Discriminating between Issue Voting Rules in Multiparty Elections Using Finite Mixture Modeling

Kirill Zhirkov

University of Michigan

kzhirkov@umich.edu

[Draft. Please don't cite or circulate]

Abstract

How do citizens of democratic polities translate their policy preferences into voting choices? Proximity and directional theories of issue voting offer different answers to this question that have strong implications for parties' strategies. Controlled scenarios in imaginary two-candidate contests recently gained popularity as a method to identify proximity and directional voters. However, they are not always applicable in comparative research where scholars often have to study multiparty elections with observational data. In the present paper, I propose a method of probabilistically discriminating between issue voting rules based on finite mixture modeling. Using a paradigmatic case in the proximity–directional debate, I demonstrate that the mixture model describes observed voting choices better than the alternatives. I also show how finite mixture modeling can be used to study individual-level characteristics of proximity and directional voters. The proposed method can be applied to study issue voting rules from the comparative perspective.

Keywords: finite mixture modeling, issue voting, proximity theory, directional theory

Translation of citizens' policy positions into voting choices is the central question in the study of political representation. It has profound implications for both normative democratic theory and empirical analysis of strategies employed by political parties. The two alternative theories of issue voting offer different mathematical models linking policy positions to electoral behavior. The proximity model is based on the assumption that a voter chooses the party that is closest in terms of policy positions (Downs 1957). The directional model, in turn, assumes that citizens have preferences with respect to issue sides and vote for the most extreme of acceptable parties on the same side (Rabinowitz and Macdonald 1989). Therefore, the dominant strategies for parties should be moderation under proximity voting and polarization under directional voting.

A heated debate between supporters of the two models was largely concluded by acknowledging that none of them could perfectly explain observable behaviors of voters or strategies of parties (Lewis and King 2000). This led some authors to propose general models of issue voting that combined both directional and proximity elements, such as the unified model (Merrill and Grofman 1999), or the policy leadership model (Iversen 1994). An important feature of such models is that they assume presence of a relatively sophisticated but uniform voting rule within the population.

Recently, a survey-experimental method to identify individual respondents' issue voting rules has been proposed (Tomz and Houweling 2008). The method employs controlled scenarios: each respondent is offered two options designed so that a choice unambiguously reveals the underlying voting rule. Results of the experiment carried out on a U.S. probability sample demonstrated that, even though proximity voting in either simple or discounting form was prevalent, directional voters constituted a non-trivial share of the electorate. These findings are intuitively appealing. They suggest that individual voters use relatively simple voging rules that, nevertheless, vary within the population.

Presence of different voting rules among voters, in turn, leads to aggregate outcomes that deviate from those predicted by the pure proximity model or the pure directional model.

Discrimination between issue voting rules on the individual level has important benefits for the study of electoral behavior. Most importantly, it can be used to identify individual-level characteristics associated with different voting rules. For instance, the lower educated and strong partisans are more likely to use the directional voting rule (Tomz and Houweling 2008). At the same time, the experimental method based on controlled scenarios is not universally applicable. In multiparty elections that are prevalent in modern democracies the numbers of choices are too large to design feasible scenarios that would unambigously discriminate between different voting rules. Additionally, the main source of data on multiparty elections around the world is national and international electoral surveys. The only way to discriminate between voting rules with such (observational) data is to rely on statistical modeling.

In this paper, I propose a method that probabilistically discriminates between issue voting rules in multiparty elections based on finite mixture modeling (Imai and Tingley 2012). I apply the proposed method to the Norwegian election survey data that have been the main testing ground for the proximity and directional voting models. I demonstrate that the mixture model is superior to alternatives in terms of fit to the data. I also show how finite mixture modeling can be used to explore individual-level characteristics of proximity and directional voters using education as an example. In conclusion, I discuss how the proposed method can be used in comparative research on electoral behavior.

Model Setup

Suppose, there are survey data describing electoral choices of voters indexed $i = 1, \ldots, I$. Political parties contesting in the election under consideration are indexed $j = 1, \ldots, J$. Voters and parties are positioned on policy issues indexed $k = 1, \ldots, K$. Let v_{ik} be voter i

self-placement on the issue k. Let c_{ijk} be voter i evaluation of party (candidate) j placement on the issue k. The neutral point on all issue dimensions is defined as zero.

