

2019 EES Voter Study - Stacked Data Matrix Codebook

Giuseppe Carteny, Wilhelmine Häußling, Matthias Körnig, Julian Leiser

11.07.2022

Contents

1	Preface	2
2	Introduction	3
2.1	The stacked data matrix	3
2.2	Original data file	3
2.3	Overview of the SDM data file	3
3	Variables	5
3.1	Identification variables	5
3.2	Recoded variables	5
3.3	Generic categorical variables	9
3.4	Generic proximity variables	10
3.5	Generic synthetic variables	11
	Bibliographical References	13

1 Preface

This document consists in the codebook of a stacked data matrix based on the dataset of the 2019 European Election Studies (EES) voter study. The creation of this SDM is part of the research activities of [ProConEU](#), a research project aiming to analyse the enlarging gaps between proponents and opponents of the European Integration in terms of party politics, citizen politics, and social media communication. The project is funded by the German Federal Ministry of Education and Research (BMBF), and it involves the Mannheim Centre for European Social Research (MZES) of the University of Mannheim, the Ludwig Maximilian University of Munich, the University of Thessaloniki, and the University of Newcastle.

More specifically, this dataset is the product of the efforts of the ProConEU working package based at the MZES. The preparation of the 2019 EES SDM set was led by Hermann Schmitt and coordinated by Giuseppe Carteny. Wilhelmine Häußling, Julian Leiser, and Matthias Körnig actively participated to the realisation of both dataset and documentation. The data pipeline and workflow were completed between July 2021 and January 2022 making use of R (R Core Team, 2021), and are deposited in a online public repository available at <https://github.com/giucarny/EESstacked>.

2 Introduction

2.1 The stacked data matrix

The stacked data matrix (hereinafter, SDM) consists of a long format data matrix in which each row represents the (dyadic) relationship between two sets of relevant elements.

Among its applications, this data matrix has been extensively used for the study of voting behaviour, and in particular voters' propensity to vote and vote choice (**citations**). In this setting, the SDM observations are usually voter-party dyads, namely dyadic relationships between individual voters and the relevant vote choices available to each individual voter in a given election.

These new observations allow, then, a *shift downward* in terms of unit of analysis (the new units are analytically nested within the original sets of elements), and a *shift upward* in terms of conceptual generalisation, as explained below.

The reason behind the development of the SDM for voting behaviour studies is that it allows to go beyond problems related to the comparability of vote choice across different party systems, especially multi-party ones.

By relying on party-voted dyads the SDM allows to address research questions concerning *entire* party systems, thus enhancing the possibility to develop longitudinal and/or cross-national comparative analyses without:

1. Arbitrarily reducing the number of relevant vote choices (parties) of the system;
2. Reducing the vote alternatives available in a given election to a single property of said alternatives (e.g., party positions on the Left-Right continuum).

Hence, the SDM allows to include in the analyses all the relevant individual-, party-, and context-level factors that might affect the vote choice.

2.2 Original data file

The dataset from which the SDM (version 1) is computed is the 2019 European Election Study (EES) voter study (Schmitt et al. 2020). This study consists in a cross-national post-election survey, conducted by Gallup International in all 28 EU member states after the 2019 European Parliament elections. Respondents were selected randomly from access panel databases using stratification variables, with the exception of Malta and Cyprus where a multi-stage Random Digit Dialing (RDD) approach was used. In all countries, the samples were stratified by gender, age, region, and type of locality, and the sample size is roughly 1000 interviews in each EU member state (except Cyprus, Luxembourg, and Malta where the sample size is 500), with a total number of observations equal to 26,538.

2.3 Overview of the SDM data file

The variables of the dataset are grouped first according to their relationship with the set of variables available in the 2019 EES voter study. The first 131 variables consist in the original variables of said dataset, while the

remaining 27 are variables computed from the former ones or, in a few cases, original ones. This codebook refers to the latter set.

