Consideration Sets and Finite Mixtures: A New Approach to the Analysis of Strategic Voting

Martin Elff

July 2019

Abstract

A new approach to estimating the amount of strategic voting is proposed, which improves over prior model-based approaches in two ways: (1) Being based on a finite-mixture model, it avoids bias created by neglecting genuinely non-strategic voters. (2) By conceiving strategic voting as choosing from a restricted set of alternatives, it overcomes the requirement of untested assumptions about the influence of strategic incentives. It also avoids some paradoxical implications, such as the reversal of preferences. When applied to the UK House of Commons election of 2010, the new method agrees with the stated-reasons method in terms of the amount and the patterns of strategic voting, which underlines the convergent validity of both methods.

1 Introduction

Duverger's Law, which states that electoral systems with single-member districts and plurality rule tend to favour two-party systems, is perhaps one of the most well-known regularity of political science, (Duverger 1965; Cox 1997) even if it may not be a strict one. According to Duverger, there are two kinds of effects of electoral systems that lead to this regularity, the "mechanical effect" of electoral systems, the way in which votes are translated into seats, and the "psychological effect", a tendency of voters to avoid wasting their votes for hopeless candidates or parties. While the "mechanical effect" arises from the way the electoral system translates votes into seats, the "psychological effect" arises from a variation in voters decision making that nowadays would referred to as *strategic voting* or *tactical voting*: It arises because voters avoid "wasting their votes" for candidates or parties that are unlikely to win a seat in the election. This is an instance of strategic voting, in so far as these voters deviate in their choice from their preferred alternatives (Fisher 2004). While it should be possible to assess the "mechanical effect" of an electoral system by way of simulation, because the rules of translation of votes into seats are usually well-defined and known, the assessment of the "psychological effect", i.e. how it provides incentives for strategic voting and how much strategic voting occurs at all, is considerably more difficult.

The main challenge in assessing the distribution of strategic voting is that it is defined as voting behaviour that deviates from the voter's preferences (Fisher 2004). If actual choices can deviate from one's preferences, it is no longer possible to safely assume that choices are "revealed preferences" – a situation one could call the "dilemma of non-revealed preferences". This in turn means that voters' preferences cannot be inferred from their choices alone. Instead, one needs other ways to uncover voters' preferences. There are at least three different ways to do so in the empirical literature. The first is to categorise individuals as strategic voters if they explicitly state strategic considerations as the motive of their choice. For example, respondents of British Election Study surveys are categorised as strategic voters if they explicitly state to have "voted tactically" or that the party "they really prefer" did not have a chance to win the seat in the relevant constituency (Clarke et al. 2010). A second approach, known as the *direct* approach, (Blais et al. 2005) is based on the comparison of individuals' (self-reported) voting choices with their (self-reported) preference order over the alternatives (i.e. parties or candidates). Voters are categorized as strategic if their choices not only deviate from their preferences, but do so in such a way so that the choice increases the chance of affecting the electoral outcome.[1] A third approach uses statistical modelling to reconstruct voters' preferences based on a set of predictor variables, such as social class, religion, and issue positions and to distinguish actual votes from these preferences. The most widely known example of this *model-based* approach was introduced by Alvarez and Nagler (2000). Alvarez and Nagler start with a model that combines predictors of preferences with predictors of strategic choices. After fitting this model to observed voting data, voters are considered as strategic whenever the predictions with strategic choice predictors set to zero are different from predictions based on the full model.[2]

A model-based technique will be inevitable if no data is available that allows to identify strategic voters based on the reasons that voters, when asked, give for the vote. However, the model based-technique put forward by Alvarez and Nagler (2000) and Alvarez et al. (2006) has a few drawbacks: First, the model on which the Alvarez-Nagler method is built presupposes the same decision rules for all voters, thereby ruling out the possibility that voters are genuinely non-strategic and cast a sincere vote even if strategic incentives suggest otherwise. Secondly, the model rests on particular assumptions about the variables relevant for strategic incentives and the functional form of their influence. This makes it hard, if not impossible

1 Blais and Nadeau (1996) however do not use voters' explicit statements about preferences, but reconstruct preference orders on the basis of a logistic regression of voting choices on a range of predictors.

2 Artabe and Gardeazabal (2014) propose a method of inferring strategic voting that does not easily fit in either of these categories as it combines ideas from all of them.

to test whether these variables adequately explain strategic voting. Thirdly, the additive construction of the utility function may lead to paradoxical preference reversals, as explained later in this paper. Fourthly, the utility function and the discrete choice model used by Alvarez and Nagler (2000) and Alvarez et al. (2006) restrict the application of their method to situation where the choice set is the same for all voters, in contrast to the situation in the United Kingdom, where e.g. the SNP runs candidates in Scotland but not in England.

The present paper proposes a new model-based technique for the estimation of strategic voting that addresses these problems. It rests on two core ideas: The first idea is that each voter makes his or her choice in one of two modes, one that leads to a strategic vote when the appropriate incentives are present and one that leads to a sincere vote independent of the presence of strategic incentives. One could call the two modes a "sophisticated" and an "expressive" mode. The second idea is the strategic mode is characterised simply by a decision pattern in which non-viable alternatives are disregarded, i.e. eliminated from the menu of alternatives considered to be chosen. This allows to separate the estimation of the amount of strategic voting from particular assumptions about the impact of strategic incentives.

While Alvarez and Nagler (2000) claim that a model-based technique is inherently superior to a technique based on voters' stated reasons, the present paper does not include such a claim. Rather it demonstrates that the new method can be used to validate results obtained with another well established method, the stated-reasons technique (Evans 2002).

An archetypal instance of this conceptualisation of strategic voting is a supporter of the Labour Party living in the electoral district of Colchester during the 2010 general election of the UK House of Commons. He or she may judge that only the Liberal Democrat and the Conservative candidate have a chance to get elected, but not the candidate of his or her most preferred Labour Party. If this voter chooses in expressive mode, he or she will cast a sincere vote for the Labour Party, the hopelessness of the Labour candidate notwithstanding. If the voter chooses in sophisticated mode he or she will consider only the Liberal Democrat and the Conservative candidate as electable. Intent on preventing the Conservative Party to win the seat, he or she voters for the Liberal Democrat candidate as the "lesser evil". The first core idea means that choosing in expressive or in sophisticated mode are both options for this voter and that from the voting decision alone it cannot be inferred which mode was used. The second core idea is that a sophisticated mode of choice does not alter the preference order among the alternatives. When the Conservative Party remains in the consideration set, this is not because the voter prefers the Conservative Party to the Labour Party. To the contrary, in strategic mode the voter refrains from voting for the Labour Party, because he or she wants to *prevent* the Conservative party from winning. It is the comparison between the alternatives that remain in the consideration set of viable alternatives, that lets voting for the "second-best" alternative seem reasonable.

The next two sections describe in somewhat more detail the two core ideas of the proposed method and how they avoid the problems just explained. In the latter of these two sections a simulation study is briefly discussed that examines the performance of the proposed method. Another section describes the application of the new method to data from the 2010 British Election Study (Whiteley and Sanders 2014) and demonstrates how it agrees with results obtained with the "stated reasons" approach, thus demonstrating the convergent validity of the two approaches. The last section summarises the findings and reflects on limitations and potential extensions of the proposed method. An online appendix gives the technical details of the proposed method and some additional simulation results. The paper is accompanied by an *R* package, which will be made available as open source software once the paper has been published.

2 Voting as a mixture of sincere and sophisticated choices

The first core idea of the method proposed in this paper is simply that strategic voting is optional, i.e. voters may either take strategic considerations into account or refrain from doing so. As a consequence there are voters who vote in accordance to their preferences whatever the competitive situation may be, i.e. whatever strategic incentives it provides. One could call such voters "expressive voters" and the mode in which they make their voting decisions an "expressive mode". The other group of voters who make their decisions in a "strategic" or "sophisticated mode" are the potential strategic voters who, when faced with the appropriate strategic incentives, will deviate in their voting decision from their preferences. It should be noted, however, that voting in strategic mode is not the same as voting strategically. According to the definition put forward by Fisher (2004), a strategic vote deviates from the voter's preferences, i.e. *is different from* a (then counterfactual) sincere vote. It is however quite possible and in fact not uncommon that a choice in expressive mode and a choice in strategic mode lead to the same outcome, that is, when the voter's preferred candidate or party has the best chances to actually get elected. A typical example is a supporter for the Conservative party who in 2010 cast his or her vote in a rural district in southern England.

While the idea introduced here may seem trivial, its implications for the estimation of the amount of strategic voting are not. A formalisation of the idea in terms of random variables helps to clarify the implied complications. Let V_i denote the actually observed vote of individual i, M_i the mode in which he or she makes her choice, $U_{i|1}$ the vote choice made in sincere mode, and $U_{i|2}$ the vote choice made in strategic mode. The probability that individual i chooses alternative j from the set of available alternatives, or choice set $\mathcal{S}_i$ then becomes

$$\begin{aligned} \Pr(V_i = j) &= \Pr(V_i = j|M_i = 1)\Pr(M_i = 1) + \Pr(V_i = j|M_i = 2)\Pr(M_i = 2) \\ &= \Pr(U_{i|1} = j)\Pr(M_i = 1) + \Pr(U_{i|2} = j)\Pr(M_i = 2). \end{aligned} \tag{1}$$

By its definition, a strategic vote deviates from a vote made in sincere mode. That is, a strategic vote occurs if and only if $V_i \neq U_{i|1}$, which implies that $U_{i|1} \neq U_{i|2}$ and $M_i = 2$.

If the probabilities $\Pr(M_i = h)$ and $\Pr(U_{i|h} = j)$ are known for $h \in \{1, 2\}$ and $j \in \mathcal{S}_i$ or if one has estimates of these probabilities, one can derive the probability that individual i will vote strategically. The probability of a strategic vote then is:

$$\begin{aligned} \Pr(V_i \neq U_{i|1}) &= \Pr(U_{i|1} \neq U_{i|2} \wedge M_i = 2) \\ &= \sum_{j \in \mathcal{S}_i} \big(1 - \Pr(U_{i|1} = j)\big)\Pr(U_{i|2} = j)\Pr(M_i = 2) \end{aligned} \tag{2}$$

If one observes the voting choice $V_i = j$ one can also derive the probability that i has chosen strategically, which is the conditional probability

$$\begin{aligned} \Pr(V_i \neq U_{i|1}|V_i = j) &= \frac{\Pr(V_i \neq U_{i|1} \wedge V_i = j)}{\Pr(V_i = j)} \\ &= \frac{(1 - \Pr(U_{i|1} = j))\Pr(U_{i|2} = j)\Pr(M_i = 2)}{\Pr(U_{i|1} = j)\Pr(M_i = 1) + \Pr(U_{i|2} = j)\Pr(M_i = 2)} \end{aligned} \tag{3}$$

The probability in equation (3) can be viewed as a posterior probability of a strategic vote with (2) as a prior. If the probabilities $\Pr(M_i = 2)$ and $\Pr(U_{i|h} = j)$ are estimated from empirical data, (3) is a posterior in the *empirical Bayes* sense.

In order to obtain the probabilities stated in equations (2) and (3), one needs estimates of the probabilities $\Pr(U_{i|1} = j)$, $\Pr(U_{i|2} = j)$, and $\Pr(M_i = 2)$. However only the value of V_i is observed, which may be equal to the value of $U_{i|1}$, of $U_{i|2}$, or to both. This ambiguity notwithstanding it is possible to estimate the required probabilities, provided that they are parameterized in a way that leads to an identified model. In that case one can interpret the probabilities in equation (1) as the components of a likelihood function of a *finite mixture* model. A finite mixture discrete choice model has already been used e.g. by Duch et al. (2010) for the estimation of coalition-directed voting. Yet these authors use a fully Bayesian setup with prior distributions for the parameters in $\Pr(U_{i|1} = j)$, $\Pr(U_{i|2} = j)$, and $\Pr(M_i = 2)$ and estimate their models using a MCMC technique. Yet it is also possible to obtain maximum likelihood estimates of the parameters of the finite mixture model using an EM-algorithm (Little and Rubin 2002; McLachlan and Krishnan 2007) similar to the one common in the estimation of latent class models (Vermunt and Magidson 2004). In particular, the probability that individual i is in choice mode h can be compared with a latent class probability and is treated in the same way by the EM-algorithm. Details of the EM-algorithm are described in the web appendix of this paper.

To understand how the first core idea of the method proposed in this paper improves on the Alvarez-Nagler method, note that their method starts with a discrete choice model that includes strategic considerations and which is fitted to *all* votes in the data set at hand. Because of the inclusion of strategic considerations it is essentially a model of sophisticated choices $U_{j|2}$ (i.e. choices in sophisticated mode), but fitted to the observed choices V_j. Results of a simulation study discussed in the web appendix of this paper indicate that this leads to negatively biased estimates of the influence of strategic incentives. In this simulation study, the Alvarez-Nagler method was applied to artificial data sets with various compositions of votes in expressive and sophisticated mode, including data sets where *all* artificial voters used a sophisticated mode of choice, i.e. conformed to the model at the core of the Alvarez-Nager method. Even in those settings were the estimates of the amount of strategic voting considerably biased. This points to some problems inherent in the Alvarez-Nagler method beyond those problems that originate from ruling out the possibility of expressive voting.[3]

3 Choosing from a restricted set of alternatives: A parsimonious conceptualisation of strategic voting

The second core idea of the method proposed in this paper is that voters in strategic mode, in order to avoid wasting their vote, restrict their consideration of parties or candidates to those who are electorally viable. That is, they restrict their attention to those alternatives that have a genuine chance to win a seat, without changing the order of preference between each pair of parties that remain under consideration. In electoral systems where parties or candidates compete for one of m seats that represent the district in parliament, the $m + 1$ alternatives with the largest relative expected vote share can be considered as electorally viable (Cox 1997, 1994). In case of a single-member plurality system, such as the electoral system used for the UK House of Commons, $m = 1$ so that the two largest parties in the district (in terms of expected vote share) can be considered electorally viable.

In more formal terms, each of the modes of choice is characterised by a *consideration set* C_{hi} which is a (not necessarily proper) subset of the choice set $\mathcal{S}_i$ of the voter, the entire set of alternatives from which he or she can choose. In expressive mode ($h = 1$, say) the consideration set C_{1i} equals the full choice set. In sophisticated mode ($h = 2$, say), the consideration set C_{2i} is a proper subset of the choice set, so that some alternatives are not included in the consideration set. Apart from the set of alternatives taken into consideration, the two modes of choice do

[3] That the simulation study uncovered such surprising results underlines the value of simulation studies for the assessment of old and new methods to measure or estimate quantities of interest.

not differ. In particular, the preference order among the alternatives remains the same.

For illustrative purposes, consider again the case of the hypothetical Labour supporter in Colchester: His or her first preference is the Labour Party, the second preference is the Liberal Democratic Party, while the Conservative Party has a lower preference (perhaps even with a lower position in the preference order than the Green Party). In expressive mode his or her consideration set will be the full set of party candidates listed on the ballot paper. In comparison to all other parties, the Labour Party is the best evaluated so that he or she casts her vote for the candidate of this party. In sophisticated mode he or she considers only the candidates of the Conservative and the Liberal Democratic Party. for being chosen. He or she will then vote for the candidate of the Liberal Democratic Party, because in comparison to the other party in the consideration set it is the preferred one. This outcome notwithstanding, the fact that he or she does not consider Labour to have a chance to win the seat does not give them a lower place in his or her order of preference, which remains unchanged.

The idea discussed in the previous section, that strategic voting is a potential consequence of restricting one's attention only to electorally viable alternatives is relatively simple e.g. in comparison of a decision-theoretic construction such as Alvarez and Nagler's (2000). Yet it does require information about voters' preferences or information needed to infer these preferences – no model-based technique can do without this – and information that allows to distinguish, from the voters' point of view, between electorally viable and nonviable alternatives. What it does not require are variables that describe strategic incentives beyond this distinction nor does it require to specify the functional form of their influence on electoral choices, in contrast to Alvarez and Nagler's method. That the method does not take into account such information is however not a drawback, but an advantage, because it allows to separate the estimation of the amount of strategic voting from explaining it and thus makes possible at all to empirically test such explanations. As a simulation study reported further below in this section shows, not knowing the relevant factors that lead individuals to choose in strategic mode does not lead to a bias in the estimated amount of strategic voting.

Like a model-based method such as Alvarez and Nagler's, the proposed method does not require information about voters' exact preference order over the parties or candidates running for office. Electoral study Survey data rarely include such information. What the method requires are good predictors for voters' party or candidate preferences. Such predictors could be voters' social position, their policy positions or their feelings towards the parties and their leaders. Of course, such predictors should be exogenous with respect to voting choices. Otherwise the predictions about the conditional – and potentially counterfactual – choices $U_{i|h}$ would be contaminated by the actual choices, with the consequence that the frequency in which voters depart from sincere choices will be biased downward.

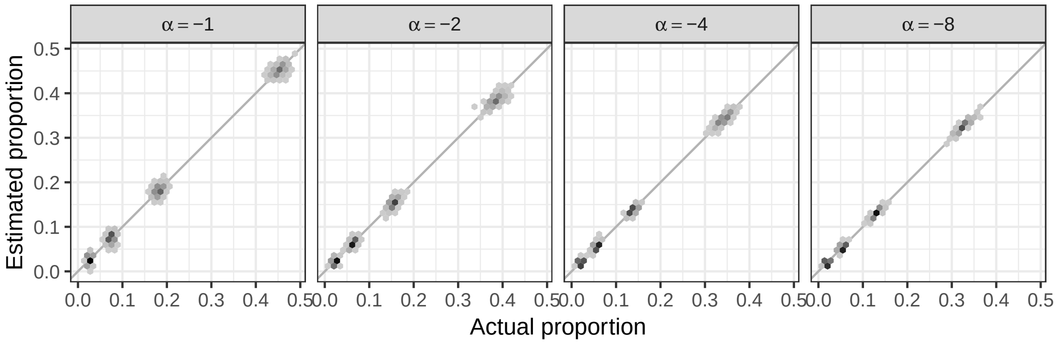
Given that the predictions about the conditional choices are unlikely to be perfect, the link between predictor variables and these choices should be modelled as probabilistic. A probabilistic link that accommodates varying consideration sets very well is McFadden's conditional logit (McFadden 1974; for a more accessible discussion of this model see e.g. Agresti 2002, 298ff): Let i denote a number that identifies an individual, j denote a number that identifies an alternative in the individual's consideration set C_{hi} for choice mode h, $\boldsymbol{x}_{ij}$ denote a vector of predictor values and $\boldsymbol{\alpha}$ a vector of coefficients. In the adaptation of the conditional logit model, the probability that individual i chooses alternative j from consideration set C_{hi} when in choice mode h is

$$\Pr(V_i = j|M_i = h) = \Pr(U_{i|h} = j) = \frac{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\alpha})}{\sum_{k\in C_{hi}} \exp(\boldsymbol{x}'_{ik}\boldsymbol{\alpha})}. \tag{4}$$

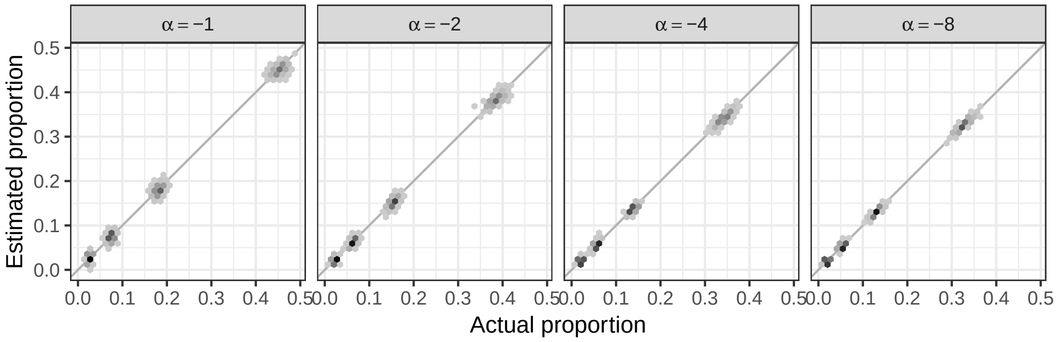
Here $\boldsymbol{x}'_{ij}\boldsymbol{\alpha}$ is a scalar product, so that $\boldsymbol{x}'_{ij}\boldsymbol{\alpha} = \alpha_1 x_{ij1} + \cdots + \alpha_1 x_{ijq}$, where q is the number of predictors for the conditional choices. The value x_{1ij} could be, for example, the perceived policy distance between individual i and party or candidate j or the feelings that individual i has about party or candidate j.

