CESifo WORKING PAPERS

10855
2023

December 2023

A Bayesian Networks Approach for Analyzing Voting Behavior

Miguel Calvin, Pilar Rey del Castillo

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10855*

A Bayesian Networks Approach for Analyzing Voting Behavior

Abstract

The problem of finding the factors influencing voting behavior is of crucial interest in political science and is frequently analyzed in books and articles. But there are not so many studies whose supporting information comes from official registers. This work uses official vote records in Spain matched to other files containing the values of some determinants of voting behavior at a previously unexplored level of disaggregation. The statistical relationships among the participation, the vote for parties and some socio-economic variables are analyzed by means of Gaussian Bayesian Networks. These networks, developed by the machine learning community, are built from data including only the dependencies among the variables needed to explain the data by maximizing the likelihood of the underlying probabilistic Gaussian model. The results are simple, sparse, and non-redundant graph representations encoding the complex structure of the data. The generated structure of dependencies confirms many previously studied influences, but it can also discover unreported ones such as the proportion of foreign population on all vote variables.

JEL-Codes: C460, D310, D550, D720, D910.

Keywords: Bayesian networks, Gaussian distributions, voting behaviour, elections, voter turnout, political participation.

Miguel Calvín
ATM Grupo Maggioli S. L.
Avda. Poeta Joan Maragall, 1
Spain - 28020 Madrid
miguel.calvin@atm-maggioli.es

*Pilar Rey del Castillo**
Institute for Fiscal Studies
Avda. Cardenal Herrera Oria, 378
Spain - 28035 Madrid
mpilar.rey.castillo@ief.hacienda.gob.es

*corresponding author

December 2023

1. Introduction

Free and fair elections are fundamental pillars for well-functioning representative democracies. It is widely accepted that, given the possible difficulties of direct democracies, people must choose political mediators to act on their behalf (Berganza, 2000).

The fact is still that people do vote although in large electorates there is a small chance for an individual vote to be pivotal, what has been called “the paradox of voting” (Bendor et al., 2003; Downs, 1957). The reasons for electors, first, to take part in casting their vote and then, to make their choice between the different vote options, become thus an interesting subject to study. For both the citizens –to participate more effectively in the political process– and the political parties –to plan their campaign strategy–, it is therefore crucial to know about the factors determining voting behavior. There are countless models trying to explain the factors influencing citizens' behavior when they have to confront these decisions (Antunes, 2010; Mahsud & Amin, 2020; Smets & van Ham, 2013).These models are commonly grouped into three categories: the sociological models, emphasizing the influences of social factors; the psycho-social models, focusing on voters' identification with political parties; and the rational-choice models, referred to as models of economic vote. The issue is that all these theoretical explanations have found their own empirical support. Consequently, the final decision of an individual in a particular electoral occasion is the result of complex interactions among several diverse kinds of factors.

Most of the studies on the subject discuss the first of the questions, i.e., what factors are determining electoral participation. Without any attempt to be exhaustive, some of the principal determinants at micro level are demographic (age, gender, ethnicity, religion), economic (employment, income levels), cultural (family and community values, cultural norms), psychological (cognitive biases, emotions, personal beliefs), and political (debates, campaign strategies). Stockemer (2017) synthesized the results of more than 130 articles published in peer reviewed journals using electoral participation (also called turnout) as the dependent variable. The study suggests that the factors determining turnout might be more complex and more context dependent than the current theory suggests. Similarly, Ansolabehere and Hersh (2017) found, from respondents of surveys, that standard determinants of participation such as demographic characteristics or measures of political engagement explain at most between a third and a half of voting participation; and a more recent study, using millions of regressions with varying model specifications, established which social, institutional, and political factors driving national-level voter turnout are empirically robust (Frank & Martínez i Coma, 2023). There are also other studies analyzing the role of one or more particular determinants of electoral participation: of age (Bhatti & Hansen, 2012; Hannan-Morrow & Roden, 2014; Zeglovits & Aichholzer, 2014); economic inequalities (Bhatti et al., 2019; Freeman, 2005; Matsubayashi & Sakaiya, 2021); income levels (Akee et al., 2020; Piketty, 2021); race, sex and party of affiliation (Barber & Holbein, 2022);

local weather (Arnold & Freier, 2016; Artés, 2014; Damsbo-Svendsen & Hansen, 2023); economic perceptions (Green, 1985; Wlezien et al., 1997); taxes (Bierbrauer et al., 2018; Brett & Weymark, 2017); educational attainment (Dawes et al., 2021; Filer et al., 1989); and political and institutional variables in different countries (Cancela & Geys, 2016; Denny & Doyle, 2008; Geys, 2006; Goldberg & Sciarini, 2023; Petitpas et al., 2021). There are even studies in the opposite direction, that is, analyzing the effects of electoral participation on other variables, e. g. the effects of electoral participation on economic growth (Mueller & Stratmann, 2003).

There is also a vast literature on the other question, the determinants of the vote choice or vote for the different options. For example, the theory that voters cast their ballots to the most credible party proponent of a particular issue is refined by Bélanger & Meguid (2008). Other related topics in the context of multiparty systems have also been studied. A few examples are: the affective polarization between the supporters of opposing political parties (Comellas & Torcal, 2023; Wagner, 2021; Harteveld, 2021); the factors influencing the vote for populist parties in Germany, the United Kingdom and other European countries (Aronsson et al., 2020; Arin et al., 2023) ; the association of changes of voting patterns with moral values (Enke, 2020); the relationship between inequality and partisan voting in the United Stated (Gelman et al., 2010); and the link between individual deprivation and the support to the protest party (Altomonte et al., 2019) .

Another relevant issue refers to the supporting data used to validate the models proposed in the different analysis. The confidentiality of the individual votes in elections makes it difficult to discover their relationship with other factors. Sometimes, experimental designs are used to validate the models, usually trying to mimic the functioning of an electoral occasion where voters choose between specific options and are confronted with given requirements (Sondheimer & Green, 2010; Wiese & Jong-A-Pin, 2017). However, more often the supporting information comes from results of electoral opinion surveys, or from validated official records. The surveys are usually conducted before or after an election and collect the opinions and attitudes of representative samples of citizens with the right to vote. Thus, they may include, as well as the questions on the participation on the election and which candidate or list they will vote or voted for (depending on being a pre- or a post-electoral survey), other questions reflecting the values of some variables related to the first ones. Studies based on these surveys are countless; just a few examples are Bélanger and Meguid (2008), Gelman et al. (2010), Geys et al. (2022), and Maldonado & Hernandez (2020). But the electoral surveys suffer from some problems (Barber & Holbein, 2022; Healy et al., 2016; Stocké & Stark, 2007; Wiese & Jong-A-Pin, 2017). The first –and the most important– is that the answers to the questions are self-reported, what makes them affected by the social desirability bias (Berinsky, 1999). Besides, the data are only elicited at a single point in time and are thus influenced by limited human memory. And lastly, most of the surveys are designed to be representative at the aggregate level, being rarely representative at more detailed levels for the variables of interest. An interesting study to verify and validate a large number of electoral surveys in the United States found, first, that over-estimation of electoral turnout is now primarily due to misreporting rather than to sample selection bias, and second, that respondents are found to misreport only on survey items associated with socially desirable outcomes (Ansolabehere & Hersh, 2017).

