Separating Tactical from Sincere Voting: A Finite-Mixture Discrete-Choice Modelling Approach to Disentangling Voting Calculi

Martin Elff*

April 3, 2014

Abstract

Voters' propensities to vote tactically is the main component of what Duverger calls the "psychological effect" of electoral systems. Electoral research has therefore given considerable attention to the amount of tactical voting and its consequences for electoral results. Despite the importance of the topic, research on tactical voting has not yet reached a consensus on how to measure it. The present paper proposes a finite-mixture discrete-choice model for the reconstruction of voting calculi, which makes possible to explicitly specify the calculi of different voting types and to derive individuals' posterior probabilities to have used any of these. It is further shown how maximum marginal likelihood estimates of the model parameters can be computed via an expectation-maximization (EM) algorithm. The model is applied to the case of the 2010 UK general election. It is shown that the proportion of tactical votes reconstructed from constituency data and observed choices is close the proportion of tactical votes reconstructed in terms of voters' stated motives. The paper also briefly discusses the application of the model to ticket-splitting in mixed electoral systems.

1 Introduction

Duverger's Law, which states that electoral systems with single-member districts and plurality rule tend to favor two-party systems, is perhaps one of the most well-known regularity of political science, despite not being a strict one (Duverger, 1965; Cox, 1997). According to

*Department of Politics and Public Administration, University of Konstanz, Box 84, 78457 Konstanz, Germany. martin.elff@uni-konstanz.de

Duverger, two kinds of effects of electoral systems lie behind this regularity, the "mechanical effect" of electoral systems, which stems from the way that votes are translated into seats, and the "psychological effect", which stems from a supposed tendency of voters to avoid wasting their votes for hopeless candidates or parties.

While establishing the "mechanical effect" is in principle straightforward – it is merely a matter of computation to translate the distribution of votes in voting districts into parliamentary seats – this is less so in the case of the "psychological effect". From the point of view of rational-choice theory, the chief mechanism behind this effect is just a case of strategic voting, so calling it "psychological" may be a misnomer: A voter is envisaged to avoid wasting her vote if she refrains from giving it to the party or candidate that she otherwise most prefers but perceives it/her as hopeless and instead gives it to a party or candidate that has a real chance of winning the constituency seat, thus maximizing utility by maximizing her influence on the eventual electoral outcome (Downs, 1957; Cox, 1997). Nevertheless it is quite difficult to establish empirically whether a particular voter has or how many voters have voted strategically in this sense. Hence it is no surprise that no consensus has been reached yet on how to measure this kind of strategic voting.

The extend to which voters engage in this kind of strategic decision-making has been researched with particular intensity with regards to the case the United Kingdom, where the phenomenon is better known as "tactical voting". One could discern two major strands in this literature. The first strand focuses on voters' stated reasons for their vote. British Election studies have often included items in their surveys asking respondents about their reasons for their vote intention or voting decision. And if respondents answered explicitly to have voted tactically or to have rather preferred another party that did not have a chance to win, they are categorized as tactical voters and otherwise they are categorized as non-tactical or "sincere" voters, and there has been some debate about whether respondents should be encouraged to give such reasons by providing them as explicit options in a semi-open question format (Heath et al., 1991; Niemi et al., 1992; Evans and Heath, 1993; Franklin et al., 1994). The second strand focuses on the influence of the distribution of the votes in a voting district on her vote intention or voting decision. This approach mainly goes back to Alvarez and Nagler (2000), who argue that tactical voting are those departures from a voting decision as predicted by common covariates of the voting decision, such as class, party identification, etc, that can be captured by the effects of the district-level vote shares of the parties or candidates.

The approaches of both strands in the literature have their advantages and disadvantages. The approach resting on respondents' given reasons leads directly to a categorization into tactical and non-tactical voters. Yet this approach is limited in so far as giving a reason indicating a tactical vote is not a necessary condition for having voted tactically (Alvarez and Nagler, 2000; Alvarez et al., 2006). Furthermore, the stated-reason approach rests on the relevant items being included in electoral surveys. Yet scholars may also be interested in uncovering the amount of tactical voting in countries where such items are not routinely included in electoral surveys (or where reasons indicating a tactical vote are not explicitly given as a response option). The approach that rests on district-level vote shares as predictors does not depend on such survey items, yet incorporating the competitive situation in a district into voters' utility functions (Alvarez and Nagler, 2000; Alvarez et al., 2006) does not by itself lead to a categorization of survey respondents in tactical or non-tactical voters.

