Final Data Essay - Who Votes For Whom?

Statistical Models in Political Analysis (Fall 2019)

Jingwen Zhang (17-746-487)

3/1/2020

1 Introduction

The aim of this essay is to empirically test hypotheses regarding the relationship between vote choice and issues like income, attitudes toward social welfare, political positions, evaluation of government performance, etc. Hypotheses are formulated based on different theories, cleavage theory suggests that vote choice is determined by structural and social factors (Lipset and Rokkan, 1976), while Downs (1957) argued for an understanding of a voter as a rational actor, and Ferejohn (1986) and Fiorina (1981) think voters can evaluate politicians retrospectively.

Hypotheses based on these theories were tested in this paper with the help of data of the British Election Study of 1987 across Scotland, Wales, and England including 6000 individuals. The findings show all three theories explain people's vote choice, but the effects are different and the rational actor theory is the most convincing.

The structure of the essay is as follows, in Part 2 different hypotheses based on theories are formulated, and the description of data is given in Part 3 and details of models are in Part 4. The results of models are discussed in Part 5, and the final part includes conclusions on findings and potential improvement for further research.

2 Hypotheses

Cleavage theory suggests that vote choice is determined by structural and social factors (Lipset and Rokkan, 1976). Social cleavages drive voters to make their political choices, they will vote for parties that best represent their positions among cleavages. Since the UK has a two-party system, the Conservative and the Labour, it represents one basic pair of cleavages, owners vs. workers. However, the distinction between workers and owners nowadays are no long clear-cutting as before, instead, high- vs. low- income individuals should be considered. Hereby I assume that *voters with higher income are more likely to vote for the Conservative Party* (H1).

In comparison, **rational choice theory** argued that a voter is a rational actor (Downs, 1957), instead of constrained by structural or social factors, an individual is free to vote for the party that best represents her or his political interests or positions since an individual's voting behaviors are motivated by self-interest and utility maximization. Therefore, the second hypothesis is that *voters are more likely to vote for parties that represent their political positions better* (H2).

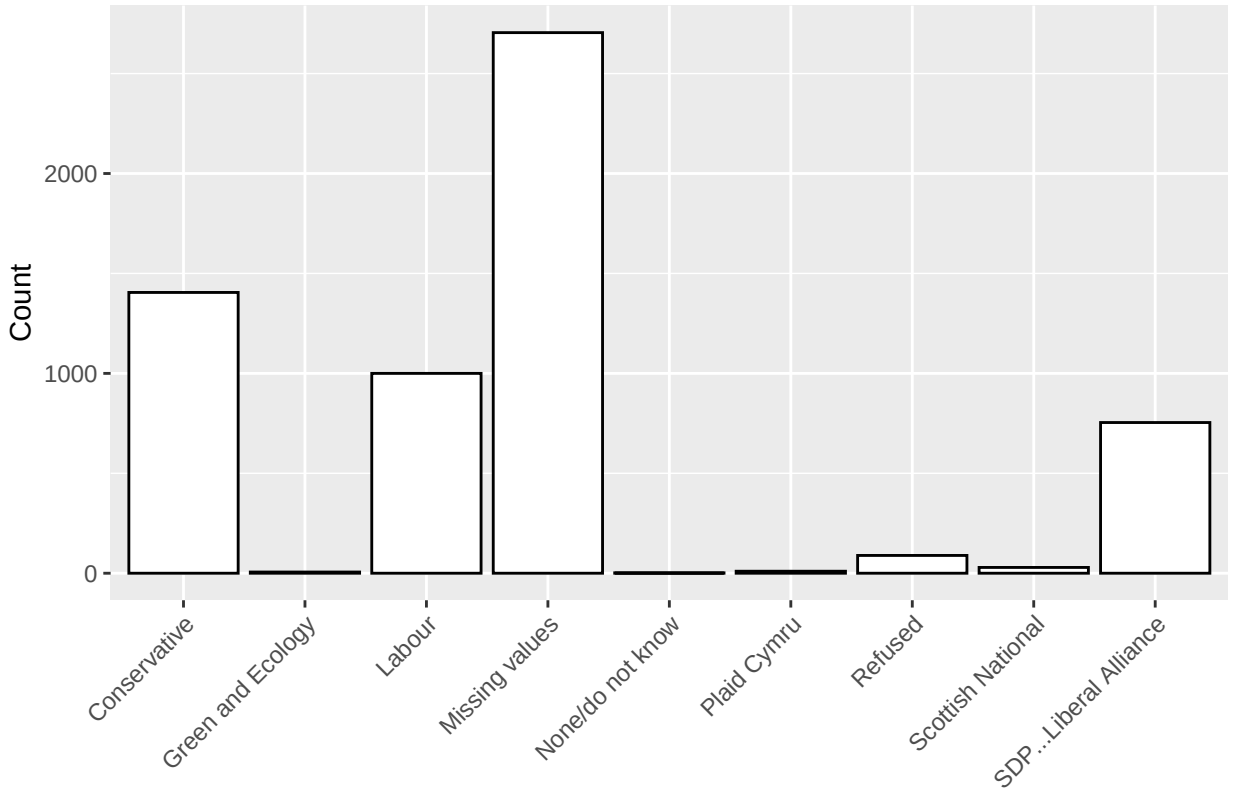
According to Ferejohn (1986) and Fiorina (1981), rational choice theory can not hold since it is even difficult for the electorate to position itself. Instead, the **retrospective voting theory** suggests that voters make choice by evaluating the parties retrospectively. The third hypothesis can be derived that *if the government performed better when the party is in power, voters are more likely to vote for the party in the next election* (H3).