The variables computed for the SDM are then grouped as it follows:

- **Identification variables:** A set of variables computed in order to identify EES 2019 respondents', their national contexts, the relevant parties of said contexts, and the dyadic relationships between respondents and relevant parties. Said variables do not share a common suffix;
- **Recoded variables:** These variables consist in the building blocks of the generic variables presented below. More specifically they are recoded versions of a subset of variables included in the original 2019 EES voter study dataset¹. Said variables are identified by the suffix `_rec`;
- **Generic variables:** The variables represent the specific variables of the SDM. They concern the unit of analysis of the SDM approach, namely the dyadic relationship between each individual observation of the original data matrix (the 2019 EES voter study dataset) and each relevant party of a given party system. These variables share the suffix `_gen`. Generic variables are then subset in three distinct groups, namely *categorical*, *proximity*, and *synthetic* variables.

Political parties are considered “*relevant*” according to two criteria. First, if a 2019 EES voter study includes a propensity to vote (PTV) measure for a given party, then said party is considered relevant. Second, if said party obtained at least one seat in the 2019 EP elections, then it will be considered relevant.

¹The original 2019 EES voter study variables' coding is available on the [2019 EES voter study Master Questionnaire](#) and the [2019 EES voter study codebook](#).

3 Variables

3.1 Identification variables

3.1.1 party

Unique identifier of the relevant parties participating to the European Parliament (EP) elections of 2019. Only parties for which the EES 2019 voter study propensity to vote (PTV) variable is available have been selected. Values equate to those defined in the original EES 2019 vote choice variable referring to the 2019 EP elections (Q7; See the [2019 EES voter study codebook](#)).

3.1.2 stack

Unique identifier combining the individual respondent identification code as assigned in the EES 2019 voter study (`respid`; See the [2019 EES voter study codebook](#)) (`respid`) and party codes (`party`).

3.1.3 countryname and countryshort

The first variable (`countryname`) consists in the complete name of the European Union member states in 2019, whereas the second variable (`countryshort`) consists in the two-letter country code of said states as defined by [Eurostat](#).

3.2 Recoded variables

3.2.1 D1_rec

Variable measuring whether the respondent is a member of a trade union or not (Recoded from the 2019 EES variable D1).

Values:

- 0 - Not a member of a trade union
- 1 - Member of a trade union
- 98 - Don't know
- 99 - No answer

3.2.2 D3_rec

Respondent's sex (Recoded from the respondent 2019 EES sex variable, D3).

Values:

- 1 - Male
- 2 - Female
- 3 - Other

3.2.3 D4_1_rec

Respondent's age in 2019 (Recoded from the respondent 2019 EES year of birth variable, D4_1).

3.2.4 D5_rec

Respondent's marital status (Recoded from the 2019 EES variable D5).

Values:

- 0 - Single
- 1 - Married/Remarried/Single living with a partner
- 98 - Don't know
- 99 - No answer

3.2.5 D6_rec

Respondent's occupational status (Recoded from the 2019 EES variable D6).

Values:

- 1 - Self-employed
- 2 - Employed
- 3 - In school
- 4 - Working in the household
- 5 - Retired
- 6 - Unemployed
- 7 - Other
- 99 - No answer

3.2.6 D6_std_rec

Variable measuring whether the respondent is a student or not (Recoded from the 2019 EES variable D6).

Values:

- 0 - Student
- 1 - Not a student
- 99 - No answer

3.2.7 D6_une_rec

Variable measuring whether the respondent is unemployed or not (Recoded from the 2019 EES variable D6).

Values:

- 0 - Not Unemployed
- 1 - Unemployed
- 99 - No answer

3.2.8 D7_rec

Respondent's subjective social class (Recoded from the 2019 EES variable D7).

Values:

- 0 - Working or lower middle class
- 1 - Middle class
- 2 - Upper middle or upper class
- 97 - Other
- 98 - Don't know
- 99 - No answer

3.2.9 D8_rec

Respondent's area of residency (Recoded from the 2019 EES variable D8).