Such data can be used to define and estimate the models of issue voting. The utility of voter i from party j being elected under the proximity model is defined as:

$$u_{ij}^p = \alpha_j^p + \beta^p \frac{1}{K} \sum_{k=1}^{K} |v_{ik} - c_{ijk}| \,, \tag{1}$$

where α_j is the non-issue utility from voting for party j and β^p is the proximity coefficient.[1] Under the directional model, the utility of voter i from party j being elected is:

$$u_{ij}^d = \alpha_j^d + \beta^d \frac{1}{K} \sum_{k=1}^{K} (v_{ik} \times c_{ijk}) \,, \tag{2}$$

where β^d is the directional coefficient. The utility function that assumes a combination of directional and proximity logics in individual voting decisions (general or unified model) takes the following form:

$$u_{ij}^g = \alpha_j^g + \beta_p^g \frac{1}{K} \sum_{k=1}^{K} |v_{ik} - c_{ijk}| + \beta_d^g \frac{1}{K} \sum_{k=1}^{K} (v_{ik} \times c_{ijk}) \,. \tag{3}$$

The utilities defined above can be translated into choice models using the logistic link function. Let y_i be a variable denoting voter i choice in election under consideration such that $y_i = j$ means that the respondent i voted for the party j. Assume that, besides fixed components presented above, the utility functions also include random components that follow the type I generalized extreme value distribution (also known as the Gumbel distribution). Then, the probability that respondent i votes for party j becomes:

$$\Pr(y_i = j) = \frac{\exp(u_{ij})}{\sum_j \exp(u_{ij})}, \tag{4}$$

where the formula for u_{ij} can correspond to Equation 1, Equation 2, or Equation 3, depending on the voting rule.

[1] Following Berinsky and Lewis (2007), I explicitly estimated power of distance function under the proximity rule and found that it was not significantly different from one (i.e., absolute distance specification). See Online Appendix for details.

The mixture model, under which a voter can use either proximity or directional logic in their electoral choices, is more complicated. Let π_i be the probability that voter i uses the directional voting rule. Then, it is possible to derive an expression for the total probability that respondent i votes for party j:

$$\Pr(y_i = j) = (\pi_i) \frac{\exp(u_{ij}^d)}{\sum_j \exp(u_{ij}^d)} + (1 - \pi_i) \frac{\exp(u_{ij}^p)}{\sum_j \exp(u_{ij}^p)}, \tag{5}$$

where utility calculations under the proximity and directional rules correspond to Equation 1 and Equation 2 respectively.

An important advantage of the mixture model is a possibility to predict probability of the voting rule applied. Assume logistic functional form linking probability of voter i using the directional voting rule to the linear predictor w_i:

$$\pi_i = \frac{1}{1 + \exp(-w_i)}. \tag{6}$$

The linear predictor, in turn, uses individual-level covariates indexed $h \in 1, \ldots, H$. Let x_{ih} be voter i's value of covariate h. Then:

$$w_i = \gamma + \sum_{h=1}^{H} \delta_h x_{ih}, \tag{7}$$

where γ is the constant and δ_h is the covariate-specific coefficient. Conceptually, the mixture model can be seen as a compound process in which an individual first chooses the voting rule and then applies it to cast a ballot for a party.

Model Comparison

It is necessary to note that the combined or unified model of issue voting on the one hand and the mixture model on the other hand cannot be considered nested with respect to each other. Therefore, the standard likelihood-ratio test cannot be used to compare them. To overcome this problem, I employ a method of comparison between non-nested discrete choice models proposed by Horowitz (1983). He defined the adjusted likelihood ratio index criterion as:

$$\bar{\rho}^2 = 1 - \frac{L_m - k/2}{L_0}. \tag{8}$$

where L_m is the model log-likelihood, L_0 is the null log-likelihood, and k is the number of model parameters. Horowitz's criterion effectively penalizes over-parametrization when equivalent fit can be achieved by a more parsimonious model.

There is a formal test for comparison of discrete choice models based on Horowitz's likelihood ratio indices. Let $\bar{\rho}_l^2$ be the lower adjusted likelihood ratio index and $\bar{\rho}_h^2$ the higher adjusted likelihood ratio index. Then the test statistic z that follows the standard normal distribution is defined as:

$$z = \sqrt{-2L_0(\bar{\rho}_h^2 - \bar{\rho}_l^2)}. \tag{9}$$

Corresponding p value for the null hypothesis that the model with the higher adjusted likelihood ratio index does not fit the data better can be calculated as follows:

$$p = 1 - \Phi(z), \tag{10}$$

where Φ is the standard normal cumulative distribution function.