In contrast to data from surveys, the data from validated official records are not affected by these problems and its information is considered more reliable. The official data on participation, in most countries, typically only include other simple variables such as age or sex. For this reason, it is common to find studies matching the administrative register data to other data files including some variables of interest (Bellettini et al., 2017; Bratsberg et al., 2019; Goldberg & Sciarini, 2023; Zeglovits & Aichholzer, 2014).

All the earlier examples use individual data on participation. Official data on voting for parties or candidates are only released in aggregate form for geographical areas, to protect the privacy of individual voters (and because, usually, information on an individual's specific vote is not accessible). These aggregated data can also be matched with other datasets to analyze the connection with other features: for example, data on vote for parties at the county level in Poland (380 counties as of December 2023) is matched with administrative information on the number of emigrants by county to study the effect of emigration rates on voting for right-wing parties (Giesing et al., 2023).

The study presented on this paper makes use of data from official registers, specifically the records on the distribution of votes to parties in the November 2019 General Elections by census sections in Spain (more than 36000 as of December 2023). These data are matched to other official files including the values of some possible predictors of voting behavior to analyze their relationships. Apart of using entire datasets −with information of all and not of a sample of the census sections− in this work we learn the statistical relationships among the participation, the vote for parties and some socio-economic variables, by means of a method not used before to our knowledge for the purpose, i.e., Gaussian Bayesian Networks (Geiger & Heckerman, 1994; Shachter & Kenley, 1989).

The remainder of this paper is organized as follows: Section 2 presents the information on electoral data and possible predictors obtained from administrative records; Section 3 summarizes the features of Gaussian Bayesian Networks and the methods used; Section 4 shows the more important results referring to the interactions between the vote and some factors; and finally, Section 5 presents some comments and remarks.

2. The data

The census sections in Spain are geographical areas designed for the control and management of electoral processes. These areas (more than 36000) result from the union of polling stations, the level at which the ballots are counted. Each census section includes between 500 and 2000 residents. There was no official statistical information at this level of detail until recently. But since 2019 the Spanish National Statistical Institute publishes a new statistic called Household Income Distribution Atlas (HIDA)[1] based on administrative registers. It contains information about statistical indicators, at geographically detailed levels, on the level and distribution of the income per capita and per household, along with other demographic and social indicators.

Our aim is to study the interactions between voting behavior and some predictors. It may be assumed that the variables are quite similar within the census sections, something that may help to improve the estimation of their interactions with the vote. The data obtained from the HIDA 2020 (the data are more reliable from this year), and referring to census sections, are the following:

- Variables of incomes. Income indicators are obtained from the collaboration between the Spanish National Statistical Office and the Spanish Tax Agency. The variables used in this work are:
    - Average annual net income per person (€ in thousands) (*AvNPI*)
    - “ “ “ “ “ household (€ in thousands) (*AvNHhI*)
    - Average personal tax rate (*AvPT*). Following the methodology of the HIDA 2020, the differences between the gross income per person and the net income per person for each census section are the taxes and contributions. Thus, the rate $(Gross\ income\ per\ person - Net\ income\ per\ person) * 100/Gross\ income\ per\ person$ can be used as an estimation of the average personal tax rate (%) in each census section.
    - Average household tax rate (*AvHhT*). Similarly, for the estimation of the average household tax rate (%).
    - Median after-tax income per consumption unit (€ in thousands) (*MdICU*)
    - Average percentage of household income from wages and salaries (*Wages*)
    - “ “ “ “ “ “ pensions (*Pensions*)
    - “ “ “ “ “ “ unemployment benefits (*Unempl*)
    - “ “ “ “ “ “ other benefits (*OthBen*)
    - “ “ “ “ “ “ “ sources (*OthSour*)
- Demographic information. There are six main items coming from the population census. These variables are:

[1] https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177088&menu=ultiDatos&idp=1254735976608

- Average age (*AvAge*)
- Percentage of population under 18 years old (*Pop<18*)
- “ “ “ over 64 years old (*Pop>64*)
- Average size of households (*AvHhS*)
- Percentage of one-person households (*1PHh*)
- “ “ foreign population (*Foreign*)

- Inequality measures. The matching of demographic and tax data at the census sections level allows for calculating some measures of inequality. In particular:
  - The Gini coefficient, a single statistic of inequality summarizing the distribution of income or wealth across the population (*Gini*). A Gini coefficient of 0 reflects complete equality, while a Gini coefficient of 100 signifies maximal inequality.
  - The ratio of household income at the 80^{th} percentile of income distribution to that at the 20^{th} percentile, one of the most frequent references in analyses of inequality (*P80/P20*) (Bartels, 2017).

These 18 variables from HIDA are considered of interest to check their influence on the electoral results. There exist obvious relationships among some of them. However, they are kept because they may capture different nuances of complex interactions. Also, we must have in mind that these are not all the factors influencing the vote, but there may be other variables confounding the relationships between them.

On the other hand, once the results of an election are officially declared, the Spanish Interior Department handles publishing the data, which can be downloaded from its website[2]. The interest of this work is focused on voting for national parties. Therefore, for each census section, the variables obtained from the Interior Ministry and referred to the 2019 General Elections in Spain are the following:

- Percentage of turnout (*Turnout*)
- “ “ votes for Ciudadanos (*Ciudadanos*)
- “ “ “ “ PP (*PP*)
- “ “ “ “ Podemos (*Podemos*)
- “ “ “ “ PSOE (*PSOE*)
- “ “ “ “ VOX (*VOX*)

Consequently, there are 18 variables (from income, demographic and inequality measures) to study as predictors of participation and vote for 5 national parties. The procedures used are explained in the following section.

[2] https://infoelectoral.interior.gob.es/opencms/en/inicio

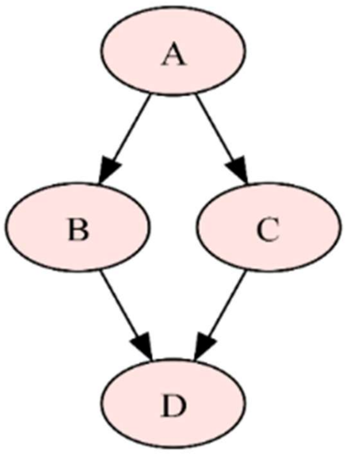
3. The methods

The essential tool exploited in this work to analyze the relationships between the variables is the generation of Gaussian Bayesian Networks from the data available.

To briefly introduce the Gaussian Bayesian Networks, first we remember the concept of independence between random variables: two discrete random variables X, Y are said to be independent if and only if $P(X = x, Y = y) = P(X = x).P(Y = y), \forall x, y$. Another way to say this, using conditional probabilities, is that $P(X = x|Y = y) = P(X = x), \forall x, y$, or else another way, symmetrically, that $P(Y = y|X = x) = P(Y = y), \forall x, y$ (Kingman & Feller, 1972). Intuitively, what this means is that the occurrence of $Y = y$ does not give any information on the distribution of the occurrence of X (and symmetrically).