Values:

- 0 - Rural area or village
- 1 - Small, middle, or large town

3.2.10 D9_rec

Respondent's religious denomination (Recoded from the 2019 EES variable D9).

Values:

- 0 - Non believer/Atheist/Agnostic
- 1 - Catholic
- 2 - Orthodox
- 3 - Protestant
- 4 - Other Christian
- 5 - Other
- 99 - No answer

3.2.11 D10_rec

Respondent's frequency of religious service attendance (Recoded from the 2019 EES variable D10).

Values:

- 0 - Never/About once a year
- 1 - Less often
- 2 - About once a year
- 3 - Only on special holy days
- 4 - About each 2 or 3 month
- 5 - Once a month

- 6 - Once a week
- 7 - More than once a week
- 98 - Don't know
- 99 - No answer

N.B.: 0 includes “*Non believer/Atheist/Agnostic*” in D9_rec if and only if “*No answer*” in D10.

3.2.12 EDU_rec

Respondent's level of education (Recoded from the 2019 EES variables EDU and D2).

Values:

- 1 - Low (15 or less years of schooling)
- 2 - Medium (16-19 years of schooling)
- 3 - High (20+ years of schooling)
- 99 - No answer

3.2.13 Q25_rec

Variable measuring whether the respondent feels close to any political party or not. Differently from the original variable (Q25) party codes have been recoded in order to be line with those of the 2019 EP vote choice variable (Q7, see the [2019 EES voter study codebook](#)).

Values:

- 0 - Respondent does not feel close to a political party
- 90 - Respondent feels close to a party not among the answer categories or a non-relevant party
- 101-2807 Respondent feels close to the party [Q25_rec value]

3.2.14 Q26_rec

Variable measuring the strength of the respondent closeness to the political identified in Q25_rec.

Values:

- 0 - Respondent is merely a sympathiser of the party [Q25_rec value]
- 1 - Respondent is fairly close to the party [Q25_rec value]
- 2 - Respondent is very close to the party [Q25_rec value]
- 3 - Not asked (Respondent does not feel close to any party or doesn't know)
- 99 - Respondent does not remember/No answer

3.2.15 Q9_rec

Respondent's (recalled) vote choice at the last national elections prior to 2019. Differently from the original variable (Q9) party codes are in line with those of the 2019 EP vote choice variable (Q7, see the [2019 EES voter study codebook](#)).

Values:

- 0 - Respondent did not vote
 - 90 - Respondent voted for another party
 - 96 - Respondent did vote blanc or nil
 - 98 - Respondent does not remember
 - 99 - No answer
 - 101-2814 Respondent voted for the party [Q9_rec value]
-

3.3 Generic categorical variables

3.3.1 Q2_gen

Variable measuring whether the respondent believes that the stack party would be the best at dealing with the most important issue (as identified by the respondent herself) faced by the respondent's country (Recoded from the 2019 EES variables Q2).

Values:

- 0 - Respondent does not consider the stack party the best at dealing with the most important issue
- 1 - Respondent considers the stack party the best at dealing with the most important issue
- 96 - Not applicable (Answer to Q1 = Don't know)
- 98 - Respondent does not know
- 99 - No answer

3.3.2 Q7_gen

Variable measuring whether the respondent (recalls to have) voted for the stack party at the 2019 European Parliament (EP) elections (Recoded from the original 2019 EP vote choice variable of the EES voter study, Q7; see the [2019 EES voter study codebook](#)).

Values:

- 0 - Respondent did not vote for the stack party
- 1 - Respondent voted for the stack party
- 98 - Respondent does not remember

N.B.: 0 includes all the cases in which the respondent voted for another party, did not vote, voted blank or nil.

3.3.3 Q9_gen

Variable measuring whether the respondent (recalls to have) voted for the stack party at the last national general elections (Recoded from Q9_rec).

Values:

- 0 - Respondent did not vote for the stack party
- 1 - Respondent voted for the stack party
- 98 - Respondent does not remember

N.B.: 0 includes all the cases in which the respondent voted for another party, did not vote, voted blank or nil.

