

Author: Levente Kocsis

The Extent of Minilateralism Datasets

# **User's Manual**

## 1. Introduction

This document is intended to serve as a user's guide to the datasets, including the source and description of raw data and derived quantities. Steps necessary to retrieve the original datasets from their respective sources are discussed. In addition, the methodology underlying derived quantities as well as their most likely interpretation is explained.

## 2. Original datasets

In the following paragraphs the source and description of the original datasets are presented. While the use of a single dataset is preferred, the fact that no official database is complete – with regards to timeframe – necessitates their combined use.

### a. The Council's Open Data

The original – raw – dataset on EU Member State representatives' votes in the Council of the EU are retrieved from the Council's [Open Data portal](#). The dataset was released in April 2015 within the bounds of the Council's Open Data Initiative (CODI) with the aim of releasing the results of voting sessions on legislative acts in the Council to the public. Since then, several improvements have been made and the scope of the project has been expanded; however, they are of no interest to the task at hand.<sup>1</sup> In addition to the votes of the Council, the dataset contains various other variables such as:<sup>2</sup>

- Form of adoption
- Council meeting session
- Council configuration
- Date
- Policy area
- Act type
- Act number
- Document number
- Interinstitutional code
- Action by the council (according to the steps of the procedure)

---

<sup>1</sup> For more information on available CODI datasets, please consult the corresponding [User's Guide](#).

<sup>2</sup> A more in-depth description of the above variables might be found in a [supplementary document](#) concerned with CODI datasets.

- Legislative procedure (ordinary, or special legislative procedure)
- Voting rule (qualified majority or unanimity)

The dataset was obtained using the SPARQL [endpoint](#) found on the Open Data portal. The dataset contains data from December 2010, and is updated daily. This means that several months of data are missing, hence the necessity to complete the database with the use of other – yet official – sources.

### **b. Monthly summaries of Council acts**

Votes of the council prior to December 2010 were gathered from [monthly summaries](#) of the Council's acts. In addition to votes, these documents contain information such as:

- Act title
- Council meeting session
- Council configuration
- Date
- Document number
- Statements by the Commission and/or Member States

Evidently, the scope of variables listed in these documents is more limited compared to those in the CODI dataset. As the purpose of the research does not require the whole list of variables available in the CODI database this does not pose much of a problem. Nevertheless, the variables deemed necessary were obtained from [EUR-Lex](#) and the EP's [Legislative Observatory](#) or by our own assessment.<sup>3</sup>

### **c. The combined dataset**

The combined dataset has been constructed so it contains data obtained from the monthly summaries from 08.06.2009 to 11.31.2009, and from the Open Data set from 12.01.2009 to 27.05.2019. Furthermore, the dataset is split into two periods that match the 7<sup>th</sup> and 8<sup>th</sup> terms of the European Parliament.

---

<sup>3</sup> In case of policy areas.

The number of distinguished policy areas in the original dataset was unfeasibly high, so they were consolidated into 12 policy areas. For the list of these, see the “*Consolidated policy areas*” sheet of the related spreadsheet.

Each of the combined raw databases for the groups contains the following variables:

- Date
- Council meeting session
- Council configuration
- Policy area
- Interinstitutional code (Procedure ID)
- Votes of Member State representatives

The file included in this package containing these data is as follows:

- “*IVF\_Council\_votes\_DATA\_FINAL.xlsx*”

### **3. Derived datasets**

#### **a. Co-voting database**

The co-voting database contains two primary indicators that are frequently investigated during the study of co-voting behaviour:

- Vote similarity (in per cents)
- Frequency of synchronous opposing votes (number)
- The above two by policy area