Now, we present the Bayesian Networks (BNs), a class of graphical models (D. E. Heckerman, 1999; Jensen & Nielsen, 2007; Lauritzen, 1992; Pearl, 1988a). Formally, a Bayesian Network $\mathcal{B} = (\mathcal{G}, \mathcal{P})$ consists of (a) a directed acyclic graph $\mathcal{G}$, composed of nodes or variables and directed edges, also called arcs, encoding conditional independence relationships among them, and (b) a set of parameters $\mathcal{P}$. These parameters constitute the conditional probability distributions over the nodes, i.e., the quantitative information on the dependencies among them.

Figure 1
Bayesian network graph $\mathcal{G}$



An example with four discrete random variables A, B, C and D is the Bayesian Network $\mathcal{B}$ where the directed acyclic graph $\mathcal{G}$ is shown in Figure 1, and $\mathcal{P}$ is the set of values of the conditional probability distributions $\{P(A), P(B|A), P(C|A), P(D|B, C)\}$. The arcs in $\mathcal{G}$ are the set $\{(A, B), (A, C), (B, D), (C, D)\}$ where (x, y) represents a directed arc from x to y. These arcs encode direct relationships of dependence. The independence between two variables is characterized by the lack of a direct or indirect path between them, e.g., from $\mathcal{G}$ it follows that B and C are independent.

The construction of a BN allows for decomposing the joint probability distribution of a set of random variables into a group of conditional probability distributions. The joint probability distribution may then be represented as the product of these conditional probabilities, implying a significant reduction in the number of parameters necessary to completely specify it. BNs constitute a potent multivariate technique, allowing the search of conditional probability dependencies over a set of random variables. If $\{X_1, \dots, X_n\}$ is such a set, the joint probability distribution is then $P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i|\pi_i)$, where π_i, $i = 1, \dots, n$ are the parents of node X_i, i.e., the nodes from which exist a direct arc to node X_i. The inferred simplification in the calculation of all conditional and unconditional distributions between subsets of variables allows us to predict one or more variables from the values of others, and to answer probabilistic queries. The procedure of computing the probability distributions of some variables given the values of others is called *Evidence Propagation*.

The *Markov blanket* of a node includes all its parents, children and children's parents (Pearl, 1988b). Given values for the nodes in its *Markov blanket*, a node results conditionally independent of all other nodes. BNs have been developed by the machine learning community and constitute an appealing and complete approach to build complex data-driven networks within a probabilistic framework.

A BN can be generated, or learned in machine learning terms, from experts, but also from data on a set of variables. What must be learned is, first, the structure of dependencies or best directed acyclic graph representing the dependencies and then, the distributions conditioned on their corresponding parents. There are two basic approaches to learn the structure: first, *constrained-based* methods, using conditional independence tests to identify the relationships among variables; and second, *score + search* methods, establishing the learning task as an optimization problem where, within a set of possible structures, each candidate is assigned a score —measuring its quality with respect to the data— which the algorithm attempts to maximize.

On the other hand, once the structure is learned, the conditional distributions can also be estimated from data (Heckerman et al., 1995). Both learning (the structure and the conditional distributions) and evidence propagation are compute-intensive tasks, but many methods have been developed since the appearance of BNs in the 1980s, using exact and approximate algorithms (Fenton & Neil, 2012).

All the previous concepts for BNs of discrete random variables can be extended to continuous random ones. This gives rise to the concept of interest in this paper, that of Gaussian Bayesian Networks (GBNs), which are Bayesian networks whose associated joint probability distribution is a multivariate Gaussian distribution. It can be shown that the conditional probability density functions $f(X_i|\pi_i), i = 1, \dots, n$ contributing to the factorization of the joint probability distribution, follow distributions $N\left(\mu_i + \sum_{X_j \in \pi_j} \beta_{ij}(X_j - \mu_j), \nu_i\right)$, where μ_i is the unconditional mean of X_i, β_{ij} is the regression coefficient of X_j in the regression of X_i on its parents π_i , and ν_i is the conditional variance of X_i given its parents. The Markov Blanket of a node in a GBN is constituted exclusively by its parents (Murphy, 2012). In this case, the evidence propagation can be computed using a method presented by Castillo et al. (2012). Given the structure of the GBN, i.e., the set of parents for each variable, the parameters can be easily learned as a maximum likelihood fit of the linear regression of each variable on its parents. The GBN approach provides an optimal and non-redundant probabilistic Gaussian model of the complex system of interactions among variables using a network support describing the relevant dependencies (Graafland et al., 2019).

Once a BN has been learned from data, the direction of the arcs indicates the direction of the probabilistic influences, do not necessarily represent causal relationships. The existence of latent variables –unobserved variables that may influence the variables in the network– can produce spurious correlations between the observed variables and thus introduce bias in the causal relationships (Gillies, 2002; Pearl, 2009). In the case of the network to build using the variables in Section 2, the earlier references presented in Section 1 suggest many variables that can play the role of latent ones.

A GBN learned from data can be represented as a set of multivariate linear relationships, where the arcs are assigned the regression coefficients as connection weights from parent node-set π_i to node X_i indicating the strength of the association. In the following section a GBN is learned from the data described in Section 2.

4. Interactions among the census section variables

As a first introduction, to understand the country's political party's context, Table 1 shows the average of the parties' ideology assessment from 1 to 10 (from extreme left to extreme right) from a 2019 electoral survey (Study number 3263 from Sociological Research Center, Madrid)[3].

[3] https://www.cis.es/cis/opencm/ES/2_bancodatos/estudios/ver.jsp?estudio=14473&cuestionario=17452&muestra=24446