3 Data

The whole dataset contains information of a total of 6000 individuals surveyed by the British Election Study of 1987 across Scotland, Wales, and England. The unit of analysis is the vote choices of parties of adults on the electoral register in 1987 in the UK. This essay will work on

a subset of the original dataset. As shown in Figure 1, the major votes lie in the Conservative Party (the incumbent party during the election period), followed by the Labour Party, and the SDP-Liberal alliance held a considerable vote share. Since the Conservative won 376 seats, the Labour 229 seats while the SDP-Liberal alliance only 22 seats in the election, the analysis will focus on these two parties and a binary variable *conservative* is generated.

Figure 1: Distribution of Vote Choice



Three indices have been generated in accordance with 3 hypotheses. The index *opinions on social welfare issues* is calculated by adding up respondents' scores on 7 survey questions relevant to redistribution, welfare state, welfare benefits, etc, each was measured by a five-point Likert scale. The lower the value of the index, the more likely the respondent holds a left political attitude since the welfare state is one of the policies left-wing parties advocate. The index *respondents' political positions* is the average of the respondent's political positions on 7 salient topics concerning defense, unemployment, inflation, taxation, redistribution, law,

welfare, etc, measured by an interval scale from 1 to 11. The higher the value of the index, the more likely the respondent has the right political position. The index *evaluations on government performance* is calculated the same as the first index, it is based on individuals' evaluations on government performance since the last election on, including aspects like prices, unemployment, taxes, education, health, social services, etc, measured by five-point Likert scales from increase a lot (=1) to fallen a lot (= 5). The higher the score of this index, the more positive the evaluation of government performance is.

Table 1: Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Age	2,405	46.98	18.08	17	32	61	99
Education	2,405	16.91	8.52	0	15	17	99
Welfare	2,405	20.59	5.41	7	17	24	35
Government	2,405	15.82	3.42	7	13	18	28
Female	2,405	0.52	0.50	0	0	1	1
Conservative	2,405	0.58	0.49	0	0	1	1
Position	2,405	5.33	1.82	1	4	6.6	11
income	2,405	3.57	1.54	1	2	4	7

Table 1 shows the descriptive statistics of these variables. The number of observations is smaller because the given subset dataset itself has 2174 missing values for all main variables, moreover, we only consider vote choice for the 2 main parties in the research. If only refer to numbers in Table 1 retrieved so far, they indicate that respondents are weakly unsupportive of the welfare state, somewhat satisfied with the government's performance after June 1983 general election and their political positions are slight to the left.

