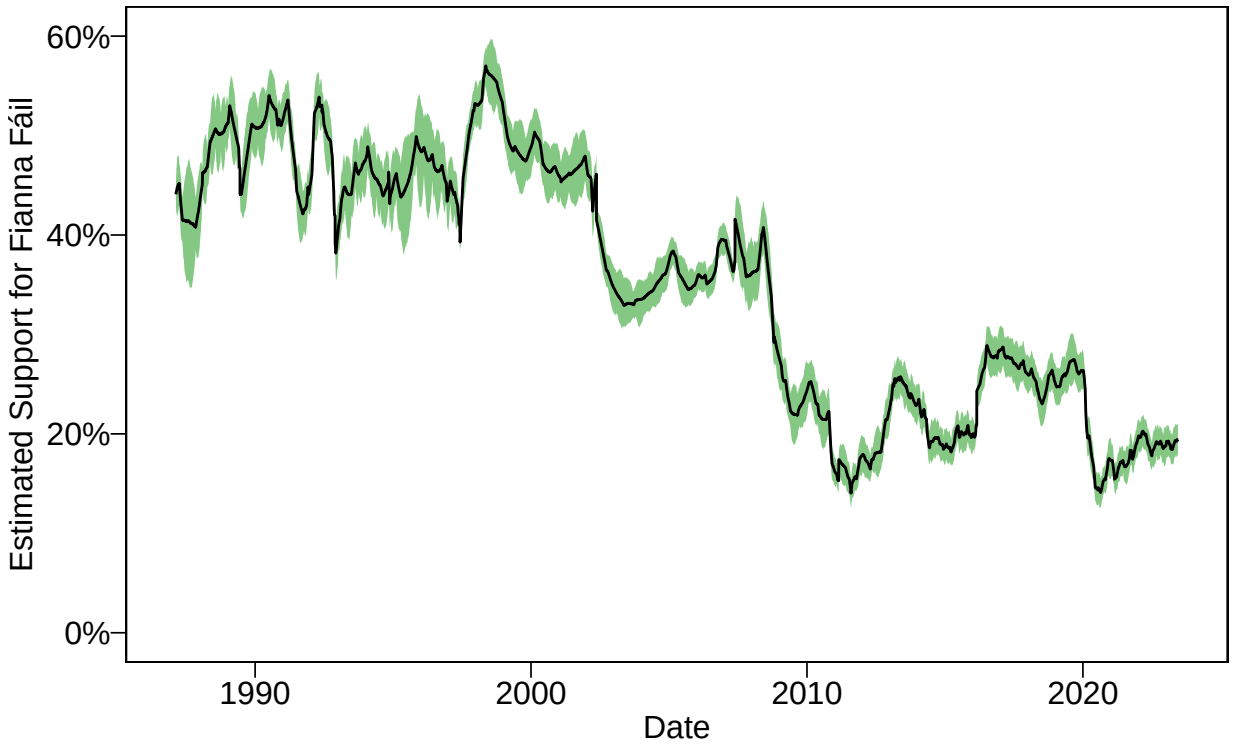
The code below shows how to import the dataset, describes the structure and visualises the daily estimates and uncertainty intervals for one party.

```
# load dataset of raw polls
data_polllingindicator <- read.csv("data_pollingindicator.csv")

# inspect variable names
names(data_polllingindicator)
```

```
##  [1] "date"    "cycle"   "FF"      "FF_lo"   "FF_hi"   "FG"      "FG_lo"
##  [8] "FG_hi"   "LAB"     "LAB_lo"  "LAB_hi"  "PD"      "PD_lo"   "PD_hi"
## [15] "WP"      "WP_lo"   "WP_hi"   "OTH"     "OTH_lo"  "OTH_hi"  "GP"
## [22] "GP_lo"   "GP_hi"   "DL"      "DL_lo"   "DL_hi"   "SF"      "SF_lo"
## [29] "SF_hi"   "SD"      "SD_lo"   "SD_hi"   "SPBP"    "SPBP_lo" "SPBP_hi"
## [36] "AU"      "AU_lo"   "AU_hi"
```

```
# plot estimates for Fianna Fáil (FF)
ggplot(data_polllingindicator,
       aes(x = as.Date(date),
           y = FF,
           ymin = FF_lo, ymax = FF_hi)) +
    geom_ribbon(fill = "#66BB66", alpha = 0.8) +
    geom_line()  +
    scale_y_continuous(limits = c(0, 0.6),
                       labels = scales::percent_format(accuracy = 1)) +
    labs(x = "Date", y = "Estimated Support for Fianna Fáil")
```



6 About the Maintainers

Tom Louwerse is an Associate Professor in Political Science at Leiden University, the Netherlands. Tom's research and teaching focuses on elections, political representation and parliamentary politics in the Netherlands and other established democracies.

- Personal website

Stefan Müller is an Assistant Professor and Ad Astra Fellow in the School of Politics and International Relations at University College Dublin. Stefan's research focuses on political representation, party competition, political communication, public opinion, quantitative text analysis, and the application of computer vision techniques.

- Personal website

7 Questions

If you have any questions or spot errors, please do not hesitate to contact the maintainers.

- Tom Louwerse: t.p.louwerse@fsw.leidenuniv.nl
- Stefan Müller: stefan.mueller@ucd.ie

8 Supporters and Partners

The Irish Polling Indicator is hosted at the Institute of Political Science at Leiden University and the Connected_Politics Lab at University College Dublin. The project received financial support from the 2021 Strategic Funding Scheme of the UCD College of Social Sciences and Law.

References

Louwerse, Tom. 2016. "Improving Opinion Poll Reporting: The Irish Polling Indicator." *Irish Political Studies* 31 (4): 541–66. https://doi.org/10.1080/07907184.2016.1213719.