Earlier in this paper it was stated that the probability $\Pr(M_i = h)$ of individual i making his or her choice in mode h can be estimated in a way similar to the estimation of a latent class probability in a latent class model. That way it does not require to make any assumptions about what influences this probability. Yet there are scholars who are interested in what these factors are or how rational individuals use strategic incentives to vote strategically (e.g. Myatt 2007; Kselman and Niou 2010). That is, they are interested in the impact of strategic incentives and other factors on the probability $\Pr(M_i = h)$. For example, one may ask what the costs of ignoring strategic incentives – such as the closeness of competition between parties and the distance from contention which play a central role for the reconstruction of strategic voting in the Alvarez-Nagler approach – are for the estimation of strategic voting. That is, does ignoring (or simply not knowing) relevant predictors lead to biased or inconsistent estimates of the frequency of strategic voting? Since it is possible to model the influence of various factors on $\Pr(M_i = h)$ within the framework of the proposed method (see the web appendix of this paper for details), this question is addressed by a simulation study, which is summarized in the following.

In each of 1,000 replications of the simulation study, artificial voting data are created as a mixture of votes in sincere and in sophisticated mode, where the propensity to vote in sophisticated vote is influenced by the closeness of the race between the two strongest parties and the other parties' distance from contention. In each replication of the simulation study, two variants of the finite mixture model are fitted to this artificial data, (1) a variant in which the influence of the competitive



(a) First variant: Fully specified finite mixture model



(b) Second variant: Finite mixture model without information about predictors of strategic incentives

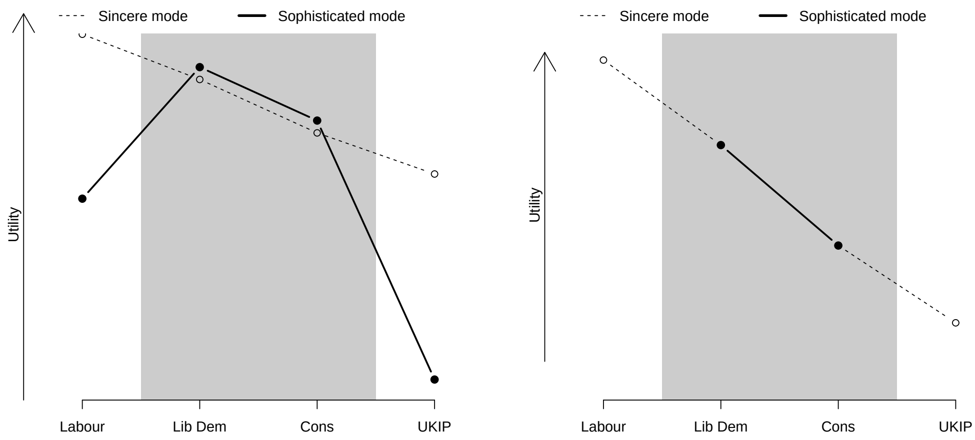
Figure 1 Hexbin plots of the difference between actual and estimated rate of strategic votes (based on posterior probabilities) for various settings for the influence of party evaluations (α) and the influence of strategic incentives on sophisticated votes (β). The different settings of the parameter β are reflected in the location of the different "clouds" that appear in the diagrams, the shading of the hexagons indicates how many data points are contained in them. Each of the panel corresponds to 4,000 simulated data points.

situation on the propensity to vote in sophisticated mode is correctly specified and (2) a variant in which this influence is ignored.[4]

Figure 1 summarizes the results of the simulation study. It shows "hexbin" plots which are a useful alternative to scatter-plots when the number of data points is as large (Carr et al. 1987). The panel on top (Figure 1a) shows differences between estimated and actual proportions of strategic voting, where the estimates are based on a model that includes the correctly specified influence of strategic incentives. The panel at the bottom (Figure 1b) does the same, but here estimates are based on a model that ignores the influence of strategic incentives and has only a single parameter for the probability of the selection of a sophisticated mode of choice.

The summary of the simulation study provided by Figure 1 appears quite encouraging: Whatever the size of the parameters α and β, hardly any systematic departures of the average of the estimated rates from the true rates of strategic

[4] Details about the simulation study can be found in the online appendix of this paper.



(a) Choice based on a utility function with strategic incentives as an additive component

(b) Choice from a restricted consideration set

Figure 2 An illustration of two conceptualisations of the difference between sincere and sophisticated voting. Both conceptualisations are used to describe the defection of a Labour supporter to the Liberal democrats. In the left-hand panel the "sophisticated" utility function gives the Conservatives a higher utility than the Labour party. In the right-hand panel the restriction of the consideration set to the viable alternatives leaves the preference for Labour relative to the Conservative Party unaltered.

voting are discernible, and this applies for both variants of the model used for estimation – with or without taking into account the information about the strategic incentives. While the size of β has an impact on the actual rate of strategic voting (this is because the independent variable has a non-zero mean), the size of α does not. It has however an impact on the variance of the estimates of the rate of strategic voting, an outcome that is easy to understand: The more information we have about voters' preferences (via the size of α), the more precisely we are able to estimate the rate of strategic voting. The more important lesson to draw from Figure 1 is, however, that it is not necessary to identify the factors that influence strategic voting to estimate its frequency accurately.[5] Therefore it is possible to test hypotheses about which factors are relevant for strategic voting and what the functional form of their influence is.

Apart from disentangling the estimation of strategic voting from explaining it, the method proposed in this paper has the additional advantage over the Alvarez-Nagler method that it avoids its paradoxical implications. The Alvarez-Nagler method rests on a model that involves a utility function that combines predictors of (sincere) preferences for parties with strategic incentives, where the inclusion of the latter potentially leads to strategic voting. A conceptual dilemma raised by this construction is that when voting occurs according to such "sophisticated" preferences it is no longer possible make use of the definition of strategic voting

5 In hindsight this does not look too surprising. It is also not necessary to estimate a correctly estimate a regression model in order to obtain an unbiased estimate of a population average.

as deviating from ones' preferences – unless one makes the distinction between two kinds of preferences based on the variables involved the utility functions that guide these preferences. But this begs the question why certain variables are so special that their inclusion into a utility function makes the resulting choices "sophisticated". Another concern that may appear less like hair-splitting is that the modified utility function can lead to preference orders that upset the original motivation to vote strategically. One could call this change of the preference order a "paradoxical preference reversal".

Figure 2 illustrates how the additive conceptualization of a utility function that describes a sophisticated mode of choice can leads to a paradoxical preference reversal and how the conceptualisation of a sophisticated mode of choice put forward in this paper avoids such reversals. Both diagrams in the figure describe the choices of a voter who (originally) prefers Labour over the Liberal Democrats, the Liberal Democrats over the Conservatives, and the Conservatives over UKIP but is faced with a competitive situation that leads him or her to a strategic deviation from Labour to the Liberal Democrats.

In the left-hand diagram, the utility of all candidates except for those of the two largest parties in the district (in terms of expected vote share), the Liberal Democrat candidate and the Conservative candidate, is reduced because of their distance from contention because of the closeness of race between the Conservative and the Liberal Democrat candidate. This utility function leads the voter to strategically desert the Labour party in favour of the Liberal Democratic candidate. However, an implication of this modified utility function is that the Conservative candidate now has a higher utility for the voter than the Labour candidate, in contrast to the intuitive notion that the voter chooses to vote for the Liberal Democratic candidate in order to *prevent* the Conservative candidate from winning the seat. It is not easy to see how to construct a sophisticated utility function in a way that avoid such paradox preference reversals. In fact, they are a frequently occurring consequence of the model on which the Alvarez-Nagler method rests, as a simulation study indicates that is reported in the web appendix of this paper.

The right-hand panel of Figure 2 illustrates how a restriction of the consideration set leads to desert the voter to desert the Labour candidate. While the voter still prefers the Labour party to win the seat (and the government majority) he or she does not consider this as a possible outcome. Since only Conservative or a Liberal Democratic win of the seat is perceived as a possible outcome, what counts for the electoral choice is therefore only the utility difference between the Liberal Democrat and the Conservative party. Considering only the candidates of these two parties as electorally viable (where the electoral viability of these two party candidates is indicated by the grey area in the diagram), he or she chooses the Liberal Democrat candidate.

It should be noted that, while the method proposed here allows to avoid the need to commit to a particular assumption about the influence of strategic incentives on the utility functions of sophisticated voters, it does not rule out that the decision to vote strategically can be reconstructed as instrumentally rational. It is still possible that the “second-order decision” to vote either in sincere or in strategic mode is influenced by a weighing of the “expressive utility”, coming from voting for the most-liked alternative even while seeing a disliked alternative winning the seat, against the more “instrumental utility”, coming from voting for an alternative different from the most-liked one in order to prevent a disliked alternative from winning the seat. The influence of strategic incentives on this second-order decision indeed can be modelled within the framework of this paper, as the simulation study indicates, the results of which are already presented in Figure 1a.

4 Convergent validity: Two approaches to strategic voting and the general election for the House of Commons in 2010

The previous sections presented the core ideas on which the new approach of this paper to the estimation of strategic voting is based on. While the approach performs quite well if its assumptions are satisfied and while these assumptions may be plausible, their plausibility does not guarantee that these assumptions are empirically valid and the method is applicable to empirical data. In the present section the proposed method is applied to the empirical case of strategic voting during the 2010 general election of the UK House of Commons. Furthermore, results of the method are compared to results obtained by a corrected stated-reasons method.

The 2010 election is chosen for this empirical application, because the 2010 British Election Study (Whiteley and Sanders 2014) is the most recent one that allows the application of the improved stated-reasons method: In the 2010 BES respondents were asked about the reasons for their (intended or recalled) votes. Only respondents who explicitly gave as the reason for their vote or vote intention, that they vote “tactically” or that the party they prefer has no chance of winning *and* who gave the name of a different party when asked about the party they “really preferred” where counted for the corrected rate of strategic voting (Fisher 2004). The 2015 BES data for example do not include a question about such “really preferred” parties and therefore do not allow such a correction of the rate of strategic voting (Fieldhouse et al. 2016).

Even though the stated-reasons approach has been criticised from proponents of a model-based approach (Alvarez and Nagler 2000), there are good reasons to compare results obtained by the model-based method proposed in this paper with results by the improved stated-reasons approach. Firstly, model-based approaches are not free from biases as stated earlier in this paper and demonstrated in the web appendix, so it not or no longer evident that model-based approaches are superior to the state-reasons approach. Secondly, if a model-based method and the stated-reasons method both lead to similar results, this indicates the *convergent validity* of the two methods of measuring strategic voting (Evans 2002). If on the other hand results obtained from two different methods diverge, there is no easy way to decide which is the most accurate one, in the absence of any independent yardstick with self-evident validity. If it is impossible to obtain any convergence in measures based on different methods, one may very well doubt whether the phenomenon these methods are supposed to measure exists at all that or whether these measures are anything more than artefacts.

Two kinds of data are needed for the analysis of strategic voting in the 2010 general election using the method proposed in this paper: Firstly, individual-level data on voters' choices and predictors of their preferences and, secondly, district-level electoral results that allow to distinguish between viable and non-viable party candidates, i.e. party candidates that do or do not have a chance to win a parliamentary seat. The data on voters' choices and predictors of preferences come from the 2010 British Election Study (Clarke et al. 2010). They include not only respondents' party choices, but also a variety of predictors for party preferences, such as social class, parties' perceived an respondents' own positions on political issues, party identification, and respondents' feelings about the major parties and their leaders.

A natural base for assessing the competitive situation in the electoral districts in the 2010 general elections are the vote share that the parties obtained in the preceding general election of 2005. However, these cannot be used without modification, because the number of constituencies and the constituency boundaries have changed between 2005 and 2010. Fortunately, *notional* 2005 party vote shares for the 2010 constituencies are available from Norris (2010). These notional results were computed by a method reported in Borisyuk et al. (2010). To take into account the information that voters could obtain from opinion poll results, these notional 2005 results were updated by aggregated opinion polls published immediately prior to the 2010 election (Wells 2010). The technique used for updating the district results is based on the "uniform swing assumption", which is also used, for example, in the "Swingometer" published at the BBC website (BBC 2010a,b). This updating technique is described in the web appendix of the paper.

The following variables were considered as predictors of the preference from which strategic voters deviate: (1) Social class, (2) religion, (3) region, (4) the policy distances between each party and themselves on the issue dimension of low taxation *vs* welfare spending and on the issue dimension of fighting crime *vs* protecting the rights of those who are accused of crimes, (5) voters' feelings towards each party, and (6) feelings towards each party's leader. Likelihood ratio tests of conditional logit discrete choice models showed that sequentially adding each of the six groups of variables to a simple baseline model improved goodness of fit significantly. Dropping the first three variable groups from the full model does not lead to a statistically significant loss in goodness of fit. This indicates that while structural variables are important for vote choice in 2010, their

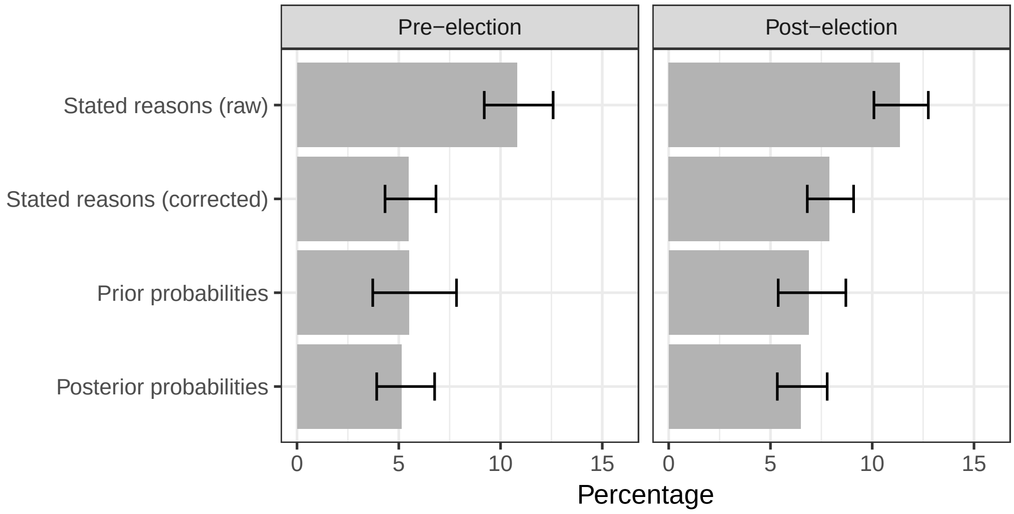


Figure 3 Estimated rate of strategic voting based on (two variants of) the stated reasons method and based on the method proposed in the paper

influence is mediated by policy distances and affective evaluations. Therefore only policy distances and affective party and party leader evaluations are used in finite mixture discrete choice model used to estimate the amount of strategic voting in the 2010 election. To save space, model tests that lead to the selection of the predictor variables as well as the estimates of the parameters of the finite mixture discrete choice model are moved to the online appendix of the paper. Instead of a lengthy discussion of these, the following paragraphs focus on the comparison of predictions about strategic voting obtained by the method proposed in this paper and by the stated-reasons method.

Figure 3 compares estimates of the amount of strategic voting (or intentions to vote strategically) during the UK general election of 2010. The diagram shows estimated percentages along with 95 per cent confidence intervals. Bars labelled "raw self-reports" correspond to the percentage of strategic voting estimated directly from the statements of BES respondents: Respondents who gave as reasons for their vote that "The party they really preferred did not have a chance to win the seat" or who explicitly stated to vote or have voted "tactically" are counted as strategic voters. Bars labelled "corrected self-reports" correspond to estimates also obtained from the stated reasons for the votes or vote intentions, but corrected in so far as voters who stated as their "real preference" the same party as the party they actually voted for are *not* counted as strategic voters. "Prior probabilities" and "Posterior probabilities" correspond to percentages estimated from the accumulated voters' prior and posterior probabilities to have voted strategically, based on the estimated finite mixture discrete choice model and equations (2) and (3). The prior probabilities are computed according to equation (2), while the posterior probabilities are computed according to equation (3).[6]

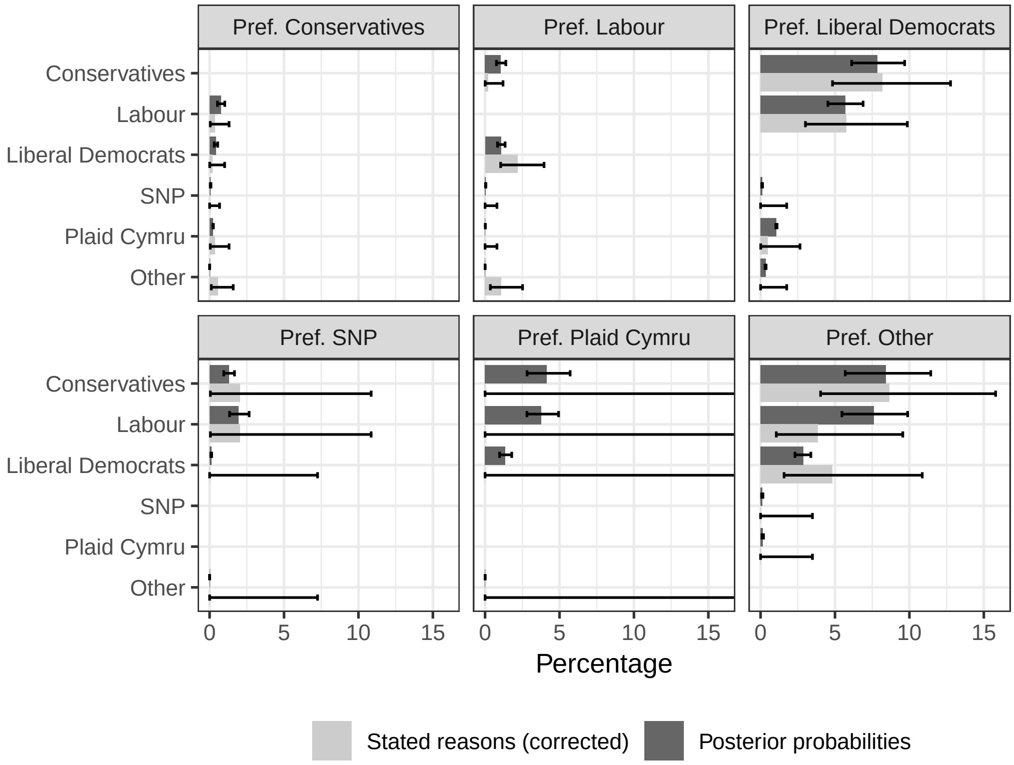
[6] The confidence intervals are based on a variant of the parametric Bootstrap proposed by King et al. (2000). See the online appendix for details.

It becomes obvious from Figure 3 that it matters which variant of the self-report method is used for the estimation of the percentage of strategic voting: The raw, uncorrected percentage of reported strategic votes is much higher than the corrected percentage. It matters much less which variant of predictions obtained from the finite mixture discrete choice model is used for estimating the percentage of strategic voting. The main difference between the estimate based on prior probabilities and the estimate based on posterior probabilities is that the latter show less variation under bootstrap re-sampling. Conditioning on the observed votes obviously increases the precision of the estimated percentages of strategic voting.