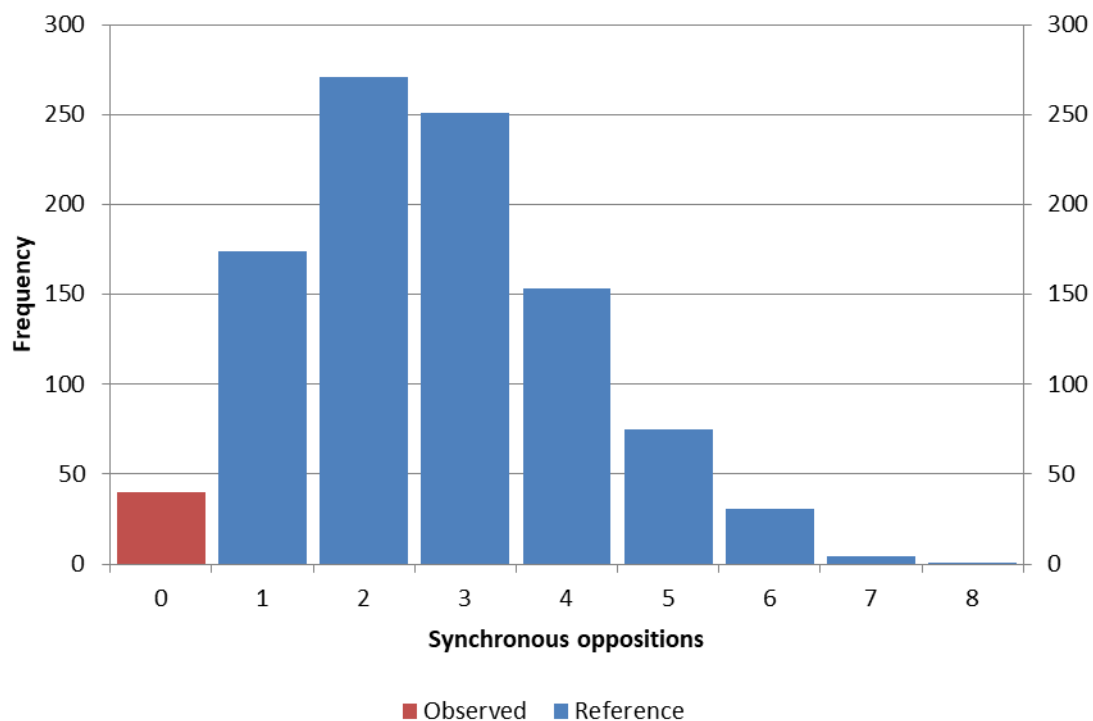
These quantities are self-explanatory, though some specifics should be pointed out. The vote similarity measure distinguishes four different types of votes (in favour, against, abstained, did not vote), while the synchronous oppositions metric differentiates between only two – support (in favour) and opposition (against, or abstentions during qualified majority votes). Non-votes are not accounted for in the second.

The relevance of the above measures might be challenged saying either similarity or synchronous opposition could be the result of likely contingencies rather than coordination. To dismantle this argument the resulting values are compared to the null hypothesis of them being coincidences.

This test is done by the following process:

1. The factual values of the relevant quantity are calculated.
2. The dataset is reshuffled, so that its characteristics are preserved.
3. The quantities are calculated on this randomised dataset.
4. Steps 2 and 3 are repeated 1000 times.

The above procedure yields a reference distribution corresponding to the null hypothesis of random contingencies (see Figure 1).

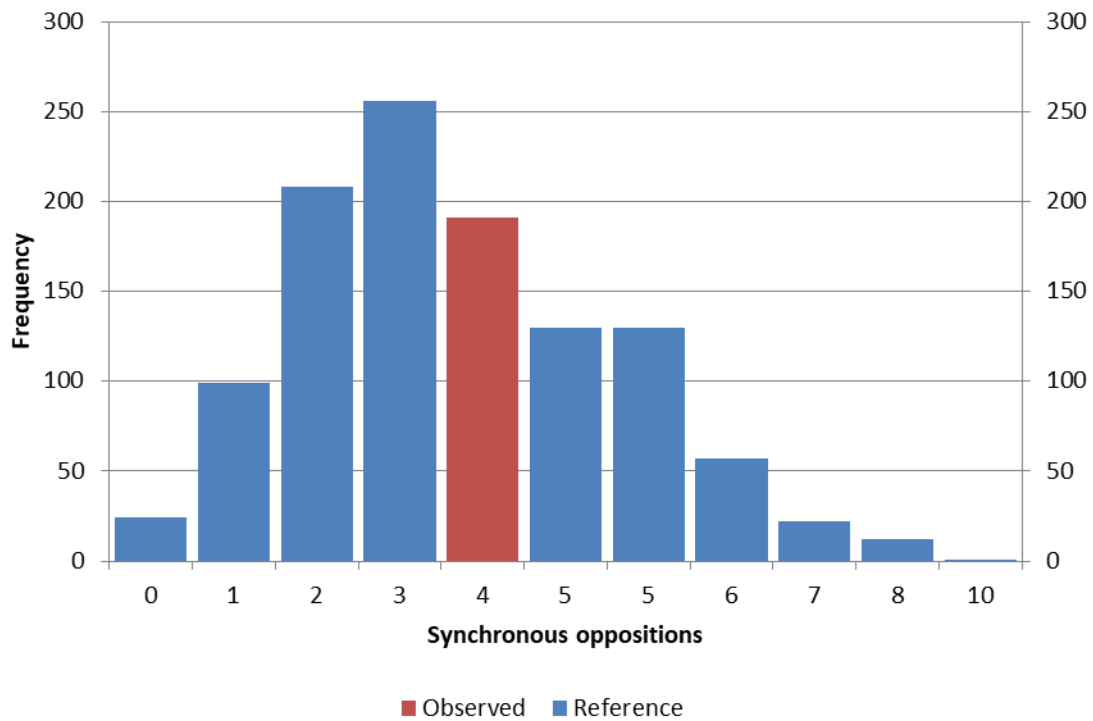


**Figure 1: The distribution of synchronous oppositions of Germany and the UK under the null hypothesis of randomness (Reference) and its observed value (Observed) during the first period of the sample.**