The present paper tries to address this situation by devising a finite mixture model of voting that leads to a method for disentangling tactical from non-tactical vote intentions or vote decisions, in so far as it (1) works without the need to take recourse to stated reasons for these intentions or decisions, (2) is explicitly derived from a clarified notion of tactical voting, (3) leads to respondents' (posterior) probabilities to have voted tactically or non-tactically, once information about vote choice predictors and parties district vote shares are known, and (4) thus allows to give estimates about the level of tactical voting in a survey sample or in an election.

The paper is composed as follows: The next section explicates the notion of tactical voting from which the finite mixture model is derived. It is followed by a formal statement of the finite mixture model and a recipe for the estimation of its parameters, which allows to obtain posterior probabilities for voters to have voted tactically or non-tactically. Another section demonstrates the application of the model and the method to the estimation of the amount of tactical voting in the 2010 General Election to the House of Commons of the UK. That section also compares the result obtained by the proposed method with results obtained using the stated-reason technique mentioned above. The applied section discusses the similarities and differences of the proposed finite mixture models with two related models, namely models inspired by Alvarez and Nagler (2000) and random consideration-set models introduced to the discipline by Steenbergen et al. (2011). This section also discusses the extension of the proposed method to the more complex situation of dual-ballot mixed-member electoral systems. The paper concludes with a summary of its findings and suggestions for future research.

2 Reconsidering Tactical Voting

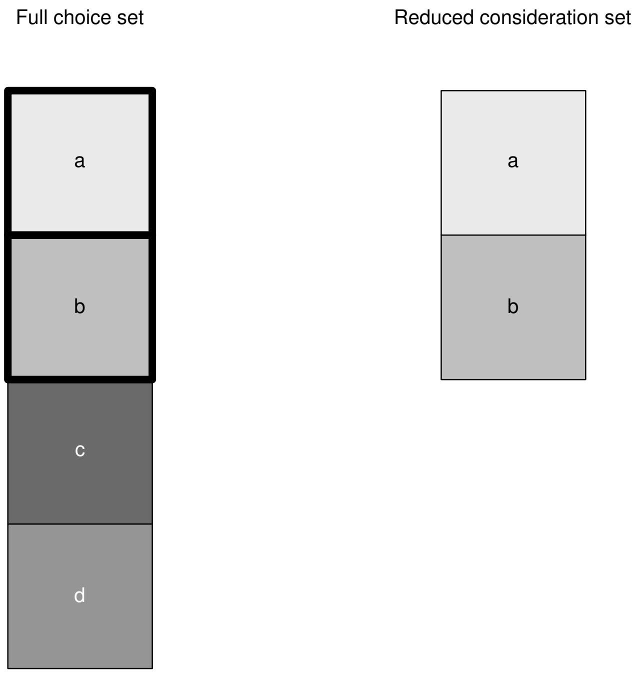
The expectation that voters try to avoid wasting their vote forms the core of the notion of tactical voting. Wasting one's vote means to make an electoral decision such that it will not affect the overall electoral result and, eventually, the recruitment or composition of the government (Downs, 1957). Of course, in large electorates the effective weight of any individual vote for the overall outcome is almost infinitesimally small, but this weight is not independent from which party or candidate one votes for and certainly not independent from the makeup of the relevant electoral system.

What wasting one's vote means is almost perfectly clear in a first-past-the-post or plurality system with single-member voting districts (such as the electoral systems relevant for the American Congress or the British House of Commons). Here, voting for a candidate that has the third-largest or smaller proportion of support in the relevant voting district means to vote for a candidate that will almost certainly not gain representation, except for the highly unlikely case that the shares in support for the three strongest parties/candidates are almost equal in the relevant constituency. In the more plausible situation, the best way for a voter who prefers the candidate second-placed in the constituency over the first-placed to bring this preference to effect is to vote for the second-placed party, even if she prefers a third candidate over the second. Giving the vote to a third candidate will not, except for unusual circumstances, increase the third candidate's chances to gain the constituency seat, and will even diminish the likelihood that the running-off candidate will successfully challenge the leading candidate, relative to a vote given to the running-off candidate instead.

If an electoral system were perfectly proportional (a situation that does not apply in the real world), the weight of an individual vote is inversely proportional to the total number of votes cast. Even in this situation one could consider a vote to be wasted if it does not affect the eventual composition of government, because it is cast for a party that has no chance of becoming the *formateur* of a government or government coalition or being considered by a *formateur* as a member of a government coalition. Arguably, the same logic that applies in first-past-the-post systems to the choice between first- and second-placed candidates in a constituency will apply in proportional systems to the choice between a party whose seat share is relevant for the formation of a government coalition and a party whose seat share is to small. Further, in real-world proportional-representation (PR) systems, there exists an effective threshold of representation, either as a consequence of explicit legal thresholds or of a limited district magnitude (i.e. the number of seats allocated to a voting district) (Cox, 1997). That is, even in proportional systems there may be incentives to vote tactically.