Data and Estimation

In the 1990s, the 1989 Norwegian Election Study became the major testing ground for the competing theories of issue voting (Macdonald, Listhaug, and Rabinowitz 1991; Westholm 1997). Therefore, it was a natural choice as the source of data for my analysis. The 1989 parliamentary election in Norway was based on party-list proportional representation in nineteen multi-member constituencies corresponding to country's administrative regions (counties). Seven major political parties, from the radical left to the radical right, had credible chances to win seats in the parliament (see Online Appendix for the full list).

As part of the election survey, respondents were asked to position themselves and the parties on five political issues: government support for agriculture, environmental protection, immigration policy, privatization of healthcare, and regulation of alcohol trade (see Online Appendix for exact question wordings). All issue scales ranged from 1 to 10, so that the centered scales took values from -4.5 to 4.5.

As it often happens in survey studies, some voters did not take positions on all five issues. Excluding all such voters from the analysis would decrease the sample size and statistical power. To avoid this problem, I defined a number of issue dimensions that voter i cared about, denoted K_i. Then, I calculated issue-related components of the utility functions using the available issue positions (with appropriate scaling), instead of omitting all respondents with $K_i < 5$. In doing so, I assumed that, if a voter did not take positions on certain issues, these issues were irrelevant for that person's voting choices.

Similarly, there were voters in the data who did not place some of the parties on the five issues in question. Again, instead of omitting such people from the analysis, I estimated their utility functions using existing choice options (parties) if at least two alternatives were available.

All models reported and discussed in the paper were estimated using PythonBiogeme (Bierlaire 2016), a free open source software designed for the maximum likelihood estimation of parametric and semi-parametric discrete choice models.

Results

Comparison of the four models—pure proximity, pure directional, general/unified, and the mixture model—in terms of fit is presented in Table 1 (for the full set of parameter estimates, see Online Appendix). The mixture model clearly outperforms all alternatives according to both log-likelihood and Horowitz's criterion. Tests based on the $\bar{\rho}^2$ statistic support this conclusion as, in all cases, they allow rejecting the null hypothesis about models' equally good fit.

[Table 1 about here]

Table 1: Model comparison

	Pure proximity	Pure directional	General/unified	Mixture
Proximity coefficient	-1.08^{***}		-1.06^{***}	-1.66^{***}
	(0.04)		(0.06)	(0.15)
Directional coefficient		1.16^{***}	0.02	0.34^{***}
		(0.11)	(0.07)	(0.11)
Log-likelihood	−2153.6	−2332.0	−2153.5	−2123.8
Parameters	7	7	8	9
Horowitz's $\bar{\rho}^2$	0.352	0.299	0.352	0.360
z statistic	7.59	20.36	7.65	
p value[a]	$< .001$	$< .001$	$< .001$	

$N = 1736$. $L_0 = -3334.5$. Standard errors in parentheses
[a] H_0: mixture model does not fit better
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

It is also necessary to emphasize that, beyond demonstrating superior fit to the data, the mixture model also leads to substantively different conclusions compared to the general/unified model. According to the general model, the directional term is not significantly different from zero ($z = 0.33$, $p = .371$). This conclusion is strengthened by a likelihood-ratio test: the general model fits no better than the pure proximity one ($\chi^2_1 = 0.11$, $p = .745$). According to this result, one would not have to account for directional logic in issue voting.

The mixture model, however, offers a starkly different picture. First, both directional and proximity coefficients are significant and in the expected direction. Second, value of the constant term suggests that, on average, approximately 21% of respondents used the directional voting rule; see column (1) in Table 2. In other words, directional voting was present in the 1989 Norwegian parliamentary election, it was used by a non-trivial share of the population, and accounting for it significantly improves the prediction of voters' behavior.