The idea here is that if (for example) 90% of the values calculated on the randomised dataset are higher/lower than the observed, we can conclude that the observed value is higher/lower than expected under the null hypothesis of random contingencies with 90% certainty.

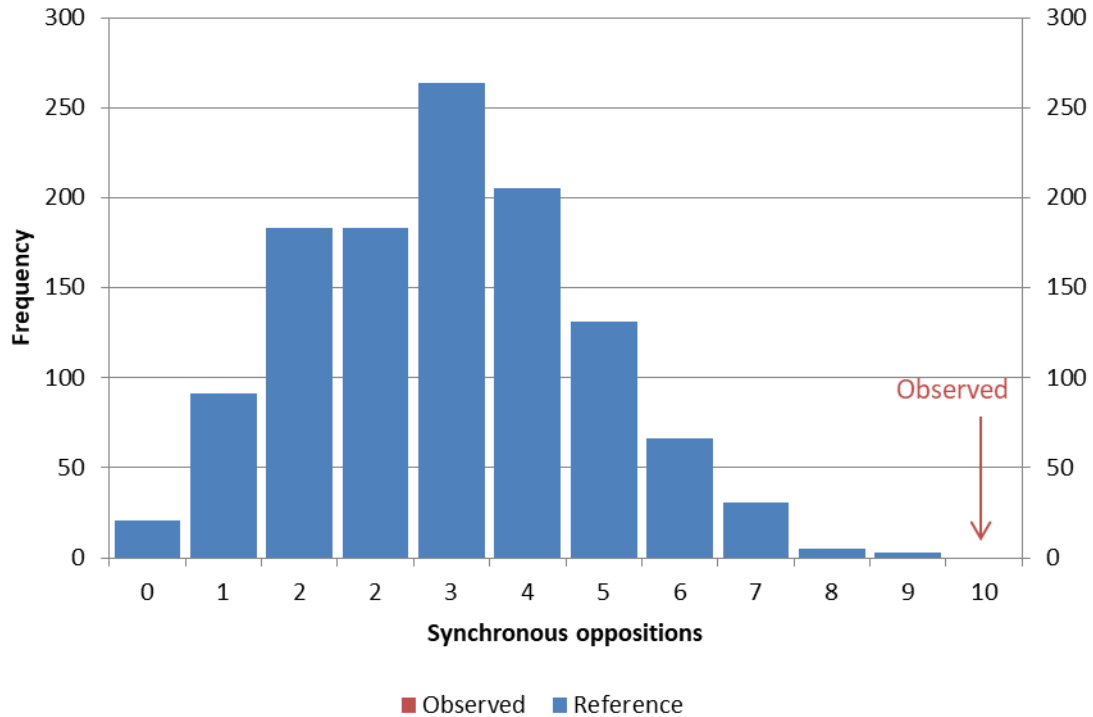
In the case of Germany and the UK, we can say that in 96% of the random cases their synchronous oppositions were higher than the observed, so the synchronous oppositions of Germany and the UK are lower than expected, and this observation is statistically significant on the 5% level. In other words, we can reject the null hypothesis of coincidence with 96% certainty.

In a similar way, Figure 2 below suggest that the 4 synchronous oppositions of Austria and the UK are not significantly smaller or greater than they would be under the null of randomness, thus we cannot reject it with any significant certainty.



**Figure 2: The distribution of synchronous oppositions of Austria and the UK under the null hypothesis of randomness (Reference) and its observed value (Observed) during the first period of the sample.**

Likewise, the below (Figure 3) example of the Netherlands and the UK illustrates a case where we can conclude that the number of synchronous oppositions is significantly higher than expected under the null hypothesis of random contingencies. In the 1000 random cases, there was not a single case when the number of synchronous oppositions was or exceeded the observed 10.



**Figure 3: The distribution of synchronous oppositions of the Netherlands and the UK under the null hypothesis of randomness (Reference) and its observed value (Observed) during the first period of the sample.**

The calculations are repeated for every policy area, though they should be treated with caution. As the number of votes in certain policy areas is small, the frequency of non-favourable votes even less there is little variation in the outcome of the randomisations – most outcomes correspond to or are around the observed. This results in that very few co-voting relations are found to be statistically significant.