Table 1
Average parties' ideology assessment

Party	*Ideology assessment*
Ciudadanos	7.1
Podemos	2.3
PP	7.9
PSOE	4.2
VOX	9.4

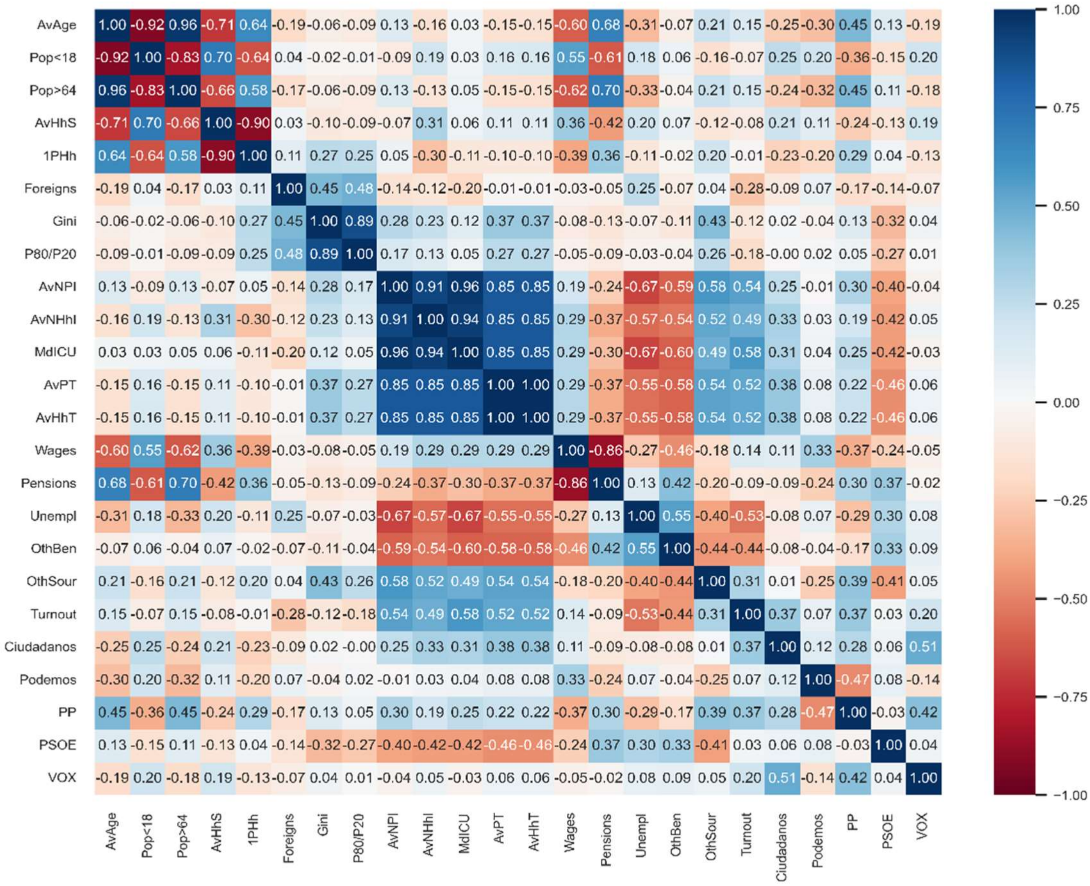
There are some references in the literature with respect to determinants of vote in Spain, most of them using data from electoral surveys (Fraile & Hernández, 2020). The present study uses data from official registers as explained before. To give an overview of the official data from Section 2, Table 2 shows a statistical summary for each variable.

Table 2
Statistical figures of census section variables

Variable	*Mean*	*Standard deviation*	*Min*	*First quartile*	*Median*	*Third quartile*	*Max*
AvAge	45.2	5.7	20.5	41.6	44.6	48.0	75.4
Pop<18	15.5	5.5	0.0	12.5	15.4	18.6	50.0
Pop>64	22.5	9.3	0.0	16.4	21.4	27.4	80.0
AvHhS	2.5	0.3	1.0	2.3	2.5	2.7	7.0
1PHh	29.9	9.5	0.0	23.7	28.8	34.8	100.0
Foreigns	10.7	9.6	0.0	3.8	8.0	15.0	89.0
Gini	29.7	3.9	20.4	27.0	29.2	31.9	44.4
P80/P20	2.6	0.4	1.8	2.3	2.5	2.8	4.4
AvNPI	12.6	3.8	4.2	10.1	11.9	14.3	31.7
AvNHhI	31.8	10.6	13.6	25.1	29.3	35.5	89.7
MdICU	16.8	4.9	5.9	13.6	15.8	19.3	37.5
AvPT	15.17	4.33	3.78	12.28	14.63	17.49	75.43
AvHhT	15.17	4.31	2.67	12.27	14.63	17.49	72.56
Wages	57.2	11.1	19.3	50.2	56.4	62.9	96.3
Pensions	22.7	9.5	0.0	17.3	22.9	28.6	61.6
Unempl	5.2	2.9	0.0	3.4	4.8	6.5	26.6
OthBen	4.6	2.3	0.0	3.3	4.4	5.8	32.2
OthSour	10.3	6.1	0.2	6.1	8.9	12.8	55.7
Turnout	70.1	8.1	0.0	65.6	70.7	75.5	100.0
Ciudadanos	4.5	2.4	0.0	2.9	4.4	5.9	40.7
Podemos	7.9	4.3	0.0	5.1	7.9	10.6	55.1
PP	15.6	9.5	0.0	8.5	14.6	20.9	93.8
PSOE	19.6	7.1	0.0	15.2	19.4	23.7	80.0
VOX	10.4	5.7	0.0	5.9	10.3	14.1	55.6

In Figure 2 the correlation coefficients between pairs of variables can be seen. Most of the variables grouped within each one of the three theoretical categories of predictors (demographic, inequality and incomes) are measures of similar or connected phenomena, being clearly correlated among themselves. These factors also have various levels of correlation with participation (*turnout*) and vote, ranging from *-0.42* between *average households' income* and *PSOE*, and *0.58* between *median income per consumption unit* and *turnout*.

Figure 2
Correlation coefficients between census section variables



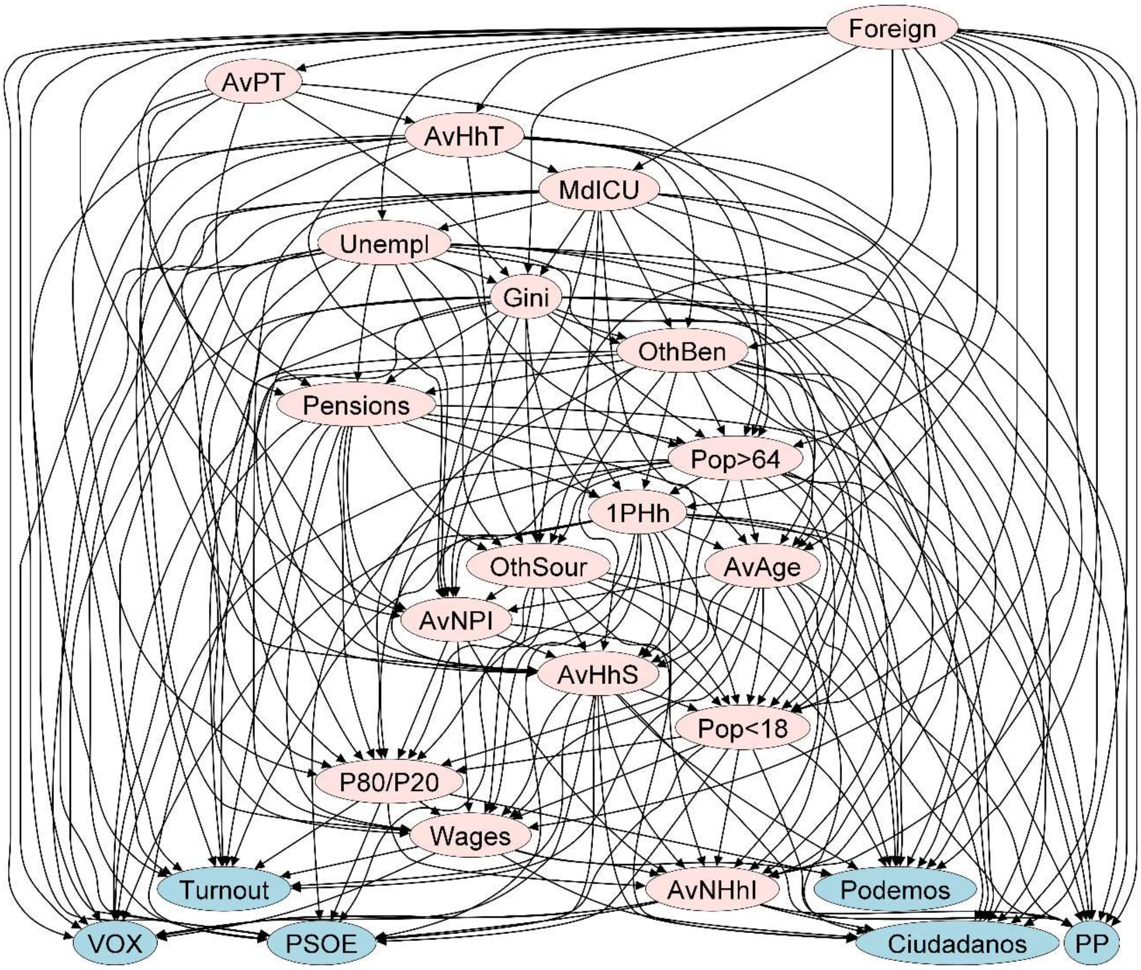
The relationships among the 24 variables are modeled in a more comprehensive way by building a GBN. This approach provides an optimal Gaussian model of the complex system of interactions using a network support characterizing the relevant dependencies. Statistical processing and analysis were performed using the software packages *Python* (Python Core Team, 2019) and *pyBNesian* (Atienza et al., 2022).