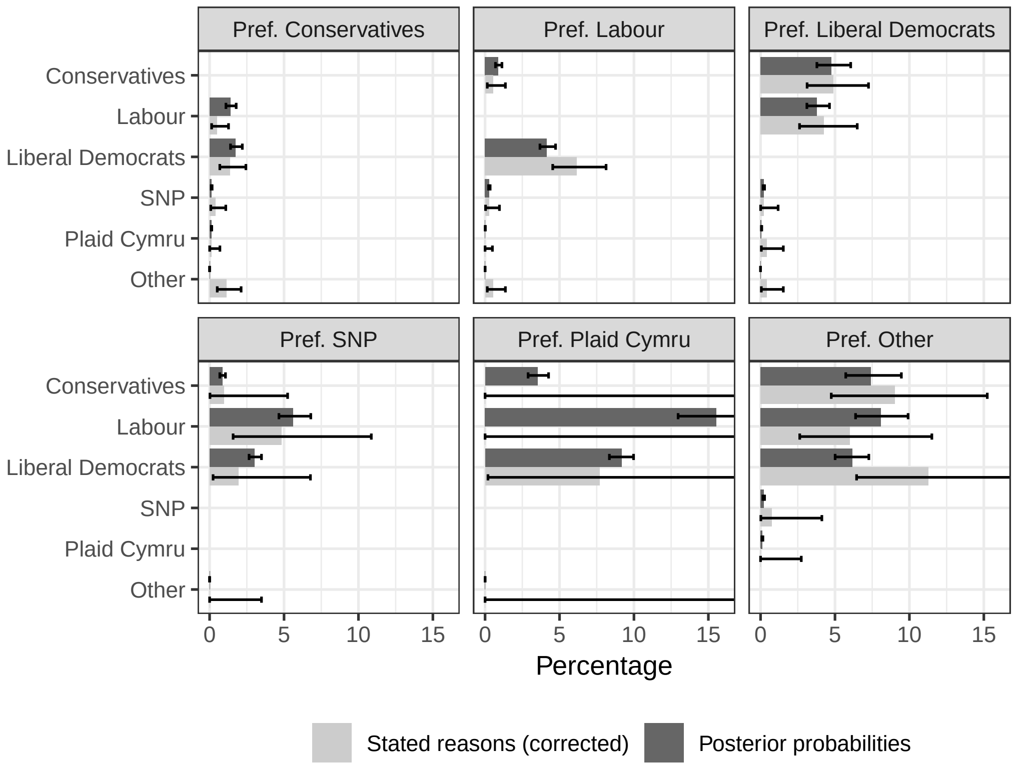
Figure 4 gives another opportunity to examine the convergence between the stated-reasons approach and the model-based approach introduced in this paper. It shows the "flow" from voters' party preferences to their actual votes or vote intentions, estimated, on the one hand, using respondents' statements about which party they voted for and which party they "really preferred", and on the other, using posterior probabilities of a strategic vote and the reconstructed party preferences based on the finite mixture discrete choice model. The latter estimates are based on a generalisation of equation (3) that is discussed in the online appendix of this paper. In order to facilitate the comparison of both approaches in terms of their predictions about strategic deviations, all instances where actual votes and reconstructed preferences coincide are dropped.[7]

The agreement between the two methods is quite striking. Most of the "qualitative" features of the flow from preferences to votes are the same in the pre- and post-election sample of the BES. Both methods agree that there were more strategic deviations from Liberal Democrat preferences in the pre-election wave than in the post-election wave. They also agree that there were more deviations from the Liberal Democrats to the Conservatives than to Labour in the pre-election wave. They also agree that there were more strategic deviations from Labour to the Liberal Democrats in the post-election than in the pre-election wave. Furthermore, there is quite large "quantitative" agreement between the two methods. By and large, differences between the vote-by-preferences percentages are within a margin of sampling error (as expressed in 95 per cent confidence intervals). There is only one aspect in which the two methods seem to disagree systematically: Estimated percentages of deviation from Labour to the Liberal Democrats are lower than those based on voters stated reasons. A possible explanation of this systematic disagreement is that some voters deviated from their preferences for other reasons than avoiding a wasted vote: A government coalition of the Conservatives with the Liberal Democrats emerged as a real possibility during the campaign. This may have led some voters to desert Labour in order to give the Liberal Democrats, the lesser evil, a greater weight in the coalition. By design, the present application is not able

[7] It should be noted that even if voters chose in a strategic mode where they seek to avoid a wasted vote, preferences and votes can coincide when the preferred alternative viable in the relevant district.

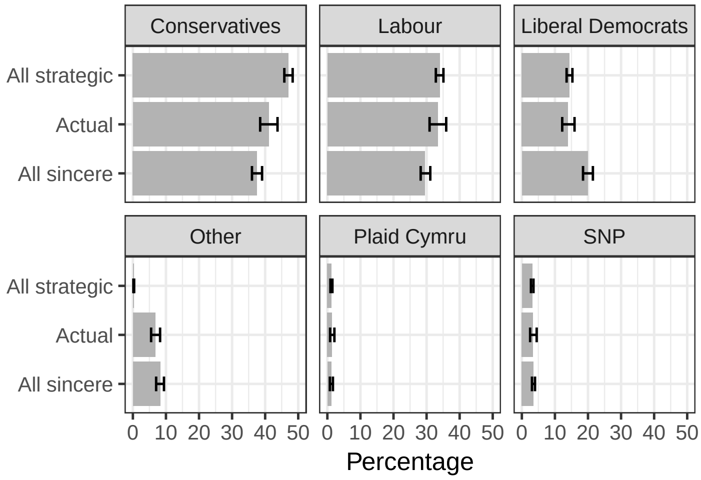


(a) Pre-election wave

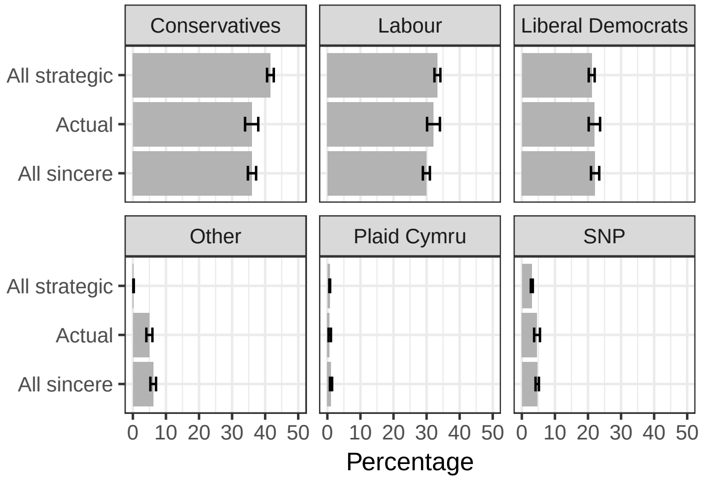


(b) Post-election wave

Figure 4 The direction of strategic vote deviations, estimated by the two methods



(a) Pre-election wave



(b) Post-election wave

Figure 5 The actual distribution of votes in the sample and the counterfactual distribution under the condition that either all or none of the voters choose in a strategic mode.

to capture this type of strategic voting, because focuses on strategic voting with the intention of wasted-vote avoidance on the district level. It is of course possible to extend the model to in this direction, but this is beyond the scope of the current paper.

With the method proposed in this paper it is not only possible to estimate the percentage of strategic voters, but also to examine the consequences if all voters or none of them vote in a strategic mode. Figure 5 compares these two scenarios with the actual distribution of vote intentions in the pre-election wave of the 2010 British Election Study and the scenarios with the actual distribution of the reported votes in the post-election wave. According to the pre-election wave data, strategic voting had cost the Liberal Democrats a good proportion of their vote share, while it benefited both Labour and the Conservatives. The Conservatives would have done even better if all voters had chosen in a strategic mode. Post-election data tell a slightly different story. Strategic voting would have hardly reduced the vote share of the Liberal Democrats, while the gains of the Conservatives from strategic voting would have been smaller and it would not have cost them in vote share if all respondents had voted sincerely.

It should be noted that the different findings based on the pre- and post-election wave are unlikely to be mere sampling fluctuations. The definition of strategic or non-strategic votes is based on the identification of the viable and non-viable alternatives in the voters' districts is based on the district results of the previous election updated by current polling results. Yet the standing of the parties has changed considerably during the 2010 electoral campaign, so that the viability of the parties, the Liberal Democrats in particular, in many districts has changed as well. Neither are the different findings an artefact created by a bias in the proposed method. The differences between the pre- and the post-election wave in terms of the flows from preferences to votes (shown in Figure 4) emerge not only when the proposed method is used, but also when the corrected self-report method is used, which does not depend on a distinction between viable and non-viable alternatives at district level.

5 Conclusion

The present paper proposed a new method for the estimating the rate of strategic voting from survey data. It rests on the following two ideas: First, a voter makes his or her choice either in a sincere or in a strategic mode, where the latter is characterised by the intention to avoid wasting one's vote for a party or candidate without a plausible chance to win a seat. Second, while a choice in a sincere mode considers all alternatives for being chosen, a choice in strategic mode is from a restricted consideration set that contains only alternatives perceived as viable. A strategic vote is a vote that differs from a choice as it would have been made in sincere mode (Fisher 2004).

Insofar as the distribution of observed choices is considered as a mixture of conditional choice distributions, the proposed method rests on a *finite mixture model*, where the mixture components are discrete choice models. It is certainly not the first time a finite mixture discrete choice model has been developed for the analysis of voting behaviour. Duch et al. (2010) construct a finite mixture model to distinguish between voting oriented on individual parties and voting oriented on the coalitions that these parties may form in order to gain government office. Their model differs from the model proposed here, in that both modes of choice involve the full choice set but differ in the utility function over the alternatives. Further, Duch et al. (2010) use a computationally demanding Bayesian approach for the estimation of the parameters of their model, while the method proposed in this paper uses a much quicker iterative algorithm to compute maximum likelihood estimates. The idea that sincere and strategic votes differ in the set of alternatives taken into consideration is also used by Kawai and Watanabe (2013), but their approach involves only aggregate-level data and does not use make use of individual-level predictors for the reconstruction of voters preferences.

The proposed method is an alternative to the stated-reasons approach in so far as it does not require survey respondents statements about the reasons they give for their votes (Heath et al. 1985; Evans and Heath 1993; Franklin et al. 1994). It can thus be used in applications with survey data where respondents are not asked about which party they "really" prefer, when they state to the intention to vote to have voted for a different party. The method differs from Alvarez and Nagler's earlier approach (Alvarez and Nagler 2000; Alvarez et al. 2006) by not requiring a particular specification of the influence of the closeness of the competition between parties or their distance from contention, by allowing for varying choice set (thus allowing to e.g. analyse strategic voting not only in England but also in other constituent countries of the United Kingdom, i.e. Scotland and Wales). Finally it takes explicitly into account that sincere and strategic voting are alternative modes of choice and that it is possible that sincere voting choices may differ from the incentives that originate in the competitive situation between the parties. By making the method independent from certain assumptions about the functional form of the impact of these incentives it makes estimating strategic voting independent from explaining it and therefore provides better opportunities for testing such assumptions. Further, it avoids certain paradoxical implications of the Alvarez-Nagler model and, more importantly, avoids the bias that comes from ignoring the distinction between alternative modes of choice.

When applied to data from the 2010 British Election Study, estimates obtained using the new method agree well with estimates obtained using the stated-reasons approach. The two methods agree not only in terms of the estimated rate of strategic voting, but also in terms of the flows from sincere preferences to strategic votes. This underlines the convergent validity of the two methods and the existence of the phenomenon of strategic voting beyond the assumptions or implications of a single method.

There is one limitation that the proposed method shares with the Alvarez-Nager approach: its dependence on information that allows to reconstruct individual voters' preferences. The quality of this reconstruction is crucial for estimates of the percentage of strategic voting. If the variables chosen for this reconstruction are only weak predictors for the preferences, the percentage of votes that deviate from these preferences is likely to be overstated. If variables are used that are endogenous with respect to actual voting choices, this is likely to lead to an understatement of strategic voting, because reconstructed preferences are "too close" the actual voting choices. In the practical application to the 2010 British Election Study data it turned out that the choice of the variables used for the reconstruction of preferences indeed has a substantial impact on the estimates of the percentage of strategic voting obtained.[8]

[8] Details are available in the online appendix of this paper.

First-past-the-post electoral systems are not the only ones that provide opportunities and incentives for strategic voting (Cox 1997). An obvious case are voting in systems with multi-member districts, where electors have one vote but m seats are to be filled. Cox (1994) has shown that in this situation the "Duvergerian" idea of strategic voting as wasted vote avoidance can be generalised: Only the first $m+1$ candidates (in terms of vote shares) are viable and any vote for the $m+k$-th candidate ($k > 1$) in the constituency would be "wasted" (Cox 1994). The generalisation of the method proposed in this paper to such cases is obvious.

Less obvious is the application of the proposed method to a type of strategic voting that has been claimed to be relevant in political systems with proportional representation voting systems and frequent government coalitions. This type of strategic voting is described by the motive of assuring the parliamentary representation of the preferred smaller coalition partner a voter supports ("threshold insurance", e.g. Gschwend 2007). If threshold insurance voting can be conceptualised in terms of consideration sets, then the finite mixture approach of this paper can also be extended to this type of strategic voting. If this succeeds, then the finite mixture approach of this paper can also be extended to split-ticket voting in mixed electoral systems such as Germany's. The finite mixture approach may then help to empirically disentangle four cases in this situation: (1) A voter chooses sincerely with both votes, (2) strategically with the first and sincerely with the second, (3) sincerely with the first vote and strategically with the second, and (4) strategically with both votes. The potential applications just mentioned can only be sketched here and would need further elaboration in specific empirical studies. But what this discussion makes clear is that the finite mixture model of strategic voting introduced in this article opens up new ways of addressing tricky research problems.

References

Agresti, Alan, 2002. *Categorical Data Analysis.* New York: Wiley, second edition.

Albert, A. and J. A. Anderson, 1984. "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models". *Biometrika* 71(1): 1–10.

Alvarez, R. Michael, Frederick J. Boehmke and Jonathan Nagler, 2006. "Strategic Voting in British Elections". *Electoral Studies* 25(1): 1–19.

Alvarez, R. Michael and Jonathan Nagler, 2000. "A New Approach for Modelling Strategic Voting in Multiparty Elections". *British Journal of Political Science* 30(1): 57–75.

Artabe, Alaitz and Javier Gardeazabal, 2014. "Strategic Votes and Sincere Counterfactuals". *Political Analysis* 22(2): 243–257.

BBC, 2010a. "Election 2010: Election Seat Calculator".
http://news.bbc.co.uk/1/hi/uk_politics/election_2010/8609989.stm

BBC, 2010b. “Election 2010 Swingometer”.
http://news.bbc.co.uk/1/hi/uk_politics/election_2010/8574653.stm

Blais, André and Richard Nadeau, 1996. “Measuring Strategic Voting: A Two-step Procedure”. *Electoral Studies* 15(1): 39–52.

Blais, André, Robert Young and Martin Turcotte, 2005. “Direct or indirect? Assessing two approaches to the measurement of strategic voting”. *Electoral Studies* 24(2): 163–176.

Borisyuk, Galina, Colin Rallings, Michael Thrasher and Ron Johnston, 2010. “Parliamentary Constituency Boundary Reviews and Electoral Bias: How Important Are Variations in Constituency Size?” *Parliamentary Affairs* 63(1): 4–21.

Carr, Daniel B., Richard J. Littlefield, W. L. Nicholson and J. S. Littlefield, 1987. “Scatterplot Matrix Techniques for Large N”. *Journal of the American Statistical Association* 82(398): 424–436.

Casella, George and Roger L. Berger, 2002. *Statistical Inference*. Pacific Grove: Duxbury, second edition.

Clarke, Harold D., David Sanders, Marianne C. Stewart and Paul Whiteley, 2010. “The British Election Study 2009-10”.
http://bes2009-10.org/

Clopper, C. J. and E. S. Pearson, 1934. “The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial”. *Biometrika* 26(4): 404–413.

Cox, Gary W., 1994. “Strategic Voting Equilibria Under the Single Nontransferable Vote”. *The American Political Science Review* 88(3): 608–621.

Cox, Gary W., 1997. *Making Votes Count: Strategic Voting in the World’s Electoral Systems*. Cambridge: Cambridge University Press.

Dempster, A. P., N. M. Laird and D. B. Rubin, 1977. “Maximum Likelihood from Incomplete Data via the EM Algorithm”. *Journal of the Royal Statistical Society. Series B (Methodological)* 39(1): 1–38.

Duch, Raymond M., Jeff May and David A. Armstrong, 2010. “Coalition-directed Voting in Multiparty Democracies”. *American Political Science Review* 104(04): 698–719.

Duverger, Maurice, 1965. *Political Parties: Their Organization and Activity in the Modern State*. New York: Wiley.

Evans, Geoffrey, 2002. “On the Validation of Measures of Strategic Motivations: A Critical Comment on Alvarez and Nagler”. *British Journal of Political Science* 32(1): 185–188.

Evans, Geoffrey and Anthony F. Heath, 1993. “A Tactical Error in the Analysis of Tactical Voting: A response to Niemi, Whitten and Franklin”. *British Journal of Political Science* 23(01): 131–137.

Fieldhouse, Ed, Jane Green, Geoffrey Evans, Hermann Schmitt, Cees van der Eijk, Jonathan Mellon and Chris Prosser, 2016. “British Election Study, 2015: Face-to-Face Post-Election Survey. [data collection].” Colchester, Essex: UK Data Service. SN: 7972.

Fisher, Steven D., 2004. “Definition and Measurement of Tactical Voting: The Role of Rational Choice”. *British Journal of Political Science* 34(01): 152–166.

Franklin, Marc N., Richard Niemi and Guy Whitten, 1994. “The Two Faces of Tactical Voting”. *British Journal of Political Science* 24(04): 549–557.

Gschwend, Thomas, 2007. “Ticket-splitting and Strategic Voting under Mixed Electoral Rules: Evidence from Germany”. *European Journal of Political Research* 46(1): 1–23.

Heath, Anthony F., Roger Jowell and John Curtice, 1985. *How Britain Votes*. Oxford: Pergamon Press.

Howat, Nicholas, Oliver Norden and Emily Pickering, 2011. “2010 British Election Study: Technical Report”. Tech. report, TNS-BMRB, London. http://bes2009-10.org/BEStechreportaug2611.pdf

Kawai, Kei and Yasutora Watanabe, 2013. “Inferring Strategic Voting”. *American Economic Review* 103(2): 624–662.

King, Gary, Michael Tomz and Jason Wittenberg, 2000. “Making the Most of Statistical Analyses: Improving Interpretation and Presentation”. *American Journal of Political Science* 44(2): 347–361.

Kselman, Daniel and Emerson Niou, 2010. “Strategic Voting in Plurality Elections”. *Political Analysis* 18(2): 227–244.

Little, Roderick J.A. and Donald B. Rubin, 2002. *Statistical Analysis with Missing Data*. Hoboken, NJ: Wiley, second edition.

Louis, Thomas A., 1982. “Finding the Observed Information Matrix when Using the EM Algorithm”. *Journal of the Royal Statistical Society. Series B (Methodological)* 44(2): 226–233.

McFadden, Daniel, 1974. “Conditional Logit Analysis of Qualitative Choice Behaviour”. In Zarembka, Paul (Ed.), *Frontiers in Econometrics*New York: Academic Press, pp. 105–142.

McLachlan, Geoffrey and Thriyambakam Krishnan, 2007. *The EM Algorithm and Extensions*. Hoboken, NJ: Wiley.

Myatt, David P., 2007. "On the Theory of Strategic Voting". *Review of Economic Studies* 74(1): 255–281.

Norris, Pippa, 2010. "Shared Data Sets". http://www.hks.harvard.edu/fs/pnorris/Data/Data.htm

R Core Team, 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Vermunt, Jeroen K. and Jay Magidson, 2004. "Latent Class Analysis". In *The Sage Encyclopedia of Social Sciences Research Methods*Volume 2. Thousand Oaks and London and New Delhi: Sage, pp. 549–553.

Wells, Anthony, 2010. "UK Polling Report". http://ukpollingreport.co.uk

Whiteley, Paul F. and David Sanders, 2014. "British Election Study, 2010: Face-to-Face Survey [computer file]". Colchester, Essex: UK Data Archive [distributor].