3.3.4 Q25_gen

Dichotomous variable, measuring whether the respondent feels close to the stack party (Recoded from Q25_rec).

Values:

- 0 - Respondent does not feel close to the stack party
- 1 - Respondent feels close to the stack party
- 98 - Respondent does not know

N.B.: 0 includes both the cases in which the respondent feels close to another party or does not feel close to any party.

3.3.5 Q26_gen

Ordinal variable, measuring the extent to which the respondent feels close to the stack party (Recoded from Q26_rec).

Values:

- 0 - Respondent does not feel close to the stack party
- 1 - Respondent is merely a sympathiser of the stack party
- 2 - Respondent feels fairly close to the stack party
- 3 - Respondent feels very close to the stack party
- 98 - Respondent does not know/No answer

N.B.: 0 includes both the cases in which the respondent feels close to another party or does not feel close to any party.

3.4 Generic proximity variables

3.4.1 Q10_gen

Variable measuring the respondent's propensity to vote for the stack party (computed from the 2019 EES variable Q10).

Values:

- 0 - Respondent has a very low propensity to vote for the stack party
- 1 - Respondent has a very high propensity to vote for the stack party
- 98 - Respondent does not know

3.4.2 Q11_Q13_gen

Variable measuring the proximity between the respondent's self-placement on the Left-Right ideological axis (Q11) and her perception of a specific party position on the same dimension (Q13).

Values:

- 0 - Respondent is very distant from the stack party
- 1 - Respondent is very close to the stack party
- 98 - Respondent does not know

3.4.3 Q23_Q24_gen

Variable measuring the proximity between the respondent's position about the EU integration process (Q23) and her perception of a specific party position about the same process (Q24).

Values:

- 0 - Respondent is very distant from the stack party
- 1 - Respondent is very close to the stack party
- 98 - Respondent does not know

3.5 Generic synthetic variables

3.5.1 socdem_synt_ptv

Variable measuring the affinity between respondent's socio-demographic characteristics and her propensity to vote for the stack party (Q7_gen). This variable is estimated using the linear predictions of an ordinary least squares (OLS) model. The list predictors for said model is presented below.

Values:

- 0 - Respondent has a very low affinity with the stack party
- 1 - Respondent has a very high affinity with the stack party
- 99 - Not available

N.B.: Values are *not* centered.

3.5.2 socdem_synt_vc

Variable measuring the affinity between respondent's socio-demographic characteristics and her generic vote choice (Q10_gen). This variable is estimated using the linear predictions (log-odds) of a binomial logistic regression model. The list predictors for said model is presented below.

Values:

−2.5 or below - Respondent has a very low affinity with the stack party
+2.5 or above - Respondent has a very high affinity with the stack party
99 - Not available

N.B.: Values are *not* centered.

3.5.3 Independent variables for socdem_synt_ptv and socdem_synt_vc estimation

3.5.3.1 Categorical independent variables:

- **D1_rec**: Variable measuring whether the respondent is a member of a trade union (1) or not (0);
- **D3_rec**: Respondent's gender (0 = Male, 1 = Female);
- **D5_rec**: Whether the respondent is married/remarried/single living with a partner (1) or single/divorced/separated/widowed (0);
- **D6_une_rec**: Variable measuring whether the respondent is unemployed (1) or not (0);
- **D7_rec**: Subjective social class (0 = working class or lower middle, 1 = middle class, 2 = upper middle or higher class);
- **D8_rec**: Whether the respondent lives in a rural (0) or urban area (1);
- **EDU_rec**: Respondent's years of formal education (1 = 15 years or less, 2 = 16-19 years, 3 = 20+).

3.5.3.2 Continuous independent variables:

- **D4_1_rec**: Respondent's age (min = 16, max = 98; ordinal treated as continuous);
- **D10_rec**: Respondent's religiosity (min = 0, max = 6; ordinal treated as continuous).

Bibliographical References

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>