The regression models include socio-demographic controls: such as age, gender, education, and income. For clearer interpretation, the variable income has been re-classed into 7 classes instead of 12. The means and standard deviations of age and education show us the representativeness of the sample included in this study.

4 Method

As the depended variable is dichotomous, all regression models are estimated with binomial General Linear Model (GLM) estimators based on logistic distribution. The first hypothesis (H1) is tested by regressing the index *social welfare issues* and the variable income as independent variables on the dependent variable *conservative*. The index of *opinions of social welfare* is included because normally voters who are not supportive of the welfare state prefer the Conservative. Since the higher-income population is less supportive of social welfare, attitudes towards the welfare state can be considered as a symbol of the cleavages between higher-income vs. lower-income populations.

We test the rational choice theory (H2) by regressing the index *political position* as the independent variable on the dependent variable, and the retrospective voting theory (H3) is tested by regressing the index *evaluations on government performance* on the dependent variable. The regression models include several control variables as shown in the last 3 rows in Table 1. They are included because they might influence the vote choice and the interpretation since some reported that elder voters prefer the Conservative, younger women or higher qualified voters are inclined to the Labour (Curtis 2017).

5 Results

As shown in the Table 2, the first model is displayed in the first column, the variable income has a positive relationship with the dependent variable (H1), its effect increases clearly when the variable income increases gradually, especially for the group with the highest income (more than 30000 GBP). The effect is significant at 0.01-level except for the income of households lower than 6000 GBP. The coefficient of the index *opinions on social welfare issues* is significant at 0.01-level, this indicates that one unit increase of the index, which means being less supportive of the welfare state, the probability of voting for the Conservative

will increase by 0.354 units. Among all sociodemographic controls, except that the coefficient of female is negative and not statistically significant, the coefficients of age and education indicate a positive relationship between these controls and the vote choice, especially age is statistically distinguishable from 0 at a 99% significance level, this indicate elder people prefer to vote for the Conservative, although the size of both effects are relatively small. In all, we can say the model supports the first hypothesis.

The relationship of the index *political position* and the dependent variable is positive and statistically significant at a 0.01-level (H2), this indicates that one unit of increase of the index, which means that average political position is more to the right, the probability voting for the Conservative will increase 1.319 units. Similar to the first model, the standard error of the control female is sizeable and not significant, coefficients for education and age are statistically significant but still small, which indicates in the second model, elder people or higher educated people are slightly inclined to the Conservative.