To avoid wasting their votes, electors will need to restrict their attention only to candidates or parties that are viable in the relevant constituency or – if the aim is to influence the composition of a government coalition – that are plausible coalition partners. Instead of choosing from all available alternatives – e.g. all candidates running in the relevant constituency – a voter trying to avoid a wasted vote will chose from only those alternatives she considers as viable. For example, consider a voter who in her voting district can choose from candidates/parties a, b, c, and d and has stable preferences among them, e.g. $c \succ d \succ b \succ a$, where the proportions of votes these parties are expected to gain in the seat of the district have the order $p(a) > p(b) > p(c) > p(d)$. This situation is illustrated by figure 1, where the left-hand column shows all available alternatives, ordered by their expected vote share and shaded according to the preference ordering of the voter (the darker, the higher the position in the preference order). Suppose that this voter realizes that c, her most preferred candidate/party, does not have a chance to gain the district seat, because in terms of expected vote share it comes third (and there is a wide enough margin between the second and third-placed candidate/party), whereas only a and b are viable contenders for it (this is reflected by the

Figure 1: An Illustration of tactical considerations in voting: A voter faces alternatives A, B, C, and D and has the preference ranking $C \succ D \succ B \succ A$, while the expected vote share proportions of the alternatives in the constituency are $p(A) > p(B) > p(C) > p(D)$. If her vote it is sincere she will vote for alternative C, while if her vote is tactical she will vote for alternative B.



thick lines in the left column of the diagram). If the voter now decides to vote tactically, that is to abandon the most preferred choice in favor of a viable one, it will of course no longer matter that she prefers c over d and it will not even matter that she prefers d over a and b. Instead all that will matter then is that she prefers b over a. Consider now an observer who has knowledge about the preference ordering of the alternatives by the voter and of the eventual voting decision. If this observer wants to recover whether the voter has voted tactically or sincerely, his task will be easy in the idealized situation of this simple example: If the voter has voted for c, the observer can conclude that the voter has voted sincerely, whereas if the voter has voted for b, the observer can then conclude that the voter has voted tactically to avoid wasting her vote.

While the discussion of the example may seem trivial, it may help to illustrate a core idea behind the method of reconstructing tactical and sincere votes proposed in this paper, the idea that voting tactically in effect means restricting one's choice set – or in other words to restrict one's attention to a *consideration set* that is a proper subset of the original choice set. The example should also highlight the notion that, to uncover sincere and tactical voting decisions, one does not need to reconstruct the weighting of utilities that reflect the original preference ordering by the viability of the alternatives – a point to which this paper returns in a later section – as long as the viability of an alternative is an either-or question.

That notwithstanding, for practical purposes the idealized setup of the example is not satisfactory. Empirical researchers usually do not have information about the voters' complete preference ordering of the alternatives from which they choose, instead they only have information about which alternative they have chosen or have intended to choose. Furthermore, whether a vote is wasted in the way discussed above depends on the *actual* distribution of votes in the relevant constituency or, respectively, on the distribution of seats that the *actual* election will bring about. This is information be available to researchers after the fact, but certainly not during an election or before it, when voters are about to make up their minds about their electoral decisions.

While it is implausible to expect voters to have perfect information about the vote distribution in advance of the election, they may form more or less accurate expectations about the electoral outcome based on various sources of information, such as the past vote shares of candidates in the relevant constituency and the popularity of candidates or parties in opinion polls published in the run-up to the election. If such information is available to the voters, it will or course also be available to researchers, who therefore should at least be able to make predictions about which of the alternatives are considered as viable by the voters.

Researchers' lack of information about voters' complete preference ordering of alternatives can compensated for to some degree, if they have indicators for such preferences or at least good predictors for voting decisions. Voter surveys conducted as part of election studies often contain questions about voters' evaluation of parties and candidates, ratings of parties and their candidates and leaders, and about assessments of parties' or candidates' positions on issues and voters' own issue positions. If voters' evaluations of parties and candidates strongly predict their voting decisions that may be so because such evaluations are from the point of view of the voters hardly distinguishable from their preferences of the parties and thus such predictions may have little genuine, non-tautological explanatory power. Yet what for an explanation of voting choices may be a drawback is in fact a considerable asset for a reconstruction of the type of the voting decision, whether tactical or sincere. And even if such predictors are not perfect, they allow at least a prediction of voting decisions in terms of choice probabilities.