A *score + search* method with the algorithm *Hill-Climbing* was used to derive the structure of the dependency graph. It is a greedy score-based learning algorithm for Gaussian Bayesian Networks (Margaritis & others, 2003). The score used is the *Bayesian Information Criterion* (BIC) (Konishi & Kitagawa, 2008). To assist the algorithm in retrieving the most proximal associations between variables, some assumptions have been incorporated in the form of black-listed arcs preventing certain relationships, i.e., limiting the number of edges that are searched for. Specifically, arcs from turnout and vote variables to other variables have been excluded.

A node's indegree is the number of arcs directed to the node. Given a maximum indegree for all nodes, the algorithm finds the direct acyclic graph with the best BIC. The final GBN learned is the result of the Hill-Climbing algorithm applied with the maximum indegree 12 that minimizes the average Mahalanobis distance (Mahalanobis, 2018) between real and predicted values for the vote variables (*1.8* points). Unlike the Euclidean distance, this solves the problem of scale and considers the correlations between all vote variables. The distance is computed by executing *k*-folds cross-validation (Stone, 1974), i.e., by splitting the data into *k* random subsets or folds and averaging the distances obtained for each fold from the model built using the rest of the folds.

Figure 3 illustrates the learned network structure using pink nodes for the predictors and blue nodes for the vote variables. This structure of dependencies provides the key for identifying as predictors of each vote variable its corresponding parents. When the values of a node's parents are known, the rest of the nodes provide no more information on its behavior. Indirect influences are denoted through the existence of a path, e. g. from *average personal tax rate* to participation (*turnout*) through some of its parent nodes. But, given the value of its parents, participation is conditionally independent from *average personal tax rate*.

Figure 3
Best GBN learned structure.



The outdegree of a node –the number of arcs that the node directs to others– may be interpreted in terms of the opportunity for influencing any other node. Considering the relevant relationships as those from determinants to vote variables, *households' size, one-person households* and *foreigners' percentage* nodes have the highest outdegree, although none of them have significant correlations with vote nodes. The lowest outdegree from determinants (0) corresponds to *personal average income* and *average tax rate*, suggesting that they have little influence on the vote variables, or that their influence is executed through others.

The probability distributions that complete the definition of the learned GBN are Gaussian (Murphy, 2012). Table 3 shows the coefficients of the parent nodes (rows) in the calculation of the mean of the distribution of the column nodes (vote variables).

Table 3

Coefficients of predictors on the linear regressions of vote variables

		Explained variable					
		Turnout	*Ciudadanos*	*Podemos*	*PP*	*PSOE*	*VOX*
Explanatory variable	*AvAge*	-0.414	-0.187	-0.353	0.415	.	.
	Pop<18	.	.	-0.348	0.254	-0.149	0.258
	Pop>64	.	-0.067	-0.050	-0.148	-0.121	-0.275
	AvHhS	-8.717	-1.509	-7.952	12.058	-10.849	1.774
	1PHh	-0.252	-0.041	-0.266	0.466	-0.262	0.060
	Foreign	-0.047	-0.023	0.047	-0.221	-0.074	-0.100
	Gini	-0.299	-0.030	-0.145	0.232	.	0.074
	P80/P20	-2.142	.	0.969	.	-1.716	.
	AvNPI	.	.	.	.	.	.
	AvNHhI	.	0.043	.	0.155	0.332	0.087
	MdICU	0.262	.	.	.	-0.775	-0.232
	AvPT	.	.	.	.	.	.
	AvHhT	0.550	0.179	0.141	.	-0.282	.
	Wages	-0.153	-0.131	0.075	.	.	-0.191
	Pensions	0.098	.	.	0.490	0.162	0.144
	Unempl	-0.772	-0.133	.	0.073	0.129	-0.222
	OthBen	-0.675	-0.185	0.152	-0.896	.	-0.531
	OthSour	.	-0.157	-0.124	0.304	-0.236	.

It is important to observe that a strong correlation between two variables does not imply an edge between them. In this way, the participation (*turnout*) has relatively high correlations with average *personal* and *households' income*, and *average personal tax rate*, but they do not appear as parents in our GBN, i.e., given the values of other variables, they provide no information on the value of turnout. On the contrary, variables with low correlation with turnout such as *households' size* and *one-person households* have been found unexpected determinants. It must be remembered that the learned network encodes dependencies indicating a probabilistic, not a causal relationship, some dependencies being indirectly explained by others. The network is the one that optimizes the structure of dependencies explaining the data available in terms of BIC and distances between real and predicted values for the vote variables.

As described in Section 3, the regression coefficients may be understood as connection weights from parent nodes to the node of interest, indicating the strength of the association between them. The magnitudes of these coefficients cannot be easily compared because the regressions have different intercepts, and the variables are measured in different scales. But something that can be studied is the direction of the association, i.e., whether it is positive –when the values of one variable tend to increase as the other variable increases– or negative, when the opposite happens.

As an example, *turnout* and vote for *Ciudadanos, PP, PSOE* and *VOX* tends to decrease when the proportion of *foreigners* increases, while it has a positive association with vote for *Podemos*.