SUPPORTING MATERIAL

A Details on the proposed method

A.1 Formal construction of the finite mixture discrete choice model

In the main text of the paper the concept of *modes of choice* was introduced and a distinction was made between an expressive mode and a sophisticated mode, where the latter differs form the former in that only the alternatives perceived as viable are considered for being chosen. This idea that modes of choice are distinguished by the relevant consideration sets is formalized in the following:

Given is a set of voters, each represented by an integer number i between 1 and n inclusive. Each voter can choose from a set of alternative parties or candidates, each represented by an integer number j between 1 and m_i inclusive. The set of integer numbers that represent the alternatives available to individual i, and for simplicity also the set of alternatives that are represented by the integer numbers, is referred to in the following as his or her *choice set* $\mathcal{S}_i$. The vote of individual i then can be represented by a random variable V_i with the choice set as sample space, so that

$$\sum_{j\in\mathcal{S}_i} \Pr(V_i = j) = 1.$$

The first core idea of this paper is that each voter makes his or her choice in one of several alternative modes of choice. This mode of choice is represented by another integer-valued random variable M_i so that

$$\Pr(V_i) = \sum_h \Pr(V_i = j|M_i = h) \Pr(M_i = h) \tag{5}$$

For example, $M_i = 1$ may correspond to the case where i chooses in "expressive" mode and (always) votes sincerely, in line with his or her evaluation of the alternatives, while $M_i = 2$ may correspond to the case where i chooses in "sophisticated" mode and will consider voting different from his or her evaluation of the alternatives, given that the appropriate strategic incentives are provided. While the values of V_i correspond to observable outcomes, the values of M_i are latent and the distribution this random variable can only be inferred with the help of auxiliary information. In order for such inferences to be possible (and to make sense at all) it is necessary to assume for some i and some $j \in \mathcal{S}_i$ that

$$\Pr(V_i = j|M_i = h) = \Pr(V_i = j|M_i = h^*) \text{ implies } h = h^*. \tag{6}$$

The second core idea of the paper is that in sophisticated mode each voter will or would choose only alternatives (parties or candidates) that are electorally viable, in order or avoid wasting his or her vote. More formally, each mode of choice h corresponds to a certain choice set $j \in C_{hi}$ so that

$$\Pr(V_i = j|M_i = h) \begin{cases} > 0 & \text{for } h \in C_{hi} \\ = 0 & \text{for } h \notin C_{hi} \end{cases} \tag{7}$$

and

$$\sum_{j \in C_{hi}} \Pr(V_i = j|M_i = h) = 1 \tag{8}$$

If $h = 1$ corresponds to voting in an expressive mode the consideration set comprises the complete choice set, i.e. $C_{hi} = \mathcal{S}_i$, while for the other modes the consideration set is a proper subset of $\mathcal{S}_i$, i.e. some elements of $\mathcal{S}_i$ are not in C_{hi}.

A.2 Derivation of prior and posterior probabilities of strategic voting

As already remarked in the main text of the paper, voting in strategic mode may be a necessary condition for a strategic vote, which deviates from a (then counterfactual) sincere vote that expresses the voters preferences. The main text also presents a formula for the computation of a prior probability that a voter will vote strategically and a formula for the computation of the posterior probability that the vote cast by a voter is a strategic vote. In the following these counterfactual choices are formalized and two propositions that state the formulae are proved.

For the statement of the propositions two types of auxiliary variables are defined. First, for each choice mode h, a random variable $U_{i|h}$ is defined with the same sample space as V_i, but with a probability distribution given by

$$\Pr(U_{i|h} = j) = \Pr(V_i = j|M_j = h)$$

This construction is possible because the conditional probabilities for each h sum to unity. Further, for all i, and $h = 1, \dots, q$ it is assumed that

$$\Pr(U_{i|1} = j_1 \wedge \cdots \wedge U_{i|q} = j_q) = \Pr(U_{i|1} = j_1) \cdots \Pr(U_{i|q} = j_q)$$

i.e. the auxiliary variables are assumed to be stochastically independent.

Second, for each value h of M_i the dummy variable D_{hi} is defined with $D_{hi} = 1$ if and only if $M_i = h$ and $D_{hi} = 0$ if and only if $M_i \neq h$. This allows us to re-express V_i as a sum:

$$V_i = \sum_h U_{i|h} D_{hi}$$

and its distribution as a finite mixture

$$\Pr(V_i = j) = \sum_h \Pr(U_{i|h} = j) \Pr(D_{hi} = 1).$$

With these auxiliary variables, the following propositions can be formulated and proved.

Let $\varphi_{hi} = \Pr(M_i = h) = \Pr(D_{hi} = 1)$ denote the probability that individual i is in choice mode h and $\pi_{ij|h} := \Pr(V_i = j|M_i = h) = \Pr(U_{i|h} = j)$ he or she chooses alternative j if choice mode h. Suppose there are two modes of choice i.e. $M_i \in \{1, 2\}$. Then the probability that individual i makes a different choice than he or she would do if in choice mode $h = 1$ is

$$\Pr(V_i \neq U_{i|1}) = \varphi_{2i}\left(1 - \sum_{j\in\mathcal{S}_i} \pi_{ij|1}\pi_{ij|2}\right) = \varphi_{2i}\sum_{j\in\mathcal{S}_i}(1 - \pi_{ij|1})\pi_{ij|2}. \tag{9}$$

First, note that $V_i = U_{i|1}D_{1i} + U_{i|2}D_{2i}$ can differ from $U_{i|1}$ only if $D_{1i} = 0$ and $D_{2i} = 1$. Therefore we have $\Pr(V_i \neq U_{i|1}) = \Pr(U_{i|2} \neq U_{i|1} \wedge D_{2i} = 1)$. Further, because of $\Pr(U_{i|2} \neq U_{i|1} \wedge D_{2i} = 1) + \Pr(U_{i|2} = U_{i|1} \wedge D_{2i} = 1) = \Pr(D_{2i} = 1)$ we have $\Pr(U_{i|2} \neq U_{i|1} \wedge D_{2i} = 1) = \Pr(D_{2i} = 1) - \Pr(U_{i|2} = U_{i|1} \wedge D_{2i} = 1)$. $U_{i|1}$ and $U_{i|2}$ are stochastic independent from each other by definition, so we obtain

$$\Pr(U_{i|2} = U_{i|1}) = \sum_{j\in\mathcal{S}_i}\Pr(U_{i|2} = j\wedge U_{i|1} = j) = \sum_{j\in\mathcal{S}_i}\Pr(U_{i|2} = j)\Pr(U_{i|1} = j) = \sum_{j\in\mathcal{S}_i}\pi_{ij|1}\pi_{ij|2}.$$

Since also D_{hi} and $U_{i|h}$ are stochastically independent from each other by assumption we get $\Pr(U_{i|2} = U_{i|1} \wedge D_{2i} = 1) = \varphi_{2i}\sum_j \pi_{ij|1}\pi_{ij|2}$ and hence

$$\begin{aligned}\Pr(U_{i|2} \neq U_{i|1} \wedge D_{2i} = 1) &= \varphi_{2i}\left(1 - \sum_{j\in\mathcal{S}_i}\pi_{ij|1}\pi_{ij|2}\right)\\ &= \varphi_{2i}\left(\sum_{j\in\mathcal{S}_i}\pi_{ij|1} - \sum_{j\in\mathcal{S}_i}\pi_{ij|1}\pi_{ij|2}\right)\\ &= \varphi_{2i}\sum_{j\in\mathcal{S}_i}(1 - \pi_{ij|2})\pi_{ij|1}.\end{aligned}$$

Let $\varphi_{hi} = \Pr(M_i = h) = \Pr(D_{hi} = 1)$ denote the probability that individual i is in choice mode h and $\pi_{ij|h} := \Pr(V_i = j|M_i = h) = \Pr(U_{i|h} = j)$ he or she chooses alternative j if choice mode h. Suppose there are two modes of choice i.e. $M_i \in \{1, 2\}$. Then if individual i has chosen alternative j, the probability that this choice has deviated from the choice he or she would have made in choice mode $h = 1$ is:

$$\Pr(U_{i|1} \neq V_i|V_i = j) = \frac{(1 - \pi_{ij|1})\pi_{ij|2}\varphi_{2i}}{\pi_{ij|1}\varphi_{i1} + \pi_{ij|2}\varphi_{i2}}. \tag{10}$$

Bayes' theorem gives

$$\Pr(U_{i|1} \neq V_i|V_i = j) = \frac{\Pr(U_{i|1} \neq V_i \wedge V_i = j)}{\Pr(V_i = j)} = \frac{\Pr(U_{i|1} \neq j \wedge V_i = j)}{\Pr(V_i = j)}.$$

The denominator on the right-hand side is $\Pr(V_i = j) = \pi_{ij|1}\varphi_{1i} + \pi_{ij|2}\varphi_{2i}$. For the numerator on the right-hand side of this equation we have $\Pr(U_{i|1} \neq j \wedge V_i = j) = \Pr(V_i = j) - \Pr(U_{i|1} = j \wedge V_i = j)$. Further

$$\begin{aligned}
\Pr(U_{i|1} = j \wedge V_i = j) &= \Pr(U_{i|1} = j \wedge U_{i|1}D_{1i} + U_{i|1}D_{1i} = j) \\
&= \Pr(U_{i|1} = j \wedge U_{i|1}D_{1i} + U_{i|1}D_{1i} = j \wedge D_{1i} = 1) \\
&\quad + \Pr(U_{i|1} = j \wedge U_{i|1}D_{1i} + U_{i|1}D_{1i} = j \wedge D_{2i} = 1) \\
&= \Pr(U_{i|1} = j \wedge D_{1i} = 1) + \Pr(U_{i|1} = j \wedge \Pr(U_{i|2} = j \wedge D_{2i} = 1) \\
&= \pi_{ij|1}\varphi_{1i} + \pi_{ij|1}\pi_{ij|2}\varphi_{2i} \\
\Pr(U_{i|1} \neq j \wedge V_i = j) &= \pi_{ij|1}\varphi_{1i} + \pi_{ij|2}\varphi_{2i} - \pi_{ij|1}\varphi_{1i} - \pi_{ij|1}\pi_{ij|2}\varphi_{2i} \\
&= \pi_{ij|2}\varphi_{2i} - \pi_{ij|1}\pi_{ij|2}\varphi_{2i} = (1 - \pi_{ij|1})\pi_{ij|2}\varphi_{2i}.
\end{aligned}$$

Hence equation (10) follows immediately.

One note of caution should be added here: If there are only two modes of choice, then it is one could re-express V_i as

$$V_i = U_{i|1}D_{1i} + U_{i|2}(1 - D_{1i})$$

because $D_{1i} + D_{2i} = 1$. This looks similar to the equation for an effect variable in the Neyman-Rubin paradigm of causal inference, but this similarity is misleading. If the aim was causal inference in this paradigm, D_{1i} would be known (being the treatment indicator) and one would try to make inferences about the difference between $U_{i|1}$ and $U_{i|2}$ without making any assumptions regarding the distributions of these variables. In the present context however, the point of departure and the aim is quite different: D_{1i} is unobserved, and additional information is used to make specify the distribution of $U_{i|1}$ and $U_{i|2}$. Therefore, the criteria for valid causal inference in the Neyman-Rubin paradigm are not applicable here.

A.3 Further Quantities of Interest

The previous subsection gave a proof for the formula for the computation of prior and posterior probabilities, which are used for estimating the proportion of strategic voters in the paper. The applied section shows not only estimates of the proportion of strategic voting in 2010 based on British Election Study data in Figure ??, but also flows from preferences to strategic votes in Figure ??. These are the column percentages of a table of frequencies with votes as rows and preferences as columns, where the cells summaries of either prior or posterior probabilities, that is:

$$m_{jk} = \sum_i \Pr(V_i = j \wedge U_{i|1} = k)$$

in case of prior probabilities and

$$m_{jk} = \sum_i \Pr(V_i = j \wedge U_{i|1} = k | V_i = j) y_{ij}$$

in case of posterior probabilities, where y_{ij} is a dummy variable which equals 1 if the observed vote of i is j and 0 if the observed choice is a different alternative than j.

The prior probabilities can, generalizing the argument of the previous subsection, be derived as

$$\begin{aligned}\Pr(V_i = j \wedge U_{i|1} = k) &= \Pr(U_{i|1}D_{1i} + U_{i|2}D_{2i} = j \wedge U_{i|1} = k)\\ &= \Pr(U_{i|1} = j \wedge U_{i|1} = k \wedge D_{1i} = 1)\\ &\quad + \Pr(U_{i|2} = j \wedge U_{i|1} = k \wedge D_{2i} = 1)\\ &= \delta_j^k \pi_{ik|1}\varphi_{1i} + \pi_{ik|1}\pi_{ij|2}\varphi_{2i}.\end{aligned}$$

where δ_j^k is the Kronecker symbol, defined as

$$\delta_j^k = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}$$

The posterior probabilities are

$$\begin{aligned}\Pr(V_i = j \wedge U_{i|1} = k | V_i = j) &= \frac{\Pr(V_i = j \wedge U_{i|1})}{\Pr(V_i = j)}\\ &= \frac{\delta_j^k \pi_{ik|1}\varphi_{1i} + \pi_{ik|1}\pi_{ij|2}\varphi_{2i}}{\pi_{ij|1}\varphi_{1i} + \pi_{ij|2}\varphi_{2i}}.\end{aligned}$$

A.4 Specification and Estimation of the Finite Mixture Model

In order to estimate conditional probabilities of choices given the mode of choice one needs to take into account auxiliary information about voter's preferences, usually in form of predictor variables that describe the attributes of the choice alternatives (e.g. their policy positions, or how well the alternatives are evaluated) or the interaction of the alternatives with the characteristics of a voter (e.g. his or her social background). Thus, the conditional probability $\pi_{ij|h}$ can be considered as a function of voter i's characteristics and the attributes of the alternatives, collected into a matrix $\boldsymbol{X}_i$, and a coefficient vector $\boldsymbol{\alpha}$, i.e.

$$\pi_{ij|h} = \pi_{ji|h}(\boldsymbol{X}_i, \boldsymbol{\alpha}).$$

A particular straightforward specification of the conditional probabilities is the conditional logit

$$\pi_{ij|h} = \frac{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\alpha})}{\sum_{k \in C_{ih}} \exp(\boldsymbol{x}'_{ik}\boldsymbol{\alpha})}, \tag{11}$$

where $\boldsymbol{x}'_{ij}$ is the j-th row of matrix $\boldsymbol{X}_i$, but other specification may also be possible.[9] A model specification by equation (11) has the advantage that, if both j and k are

[9] The transpose of a vector $\boldsymbol{a}$ is written in this article as $\boldsymbol{a}'$.

viable alternatives (i.e. $j, k \in C_{i2}$), then the odds ratio of j being chosen relative of k being chose is independent from the mode of choice:

$$\frac{\pi_{ij|1}}{\pi_{ik|1}} = \frac{\pi_{ij|2}}{\pi_{ik|2}} = \exp([\boldsymbol{x}_{ij} - \boldsymbol{x}_{ik}]'\boldsymbol{\alpha})).$$

That is, if the preference order is reflected in these odds ratios, then it is independent from whether a choice is made in sincere or tactical mode.

The probability φ_{hi} that voter i chooses in mode h can be considered either to be constant for all individuals, i.e. $\varphi_{hi} = \varphi_h$, or to depend on the characteristics of the individual voter i and the context of competition of the district in which voter i makes her choice, collected into a vector $\boldsymbol{z}_i$, and a coefficient $\boldsymbol{\beta}_h$ i.e.

$$\varphi_{hi} = \varphi_{hi}(\boldsymbol{z}; \boldsymbol{\beta}_h).$$

A natural link between these probabilities and the independent variables would be a multinomial baseline logit link:

$$\varphi_{hi} = \begin{cases} \frac{\exp(\boldsymbol{z}'\boldsymbol{\beta}_{h-1})}{1+\sum_{g>1}\exp(\boldsymbol{z}'\boldsymbol{\beta}_{g-1})} & \text{for } h > 1 \\ \frac{1}{1+\sum_{g>1}\exp(\boldsymbol{z}'\boldsymbol{\beta}_{g-1})} & \text{for } h = 1 \end{cases} \quad \text{so that:} \quad \ln\frac{\varphi_{hi}}{\varphi_{1i}} = \boldsymbol{z}'\boldsymbol{\beta}_{h-1} \tag{12}$$

In case of only two modes, sincere and tactical, this simplifies to the binomial logit link

$$\varphi_{2i} = \frac{\exp(\boldsymbol{z}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{z}'\boldsymbol{\beta})} \text{ and } \varphi_{1i} = 1 - \varphi_{2i} \quad \text{so that:} \quad \ln\frac{\varphi_{2i}}{\varphi_{1i}} = \ln\frac{\varphi_{2i}}{1-\varphi_{2i}} = \boldsymbol{z}'\boldsymbol{\beta}$$

With this specification, the model parameters are either $\boldsymbol{\alpha}$ and φ_h (for $h = 1, \ldots, q$, where m is the number of modes) or $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_h$ (for $h = 1, \ldots, q-1$). To obtain maximum likelihood estimates for these parameters, one maximizes the log-likelihood function

$$\ell = \sum_{i,j} y_{ij} \ln \pi_{ij} = \sum_{i,j} y_{ij} \ln\left(\sum_{h=1}^{q} \pi_{ij|h}\varphi_{hi}\right)$$

for $\boldsymbol{\alpha}$ and φ_h or $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_h$, where y_{ij} is a dummy indicator that equals unity if voter i has chosen alternative j and equals zero otherwise. The maximization of the likelihood function can be achieved either by an expectation-maximization (EM) algorithm (Dempster et al. 1977) or a Newton-Raphson (NR) algorithm, or a combination of both. (The software that is used to compute the model estimates in the main text uses EM updates for the first few iterations and switches to NR updates in later iterations.)

The following paragraphs derive the EM and NR steps used for maximizing the log-likelihood. To simplify the derivation of the estimation procedure, first some notation is introduced. For the contribution of the observation from individual i to the Likelihood function we write

$$\mathcal{L}_{i|h} = \prod_j \pi_{ij|h}^{y_{ij}} \quad \mathcal{L}_i = \sum_h \varphi_{hi}\mathcal{L}_{i|h} = \sum_h \varphi_{hi} \prod_j \pi_{ij|h}^{y_{ij}}$$

(where for brevity the dependence on the observed responses y_{ij} is not written out) and for the contribution to the log-likelihood we write

$$\ell_{i|h} = \ln \mathcal{L}_{i|h} = \sum_{j\in C_{hi}} y_{ij} \ln \pi_{ij|h} = \sum_{j\in C_{hi}} y_{ij}\boldsymbol{x}_{ij}'\boldsymbol{\alpha} - \ln \sum_{j\in C_{hi}} \exp(\boldsymbol{x}_{ij}'\boldsymbol{\alpha}).$$

Also note that

$$\begin{aligned}\ln \varphi_{hi} &= \begin{cases} -\ln\left(1 + \sum_{k=1}^{q-1} \exp(\boldsymbol{z}_i'\boldsymbol{\beta}_k)\right) & \text{for } h = 1 \\ \boldsymbol{z}_i'\boldsymbol{\beta}_{h-1} - \ln\left(1 + \sum_{k=1}^{q-1} \exp(\boldsymbol{z}_i'\boldsymbol{\beta}_k)\right) & \text{for } h = 2, \dots, q. \end{cases} \\ &= \boldsymbol{z}_{hi}'\boldsymbol{\beta} - \ln\left(\sum_{k=1}^{q} \exp(\boldsymbol{z}_{ki}'\boldsymbol{\beta})\right)\end{aligned}$$

with

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \dots, \boldsymbol{\beta}_{q-1})'$$

(i.e. $\boldsymbol{\beta}$ is formed by "stacking" the component vectors $\boldsymbol{\beta}_k$) and

$$\begin{aligned} \boldsymbol{z}_{hi} &= \boldsymbol{0} && \text{for } h = 1 \\ \boldsymbol{z}_{hi} &= (\boldsymbol{0}', \dots, \underbrace{\boldsymbol{z}_i',}_{h-1\text{-th position}} \dots, \boldsymbol{0})' && \text{for } h > 1 \end{aligned}$$

(i.e. by forming a "dummy-interaction" vector with $\boldsymbol{z}_i$).

The log-likelihood function for the full data set can thus be re-written:

$$\ell = \sum_i \ln \mathcal{L}_i = \sum_i \ln\left(\sum_h \varphi_{hi}\mathcal{L}_{i|h}\right)$$

The first derivatives of the log-likelihood are therefore

$$\frac{\partial \ell}{\partial \boldsymbol{\alpha}} = \sum_i \frac{\partial}{\partial \boldsymbol{\alpha}} \ln \sum_h \mathcal{L}_{i|h}\varphi_{hi} = \sum_i \sum_h \frac{\mathcal{L}_{i|h}\varphi_{hi}}{\mathcal{L}_i} \frac{\partial \ell_{i|h}}{\partial \boldsymbol{\alpha}} = \sum_i \sum_h \mathcal{P}_{hi} \frac{\partial \ell_{i|h}}{\partial \boldsymbol{\alpha}}$$

and

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_i \frac{\partial}{\partial \boldsymbol{\beta}} \ln \sum_h \mathcal{L}_{i|h}\varphi_{hi} = \sum_i \sum_h \frac{\mathcal{L}_{i|h}}{\mathcal{L}_i} \frac{\partial \varphi_{hi}}{\partial \boldsymbol{\beta}} = \sum_i \sum_h \frac{\mathcal{L}_{i|h}\varphi_{hi}}{\mathcal{L}_i} \frac{\partial \ln \varphi_{hi}}{\partial \boldsymbol{\beta}} = \sum_i \sum_h \mathcal{P}_{hi} \frac{\partial \ln \varphi_{hi}}{\partial \boldsymbol{\beta}}$$

with

$$\mathcal{P}_{hi} = \frac{\mathcal{L}_{i|h}\varphi_{hi}}{\mathcal{L}_i} = \frac{\Pr(Y_i = \boldsymbol{y}_{ij}|T_i = h)\Pr(T_i = h)}{\Pr(Y_i = \boldsymbol{y}_{ij})} = \Pr(T_i = h|Y_i = \boldsymbol{y}_i),$$

where Y_i is the random vector with elements Y_{ij} and $\boldsymbol{y}_i$ is the vector of observations with elements y_{ij}. That is, the gradient of the marginal log-likelihood takes the form of a conditional expectation given $\boldsymbol{Y}_i = \boldsymbol{y}_i$.