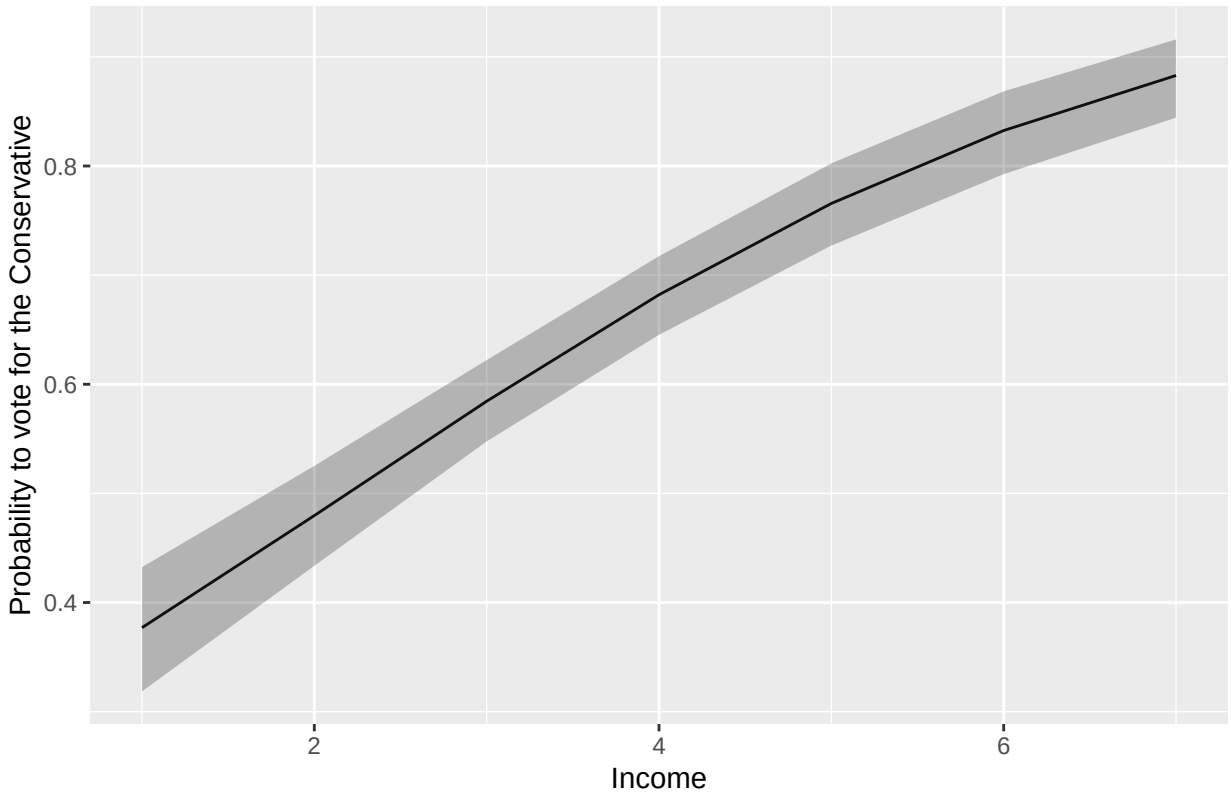
The last column shows that in the third model the relationship between the index *government performance* and the dependent variable is positive and significant at a 0.01-level (H3). One unit of increase of the index means that voters are more satisfied with the government performance since 1983 when the Conservative party was in power, so that the probability of voting for it again in 1987 will increase by 0.337 units. This is relatively small compared to the coefficient of the index about the political position in model 2 and of income in model 1. Among controls, except education, age and female are positive and significant at 0.01-level. Only in model 3, females are more likely to vote for the Conservative Party, while age is the most significant control variable in all 3 models, and education does not really have a considerable influence on the interpretation.

Table 2: GLM(binomial)- Estimation of the models: Hypothesis 1,2 and 3

	Dependent variable:		
	Conservative		
	(1)	(2)	(3)
Welfare	0.354***		
	(0.016)		
Income2	0.139		
	(0.212)		
Income3	0.688***		
	(0.225)		
Income4	1.043***		
	(0.211)		
Income5	1.626***		
	(0.260)		
Income6	1.837***		
	(0.287)		
Income7	2.617***		
	(0.460)		
Position		1.319***	
		(0.053)	
Government			0.337***
			(0.017)
Age	0.012***	0.008**	0.017***
	(0.003)	(0.003)	(0.003)
Female	−0.006	0.087	0.252***
	(0.112)	(0.116)	(0.095)
Education	0.014*	0.014*	0.008
	(0.008)	(0.008)	(0.006)
Constant	−8.486***	−7.071***	−5.948***
	(0.444)	(0.352)	(0.335)
Observations	2,405	2,405	2,405
Log Likelihood	−1,025.066	−937.802	−1,343.769
Akaike Inf. Crit.	2,072.132	1,885.604	2,697.537
Note:			*p<0.1; **p<0.05; ***p<0.01