If it is possible to derive probabilities of voting choices conditional on having decided to vote sincerely or tactically as well as probabilities for using either of these types of voting calculi, then it is possible to use Bayes' theorem to obtain probabilities of having used either of these types, conditional on the voting decisions actually observed. To return to the above example, suppose that the probability for the voter to choose alternative b conditional on a sincere voting type is $\Pr(V = b|T = 1)$ and conditional on a tactical voting type is $\Pr(V = b|T = 2)$, then her probability to have voted tactically is

$$\Pr(T = 2|V = b) = \frac{\Pr(V = b|T = 2)\Pr(T = 2)}{\Pr(V = b|T = 1)\Pr(T = 1) + \Pr(V = b|T = 2)\Pr(T = 2)}. \tag{1}$$

Suppose further that her probability to vote for c, if she votes tactically, is $\Pr(V = c|T = 2) = 0$, because c is not viable in the voting district, then her probability to have voted tactically, if her choice of c is observed, is

$$\Pr(T = 2|V = c) = \frac{0 \times \Pr(T = 2)}{\Pr(V = c|T = 1)\Pr(T = 1) + 0 \times \Pr(T = 2)} = 0. \tag{2}$$

That is, if her decision is inconsistent with a tactical voting calculus, then one can conclude with almost certainty that she has has voted non-tactically. Such a certainty is however not given if she chooses b if the probability $\Pr(V = b|T = 1)$ of a choice of b conditional on a sincere vote is different from zero (because we do not know her preference ordering exactly, but only know her evaluations of the alternatives or other choice predictors). Nevertheless, if $\Pr(V = b|T = 1)$ is small and she has voted for b, Bayes theorem leads to the conclusion that her probability to have voted tactically is high.

This reconstruction of individual probabilities to have voted tactically or sincerely seems to beg the question about where the prior probabilities, $\Pr(T = 1)$ and $\Pr(T = 2)$, come from. And indeed, these probabilities are actually looked for if one is interested in the overall distribution of voting types during an election. Certainly the size of these prior probabilities will be of little consequence if the conditional probabilities $\Pr(V = b|T = 1)$ and $\Pr(V = b|T = 2)$ are close to zero or close to one, but there is no guarantee that this is the case. Fortunately, it is possible to estimate these prior probabilities from a sample using a maximum marginal likelihood technique, which opens up the possibility to reconstruct individual voting calculi by using an empirical Bayes method. The next section gives a more formal exposition of the finite-mixture model of voting types sketched here and describes how its parameters, that is the prior probabilities and conditional probabilities, can be estimated.

3 The Finite Mixture Model and Its Estimation

For the discrete choice analysis of voting it is convenient to represent the voting decision or vote intention of an individual voter i as a series of binary random variables Y_{ij}, so that $Y_{ij} = 1$ if the voter decides or intends to vote for party or candidate j and $Y_{ij} = 0$ otherwise, where j is any of the parties or candidates from which i can choose. If the voter i can vote exactly for one alternative (party or candidate) and if the set of alternatives available – the *choice set* – is denoted as $\mathcal{S}_i$ then this implies

$$\sum_{j\in\mathcal{S}_i} Y_{ij} = 1.$$

In the discussion of the previous section the type of voting (voting calculus), was represented as a categorical random variable T_i with $\{0, 1, \ldots, m\}$, e.g. such that $T_i = 1$ if i votes non-tactically and $T_i = 2$ if i votes tactically. Equivalently, one can represent the type of voting used by i by a set of binary random variables D_{hi} such that $D_{hi} = 1$ if and only if $T_i = h$ and $D_{hi} = 0$ otherwise, such that

$$\sum_{h=1}^{m} D_{hi} = 1.$$

The considerations of the previous section now lead to a different *conditional distribution* of Y_{ij} for each possible value of T_i: If one defines $\pi_{ij|h} = \Pr(Y_{ij} = 1|T_i = h) = \Pr(Y_{ij} = 1|D_{hi} = 1)$ and $\varphi_{hi} = \Pr(T_i = h) = \Pr(D_{hi} = 1)$ then

$$\pi_{ij|h} \begin{cases} > 0 & \text{for } j \in \mathcal{S}_{hi} \text{ and} \\ = 0 & \text{for } j \notin \mathcal{S}_{hi} \end{cases} \tag{3}$$

where $\mathcal{S}_{hi} \subseteq \mathcal{S}_i$ is the consideration set that characterizes the voting type h. In case of $h = 1$ meaning sincere voting and $h = 2$ tactical voting, one has $\mathcal{S}_{1i} = \mathcal{S}_i$, while $\mathcal{S}_{2i} \subset \mathcal{S}_i$ is the subset of parties or candidates that i considers electorally viable. All this implies

$$\sum_{j\in\mathcal{S}_{hi}} \pi_{ij|h} = 1, \quad \sum_{h=1}^{m} \varphi_{hi} = 1, \quad \pi_{ij} = \sum_{h=1}^{m} \pi_{ij|h}\varphi_{hi}, \quad \text{and} \quad \sum_{j\in\mathcal{S}_i} \pi_{ij} = \sum_{j\in\mathcal{S}_{hi}} \sum_{h=1}^{m} \pi_{ij|h}\varphi_{hi} = 1, \tag{4}$$

that is, the marginal distribution of Y_{ij} (i.e. *without* conditioning on T_i) is a *finite mixture*, with *mixture components* – which are multinomial distributions – given by the probabilities $\pi_{ij|h}$ and a *mixture distribution* – again a multinomial distribution – given by the probabilities φ_{hi}.