This motivates an expectation-maximization (EM) algorithm (Dempster et al. 1977; Little and Rubin 2002; McLachlan and Krishnan 2007) that alternates between an E-step and an M-step, where the E-step consists in forming the Q-function

$$Q^{(s)} = \sum_i \sum_h \hat{\mathcal{P}}_{hi}^{(s)} \left(\ell_{i|h} + \ln \varphi_{hi}\right) \tag{13}$$

– whereby $\hat{\mathcal{P}}_{hi}^{(s)}$ is computed based on estimates $\hat{\boldsymbol{\alpha}}^{(s)}$ and $\hat{\boldsymbol{\beta}}^{(s)}$ from the previous iteration – and the M-step consists in maximizing $Q^{(s)}$ for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to obtain improved estimates $\hat{\boldsymbol{\alpha}}^{(s+1)}$ and $\hat{\boldsymbol{\beta}}^{(s+1)}$.

EM-algorithms are well-known to be numerically stable, yet slow to converge (McLachlan and Krishnan 2007). Fortunately, the information matrix of based on the marginal log-likelihood can relatively easy computed in the present set-up by

$$-\frac{\partial^2 \ell}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'} = -\sum_i\sum_h \mathcal{P}_{hi}\frac{\partial^2\ell_{i|h}}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'} - \sum_i\sum_h \mathcal{P}_{hi}\frac{\partial\ell_{i|h}}{\partial\boldsymbol{\alpha}}\frac{\partial\ell_{i|h}}{\partial\boldsymbol{\alpha}'} + \sum_i\left[\sum_h \mathcal{P}_{hi}\frac{\partial\ell_{i|h}}{\partial\boldsymbol{\alpha}}\right]\left[\sum_h \mathcal{P}_{hi}\frac{\partial\ell_{i|h}}{\partial\boldsymbol{\alpha}}\right]'$$

and analogously for $-\frac{\partial^2\ell}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}$, $-\frac{\partial^2\ell}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}'}$, etc. Therefore, the EM-algorithm can be improved upon by switching in later iterations to Newton-Raphson (or Fisher-scoring) iterations using this information matrix (Louis 1982).

The above discussion leads to the following algorithm to estimate the parameters of the finite mixture model:

1. In the first stage, initial estimates for $\boldsymbol{\alpha}$ are obtained by fitting a conventional conditional logit model to the vote decisions or vote intentions y_{ij}.

2. In the second stage, EM-iterations are performed based on the Q-function given by equation (13) with starting values for $\boldsymbol{\alpha}$ from the first stage and with zero starting values for $\boldsymbol{\beta}$.

3. After a few EM-steps, the algorithm switches to Newton-Raphson steps, which are iterated until the relative increase of the log-likelihood is smaller than $\epsilon = 10^{-7}$.

The algorithm also allows to compute standard errors from the square roots of the inverse of the information matrix, which are is computed for the Newton-Raphson steps.

For the purposes of this paper, the algorithm is implemented in the statistical programming language R (R Core Team 2013). For producing the estimates discussed in the paper, only a few Newton-Raphson steps were needed throughout and the run-time was generally just a few seconds on a contemporary desktop computer.

B Details about the simulation study about the performance of the proposed method

The main text of the paper mentions a simulation study in which examines whether not knowing or ignoring relevant strategic incentives will lead to a bias in the estimated frequency of strategic voting if the method proposed in the paper is applied. For the purpose of this simulation study artificial data sets were created that correspond to 3,000 voters each evenly distributed into 150 voting districts where they faced five alternatives to choose from. The attributes of the alternatives relevant for voters' preferences were represented by a single variable and its impact on the choices by a single coefficient $\boldsymbol{\alpha}$. This independent variable consisted of policy distances between the voters and the parties. The parties had the positions -2, -1, 0, 1, and 3 in a uni-dimensional policy space, while the distribution of voters' ideal points had a bi-modal distribution created by a mixture of two normal distributions with $\mu = -1$ and $\mu = 1$ and $\sigma^2 = .3$. The strategic incentives and the distinction between viable and non-viable alternatives in each district was created by parties' expected vote shares having a Dirichlet distribution with parameter vector $\boldsymbol{\theta} = (1, 3.4, 1.5, 0.8)$. In each of the artificial districts, the two parties with the relatively largest vote shares were designated as electorally viable and the others were designated as nonviable. Two variables representing strategic incentives were created: Closeness of competition and distance from contention. Closeness of competition was computed from the difference between the two largest parties (in terms of expected vote share) as:

$$\text{Closeness} = \frac{1}{1 + \exp(p_1 - p_2)}$$

where p_1 is the vote share of the first-placed party and p_2 the vote share of the second-placed party. Distance from contention was computed as

$$\text{Distance from Contention} = |p_1 - p_3|$$

where p_3 is the vote share of the third-placed party. For each artificial voter i the probability that he or she would vote in sophisticated mode was computed as

$$\varphi_{2i} = \Pr(M_i = 2) = \frac{\exp(\zeta)}{1 + \exp(\zeta)}, \qquad \zeta = \beta_0 + \beta_1 \text{Closeness}_i + \beta_2 \text{Distance from Contention}. \tag{14}$$

Based on these probabilities, the mode of choice for each voter was generated as a binary random number. For each voter i a party choice in expressive mode was generated as an integer between 1 and 5 with probabilities for each number given by

$$\Pr(U_{i|1} = j) = \frac{\exp(\alpha x_j)}{\sum_{k\in\{1,\ldots,5\}} \exp(\alpha x_k)}, \qquad j = 1, \ldots, 5.$$

Further, a party choice in sophisticated mode was generated for each voter as an integer from the set of a those two parties that had the largest (simulated) vote shares in the voting district of the voter

$$\Pr(U_{i|2} = j) = \frac{\exp(\alpha x_j)}{\sum_{k\in C_{2i}} \exp(\alpha x_k)}$$

where C_{2i} is the consideration set that contained these two parties. Based on the value of M_i the "observed" or "actual" simulated vote for each voter was generated according to

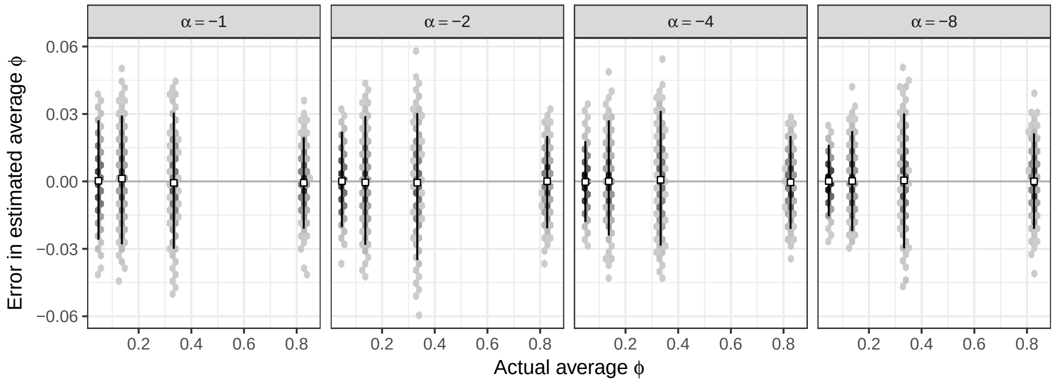
$$V_j = \begin{cases} U_{i|1} & \text{if } M_i = 1 \\ U_{i|2} & \text{if } M_i = 2 \end{cases}.$$

Those voters for which V_j was unequal to $U_{i|1}$ where counted as "strategic voters". To the generated observed choices, two variants of the finite mixture discrete choice model were fitted: First, a variant was fitted that did not take into account the strategic incentives Closeness of Competition and Distance from Contention, so that the probability of a vote in sophisticated mode was given by a single parameter:

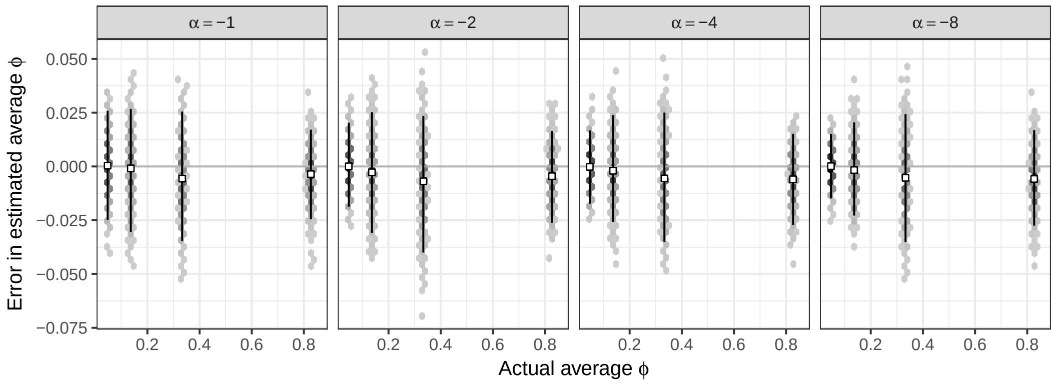
$$\varphi_{2i} = \Pr(M_i = 2) = \varphi.$$

Secondly, a variant was fitted were the probability of a choice in strategic mode was fully specified according to equation (14). From both model fits, a couple of quantities were computed for each voter: the probability of voting in sophisticated mode (i.e. $\hat{\varphi}_{2i}$), the prior probability of a strategic vote according to equation (9), and the posterior probability of a strategic vote according to equation (10). The averages of these quantities were then examined as estimates of the proportion of voters choosing in sophisticated mode and of the proportion of strategic votes. This procedure was repeated 1,000 times (one artificial data set and one application of the method in each replication), with various settings for the parameters α and β_1, namely $\alpha = -1, -2, -4, -8$, $\beta_0 = -3$ and $\beta_1 = \beta_2 = 0, 1, 2, 4$.

Figure 6 shows simulated distribution of errors in the estimates of the proportion of voters in sophisticated mode obtained from the fully specified finite mixture model and the simplified model without covariates of the mode of choice. It turns out that the estimate for φ tends to have a small downward bias if covariates are omitted and if the proportion of voters in sophisticated mode is high. This bias seems however relatively small in comparison to the overall sampling error of the estimator. In fact the intervals that contain 95% mass of the distribution of the estimation error always contain the zero line.



(a) First variant: Fully specified finite mixture model



(b) Second variant: Finite mixture model without information about predictors of strategic incentives

Figure 6 Hexbin plots of the difference between actual and estimated average probabilities of choices in sophisticated mode (φ_{2i}) for various settings for the influence of party evaluations (α) and the influence of strategic incentives on sophisticated votes (β). The different settings of the parameter β are reflected in the location of the different "clouds" that appear in the diagrams, the shading of the hexagons indicates how many data points are contained in them. The little squares correspond to the average errors of the estimates for the various settings of the true parameters. The vertical lines correspond to the 0.025 and 0.975 quantiles of the distribution of the errors, and thus correspond to 95% confidence intervals.

Figure 7 shows the distribution of errors of estimation of the proportion of strategic voting in terms of empirical Bayes prior probabilities, while Figure 8 shows the distribution of errors of estimation in terms of empirical Bayes posterior probabilities. They show so-called "hexbin" plots, which are a more effective way of representing the bivariate distribution of large amounts of data points (Carr et al. 1987). Each of the "clouds" that can be discerned in the diagrams correspond to a particular settings of the ("true") parameter values and represent 1,000 data points. That is, each diagram represents $4 \times 1,000 = 4,000$ replications of the simulation run, where artificial data are generated and the finite mixture model is fitted to the data. Note that the vertical axis has a different scale than the horizontal axis, which magnifies the estimation errors.

Both figures lead to the same conclusion: When prior or posterior probabilities from the fully specified model are used, then the distribution of errors is almost perfectly symmetrically distributed around zero, so (almost) no bias in the estimates seems to be present in the estimates. If the simplified model without covariates is used, the averages of estimation errors are a little bit more off zero, but the zero line is still enveloped by a the interval that contains 95% of the mass of the distribution. Furthermore, for values of α that are larger in size (i.e. in absolute value) the dispersion of the estimation errors is smaller and also the discrepancy of the mean values of the errors is closer to zero. This suggests that this average discrepancy is not a genuine bias but mostly a consequence of sampling error.

C Details about the application of the method to the 2010 British Election Study

C.1 Reconstruction of the District-Level Context of Competition

If voters form expectations adaptively from past constituency results one needs the vote shares of the parties in each constituency in the previous 2005 election. Unfortunately, the constituency boundaries were changed between 2005 and 2010, so any results from 2005 for the 2010 parliamentary constituencies are notional only. These results where obtained from Pippa Norris' Website on UK parliamentary elections (Norris 2010). If voters form expectations based on current information one needs both the past results in the constituencies and the current poll performance of the parties through the campaign. In case of the 2010 election this means that 2005 constituency results form a baseline which is updated by the swing exhibited in the poll results, that is, the difference between the 2005 general election results and the current poll results. The poll results were obtained from the website of the UK Polling Report (Wells 2010). For the construction of the expected district level vote shares, we use the assumption of a "uniform swing" that is commonly used in the British electoral studies literature for projecting constituency results and vote shares (for example BBC 2010a,b): Here one needs to distinguish between the three major parties – Conservatives, Labour, and the Liberal Democrats – and all other parties, since polling institutes in general reported individual proportions of voting intentions only for these three parties and collapsed the vote intentions for the other parties (SNP, Plaid Cymru, the Greens, BNP, UKIP, and various minor parties) into one category “Other”. The updated constituency results are then computed as follows: Let n_i ($i = 1, 2, 3$) be the national-level vote shares of the three parties gained in the 2005 parliamentary election and p_{it} the shares in vote intentions for these parties according to the polls at time t. Further, let n_4 be the vote shares of all other parties gained in the 2005 election put together and p_{4t} the vote intentions in the polls at t for all the other parties put together. Then the poll-based uniform swing is, for $i = 1, 2, 3, 4$,

$$s_{it} = p_{it} - n_i.$$

This uniform swing is added – without any weighting – to the constituency results of the three major parties in the previous election of 2005. Let $c_{ij,2005}$ be the vote share of any of the three major parties ($i = 1, 2, 3$) in constituency j in the election of 2005. Then the projected vote share at t is

$$\hat{c}_{ijt} = s_{it} + c_{ij,2005}.$$

The polls only give information about the support for all “other” parties taken together, therefore this procedure however reaches its limits when other parties with local strongholds are considered. This applies to the SNP, which fielded candidates only in Scotland, to Plaid Cymru, which fielded candidates only in Wales, and to the Greens, which were particularly successful in a single constituency, Brighton. One could combine all the other parties at the constituency level in the same way as in the polling data, to get more or less useful projections of the distribution of seats in Westminster, but the analysis of strategic voting, this seems less appropriate. For this reasons the swing for the “other” parties was added to the strongest of the other parties in the respective constituencies. For most Scottish constituencies this means that only the 2005 results of the SNP were assumed to be updated by the swing, for most Welsh constituencies this was applied to Plaid Cymru and in the Brighton constituency this applied to the Greens.

There are some more complications to consider here. In the Northern Ireland constituencies, none of the three major parties was electorally relevant. Thus Northern Ireland was excluded from this procedure. Further, both in 2005 and 2010, the respective Speaker of the House was not challenged by any of the other three major parties. In 2005 the Speaker was Michael Martin (now Baron Martin of Springburn) of the Labour party with Glasgow North East as constituency and in 2010 the Speaker was John Bercow of the Conservatives with Buckingham as constituency. Thus to construct accurate projected vote shares was impossible for Glasgow North East and inappropriate for Buckingham. But since Glasgow North East was a safe seat for Labour as was Buckingham for the Conservatives, the application of the uniform swing procedure was inconsequential for the projection of the seat shares in Westminster. Nevertheless, for the analysis of strategic voting, these constituencies were not considered.

C.2 The British Election Study Data Set

Data used in this paper come from the British Election Study 2009/10, led by Harold Clarke, David Sanders, Marianne Stewart, and Paul Whiteley as principal investigators. The pre-election and post-election surveys focused on here used computer assisted personal interviews. The pre-election survey was conducted by TNS-BMRB from January 23 to April 18, 2010 and had a response rate of 56% with 1935 completed interviews. The post-election survey was conducted by the same firm from May 7 to September 5, 2010, with a re-contact sample from which 1498 interviews were completed (response rate 77%) and a top-up sample from which 1577 interviews were completed (response rate 49%). For a full report see Howat et al. (2011)

As argued in the paper, uncovering strategic voting requires a way to reconstruct voters' nominal preferences, since the former is defined as departures from the latter. Such nominal preferences have, for lack of better alternatives, to be generated with the help of predictors other than those relevant for the strategic context of the choice. For the present purpose, such predictors do not need to be substantially interesting. Rather the most "tautological" predictors are best suited, simply because good predictions is all that is needed at this stage. In the 2010 British Election Study such predictors are the feelings that respondents have (or state to have) towards the parties and their leaders. Nevertheless, while it is plausible that these have a strong influence on choices, they may not the only relevant factors.

The method of selecting predictors of party preferences is oriented at the classic notion of the "funnel of causality", in which structural variables are the more distal factors, the influence of which is mediated by party identification, issue positions, and party and leader affect as more proximal factors. This method proceeds by first expanding a baseline model step-by step, moving from more distal factors to more proximate ones, where in each step it is checked whether the addition of a factor improves the prediction of party choice by means of likelihood ratio tests. In this stage, predictors are only kept in the model if the LR test indicates a statistically significant improvement. After this step-wise extension of the set of predictors the resulting maximal model is pruned, moving from more distal factors to more proximate ones, by dropping predictors if pruning does not lead to a statistically significant loss in goodness of fit.

The baseline model is a conditional logit model with dummy predictors for the Labour Party, the Liberal Democratic Party, Scottish National Party, Plaid Cymru, Greens, UKIP, BNP, and other parties. Apart from the fact that a conditional logit model allows for the choice set to vary between England, Scotland, and Wales (SNP runs only in Scotland, Plaid Cymru only in Wales), this model is equivalent to a conventional multinomial baseline logit model. This baseline model is then extended by structural predictors, namely class and religion. The influence of these structural variables is implemented in the model by interaction terms of dummy variables for their categories with dummy variables for the Conservative and the Labour party. The effect of respondents' social class is represented by three dummies, for the working class, the business, and professional classes, with all other occupational classes (i.e. mostly lower white-collar) as baseline category. It is based on respondents' current or former occupation (where only 2.3 percent of the respondents stated never to have worked). Religion is represented by three dummy variables for "Church of Scotland", "Roman Catholic", "Other Christian", "Other Non-Christian" membership, and "No Religion", with "Church of England" as baseline category. These dummy variables are constructed from two questions asked in the BES, whether respondents are member in any religious group and which religious group they are member of (if applicable). The response categories for the second question are quite numerous, but many of them sparsely populated, so that several categories where combined.

The squared issue distance is constructed from the positions that respondents state to have and view parties have on the issue dimensions concerning public spending and civil liberties. Positions were measured using an 11-point scale (from 0 to 10). The ends of the spending scale were anchored by the statements "Government should cut taxes a lot and spend much less on health and social services" (0) and "Government should increase taxes a lot and spend much more on health and social services" (10). The ends of the civil liberties scale were anchored by the statements "Reducing crime more important" (0) and "Rights of accused more important". Before construction of the distances the issue scaled were normalized to the range from zero to unity. Unfortunately, respondents were only asked about their perceptions with regards to the Conservative Party, Labour Party, the Liberal Democrats, SNP, and Plaid Cymru, but not with regards to the smaller parties Greens, UKIP, and BNP. For these parties, their perceived positions were imputed to be at the centre of the scale.