The files included in this package containing these datasets are as follows:

- “{GROUP}\_covoting\_2009-2014\_FINAL.xlsx”
- “{GROUP}\_covoting\_2014-2019\_FINAL.xlsx”

Each file comes with 13 sheets. One for the overall results and the rest for the 12 policy areas.

Each sheet is structured the same way. The columns included (see Table 1) are the following:

- *Partner Member State*: the Member State to which the row corresponds.
- *Vote similarity (%)*: vote similarity in per cents.

- *Synchronous oppositions*: the number of time the two countries opposed an act at the same time.
- *Relation to the expected*: relation to the expected under the null hypothesis of randomness. Its values can be “Smaller”, “Greater”, “No variation”. The last means that there was no variation in one of the countries’ votes (they were all favourable), so this hypothesis testing technique does not work.
- *P-value*: the estimated probability that we can reject the null hypothesis of random contingencies.

Partner Member State	Vote similarity (%)	United Kingdom		Synchronous opposition	Relation to the expected	P-value
		Relation to the expected	P-value			
Austria	74.48%	Greater	0.194	4	Greater	0.420
Belgium	74.25%	Smaller	0.525	4	Greater	0.262
Bulgaria	75.64%	Smaller	0.319	0	Smaller	0.445
Croatia	75.64%	Greater	0.193	3	Greater	0.262
Cyprus	75.41%	Smaller	0.481	0	Smaller	0.305
Czechia	74.48%	Smaller	0.517	3	Greater	0.452
Denmark	80.05%	Greater***	0.000	2	Greater	0.365
Estonia	76.33%	Greater	0.137	0	Smaller	0.550
Finland	75.87%	Greater	0.492	1	Greater	0.608
France	76.33%	Smaller	0.765	0	Smaller	0.823
Germany	73.55%	Smaller	0.122	0	Smaller**	0.046
Greece	75.64%	Greater	0.545	1	Smaller	0.707
Hungary	73.09%	Greater	0.388	2	Smaller	0.185
Ireland	81.67%	Greater***	0.000	2	Greater	0.168
Italy	76.80%	Greater**	0.010	4	Greater**	0.013
Latvia	76.33%	Greater*	0.084	1	Greater	0.637
Lithuania	76.33%	Greater*	0.061	1	Greater	0.658
Luxembourg	75.41%	Smaller	0.591	1	Smaller	0.611
Malta	75.87%	Greater	0.131	1	Smaller	0.627
Netherlands	75.64%	Greater***	0.004	10	Greater***	0.000
Poland	73.55%	Greater	0.189	3	Smaller	0.416
Portugal	75.87%	Smaller	0.512	1	Greater	0.559
Romania	75.87%	Greater	0.262	0	Smaller	0.355
Slovakia	75.17%	Greater	0.368	1	Smaller	0.485
Slovenia	75.64%	Smaller	0.640	1	Smaller	0.710
Spain	75.64%	Smaller	0.296	0	Smaller	0.424
Sweden	77.26%	Greater***	0.000	5	Greater**	0.011

Significance codes: \* - significant on 10%, \*\* - significant on 5%, \*\*\* - significant on 1%.

Table 1: Results for the United Kingdom calculated the first half of the sample.



## **b. Blocking attempts**

So far the derived datasets dealt with only pairwise coveting frequencies. However, blocking coalition attempts are also interesting. The remainder of the derived datasets is about those.

Files included in this package containing these datasets:

- “*{GROUP}\_blocking\_attempts\_FINAL.xlsx*”

Each of these files consists of 3 sheets.

“*Overall*” lists the non-favourable votes of the group members. Cases where at least two Member States opposed an act synchronously are highlighted. Supplementary information is also included such as date, meeting session, configuration, policy area and procedure ID.

“*Voting blocs*” contains information about the frequency of these constellations in both of the investigated periods.

“*Policy area*” presents the number of non-favourable votes by the representatives of the states in the group by policy area. In addition, the average number of oppositions in the group, the state with the highest/lowest number of oppositions is also listed.

## **4. Concluding remarks**

I hope this guide will make the huge load of data in this package somewhat more understandable. The discussion about the dataset might help alleviate any eventual concerns about the reliability of the data. Additionally, I am hopeful that the examples on the more statistics heavy parts included in this guide will help conducting inference on the results.

Should any questions arise, please feel free to contact me at [l.kocsis92@gmail.com](mailto:l.kocsis92@gmail.com).