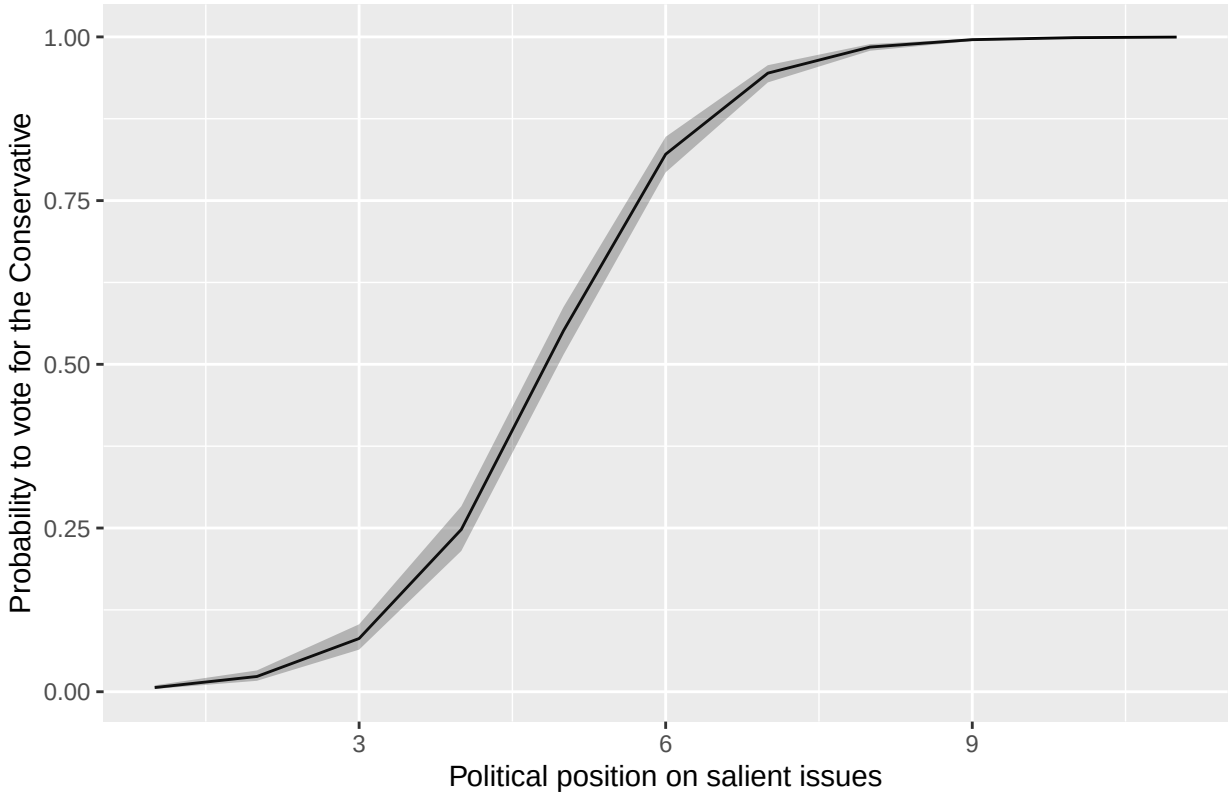
Since all 3 hypotheses are supported by their models, the quantities of interests of all models are illustrated by expected values generated from simulations of a size of 1000 females, holding variables age and education at their medians. Figure 2 presents the mean expectations with a 95% confidence interval.

Figure 2: Model 1



In Figure 2, the index of welfare is held constant at its median since H1 focuses on income, the plot shows how the probability of voting for the Conservative increases when the income increases. Figure 3 and Figure 4 visualize separately how the probability of voting for the Conservative increases when the value of political position index increases, and the value of government performance index increases. By comparison, the regression for H2 has a smaller CI range and variation, this corresponds to its bigger coefficient compared to those of other models. Although all 3 hypotheses have been supported, H2 is more convincing compared to the other models.

Figure 3: Model 2



6 Conclusions

The aim of this essay is to empirically test 3 hypotheses regarding the relationship between vote choice and aspects like income, social welfare, political position, government performance, etc, based on dataset contains information of a total of 6000 individuals surveyed by the British Election Study of 1987. The outcome finds out the relationship is actually statistically significant in all hypotheses, among which H2 has the highest significance.

The shortcomings of this analysis lie in data as we have considerable missing values, although we try to replace some of them with the median or mean of the variables. In further research, it is important to find out whether the reasons behind none-response have a relationship with our dependent variable. If not, the results of our research will be still valid.

Figure 4: Model 3