The factors determining turnout are exhaustively studied in the literature as said before. It results interesting to observe that among the age-related variables, frequently presented as relevant determinants for participation, only *average age* is found as such in our network, according to more recent studies (Ansolabehere & Hersh, 2017; Frank & Martínez i Coma, 2023). However, both variables measuring inequality (*Gini, ratio P80/P20*) have been selected as factors influencing turnout in a negative way, as other articles suggest (Bhatti et al., 2019; Freeman, 2005; Matsubayashi & Sakaiya, 2021). It seems the income inequality in the neighborhood is measured and appreciated correctly within census sections. In the same way, in our study the percentage of income from *unemployment* influences *turnout* in a negative way. Previous studies suggest that the level of unemployment benefits and the unemployment rate have a negative impact on the probability the unemployed will vote for the party expected to produce more growth (Grafstein, 2005). On the other side, the percentage of income from *pensions*, the *households tax rate* and the *median income per consumption unit* influence turnout in a positive way, while the *average personal* and *household income* provides no information on the value of participation when we know the values of its parents in the GBN.

As for interpreting the predictors of the vote for parties, the directions of the relationships are discussed below:

- The age-related variables show different behaviors. While the *average age* is positively associated with the right party *PP*, the relationship is negative with center and left parties (*Ciudadanos, Podemos*). Although in Spain people must be at least 18 years old to vote, the proportion of *population under 18* is negatively associated with left parties (*Podemos, PSOE*) and positively with right parties (*PP, VOX*). Instead, the proportion of *population over 64 years* has a negative relationship with vote for all parties.
- *Households' size* and *one-person households* relate positively to right parties (*PP, VOX*) and negatively to left and center-left parties (*Ciudadanos, Podemos, PSOE*).
- The proportion of *foreign* population has a positive influence exclusively in the vote for *Podemos*, and negative in the vote for the rest of the parties.
- The inequality measures (*Gini, ratio P80/P20*) have an alternative behavior, i.e., one and only one of the two appears as an influencing factor for each party, except in the vote for *Podemos* where both variables are considered determinants. The direction of the relationship is positive for *PP* and *VOX*, and it is negative for *Ciudadanos* and *PSOE*.

- Curiously, the *personal* variables *net income* and *tax rates* do not appear as determinant factor for any party, although having significant correlations. Their influence may be explained by indirect paths through other variables.
- *Households' income* variable has a positive probabilistic influence on the vote for all parties except *Podemos*, while *median income per consumption unit* influences negatively the vote for *PSOE* and *VOX*, something that requires more complex explanations than the trivial ones based on left/right parties.
- The variable estimating the paid *household taxes* influences negatively the vote for *PSOE* and positively *Ciudadanos* and *Podemos,* while *PP* and *VOX* are not influenced by *households' tax* payments.
- The percentage of household income from *wages* and salaries could be considered as an indicator of the weight of employees and salaries in the population of each census section. It has a negative influence on the vote for *Ciudadanos* and *VOX* while it has a positive influence on *Podemos*.
- *Income from pensions* influences positively the vote for *PP, PSOE* and *VOX*, while it is not determinant for the rest of the parties.
- A negative influence exists from the percentage of *unemployment benefits* on the vote for *Ciudadanos* and *VOX* and a positive one on the vote for *PP* and *PSOE*, being *Podemos* the only one that is not determined by *unemployment*.
- *Incomes from other benefits* (another indicator of persons financed from the national budget) has a positive influence on the vote for the leftist party *Podemos* and negative on *Ciudadanos, PP* and *VOX*, as is generally accepted for right parties' behavior.
- Finally, considering the percentage of *other sources* of incomes as indicator of rich people in each census section, it positively influences the vote for *PP* and negatively the vote for *Ciudadanos, Podemos* and *PSOE*, in line with left/right parties' behavior according to most of the studies.

5. Conclusions

In this paper we revisited the problem of analyzing the factors determining voting behavior using a previously unexplored information from the November 2019 General Elections in Spain combined with other official social and economic figures. The data have a level of disaggregation, census sections, that is not common for studies on the subject. The application of GBNs has provided a sound approach for the study of the relationships among variables in a probabilistic framework. Hence, it is generalizable to new data and has predictive power.

Some socio-economic determinants have been analyzed by inferring their behavior through the GBN structure of dependencies. We have noticed that the structure matches many expected and formerly studied influences, e.g., that of inequality measures on voter turnout and vote for parties. Other factors do not behave as expected, i.e., their relationships are difficult to explain. For example, unemployment benefits variable has negative influence on participation and on votes for *Ciudadanos* and *VOX*, and positive on *PP* and *PSOE*.

In addition, the GBN is able to discover unreported conditional dependencies and independencies. In this way, the proportion of foreign population has an influence on all vote variables, being this relationship positive only with *Podemos*. The determinant factors for this party are generally different to those determining the vote for other parties. In this way, *Podemos* is the only party not positively influenced by average households' income, while its relationship is opposite to what other parties have when it comes to foreign population and incomes from wages, unemployment and other benefits.

The majority parties *PP* and *PSOE* are related in the same direction with many variables: population over 64 years, foreigners' proportion, average households' incomes and incomes from wages and other benefits. And they are related in opposite direction with others: population under 18 years, household's size, one-person households and incomes from other sources.

Regarding *Ciudadanos*, sometimes it has influences such as those of the right parties and other times as those of the left. Finally, *VOX* has most of the associations in the same direction as the *PP* party, except in the case of unemployment benefits which is opposite.

There are also some determinant factors that have an effect aligned with a left/right behavior, i.e., negatively influencing the vote for left parties and positively the vote for right ones: population under 18 years, household's size, one-person households, Gini inequality coefficient and incomes from other sources.

Although the factors studied have shown they can predict the vote variables with a reasonable degree of accuracy, there are other dependencies whose direction are difficult to understand, sometimes behaving in a counterintuitive way. The reason may be the existence of other psycho-social or institutional predictors not included in the GBN, such as targeted electoral campaigns, televised political debates, recent political or economic events, etc.

An important fact to remark is that, in the GBN, there is not any variable with geographical information. For example, the variable province could have been thought as an interesting one to include, given the big differences of votes, parties and candidates by province. But the 18 variables compiled from income, demographic and inequality measures provide enough information to produce more accurate predictions than the ones obtained by generating a GBN for each province.

Obviously, the results depend on the data set used. In the future, we intend to repeat the exercise when the detailed results for the 2023 Spanish general elections are available, to evaluate the robustness of the discovered predictors.

REFERENCES

Akee, R., Copeland, W., Holbein, J. B., & Simeonova, E. (2020). Human Capital and Voting Behavior across Generations: Evidence from an Income Intervention. *American Political Science Review*, *114*(2), 609–616.

Altomonte, C., Gennaro, G., & Passarelli, F. (2019). *Collective Emotions and Protest Vote* (CESifo Working Paper Series, Issue 7463). CESifo.