Party identification is based on responses to the survey question "Generally speaking, do you think of yourself as Labour, Conservative, Liberal Democrat, (Scottish National/Plaid Cymru) or what?" (where references to Scottish National Party were made respectively if the respondent lived in Scotland and to Plaid Cymru if they lived in Wales). Here not only an identification with any of the parties mentioned in the question was recorded, but also with the Greens, UKIP, or the BNP. with a particular party.

Respondents' feelings towards the parties and the leaders are measured on an 11-point scale (from 0 to 10). Questions about feelings towards the parties were introduced by the requests "On a scale that runs from 0 to 10, where 0 means strongly dislike and 10 means strongly like, how do you feel about[PARTY NAME]?" and "And how do you feel about [PARTY NAME]?", and respondents were presented with the Labour Party, the Conservative Party, the Liberal Democrats, the Green

Table 1 Selection of predictors for voters' party preferences, BES 2010 pre-election wave

(a) Improvement of fit by including causally proximal predictors

	Resid. DF	Deviance	DF	χ^2	p-value
Baseline model	4806	2912.7			
+ Class	4797	2875.9	9	36.8	0.000
+ Religion	4782	2831.7	15	44.2	0.000
+ Region	4767	2792.7	15	39.0	0.001
+ Issue distance	4765	2529.1	2	263.6	0.000
+ Party feelings	4764	1174.8	1	1354.2	0.000
+ Leader feelings	4763	1165.0	1	9.9	0.002

(b) Loss of fit by dropping causally distal predictors

	Resid. DF	Deviance	DF	χ^2	p-value
Full model	4763	1165.0			
− Class	4772	1184.0	9	19.0	0.025
− Religion	4787	1188.5	15	4.4	0.996
− Region	4802	1211.2	15	22.7	0.090
− Issue distance	4803	1219.7	1	8.5	0.004
− Party feelings	4804	1885.5	1	665.8	0.000
− Leader feelings	4805	2781.3	1	895.9	0.000

Party, UKIP, and the BNP. In Scotland they were additionally presented with the SNP and in Wales additionally with Plaid Cymru. Questions about feelings towards party leaders were asked only with respect to the leaders of Labour (Gordon Brown), the Conservatives (David Cameron), and the Liberal Democrats (Nick Clegg), additionally in Scotland with respect to the SNP leader (Alex Salmond) and in Wales with respect to the Plaid Cymru leader (Ieuan Wyn Jones). The question about party leaders was asked as "Now let's think about party leaders for a moment. Using a scale that runs from 0 to 10, where 0 means strongly dislike and 10 means strongly like, how do you feel about[PARTY LEADER]?" and "And how do you feel about [PARTY LEADER]?" In the long format of the data, feelings towards the parties and towards party leaders are represented by a single variable each. Since respondents feelings towards the leaders of the smaller parties Greens, UKIP and BNP were not were not probed in the British Election Study interviews the scale values were imputed with the value zero. For the analysis the feeling scale was normalized to the range from zero to unity.

C.3 Selecting Predictors for Voters' Party Preferences

In the application to the 2010 UK general election, the selection of predictors for nominal preferences is led by the common notion of a "funnel of causality", which leads from more distal factors, i.e. social-structural variables, via intermediate factors, such as issue positions, to the most immediate evaluation of the parties

Table 2 Selection of predictors for voters' party preferences, BES 2010 post-election wave

(a) Improvement of fit by including causally proximal predictors

	Resid. DF	Deviance	DF	χ^2	p-value
Baseline model	8249	4911.3			
+ Class	8240	4850.2	9	61.1	0.000
+ Religion	8225	4778.0	15	72.2	0.000
+ Region	8210	4709.8	15	68.2	0.000
+ Issue distance	8208	4282.2	2	427.6	0.000
+ Party feelings	8207	2274.5	1	2007.7	0.000
+ Leader feelings	8206	2250.2	1	24.3	0.000

(b) Loss of fit by dropping causally distal predictors

	Resid. DF	Deviance	DF	χ^2	p-value
Full model	8206	2250.2			
− Class	8215	2268.1	9	17.9	0.036
− Religion	8230	2287.0	15	18.9	0.219
− Region	8245	2307.9	15	20.9	0.141
− Issue distance	8246	2308.0	1	0.1	0.752
− Party feelings	8247	3064.6	1	756.6	0.000
− Leader feelings	8248	4668.2	1	1603.6	0.000

and their leaders. Tables 1 and 2 shows how this notion is applied to the search for a well-fitting, yet parsimonious model of nominal preferences: In a first round, a baseline model with only dummy predictors for the parties is extended step-wise by class, religion, issue distance, and feelings towards the parties and their leaders. The impact of class is represented by coefficients of interaction terms of dummy variables for social class and the dummies for the Labour Party and the Conservative Party.[10] Similarly, the impact of religion is represented by coefficients of interaction terms of dummy variables for religious and non-religious categories and dummies for the Labour Party and Conservative Party.[11] The impact of issue distance is represented by coefficients of the squared distance between each voter and each of the parties in his or her choice set on two 0–10 issue scales, which contrast fighting crime to protecting defendants' rights and contrast increasing government spending to tax reduction. Respondents' feelings towards each of the parties is measured by a rating scale ranging from 0 ("Strongly dislike") to 10 ("Strongly like") as is the feeling towards the respective parties' leaders. The likelihood ratio tests in the upper part of the table indicate that each more proximal factor adds to the overall goodness of fit and the improvement is greatest when

[10] Four social classes are distinguished: "Manual Workers" (e.g. factory workers etc.), "Non-Manual Workers" (office employees etc.), "Professionals" (lawyers, doctors, scientists, teachers etc.), and "Business" (managers and owners of shops and enterprises). The non-manual workers figured the baseline category of the dummy coding.

[11] The religion variable had six categories: "Church of England", "Church of Scotland", "Roman Catholic", "Other Christian", "Other Non-Christian", and "No Religion".

Table 3 Parameter estimates of the finite mixture model of strategic voting in the 2010 general election of the UK House of Commons

	Pre-election	Post-election
Dummy for 'other' parties	-0.491^{***}	-0.864^{***}
	(0.137)	(0.113)
Taxation issue distance	-2.919^{***}	-0.529
	(0.768)	(0.587)
Rights issue distance	-2.734^{***}	-1.188^{**}
	(0.482)	(0.424)
Feelings toward party	11.439^{***}	10.481^{***}
	(0.489)	(0.367)
Feelings toward party leaders	2.262^{***}	1.953^{***}
	(0.315)	(0.266)
Intercept in choice mode equation	-1.702^{***}	-1.502^{***}
	(0.220)	(0.157)
Log-likelihood	−701.7	−1253.6
Deviance	1403.3	2507.2
N	1308	2171

Significance: $*** \equiv p < 0.001$; $** \equiv p < 0.01$; $* \equiv p < 0.05$

party identification is added. The lower part of the table reports the results of likelihood ratio tests about the loss of model fit by dropping the more distal factors from the model. These results indicate that social structural factors and issue distance are redundant as predictors of party choice once feelings (towards parties and leaders) are taken into account. Of these proximal factors none is redundant, because dropping each of them leads to a statistically significant loss of model fit. This does not mean that social structural factors are irrelevant as predictors for voters' party preference of course, but that their impact is mediated by voters' policy distances from the parties, and by their feelings towards the parties and their leaders.

Based on the results shown in Tables 1 and 2, the choice predictors used in the model of strategic voting discussed in the main text are issue distances and respondents' feelings towards the parties and their leaders. The estimates of the parameters of the resulting finite mixture model of strategic voting in the 2010 general election of the UK House of Commons are shown in Table 3.

As already mentioned in the main text, the estimated proportion of strategic voting depends on who well voters' "sincere" preference about parties and/or candidates can be reconstructed with the appropriate predictor variables. This is illustrated by Figure 9, which shows the estimated proportion of strategic voting obtained from posterior probabilities of strategic votes based on the models compared in Tables 1 and 2, as well as based on the final model the parameters of which are shown in Table 3. When a baseline model (with only party dummies) is used or a model that contains only social structural variables, but not voters' (stated) feelings towards the parties and their leaders, then the estimates for the proportion of strategic voting are at least 15 per cent and can get as large as 30 per cent (in the post-election wave of the 2010 British Election Study). Clearly, when the reconstruction of voters' is less then perfect then more deviations form the reconstructed preferences can be observed. If predictors are used that are close to the actual voting decision and are most likely the predictors that are the closest to voters' preferences in the causal change from more distal predictors to the most proximal predictors, then the estimated proportion of strategic voting can fall down to about 5 per cent (in the post-election wave of the 2010 BES) or even to 2.5 per cent (in the pre-election wave of the 2010 BES). Since this "full model" contains dummy-variables for the major parties in Great Britain, it may over-fit the observed votes and therefor may be biased downward in terms of the estimated proportion of strategic vote. As also can be seen in Figure 9, a more parsimonious model, the estimates of which are shown in Table 3 leads to somewhat higher estimates of strategic voting, namely about 5 per cent in the pre-election wave and about 7 per cent in the post-election wave.

C.4 Confidence intervals for quantities of interest

Figure ?? in the main part of the paper shows estimates of the percentage of strategic voting based on the often-used stated reasons method and based on the method proposed in the paper. In addition to the estimates it shows also 95 per cent confidence intervals. For estimates based on the stated-reasons approach confidence intervals are based on the common assumption of a binomial distribution of positive outcomes (i.e. strategic voting). The confidence limits of the proportion of strategic votes in this case are based on the Clopper-Pearson interval (Clopper and Pearson 1934), which can be interpreted in terms of quantiles of two Beta distributions with parameters constructed from the sample counts. That is, if k/n is the proportion of strategic votes in the sample, the appropriate binomial distribution has the probability mass function

$$f_{\text{Bin}}(k; p, n) = \binom{n}{k} p^k (1-p)^{n-k}$$

where p is the proportion in the population. A Beta distribution with parameters α and β has the density function

$$f_{\text{Beta}}(p; \alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} p^{\alpha} (1-p)^{\beta}$$

The lower limit of a 95% confidence interval is the 2.5% quantile of a Beta distribution with parameters $\alpha = k$ and $\beta = n - k + 1$, while the upper limit is the 97.5% quantile of a Beta distribution with parameters $\alpha = k + 1$ and $\beta = n - k$.

Figure ?? also shows estimates of and confidence intervals for the proportion of strategic voting based in the method proposed in the paper. The proportion of strategic voting was estimated in two variants: by the sample average of respondents' prior probabilities of voting strategically, based on equation (9), and by the sample average of respondents' posterior probabilities of voting strategically, based on equation (10). Since these proportions are computed from model estimates the construction of confidence intervals cannot just be based on the assumption of a binomial distribution. Instead, a form of parametric bootstrap is used for the construction of confidence intervals which is inspired by King et al. (2000). The method is based on the assumption that the maximum likelihood estimator of the parameters of the finite mixture model is asymptotic normal (Casella and Berger 2002). It starts with the estimate $\hat{\boldsymbol{\theta}}$ of the parameter vector of the finite mixture model, where $\hat{\boldsymbol{\theta}}$ is composed of the maximum likelihood estimates of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ that appear in equations (11) and (12). Based on the estimate of the parameter vector, an estimate of the Fisher-Information matrix $\mathcal{I}(\boldsymbol{\theta})$ of the model (i.e. the matrix of second derivatives of the log-likelihood function with respect to $\boldsymbol{\theta}$) is computed, which in turn is inverted to obtain the variance matrix of $\hat{\boldsymbol{\theta}}$. Then parameter values $\boldsymbol{\theta}^{*(r)}$ are simulated with a multivariate normal distribution with mean $\hat{\boldsymbol{\theta}}$ and variance $\hat{\mathcal{I}}(\boldsymbol{\theta})$, i.e.

$$\boldsymbol{\theta}^{*(r)} \sim N\left(\hat{\boldsymbol{\theta}}, \hat{\mathcal{I}}(\boldsymbol{\theta})^{-1}\right)$$

From each simulated parameter vector $\boldsymbol{\theta}^{*(r)}$ prior probabilities

$$p^{*(r)}_{\text{prior},i} = \varphi^{*(r)}_{2i} \sum_{j\in\mathcal{S}_i} \left(1 - \pi^{*(r)}_{ij|1}\right) \pi^{*(r)}_{ij|2}$$

and posterior probabilities

$$p^{*(r)}_{\text{posterior},i} = \frac{\left(1 - \pi^{*(r)}_{ij|1}\right) \pi^{*(r)}_{ij|2} \varphi^{*(r)}_{2i}}{\pi^{*(r)}_{ij|1} \varphi^{*(r)}_{i1} + \pi^{*(r)}_{ij|2} \varphi^{*(r)}_{i2}}$$

of a strategic vote are computed for each individual i, where

$$\pi^{*(r)}_{ij|h} = \frac{\exp\left(\boldsymbol{x}'_{ij}\boldsymbol{\alpha}^{*(r)}\right)}{\sum_{k\in C_{ih}} \exp\left(\boldsymbol{x}'_{ik}\boldsymbol{\alpha}^{*(r)}\right)} \text{ and } \varphi^{*(r)}_{2i} = \frac{\exp\left(\beta^{*(r)}\right)}{1 + \exp\left(\beta^{*(r)}\right)}.$$

From these simulated prior and posterior probabilities, simulated proportions of strategic voting are computed for each r. Finally, the 2.5- and 97.5-percentiles are used as confidence interval limits for the estimates of strategic voting based on prior and posterior probabilities.

D A simulation study of the Alvarez-Nagler method

The main part of the paper mentions a simulation study that shows that the Alvarez-Nagler method may lead to biased estimates of the proportion or percentage of strategic votes. This simulation study is based on data generated from a model of the kind on which the Alvarez-Nagler method is based. In so far it is designed to assess the Alvarez-Nagler "by its own criteria", with the sole exception that the simulation allows, in contrast to the original Alvarez-Nagler method, for genuinely sincere or expressive voters. This is to make it possible to examine the bias in the estimated proportion of strategic voters that is created by the presence of expressive voting.

The simulation study consists of $4 \times 4 \times 5 \times 1,000$ artificial elections where three parties compete in 150 districts for votes. The parties have positions $+1$, -1, and 0 on an abstract political dimension, e.g. an ideological left-right axis. Their vote shares in the voting districts have a Dirichlet distribution with parameter values 3.4, 3.3, and 1.5, so that expected values of the vote shares are 0.415, 0.402, and 0.183. These vote shares are the basis of the strategic incentives facing the voters in the districts. For each electoral district i the vote share of party $j = 1, 2, 3$ is denoted by p_{ij}, and for each district i and party j a "distance from contention" w_{1ij} and "closeness of competition" between the other two parties is computed from the vote shares according to (see Alvarez and Nagler 2000):

$$\begin{aligned}
w_{1i1} &= |\max(p_{i2}, p_{i3}) - p_{i1}|, \quad & w_{2i1} &= \frac{1}{1 + |p_{i2} - p_{i3}|}, \\
w_{1i2} &= |\max(p_{i1}, p_{i3}) - p_{i2}|, \quad & w_{2i2} &= \frac{1}{1 + |p_{i1} - p_{i3}|}, \\
w_{1i3} &= |\max(p_{i1}, p_{i2}) - p_{i3}|, \quad & w_{2i3} &= \frac{1}{1 + |p_{i1} - p_{i2}|}.
\end{aligned}$$

Note that 1 is added to the absolute difference of the vote shares in the denominator, so that w_{1ij} will remain finite even if the vote shares are equal. In the unlikely case that all three parties' vote shares are equal, the distance from contention value w_{1ij} and the closeness of competition value w_{2ij} will be equal for all parties $j = 1, 2, 3$ so that no strategic incentive favouring any of the parties exists.

For each election, a sample 3000 artificial voters is drawn (evenly distributed into the 150 electoral districts). The voters' positions on the abstract political dimension has a bi-modal distribution, constructed from a mixture of two normal distributions with means -1 and $+1$ and variance 0.3. The squared distances between voters' and parties' positions are divided by 7 so that they are mostly between 0 and 1 and used as predictors x_{ij}. For each voter an expressive and a sophisticated vote is simulated according to

$$\Pr(V_{\text{sinc},i} = j; \alpha) = \frac{\exp(\alpha x_{ij})}{\sum_{k \in \{1,2,3\}} \exp(\alpha x_{ij})} \tag{15}$$

and

$$\Pr(V_{\text{soph},i} = j; \alpha, \beta) = \frac{\exp(\alpha x_{ij} + \beta w_{1ij} + \beta w_{2ij} + \beta w_{1ij} w_{2ij})}{\sum_{k\in\{1,2,3\}} \exp(\alpha x_{ik} + \beta w_{1ik} + \beta w_{2ikj} + \beta w_{1ilj} w_{2ikj})}, \tag{16}$$

That is, an expressive vote disregards the strategic incentives for voter i, w_{1ij} and w_{2ij}, to vote for party j, while a sophisticated vote takes these incentives into account. Actual choices are a combination of expressive and the sophisticated choices: For each voter a dummy variable was sampled with $\Pr(T_i = 1) = \varphi$ and the actual vote is computed as $Y_i = Y_{\text{sinc},i}(1 - T_i) + Y_{\text{soph},i} T_i$. As a consequence, each voter is a sophisticated voter with probability φ and a an expressive voter with probability $1 - \varphi$. Each voter for which $Y_i = Y_{\text{sinc},i}$ is counted as a sincere voter, whereas each voter for which $Y_i \neq Y_{\text{sinc},i}$ is counted as a strategic voter.[12]

The Alvarez-Nagler method of estimating the rate of tactical voting was applied to the simulated votes Y_i as follows:

1. A conditional logit model with the specification

   $$\Pr(V_i = j; \alpha, \beta_1, \beta_2, \beta_3) = \frac{\exp(\alpha x_{ij} + \beta_1 w_{1ij} + \beta_2 w_{2ij} + \beta_3 w_{1ij} w_{2ij})}{\sum_{k\in\{1,2,3\}} \exp(\alpha x_{ik} + \beta_1 w_{1ik} + \beta_2 w_{2ikj} + \beta_3 w_{1ilj} w_{2ikj})},$$

   is fitted to the data. It differs form the model that is used to generate the sophisticated votes only in so far as the coefficients β_1, β_2, and β_3 are allowed to differ.

2. Based on the model estimates α, β_1, β_2, and β_3, choice predictions are created: For each individual i the predicted choice $\hat{v}_i$ is the alternative j with

   $$\hat{\alpha} x_{ij} + \hat{\beta}_1 w_{1ij} + \hat{\beta}_2 w_{2ij} + \hat{\beta}_3 w_{1ij} w_{2ij} > \hat{\alpha} x_{ik} + \hat{\beta}_1 w_{1ik} + \hat{\beta}_2 w_{2ik} + \hat{\beta}_3 w_{1ik} w_{2ik}$$

   or equivalently,

   $$\Pr(V_i = j; \hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) > \Pr(V_i = k; \hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$$

   for $j = \hat{v}_i$ and any other alternative $k \neq \hat{v}_i$.

3. Counterfactual choices are generated by first setting w_{1ij} and w_{2ij} ($j = 1, 2, 3$), or equivalently, β_1, β_2, and β_3 to zero and creating predictions: For each individual i the predicted choice $\hat{v}_{\text{cnft},i}$ is the alternative j with

   $$\hat{\alpha} x_{ij} > \hat{\alpha} x_{ik}$$

   or equivalently,

   $$\Pr(V_i = j; \hat{\alpha}, 0, 0, 0) > \Pr(V_i = k; \hat{\alpha}, 0, 0, 0)$$

12 Note that $Y_i = Y_{\text{soph},i}$ does not imply that voter i is a tactical voter, because this does not preclude the case that $Y_{\text{sinc},i} = Y_{\text{soph},i}$ and therefore $Y_i = Y_{\text{sinc},i}$.

for $j = \hat{v}_i$ and any other alternative $k \neq \hat{v}_i$.

4. All individuals i with

$$\hat{v}_i = \hat{v}_{\text{cnft},i}$$

are counted as strategic voters. And the proportion of these individuals from among all simulated voters is taken as the estimate of the proportion of strategic voting.

This procedure is repeated 1,000 times for each combination of parameter settings, with any combination of "true" parameter values $\alpha = -1, -2, -4, -8$, $\beta = -1, -2, -4, -8$, and $\varphi = 0, 0.25, 0.5, 0.75, 1$. That is, with $\varphi = 1$ the simulation study also includes the situation none of the simulated voters is an expressive voter, as assumed by the Alvarez-Nagler method, as various situations with expressive voters present, as implicitly ruled out by the method.