[Table 2 about here]

Table 2: Predicting voting rule

	(1)	(2)	(3)	(4)
Education				
Binary		-1.09^{***}		
		(0.35)		
Interval			-0.24^{***}	
			(0.06)	
Years				-0.23^{***}
				(0.06)
Constant	-1.34^{***}	-0.93^{**}	-1.15^{***}	-1.18^{***}
	(0.33)	(0.31)	(0.34)	(0.35)
Log-likelihood	−2120.6	−2096.7	−2090.8	−2085.6

$N = 1718$. Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As mentioned previously, the experimental study by Tomz and Houweling (2008) found that the lower educated were more likely to vote using the directional logic. Attempting to replicate this finding, I re-estimated the mixture model with education as a predictor of the directional vs. proximity voting rule being used. Results are presented in columns (2), (3) and (4) of Table 2 for three different measures of education: binary (college or higher vs. less than college), interval (8-point from elementary to complete tertiary), and total years of full-time education. Non-binary measures of education were centered around median to obtain meaningful constant terms.

According to my results, education was indeed significantly and robustly related to issue voting rules: highly educated respondents were less likely to use the directional rule (vis-a-vis the proximity one). The effects were also substantial in terms of magnitude. For instance, according to the model using the binary measure of education predicted probability of using the directional voting rule was approximately 28% for a person without college education and only 12% for a college graduate (see Online Appendix for the full sets of estimates).

Conclusion

In this paper, I proposed a method of probabilistically discriminating between issue voting rules in multiparty elections based on finite mixture modeling. I applied this method to the 1989 Norwegian parliamentary election, a classic case in directional vs. proximity voting debate. I demonstrated that the mixture model fitted the data significantly better than the alternatives—including the general/unified model of issue voting. My analysis also highlighted a major benefit of finite mixture modeling, namely the ability to predict application of voting rules using individual-level covariates. Specifically, I showed that lower educated voters were more likely to cast ballots using the directional voting rule.

The key benefit of the proposed method is its applicability to observational data from national and international electoral surveys. Therefore, it can be used in comparative studies on relative prevalence of directional vs. proximity voting in a country and its relationships to electoral institutions and party systems. Additionally, it might be interesting to see whether the found relationship between education and issue voting rules is replicated across countries. Scholars can also look for other individual-level characteristics that distinguish directional voters from proximity voters.

Finite mixture modeling can be also applied to a number of other questions related to issue voting. For instance, it can be used to identify groups of voters who choose parties/candidates on the basis of issues and those who cast their ballots for some other reasons (non-issue voters). Similarly, finite mixture modeling can be used to contrast individuals who care about different policy areas, such as "economic" vs. "cultural" voters. Altogether, finite mixture modeling is a powerful and flexible analytical tool that deserves more attention from scholars of electoral behavior, especially in comparative perspective.

References

Berinsky, Adam J., and Jeffrey B. Lewis. 2007. "An Estimate of Risk Aversion in the U.S. Electorate." *Quarterly Journal of Political Science* 2 (2): 139–54.

Bierlaire, Michel. 2016. "PythonBiogeme: A Short Introduction." Technical report. http://transp-or.epfl.ch/pythonbiogeme/documentation/pythonfirstmodel/pythonfirstmodel.pdf.

Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper.

Horowitz, Joel L. 1983. "Statistical Comparison of Non-Nested Probabilistic Discrete Choice Models." *Transportation Science* 17 (3): 319–50.

Imai, Kosuke, and Dustin Tingley. 2012. "A Statistical Method for Empirical Testing of Competing Theories." *American Journal of Political Science* 56 (1): 218–36.

Iversen, Torben. 1994. "Political Leadership and Representation in West European Democracies: A Test of Three Models of Voting." *American Journal of Political Science* 38 (1): 45–74.

Lewis, Jeffrey B., and Gary King. 2000. "No Evidence on Directional vs. Proximity Voting." *Political Analysis* 8 (1): 21–33.

Macdonald, Stuart Elaine, Ola Listhaug, and George Rabinowitz. 1991. "Issues and Party Support in Multiparty Systems." *American Political Science Review* 85 (4): 1107–31.

Merrill, Samuel, III, and Bernard Grofman. 1999. *A Unified Theory of Voting: Directional and Proximity Spatial Models.* Cambridge and New York: MIT Press.

Rabinowitz, George, and Stuart Elaine Macdonald. 1989. "A Directional Theory of Issue Voting." *American Political Science Review* 83 (1): 93–121.

Tomz, Michael, and Robert P. Van Houweling. 2008. "Candidate Positioning and Voter Choice." *American Political Science Review* 102 (3): 303–18.

Westholm, Anders. 1997. "Distance versus Direction: The Illusory Defeat of the Proximity Theory of Electoral Choice." *American Political Science Review* 91 (4): 865–83.