Ansolabehere, S., & Hersh, E. (2017). Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate. *Political Analysis*, *20*(4), 437–459. https://doi.org/DOI: 10.1093/pan/mps023

Antunes, R. F. (2010). *Theoretical models of voting behaviour*. https://api.semanticscholar.org/CorpusID:62788428

Arnold, F., & Freier, R. (2016). Only conservatives are voting in the rain: Evidence from German local and state elections. *Electoral Studies*, *41*, 216–221. https://doi.org/https://doi.org/10.1016/j.electstud.2015.11.005

Artés, J. (2014). The rain in Spain: Turnout and partisan voting in Spanish elections. *European Journal of Political Economy*, *34*, 126–141. https://doi.org/https://doi.org/10.1016/j.ejpoleco.2014.01.005

Atienza, D., Bielza, C., & Larrañaga, P. (2022). PyBNesian: An extensible python package for Bayesian networks. *Neurocomputing*, *504*, 204–209. https://doi.org/https://doi.org/10.1016/j.neucom.2022.06.112

Barber, M., & Holbein, J. B. (2022). 400 million voting records show profound racial and geographic disparities in voter turnout in the United States. *PLoS ONE*, *17*(6).

Bartels, L. M. (2017). Unequal Democracy: The Political Economy of the New Gilded Age. In *The Political Economy of the New Gilded Age*. Princeton University Press. https://doi.org/doi:10.1515/9781400883363

Bélanger, É., & Meguid, B. M. (2008). Issue salience, issue ownership, and issue-based vote choice. *Electoral Studies*, *27*(3), 477–491. https://doi.org/https://doi.org/10.1016/j.electstud.2008.01.001

Bellettini, G., Ceroni, C. B., & Monfardini, C. (2017). *Ethnic Diversity and Political Participation: The Role of Individual Income* (CESifo Working Paper Series, Issue 6772). CESifo.

Bendor, J., Diermeier, D., & Ting, M. (2003). A Behavioral Model of Turnout. *The American Political Science Review*, *97*(2), 261–280. http://www.jstor.org/stable/3118208

Berganza, J. C. (2000). Politicians, voters and electoral processes: an overview. *Investigaciones Economicas*, *24*(3), 501–543. https://ideas.repec.org/a/iec/inveco/v24y2000i3p501-543.html

Berinsky, A. J. (1999). The Two Faces of Public Opinion. *American Journal of Political Science*, *43*(4), 1209–1230. https://doi.org/10.2307/2991824

Bhatti, Y., Dahlgaard, J. O., Hansen, J. H., & Hansen, K. M. (2019). Core and Peripheral Voters. *Political Studies*, *67*(2), 348–366.

Bhatti, Y., & Hansen, K. M. (2012). Retiring from Voting: Turnout among Senior Voters. *Journal of Elections, Public Opinion and Parties*, *22*(4), 479–500. https://doi.org/10.1080/17457289.2012.720577

Bierbrauer, F., Tsyvinski, A., & Werquin, N. (2018). Taxes and Turnout. *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, *111*, 1–61. https://www.jstor.org/stable/26939543

Bratsberg, B., Kotsadam, A., Lind, J. T., Mehlum, H., & Raaum, O. (2019). *Election Turnout Inequality - Insights from Administrative Registers* (CESifo Working Paper Series, Issue 7465). CESifo.

Brett, C., & Weymark, J. A. (2017). Voting over selfishly optimal nonlinear income tax schedules. *Games and Economic Behavior*, *101*, 172–188. https://doi.org/https://doi.org/10.1016/j.geb.2016.02.008

Cancela, J., & Geys, B. (2016). Explaining voter turnout: A meta-analysis of national and subnational elections. *Electoral Studies*, *42*, 264–275. https://doi.org/https://doi.org/10.1016/j.electstud.2016.03.005

Castillo, E., Gutierrez, J. M., & Hadi, A. S. (2012). *Expert Systems and Probabilistic Network Models* (1st ed.). Springer Publishing Company, Incorporated.

Comellas, J. M., & Torcal, M. (2023). Ideological identity, issue-based ideology and bipolar affective polarization in multiparty systems: The cases of Argentina, Chile, Italy, Portugal and Spain. *Electoral Studies*, *83*, 102615. https://doi.org/https://doi.org/10.1016/j.electstud.2023.102615

Damsbo-Svendsen, S., & Hansen, K. M. (2023). When the election rains out and how bad weather excludes marginal voters from turning out. *Electoral Studies*, *81*, 102573. https://doi.org/https://doi.org/10.1016/j.electstud.2022.102573

Dawes, C. T., Okbay, A., Oskarsson, S., & Rustichini, A. (2021). A polygenic score for educational attainment partially predicts voter turnout. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(50), e2022715118. https://doi.org/10.1073/pnas.2022715118

Denny, K., & Doyle, O. (2008). Political Interest, Cognitive Ability and Personality: Determinants of Voter Turnout in Britain. *British Journal of Political Science*, *38*(2), 291–310. http://www.jstor.org/stable/27568346

Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*, *65*(2), 135–150. http://www.jstor.org/stable/1827369

Enke, B. (2020). Moral Values and Voting. *Journal of Political Economy*, *128*(10), 3679–3729. https://doi.org/DOI: 10.1086/708857.

Fenton, N., & Neil, M. (2012). *Risk Assessment and Decision Analysis with Bayesian Networks* (1st ed.). CRC Press, Inc.

Filer, J. E., Kenny, L. W., & Morton, R. B. (1989). *Voting Laws, Educational Policies And Minority Turnout* (Papers, Issues 89–7). Florida - College of Business Administration.

Fraile, M., & Hernández, E. (2020). Determinants of Voting Behaviour. In *The Oxford Handbook of Spanish Politics*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198826934.013.22

Frank, R. W., & Martínez i Coma, F. (2023). Correlates of Voter Turnout. *Political Behavior*, *45*(2), 607–633. https://doi.org/10.1007/s11109-021-09720-y

Freeman, B. (2005). *Investigating Economic Inequality And Voter Turnout In The Industrialized Democracies* (2004-2019. Electronic Theses and Dissertations & University of Central Florida, Eds.). https://api.semanticscholar.org/CorpusID:54717004

Geiger, D., & Heckerman, D. (1994). Learning Gaussian Networks. In R. L. de Mantaras & D. Poole (Eds.), *Uncertainty Proceedings 1994* (pp. 235–243). Morgan Kaufmann. https://doi.org/https://doi.org/10.1016/B978-1-55860-332-5.50035-3

Gelman, A., Kenworthy, L., & Su, Y.-S. (2010). Income Inequality and Partisan Voting in the United States. *Social Science Quarterly*, *91*(5), 1203–1219. http://www.jstor.org/stable/42956457

Geys, B. (2006). Explaining voter turnout: A review of aggregate-level research. *Electoral Studies*, *25*(4), 637–663. https://doi.org/https://doi.org/10.1016/j.electstud.2005.09.002

Geys, B., Heggedal, T.-R., & Sørensen, R. J. (2022). Age and vote choice: Is there a conservative shift among older voters? *Electoral Studies*, *78*, 102485. https://doi.org/https://doi.org/10.1016/j.electstud.2022.102485

Giesing, Y., Schikora, F., & Shima, G. (2023). Emigration and Elections: The Role of Emigrants' Missing Votes. In *EconPol Forum* (Vol. 24, Issue 5). https://www.cesifo.org/de/publikationen/2023/aufsatz-zeitschrift/emigration-and-elections-role-emigrants-missing-votes