Figure 10 compares true and estimated values of the parameter beta, which represents the influence of strategic incentives on sophisticated voting choices. The diagram makes quite clear that the presence of expressive voting, manifested in values $\varphi < 1$, leads the estimates of the influence of one of the strategic incentives to be biased downwards in absolute value: On average the size (absolute value) of the estimates of β_1 are smaller than their true values if $\varphi < 1$. This downwards bias tends to be the larger the smaller φ is: The dots that represent the averages of the estimates are farther way from the "diagonal of identity". In the same way as Figure 10 it is also possible to compare the estimates and the true values of the coefficients β_2 and β_3, which correspond to the influence of some further strategic incentives: the influence of a the distance from contention of a party (β_2) and the interaction of closeness of competition and distance from contention.

Figure 11 shows hexbin plots that compare estimated proportions of strategic voting with the (artificial) "true" proportions. Each "cloud" in the hexbin plot corresponds to a particular setting of the (true) parameters α, β, and φ and represents 1,000 replications of the simulation. Consequently, each of the hexbin plots in the diagram represents 5,000 data points. They show that the Alvarez-Nagler method leads to both positively and negatively biased estimates of the proportion of strategic voting. The bias seems to be relatively small in size when votes are relatively "informative" about the variables that predict them, that is, when the values of the parameters α and $\beta_1 = \beta_2 = \beta_3 = \beta$ are relatively large in size. Surprisingly, the size of the bias does not decrease with φ, the probability that voters cast a sophisticated vote. Even if *all* of the artificial voters cast sophisticated votes, the Alvarez-Nagler method leads to estimates of the proportion of strategic voting that are biased downwards, in contrast to the findings obtained from Figure 10. Obviously, the bias in the estimated proportion of strategic voting is not (only) a consequence of a bias in the estimated values of β_1, β_2, and β_3.

A potential explanation of this divergence between the bias in the parameter estimates and the estimated proportion of strategic voting is a discrepancy between the stochastic nature of the data and the statistical model and the deterministic character of the predictions generated from the statistical model: A probabilistic discrete choice model is at the core of the Alvarez-Nagler model and the sophisticated votes in the simulation are generated according to such a model.[13] However, the predictions generated from the model are deterministic in so far as for each voter the predicted vote is the alternatives for which the predicted probability is the relatively largest. In an attempt to improve the match between this aspect of Alvarez and Nagler's method and the simulation study it was re-run with a different data generating process, where the simulated votes are not stochastic but deterministic in the same way as the predictions from Alvarez-Nagler model. That is, simulated sincere votes are those for which the probability in equation (15) is maximal, while simulated sophisticated votes are those for which the probability in equation (16) is maximal. As Figure 12 indicates, this "improvement" leads new problems.

Figure 12 compares true and estimated values of β_1 in the modified simulation study with deterministic votes. In this figure, diagrams with $\varphi = 0$ and $\varphi = 1$ are empty, because in none of the simulation runs, did the fitting algorithm converge. This non-convergence seems to be a consequence of the impossibility to find any finite parameter values for which the log-likelihood function is maximal. This is a situation analogous to the problem of "separation" that may occur in logistic regression (Albert and Anderson 1984). Apparently, the stochastic perturbation created by the mixture of sincere and sophisticate votes when $0 < \varphi < 1$ is necessary to allow for finite estimates of the model parameters. Yet these perturbations not suffice to preclude biases in the estimates of the model parameters. As can be gleaned from the figure the average size of estimates can be more then three times the true parameter values. That notwithstanding, the main conclusion that one can draw from Figure 12 is that if all voters follow the Alvarez-Nagler model in a deterministic way, it is impossible to apply the Alvarez-Nagler method.

Alvarez, Boehmke and Nagler propose as an improvement in the method of estimating the rate of tactical voting not to look at the rate in which votes deviate from counterfactually sincere votes among all voters (Alvarez et al. 2006), but at the rate in which this deviation occurs among those voters whose counterfactually sincere preference is not competitive, i.e. has third place in a constituency. This method has a different estimand, as Alvarez, Boehmke and Nagler make clear: The intended estimand is the proportion of voters that vote strategically, i.e. deviate from their sincere preference, *if* they have an opportunity or incentive to do so (Alvarez et al. 2006). For this reason, the method of Alvarez, Boehmke and Nagler has a different objective than the other methods discussed in the paper, which aim simply to estimate the proportion of strategic voters.

13 Alvarez and Nagler (2000) use a multinomial probit model to allow for correlated random utilities. The simulation study is a simplification, in so far as the generated data and the statistical model do not include correlated random utilities.

The modified estimands and modified estimates of the Alvarez-Bohmke-Nagler method require only a slight modification of the computed proportions. Only those voters are counted as strategic for whom $Y_i \neq Y_{\text{soph},i}$ and $Y_{\text{soph},i}$ is a third-placed party in terms of the vote shares p_{ij} and their number is divided by all voters for which $Y_{\text{soph},i}$ is a third-placed party. In the same vein, in the application of the Alvarez-Bohmke-Nagler model predictions are counted as strategic for whom $\hat{v}_i \neq \hat{v}_{\text{cnft},i}$ and $\hat{v}_{\text{cnft},i}$ is a third-placed party and their number is divided by all voters for which $\hat{Y}_{\text{cnft},i}$ is a third-placed party in the relevant electoral district. Figure 13 shows the simulation results regarding the method of Alvarez et al. (2006). In comparison to Figure 11 both true and estimated rates of tactical voting tend to be higher. Estimates obtained in the first variant of the simulation study appear to have a much less of a downward bias than the corresponding estimates of the original Alvarez-Nagler method (Alvarez and Nagler 2000), but still the can be substantially biased upward.

In the main part of the paper the concept of paradox preference reversals is introduced: It occurs when a sophisticated utility function which takes into account strategic incentives puts an alternative at the top of the preference order which would be third-placed without the strategic incentives and/or puts an alternative at the third or lower place in the preference order which would be first-placed without taking into account strategic incentives. Such a paradox preference reversal would occur, for example, in the case of voter in Colchester, whose "sincere" preference order puts Labour on top, the Liberal Democrats on second place, and the Conservatives on third place, but whose "sophisticated" preference order puts Labour on third place or the Conservatives on first place. One could call the case where an alternative placed on top in the "sincere" preference order ends up at third or lower-place in the "sophisticated" preference order a *partial preference reversal of the first kind*, the case where an alternative on third or lower place in the "sincere" preference order is put on top of the "sophisticated" preference order a *partial preference reversal of the second kind*, and a case where both reversals occur a *complete preference reversal.* Such a complete preference reversal would mean in the example of the voter in Colchester that the sophisticated preference order puts the Conservatives on top and Labour on third place.

In the framework of the Alvarez-Nagler model, a partial preference reversal of the first kind occurs in individual i faced with alternatives $j = 1, 2, 3$, e.g. if

$$\alpha x_{i1} > \alpha x_{i2} > \alpha x_{i3}$$

and

$$\begin{aligned}
&\alpha x_{i2} + \beta w_{1i2} + \beta w_{2i2} + \beta w_{1ij} w_{2i2} > \\
&\alpha x_{i3} + \beta w_{1i3} + \beta w_{2i3} + \beta w_{1ij} w_{2i3} > \\
&\alpha x_{i1} + \beta w_{1i1} + \beta w_{2i1} + \beta w_{1i2} w_{2i1}
\end{aligned}$$

A preference reversal of the second kind occurs in this situation if

$$\begin{aligned}
&\alpha x_{i3} + \beta w_{1i3} + \beta w_{2i3} + \beta w_{1ij} w_{2i3} > \\
&\alpha x_{i1} + \beta w_{1i1} + \beta w_{2i1} + \beta w_{1ij} w_{2i1} > \\
&\alpha x_{i2} + \beta w_{1i2} + \beta w_{2i2} + \beta w_{1i2} w_{2i2}
\end{aligned}$$

while a complete preference reversal occurs if

$$\begin{aligned}
&\alpha x_{i3} + \beta w_{1i3} + \beta w_{2i3} + \beta w_{1ij} w_{2i3} > \\
&\alpha x_{i2} + \beta w_{1i2} + \beta w_{2i2} + \beta w_{1i2} w_{2i2} > \\
&\alpha x_{i1} + \beta w_{1i1} + \beta w_{2i1} + \beta w_{1ij} w_{2i1}
\end{aligned}$$

An intuitively acceptable modification of the preference order by taking into account the strategic incentives w_{1ij} and w_{2ij} might perhaps be

$$\begin{aligned}
&\alpha x_{i2} + \beta w_{1i2} + \beta w_{2i2} + \beta w_{1i2} w_{2i2} > \\
&\alpha x_{i1} + \beta w_{1i1} + \beta w_{2i1} + \beta w_{1ij} w_{2i1} > \\
&\alpha x_{i3} + \beta w_{1i3} + \beta w_{2i3} + \beta w_{1ij} w_{2i3}
\end{aligned}$$

where the alternative which according to the "sincere" preference order ranks on top is ranked second according to the "sophisticated" preference order, and the "second best" according to the "sincere" preference order ends up on first place.

Partial preference reversals of the first and of the second kind, and in particular complete preference reversals may undermine the intuitive appeal of the Alvarez-Nagler method, but there are no logical reasons to rule out their occurrence. The simulation study discussed in this section therefore also examines their relative frequency in the simulated data. In this simulation study one can distinguish between "true" preference reversals that are implied by the parameter settings of α, β, and φ in the data generating process (which follows the Alvarez-Nagler model) and the "predicted" preference reversals as created by the application of the Alvarez-Nagler method. To avoid needless repetitions and because they turn out to occur the relatively most often, the following focuses on predicted partial preference reversals of the second kind. They are identified using the following steps:

1. The alternatives $j = 1, 2, 3$ are ranked in terms of their estimated "sincere" utilities $\hat{\alpha} x_{i1}$ or, equivalently, in terms of the sincere voting probabilities $\Pr(V = j; \hat{\alpha}, 0, 0, 0)$.
2. The alternatives $j = 1, 2, 3$ are ranked in terms of their estimated "sophisticated" utilities $\hat{\alpha} x_{ij} + \hat{\beta}_1 w_{1ij} + \hat{\beta}_2 w_{2ij} + \hat{\beta}_3 w_{1ij} w_{2ij}$ or, equivalently, by their sophisticated voting probabilities $\Pr(V_i = j; \hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$.

3. Each artificial voter for which the alternative j^* which ranks first in terms of $\Pr(V = j^*; \hat{\alpha}, 0, 0, 0)$ ranks third in terms of $\Pr(V_i = j^*; \hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ is counted as having a predicted partial preference reversal of the second kind.

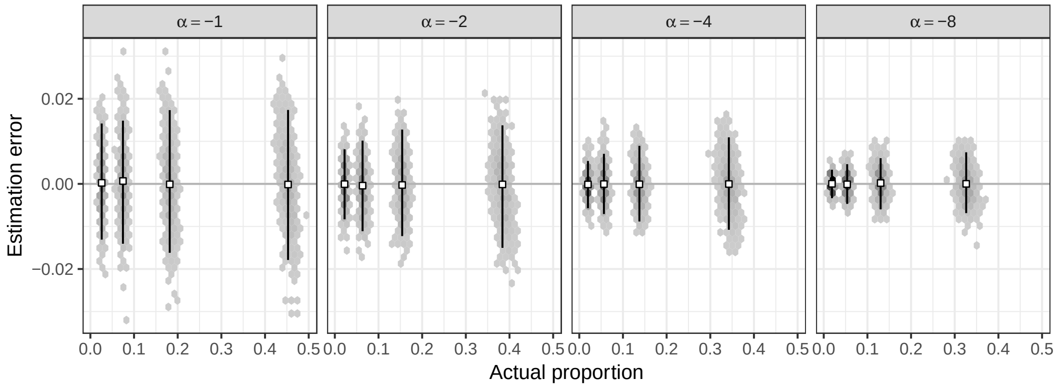
   Technically this is done by computing for $j = 1, 2, 3$ the rank differences $r_{ij} = \text{rank}(\Pr(V = j; \hat{\alpha}, 0, 0, 0)) - \text{rank}(\Pr(V_i = j; \hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3))$ and counting the instances where the rank differences are lesser or equal to -2, i.e. $r_{ij} \leq -2$.

Figure 14 compares the predicted proportions of partial preference reversals of the first kind the with estimated proportions of strategic voting. The horizontal axis corresponds to the proportion of strategic voting estimated using the Alvarez-Nagler method in the simulation runs, the vertical axis corresponds to the *ratio* of the proportion of predicted partial preference reversals to the estimated proportion of strategic voting. That is, it describes how many of the strategic votes implicate such preference reversals. It makes quite obvious that the predicted proportion of partial preference reversals of the first kind increases with the estimated proportion of strategic voting, and that the proportion of these preference reversals can make up close to one half of the reconstructed strategic votes. Further, preference reversals also occur the more often the stronger the influence of strategic incentives is, and can make up to 50 per cent of the estimated share of strategic votes.

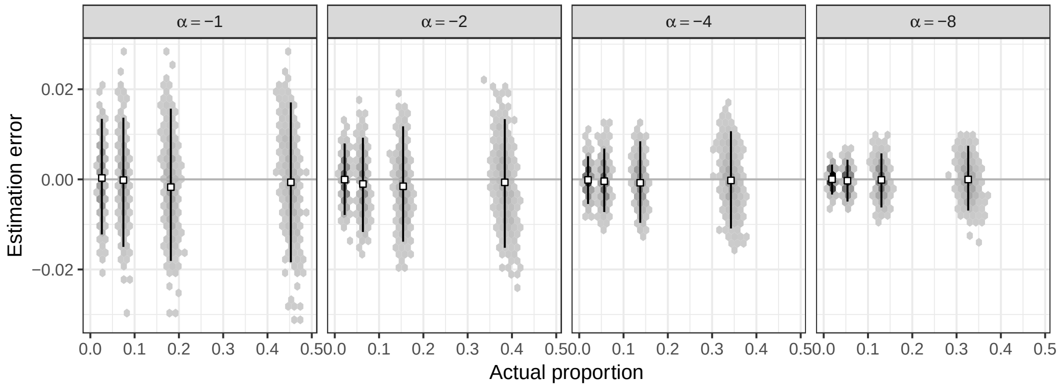
The simulation study shows that the Alvarez-Nagler method suffers from at least two quite serious problems. The first problem is that the model providing the basis of this model does not admit genuine expressive voters who vote sincerely even in the presence of strategic incentives to vote differently. This is because it is a model of sophisticated voting that is fitted to *all* voters in a sample. Votes can only be sincere in this context if taking strategic incentives into account lead to the same choice as one without these incentives. As a consequence, when a sophisticated voting model is fitted to all voters, among which there may also expressive voters, the parameters that describe the influence of strategic incentives can be estimated as too small in absolute size, as demonstrated by Figure 10. Yet surprisingly, the estimated proportion of strategic voting is too small even when there are no expressive voters in the simulated data, so that the estimate of the influence of strategic incentives is more or less correct on average. The root of the bias in estimated proportion of strategic voting seems to be the mismatch between the probabilistic nature of the model *and* of the data generating process in the simulation study and the deterministic nature of the predictions that are used to obtain the proportion of strategic votes from the estimated model. (That is, the bias does *not* come from a mismatch between the model and the data generating process.) Yet if the simulation study is modified so that the generated votes behave more like the kind of model-based predictions that Alvarez and Nagler (2000) use, then fitting the model to these generated data leads to grossly over-estimated parameters or no finite model estimates can be found at all.

The second problem is that the construction of the Alvarez-Nagler model (the model on which the Alvarez-Nagler method is built) allows what is called in this paper "preference reversals". A kind of preference reversals which can occur quite often when the Alvarez-Nagler method is applied (and it is assumed that voters behave according to its core model) is that an alternative ranked first in terms of voters' "sincere" preferences ranks third in terms of voters' sophisticated. As a consequence, an alternative that a voter may want to prevent from winning a seat by voting strategically may get a higher position in the sophisticated-utility based preference order than his or her originally most liked alternative. In so far, the sophisticated utility function may contradict the original motive of strategic voting: Choosing the "second-best" alternative to prevent a disliked alternative from winning a seat or even the election.

Of course, the simulation study cannot rule out that for *certain particular settings* of the true parameter values the bias in the estimated proportion of strategic voting vanishes or at least becomes negligible. Also, the simulation study does not rule out that there are other ways to specify the influence of strategic incentives that avoid the implication of preference reversals. Yet it is not easy to see how this can be achieved. The difficulty to construct a "sophisticated" utility function that avoids preference reversals was one of the original motivations of the author of the paper to attempt to find a different way to estimate strategic voting.

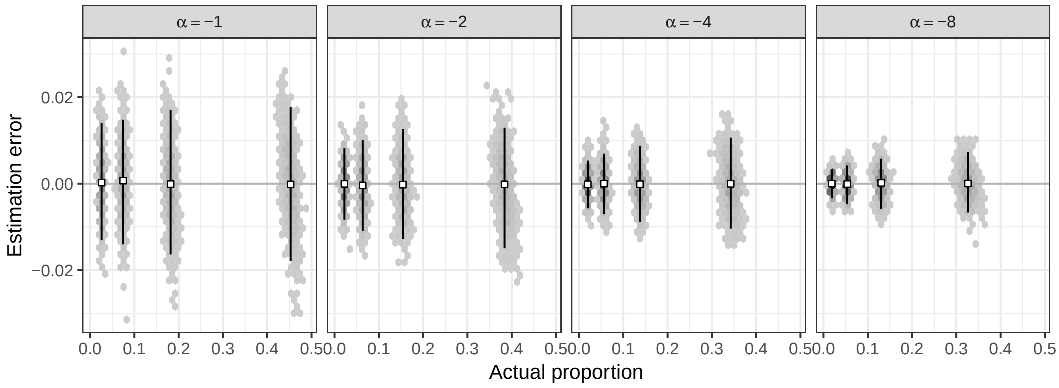


(a) First variant: Fully specified finite mixture model

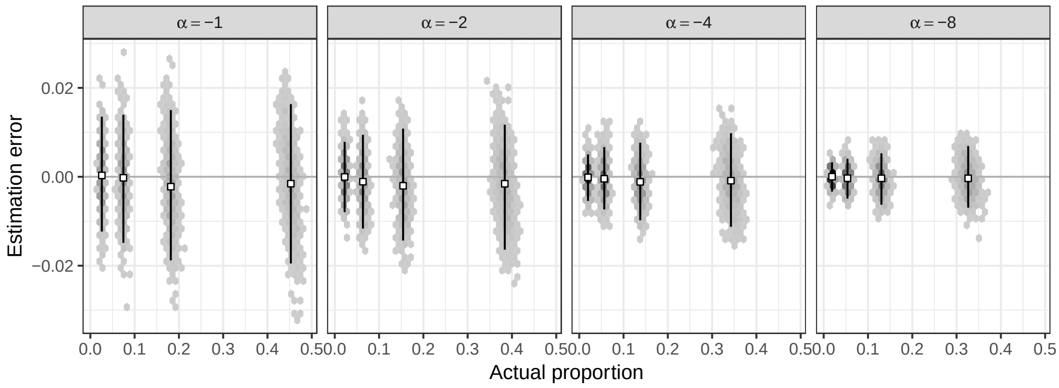


(b) Second variant: Finite mixture model without information about predictors of strategic incentives

Figure 7 Hexbin plots of the difference between actual proportions of strategic voting average empirical Bayes prior probabilities of strategic voting for various settings for the influence of party evaluations (α) and the influence of strategic incentives on sophisticated votes (β). The different settings of the parameter β are reflected in the location of the different "clouds" that appear in the diagrams, the shading of the hexagons indicates how many data points are contained in them. The little squares correspond to the average errors of the estimates for the various settings of the true parameters. The vertical lines correspond to the 0.025 and 0.975 quantiles of the distribution of the errors, and thus correspond to 95% confidence intervals.



(a) First variant: Fully specified finite mixture model



(b) Second variant: Finite mixture model without information about predictors of strategic incentives

Figure 8 Hexbin plots of the difference between actual and average posterior probabilities of votes being strategic for various settings for the influence of party evaluations (α) and the influence of strategic incentives on sophisticated votes (β). The different settings of the parameter β are reflected in the location of the different "clouds" that appear in the diagrams, the shading of the hexagons indicates how many data points are contained in them. The little squares correspond to the average errors of the estimates for the various settings of the true parameters. The vertical lines correspond to the 0.025 and 0.975 quantiles of the distribution of the errors, and thus correspond to 95% confidence intervals.