The mixture components of the choice distribution and the mixing distribution of the choice types can be considered as depending on certain covariates. In the simplest case there is a covariate matrix $\boldsymbol{X}_i$ with row vectors $\boldsymbol{x}_{ij}$ and a coefficient vector $\boldsymbol{\alpha}$, such that for $j \in \mathcal{S}_{hi}$

$$\pi_{ij|h} = \frac{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\alpha})}{\sum_{k\in\mathcal{S}_{hi}} \exp(\boldsymbol{x}'_{ik}\boldsymbol{\alpha})}, \tag{5}$$

and a covariate vector $\boldsymbol{z}_i$ and coefficient vectors $\boldsymbol{\beta}_h$, such that for $h = 1$

$$\varphi_{1i} = \frac{1}{1 + \sum_{g=2}^{m} \exp(\boldsymbol{z}'_i\boldsymbol{\beta}_g)} = \frac{1}{1 + \sum_{g=2}^{m} \exp(\boldsymbol{z}'_{ig}\boldsymbol{\beta})}, \tag{6}$$

and for $h = 2, \ldots, m$

$$\varphi_{hi} = \frac{\exp(\boldsymbol{z}'_i\boldsymbol{\beta}_h)}{1 + \sum_{g=2}^{m} \exp(\boldsymbol{z}'_i\boldsymbol{\beta}_g)} = \frac{\exp(\boldsymbol{z}'_{ih}\boldsymbol{\beta})}{1 + \sum_{g=2}^{m} \exp(\boldsymbol{z}'_{ig}\boldsymbol{\beta})}, \tag{7}$$

with $\boldsymbol{\beta} = (\boldsymbol{\beta}'_2, \ldots, \boldsymbol{\beta}'_m)'$ and $\boldsymbol{z}_{ig} = \boldsymbol{e}_{ig} \otimes \boldsymbol{z}_i$, where $\boldsymbol{e}_{ig}$ is a vector of dimension (length) $m - 1$ with the $(g - 1)$th element equal to unity and all other elements equal to zero and $\otimes$ is the

Kronecker product. That is, the conditional choice probabilities follow a (modified) conditional logit model whereas the prior probabilities of the types of choice follow a baseline multinomial logit model with baseline category $h = 0$. To arrive at estimates of $\pi_{ij|h}$ and φ_{hi} one can use the ML estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}_h$ of the coefficient vectors. Of course, if $m = 2$ then $\boldsymbol{z}_{ih} = \boldsymbol{z}_i$, $\boldsymbol{\beta}_h = \boldsymbol{\beta}$ and equations (6) and (7) specify a binary logit model.

To understand the estimation of the finite mixture model, it is helpful to consider the (hypothetical) case where the voting types are actually observed. The corresponding *complete-data* likelihood (Little and Rubin, 2002), the joint likelihood of observations $Y_{ij} = y_{ij}$ and $D_{hi} = d_{hi}$, then is

$$\begin{aligned}
\mathcal{L}_{\text{cpl}} = \prod_i \mathcal{L}_{\text{cpl},i}, \quad \mathcal{L}_{\text{cpl},i} &= \prod_h \prod_{j \in \mathcal{S}_{hi}} \Pr(Y_{ij} = y_{ij} | D_{hi} = d_{hi}) \Pr(D_{hi} = d_{hi}) \\
&= \prod_h \prod_{j \in \mathcal{S}_{hi}} \pi_{ij|h}^{y_{ij}} \varphi_{hi}^{d_{hi}} = \prod_h \mathcal{L}_{i|h} \varphi_{hi}^{d_{hi}}, \qquad \text{say,}
\end{aligned}$$

where

$$\mathcal{L}_{i|h} = \begin{cases} \prod_{j \in \mathcal{S}_{hi}} \pi_{ij|h}^{y_{ij}} & \text{if } y_{ij} = 1 \text{ for } j \in \mathcal{S}_{hi} \text{ and} \\ 0 & \text{if } y_{ij} = 1 \text{ for } j \notin \mathcal{S}_{hi}. \end{cases}$$