Gillies, D. (2002). Causality, Propensity, and Bayesian Networks. *Synthese*, *132*(1), 63–88. https://doi.org/10.1023/A:1019618817314

Goldberg, A. C., & Sciarini, P. (2023). Voter turnout in direct democracy: A joint analysis of individual, referendum and community factors. *European Journal of Political Research*, *62*(1), 25–46. https://doi.org/https://doi.org/10.1111/1475-6765.12493

Graafland, C. E., Gutiérrez, J. M., López, J. M., Paz'o, D., & Rodríguez, M. Á. (2019). The probabilistic backbone of data-driven complex networks: an example in climate. *Scientific Reports*, *10*. https://api.semanticscholar.org/CorpusID:208909951

Grafstein, R. (2005). The Impact of Employment Status on Voting Behavior. *The Journal of Politics*, *67*(3), 804–824. https://doi.org/10.1111/j.1468-2508.2005.00340.x

Green, J. C. (1985). Macroeconomics and Micropolitics: The Electoral Effects of Economic Issues. D. Roderick Kiewiet Political Economics. James E. Alt , K. Alec Chrystal. *The Journal of Politics*, *47*(3), 1006–1008. https://doi.org/10.2307/2131225

Hannan-Morrow, S., & Roden, M. (2014). Gender, Age and Generational Effects on Turnout in Australian Federal Elections. *The Australian Political Studies Association Annual Conference*. https://doi.org/10.2139/ssrn.2440487

Healy, A. J., Persson, M., & Snowberg, E. (2016). *Digging into the Pocketbook: Evidence on Economic Voting from Income Registry Data Matched to a Voter Survey* (CESifo Working Paper Series, Issue 6171). CESifo.

Heckerman, D. E. (1999). A Tutorial on Learning with Bayesian Networks. *Innovations in Bayesian Networks*. https://api.semanticscholar.org/CorpusID:1080631

Heckerman, D., Geiger, D., & Chickering, D. M. (1995). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning*, *20*(3), 197–243. https://doi.org/10.1023/A:1022623210503

Jensen, F. V, & Nielsen, T. D. (2007). *Bayesian Networks and Decision Graphs* (2nd ed.). Springer Publishing Company, Incorporated.

Kingman, J. F. C., & Feller, W. (1972). An Introduction to Probability Theory and its Applications. *Journal of the Royal Statistical Society. Series A (General)*, *135*(3). https://doi.org/10.2307/2344620

Konishi, S., & Kitagawa, G. (2008). Bayesian Information Criteria. In S. Konishi & G. Kitagawa (Eds.), *Information Criteria and Statistical Modeling* (pp. 211–237). Springer New York. https://doi.org/10.1007/978-0-387-71887-3_9

Lauritzen, S. L. (1992). Propagation of Probabilities, Means, and Variances in Mixed Graphical Association Models. *Journal of the American Statistical Association*, *87*(420), 1098–1108. https://doi.org/10.2307/2290647

Mahalanobis, P. (2018). Reprint of: Mahalanobis, P.C. (1936) "On the Generalised Distance in Statistics." *Sankhya A*, *80*(1), 1–7. https://doi.org/10.1007/s13171-019-00164-5

Mahsud, N. H. K., & Amin, H. (2020). *Theoretical Approaches to the Study of Voting Behaviour: A Comparative Analysis*. https://api.semanticscholar.org/CorpusID:225185473

Maldonado, M. F., & Hernandez, E. (2020). *Determinants of voting behaviour*. https://doi.org/https://doi.org/10.1093/oxfordhb/9780198826934.013.22

Margaritis, D., & others. (2003). *Learning Bayesian network model structure from data*. School of Computer Science, Carnegie Mellon University Pittsburgh, PA, USA.

Matsubayashi, T., & Sakaiya, S. (2021). Income inequality and income bias in voter turnout. *European Journal of Political Economy*, *66*, 101966. https://doi.org/https://doi.org/10.1016/j.ejpoleco.2020.101966

Mueller, D. C., & Stratmann, T. (2003). The economic effects of democratic participation. *Journal of Public Economics*, *87*(9), 2129–2155. https://doi.org/https://doi.org/10.1016/S0047-2727(02)00046-4

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*.

Pearl, J. (1988a). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc.

Pearl, J. (1988b). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (Cambridge University Press, Ed.; 2nd ed.).

Petitpas, A., Jaquet, J. M., & Sciarini, P. (2021). Does E-Voting matter for turnout, and to whom? *Electoral Studies*, *71*, 102245. https://doi.org/https://doi.org/10.1016/j.electstud.2020.102245

Piketty, T. (2021). Brahmin Left versus Merchant Right: Rising Inequality and the Changing Structure of Political Conflict in France, the United States, and the United Kingdom, 1948–2020. In A. Gethin, C. Martínez-Toledano, & T. Piketty (Eds.), *A Study of Fifty Democracies, 1948–2020* (pp. 85–135). Harvard University Press. https://doi.org/doi:10.4159/9780674269910-003

Python Core Team. (2019). *Python: A dynamic, open source programming language*. https://www.python.org/

Shachter, R. D., & Kenley, C. R. (1989). Gaussian Influence Diagrams. *Management Science*, *35*(5), 527–550. http://www.jstor.org/stable/2632102

Smets, K., & van Ham, C. (2013). The embarrassment of riches? A meta-analysis of individual-level research on voter turnout. *Electoral Studies*, *32*(2), 344–359. https://doi.org/https://doi.org/10.1016/j.electstud.2012.12.006

Sondheimer, R. M., & Green, D. P. (2010). Using Experiments to Estimate the Effects of Education on Voter Turnout. *American Journal of Political Science*, *54*(1), 174–189. http://www.jstor.org/stable/20647978

Stockemer, D. (2017). What Affects Voter Turnout? A Review Article/Meta-Analysis of Aggregate Research. *Government and Opposition*, *52*(4), 698–722. https://doi.org/DOI: 10.1017/gov.2016.30

Stocké, V., & Stark, T. (2007). Political involvement and memory failure as interdependent determinants of vote overreporting. *Applied Cognitive Psychology*, *21*(2), 239–257. https://doi.org/https://doi.org/10.1002/acp.1339

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(2), 111–133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x

Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies*, *69*, 102199. https://doi.org/https://doi.org/10.1016/j.electstud.2020.102199

Wiese, R., & Jong-A-Pin, R. (2017). Expressive voting and political ideology in a laboratory democracy. *European Journal of Political Economy*, *50*, 54–74. https://doi.org/https://doi.org/10.1016/j.ejpoleco.2016.12.001

Wlezien, C., Franklin, M., & Twiggs, D. (1997). Economic Perceptions and Vote Choice: Disentangling the Endogeneity. *Political Behavior*, *19*(1), 7–17. http://www.jstor.org/stable/586547

Zeglovits, E., & Aichholzer, J. (2014). Are People More Inclined to Vote at 16 than at 18? Evidence for the First-Time Voting Boost Among 16- to 25-Year-Olds in Austria. *Journal of Elections, Public Opinion and Parties*, *24*, 351–361. https://api.semanticscholar.org/CorpusID:44707975