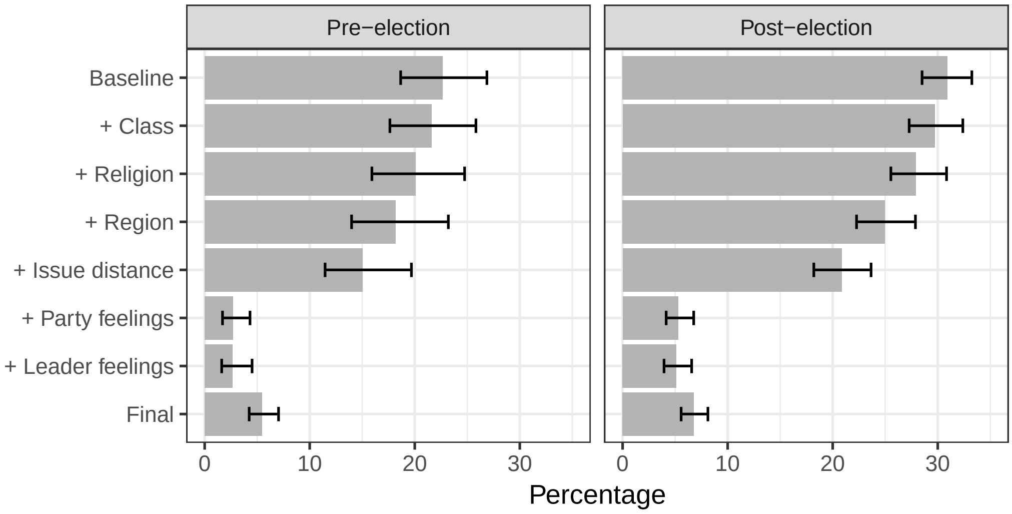


Figure 9 The inclusion of predictors for voters' preferences and/or parties into the finite mixture model and the estimated proportion of tactical votes

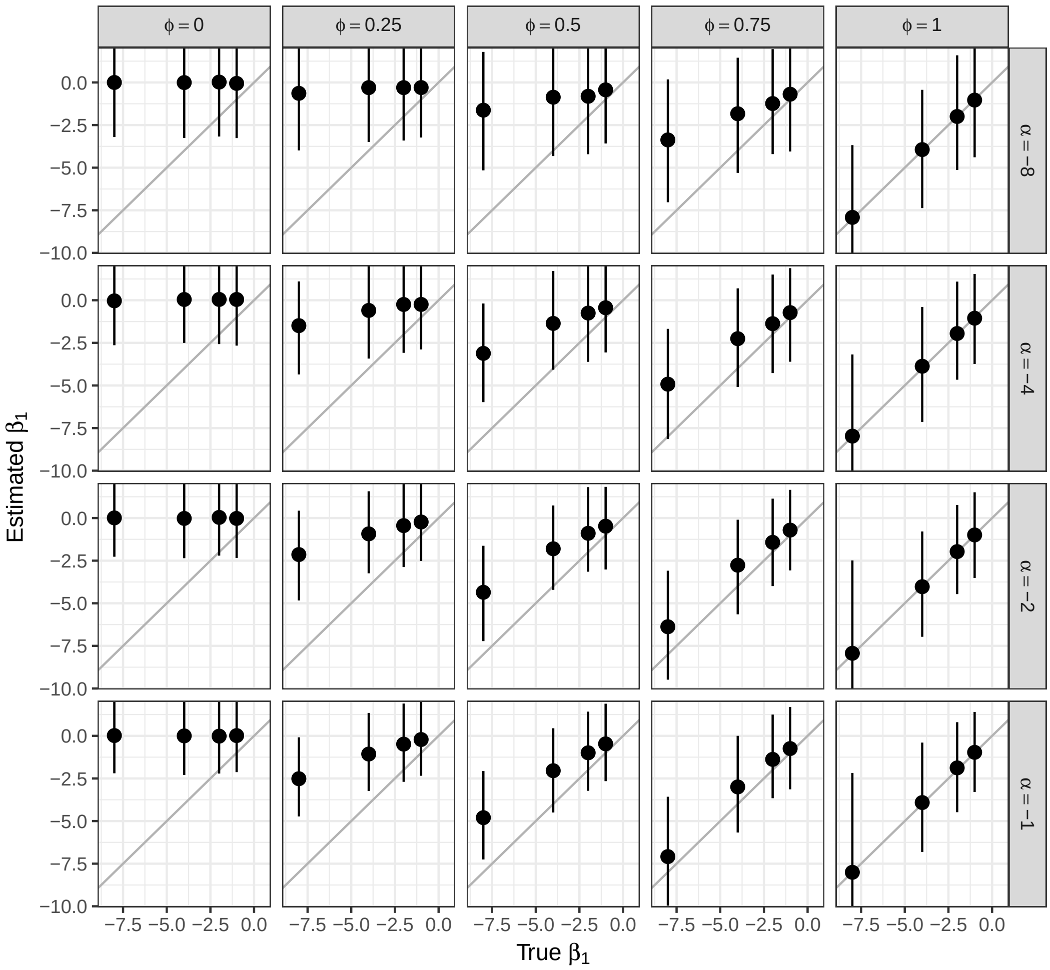


Figure 10 True and estimated values of the parameter β_1, which represents the influence of one of the strategic incentives (the closeness of competition between the "other" two parties) on sophisticated voting choices in the Alvarez-Nagler model. The round dots represent the average of the estimates for each true parameter value setting. The vertical line segment correspond to the range that covers 95% of the estimates. The diagonal line is a "diagonal of identity": it connects all locations in the coordinate system where the estimated parameter value equals the true parameter value.

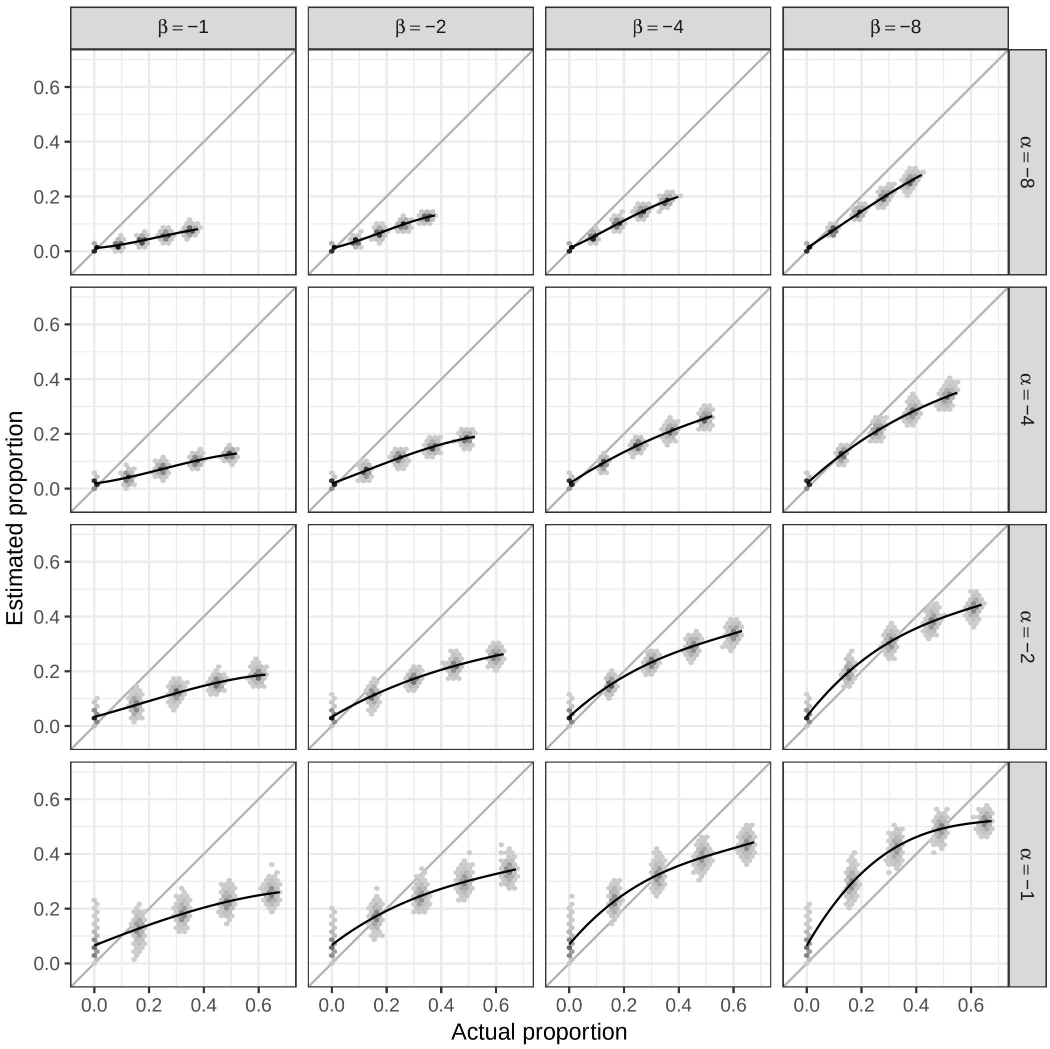


Figure 11 True rates of strategic voting and estimated rates based on the Alvarez-Nagler method. The different settings of the latter parameter is reflected in the location of the different "clouds" that appear in the diagrams.

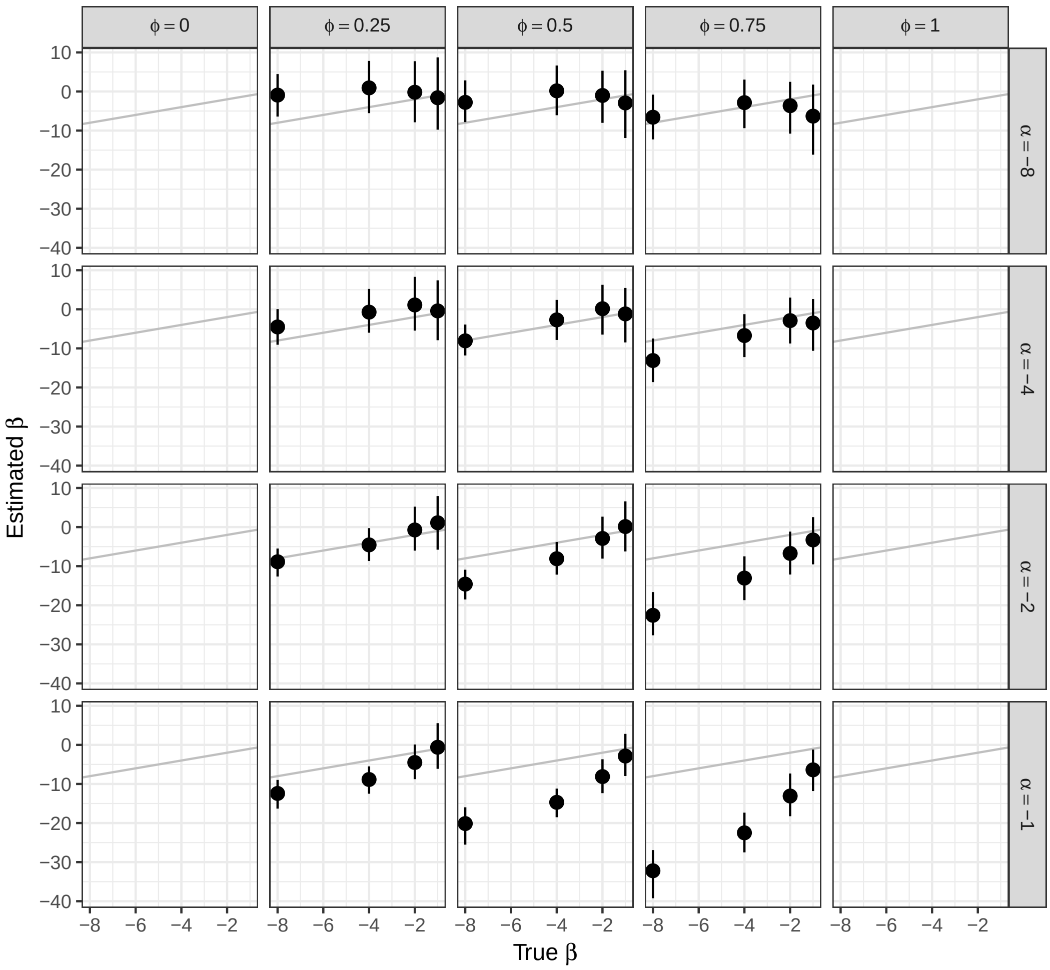


Figure 12 The modified simulation study with deterministic choices: True and estimated values of the parameter β_1, which represents the influence of one of the strategic incentives (the closeness of competition between the "other" two parties) on sophisticated voting choices in the Alvarez-Nagler model. The round dots represent the average of the estimates for each true parameter value setting. The vertical line segment correspond to the range that covers 95% of the estimates. The diagonal line is a "diagonal of identity": it connects all locations in the coordinate system where the estimated parameter value equals the true parameter value. The diagrams for $\varphi = 0$ and $\varphi = 1$ because none of the simulation runs yielded finite parameter estimates.

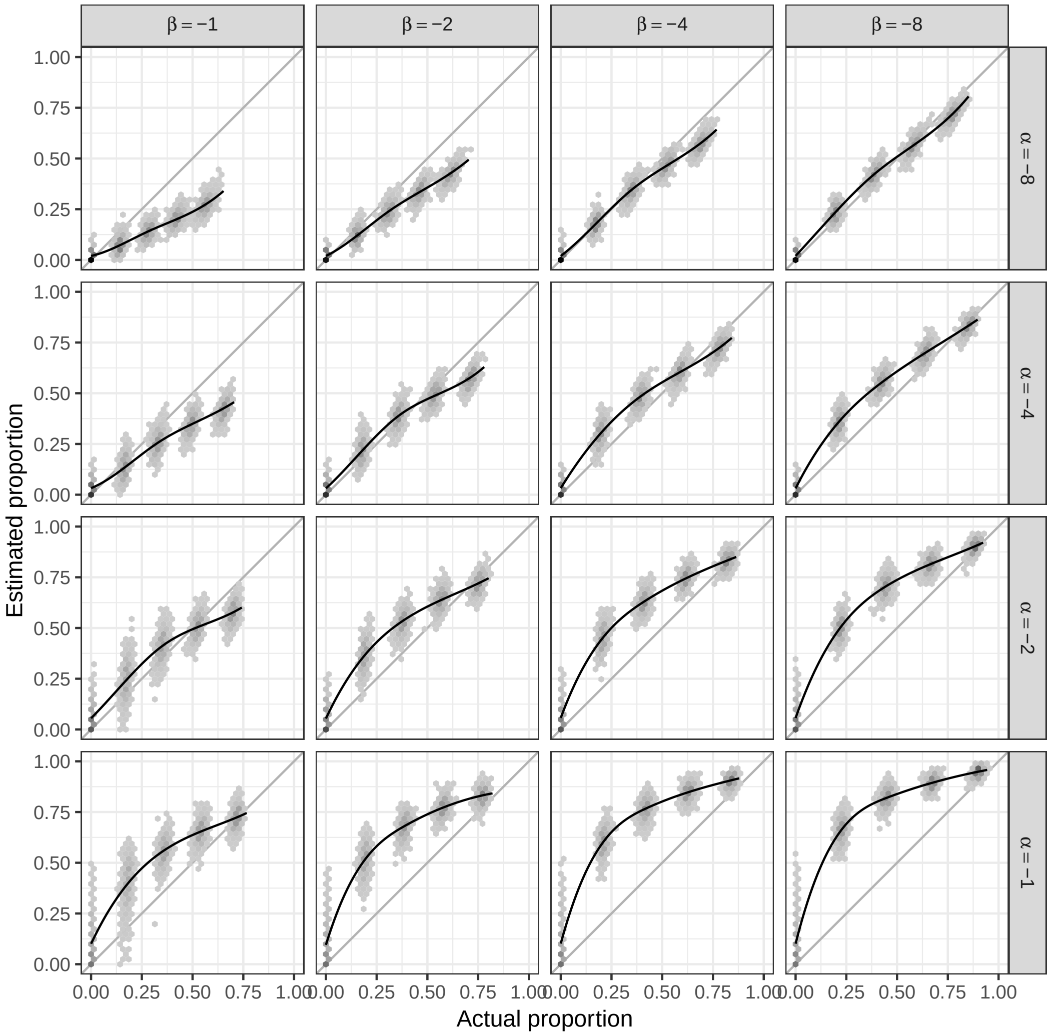


Figure 13 True rates of strategic voting and estimated rates based on the method proposed by Alvarez et al. (2006).

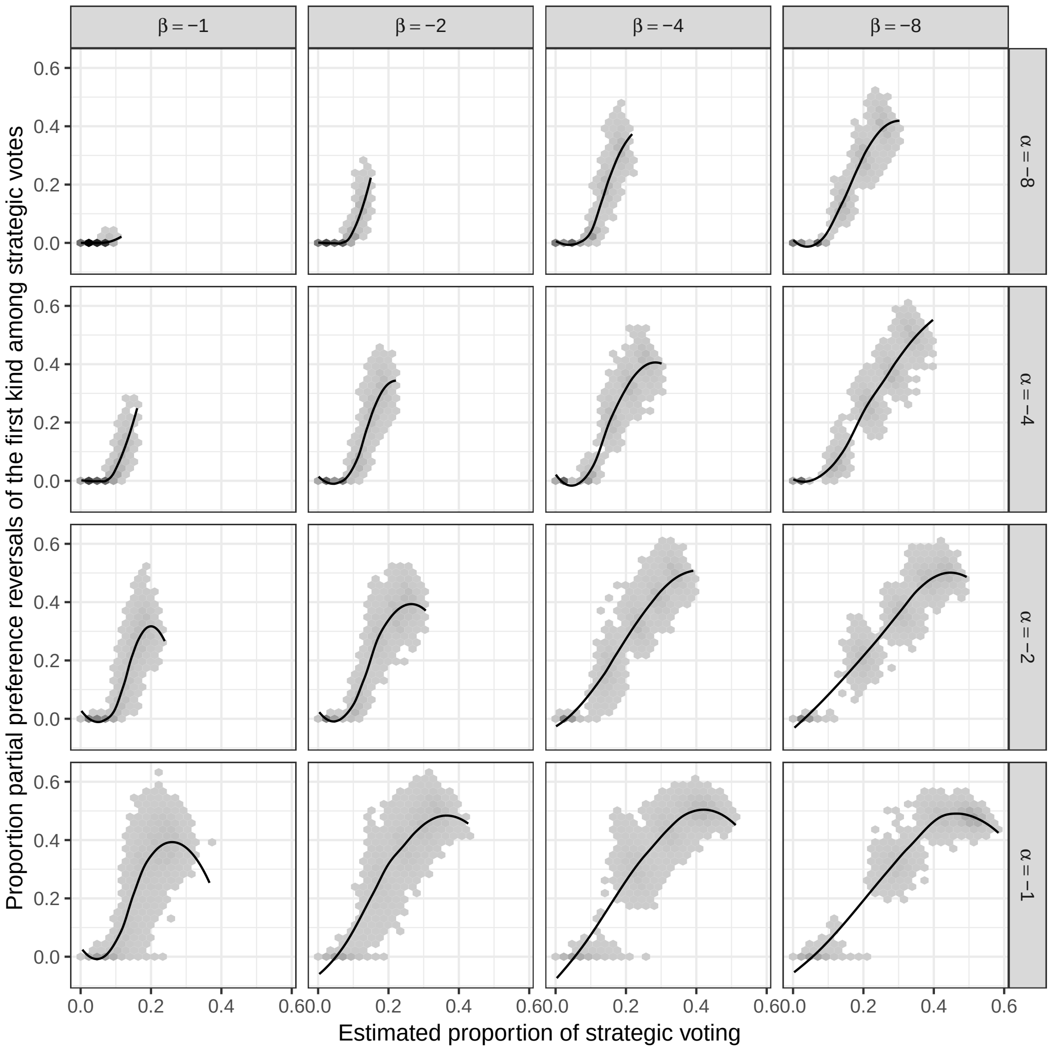


Figure 14 Proportion of strategic voting estimated by the Alvarez-Nagler-Method and predicted proportion of partial preference reversals of the first kind. The vertical axis is the ratio of the proportion of these preference reversals to the estimated proportion of strategic voting.