The observed-data or marginal likelihood, which does not depend on the unobserved data d_{hi} (Little and Rubin, 2002), can then be obtained by summing each of the complete-data contributions of each individual i over all m possible dummy vectors $(d_{1i}, \ldots, d_{mi})'$ – which is the discrete analogue to integrating out an unobserved continous variable – noting that exactly one of their elements equals unity:

$$\mathcal{L} = \prod_i \mathcal{L}_i = \prod_i \sum_{d_{1i}+\ldots+d_{mi}=1} \prod_h \mathcal{L}_{i|h} \varphi_{hi}^{d_{hi}} = \prod_i \sum_{h=1}^{m} \mathcal{L}_{i|h} \varphi_{hi}$$

Equivalently (and more conveniently) one can maximize the log-likelihood

$$\ell = \sum_i \ln \mathcal{L}_i = \sum_i \ln \sum_{h=1}^{m} \mathcal{L}_{i|h} \varphi_{hi} = \sum_i \ln \sum_{\substack{h=1 \\ \mathcal{L}_{i|h} > 0}}^{m} \exp(\ell_{i|h}) \varphi_{hi}$$

where (only for $\mathcal{L}_{i|h} > 0$, i.e. $y_{ij} = 1$ with $j \in \mathcal{S}_{hi}$)

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

Note that for the first derivatives of the log-likelihood one obtains

$$\frac{\partial \ell}{\partial \boldsymbol{\alpha}} = \sum_i \frac{\partial}{\partial \boldsymbol{\alpha}} \ln \sum_{\mathcal{L}_{i|h}>0} \mathcal{L}_{i|h}\varphi_{hi} = \sum_i \sum_{\mathcal{L}_{i|h}>0} \frac{\mathcal{L}_{i|h}\varphi_{hi}}{\mathcal{L}_i} \frac{\partial \ell_{i|h}}{\partial \boldsymbol{\alpha}} = \sum_i \sum_{\mathcal{L}_{i|h}>0} \mathcal{P}_{hi} \frac{\partial \ell_{i|h}}{\partial \boldsymbol{\alpha}} \qquad \text{and}$$
$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_i \frac{\partial}{\partial \boldsymbol{\beta}} \ln \sum_{\mathcal{L}_{i|h}>0} \mathcal{L}_{i|h}\varphi_{hi} = \sum_i \sum_{\mathcal{L}_{i|h}>0} \frac{\mathcal{L}_{i|h}\varphi_{hi}}{\mathcal{L}_i} \frac{\partial \ln \varphi_{hi}}{\partial \boldsymbol{\beta}} = \sum_i \sum_{\mathcal{L}_{i|h}>0} \mathcal{P}_{hi} \frac{\partial \ln \varphi_{hi}}{\partial \boldsymbol{\beta}}$$

with

$$\mathcal{P}_{hi} = \frac{\mathcal{L}_{i|h}\varphi_{hi}}{\mathcal{L}_i} = \prod_{j\in\mathcal{S}_{hi}} \Pr(Y_{ij} = y_{ij}|D_{hi} = d_{hi}) \Pr(D_{hi} = d_{hi}) \Big/ \prod_{j\in\mathcal{S}_i} \Pr(Y_{ij} = y_{ij})$$
$$= \Pr(D_{hi} = d_{hi}|Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots) = \Pr(D_{hi} = d_{hi}|Y_i = \boldsymbol{y}_i),$$

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 $Y_i = \boldsymbol{y}_i$. 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

$$Q^{(s)} = \sum_i \sum_{\mathcal{L}_{i|h}>0} \hat{\mathcal{P}}^{(s)}_{hi} \left(\ell_{i|h} + \ln \varphi_{hi}\right) \tag{8}$$

– whereby $\hat{\mathcal{P}}^{(s)}_{hi}$ 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 setup by

$$-\frac{\partial^2\ell}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'} = -\sum_i \sum_{\mathcal{L}_{i|h}>0} \mathcal{P}_{hi}\frac{\partial^2\ell_{i|h}}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'} - \sum_i \sum_{\mathcal{L}_{i|h}>0} \mathcal{P}_{hi}\frac{\partial\ell_{i|h}}{\partial\boldsymbol{\alpha}}\frac{\partial\ell_{i|h}}{\partial\boldsymbol{\alpha}'} + \sum_i \left[\sum_{\mathcal{L}_{i|h}>0} \mathcal{P}_{hi}^{(s)}\frac{\partial\ell_{i|h}}{\partial\boldsymbol{\alpha}}\right]\left[\sum_{\mathcal{L}_{i|h}>0} \mathcal{P}_{hi}^{(s)}\frac{\partial\ell_{i|h}}{\partial\boldsymbol{\alpha}}\right]'$$

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 (8) 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, that 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 next section, fewer than 10 Newton-Raphson steps were needed throughout and the run-time was generally just a few seconds on a contemporary desktop computer.

4 An Example Application: The 2010 Election to the UK House of Commons

4.1 Fitting the Finite Mixture Model to British Election Study Data

The UK is for several reasons to be suited for the examining the applicability of a method to disentangle tactical from non-tactical voting for several reasons: Its electoral system is the purest form of a first-past-the-post system, thus leading to situations that unambiguously give incentives to vote tactically. Nevertheless, its party system is sufficiently complex to allow for a distinction between tactical and non-tactical voting. In the US the dominance of the two main parties is so strong that interesting patterns of tactical or non-tactical voting are hard to observe, whereas in the UK some of the minor parties have regional strongholds, so that parties that are – on a national level – third, fourth or even fifth-placed in terms of voting share can gain representation in the House of Commons. As a consequence, it will not always be the same parties that can profit from tactical voting decisions. Finally, tactical voting has time and again been subject to public and academic debate, including a debate about how to measure it. The present section discusses the application of the method developed in this paper to the General Election to the House of Commons in 2010. The data used for this purpose comes from the post-election wave of the British Election Study (BES) (Clarke et al., 2010). Throughout the analyses of this chapter, only respondents from England, Scotland and Wales are included, because of the radically different shape of the party system in Northern Ireland, where none of the three major parties run candidates. The analyses also exclude respondents from the uncontested constituencies of the current Speaker of the House of Commons and the previous speaker, that is from Buckinghamshire and Glasgow North West. Finally, in preparing the data it has been made sure that for each respondent only those parties or candidates appear in the choice set that actually ran in the respondent's voting district. (Information on which parties run candidates in which districts was obtained from the website of The Electoral Commission of the UK [2014].)

As discussed earlier in the paper, tactical voting is a deviation in the voting decision from the most preferred alternative to one that may not be the most preferred one, but is the preferred alternative among the electorally viable ones. It was also stated that such preferences are not directly available to the empirical investigator, but that factors predicting choices can be used as a proxy for these preferences. Hence to reconstruct tactical voting one first needs to identify some good predictor variables for voting choices. In the present paper this is done based on a conditional logit model – of the same type as it is used to model choices *conditional* on using a tactical or non-tactical voting calculus. Fortunately, the BES data makes available a variety of potential predictor variables, such as parties' and respondents' positions on the issues of taxation and civil rights, occupation (from which a measure of social class can be constructed), party identification, and respondents' feelings about the parties and their leaders. Taking into account all these potential predictors, adequately modeling their influence means navigating between the Skylla of omitted-variable bias and the Charybdis of overspecification and multicollinearity.

Table 1 documents the search for an appropriate predictive model for voting choice. The first step in this search is the construction of a simple issue-based model motivated by the spatial theory of voting. This model, V_1, posits an influence of the squared distance between voters' on positions and the respective parties' positions on the question whether taxes should be reduced or social spending should be increased and on the question whether protecting defendants' rights is more/less important than effectively fighting crime. That is, model V_1 assesses the role of two issues that play a prominent role for ideological distinctions along the economic left-right axis and the libertarian-authoritarian axis, respectively. Note that the issue-variables where rescaled from the range from 0 to 10 to the range from -0.5 to 0.5 in order to obtain coefficient, that appear legible in the table. In line with expectations coming from a spatial theory of voting, voters are the less likely to vote for a party the higher the (squared) distance between them and the party in question, as the clearly statistically significant negative coefficients of the square distances indicate.

Unfortunately, the BES data contain information about respondents' perceived issue positions only with respect to the Conservatives, Labour, the Liberal Democrats, the Scottish National Party (SNP), and the Welsh nationalist party Plaid Cymru. So the minor parties, the Green Party, the British National Party (BNP), and the United Kingdom Independence Party

Table 1: Constructing the basic choice model

	V_1	V_2	V_3	V_4	V_5	V_6
Δ^2 Tax/spend	-7.016^{***}	-6.572^{***}	-6.571^{***}	-3.078^{***}	-0.640	
	(0.535)	(0.502)	(0.507)	(0.607)	(0.812)	
Δ^2 Fight crime/rights	-3.319^{***}	-3.186^{***}	-3.139^{***}	-1.887^{***}	-0.954	
	(0.340)	(0.317)	(0.323)	(0.426)	(0.537)	
Minor parties		-3.362^{***}	-3.194^{***}	-1.801^{***}	-1.611^{***}	-1.569^{***}
		(0.108)	(0.111)	(0.131)	(0.142)	(0.129)
Labour × Working class			0.814^{***}	-0.055	0.145	
			(0.087)	(0.148)	(0.174)	
LibDem × Professional			-0.153	0.631^{***}	0.390^{*}	0.265
			(0.115)	(0.174)	(0.187)	(0.160)
LibDem × Business			0.110	0.879^{***}	0.687^{**}	0.565^{***}
			(0.133)	(0.190)	(0.213)	(0.169)
Cons × Professional			0.325^{**}	0.432^{*}	0.254	
			(0.106)	(0.175)	(0.199)	
Cons × Business			0.622^{***}	0.423^{*}	0.184	
			(0.120)	(0.193)	(0.219)	
Party identification				3.014^{***}	1.794^{***}	1.806^{***}
				(0.077)	(0.091)	(0.089)
Feeling party					5.388^{***}	5.445^{***}
					(0.392)	(0.389)
Feeling party leader					1.938^{***}	2.044^{***}
					(0.342)	(0.337)
Log-likelihood	−2291.1	−2804.7	−2729.9	−1359.5	−1090.5	−1094.2
Deviance	4582.3	5609.4	5459.9	2719.1	2181.0	2188.4
N	2149	2252	2246	2177	2177	2177

Notes: Maximum likelihood estimates with standard errors in parentheses. p-value symbols: $***$: $p < .001$, $**$: $p < .01$, $*$: $p < .05$.

(UKIP) and all other parties had to be excluded to from the data to which model V_1 was fitted to. In order to integrate potential preferences for these minor parties into the analysis, some adjustments were made in the construction of model V_2: The distances between voters and these minor parties were set to zero and to compensate for this, a dummy variable was included to represent the absence of information about the perceived positions of these parties. This adjustment seems to work, since the coefficients of the squared distances do not change much (the change of the coefficient of the squared tax/spend distance is below one standard error of the estimate). Comparing the deviance values for models V_1 and V_2 should not mislead: the deviance of model V_2 is higher because more data – more respondent-party-pairs are used for fitting model V_2.

Class was in the past often considered as one of the most important predictors of voting in Britain (Goldthorpe, 1999). This motivates the extension of model V_2 by some party-class dummy interactions: Since the Labour Party traditionally appealed to the member of the industrial working class, model V_3 includes a interaction term of a Labour-Party dummy and a working class dummy. In addition model V_3 includes interaction terms of dummy variables for the Liberal Democrats and the Conservatives with dummy variables for the business and professional classes. The class dummies were created based on respondents' statements about their occupational group, that was asked about with ten response categories, viz. "Professional or higher technical work", "Manager or senior administrator", "Clerical", "Sales or services", "Small business owner", "Foreman or supervisor of other workers", "Skilled manual work", "Semi-skilled or unskilled manual work", "Never worked", and "Other". The dummy variable for the professional group was set to unity for all respondents falling into the first category of the occupational variable, the dummy variable for the business class was set to unity for all respondents falling into the category "Manager or senior administrator" and "Small business owner", whereas the working class dummy was set to unity for respondents falling into the categories "Foreman or supervisor of other workers", "Skilled manual work", or "Semi-skilled or unskilled manual work". The coefficient estimates for the interaction terms involving the Labour Party and Conservative Party dummies are coherent with traditional concepts of class voting in Britain: They indicate that, even after taking into account parties' and voters' issue positions, Labour is more often chosen by members of the working class whereas the Conservatives are more often chosen by members of the business and professional groups. (The coefficients of interaction terms involving the Liberal Democrats dummy remain statistically insignificant and very small, however.)

Moving forward in the "funnel of causality" leading to vote decisions model V_4 includes the party identification of respondents. This is represented by a dummy variable that attains unity for respondent-party combinations where the respondent identifies with the respective party. The coefficient of party identification is clearly statistically significant. Also, the inclusion of party identification affects the coefficients of several other coefficients in the model. The coefficients of the squared issue distances become considerably smaller, suggesting that issue distances are associated with party identification – either influenced by party identification or reflecting it. Furthermore, the coefficient of the Labour Party–working class interaction term now is close to zero, whereas the interaction terms involving the Liberal Democrats dummy now are positive and statistically significant. This suggests that party identification is a main mediating factor linking working class membership and Labour Party support, whereas it seems to play less of a role in mediating membership in the business and professional classes with the support for the Liberal Democrats and the Conservatives. Including party identification costs about 90 observations, yet the improvement of model fit is considerable: The deviance is more than halved from 5459.9 to 2719.1.

Model V_5 adds two variables that may be considered to be the most proximal causes of voting decisions: respondents' stated feelings about the parties and their leaders. Originally in the BES data the response categories range from 0 to 10, for the analysis the feeling variables where rescaled to the range from -0.5 to 0.5. The inclusion of the feeling variables leads to another improvement of model fit, reducing the model deviance from 2719.1 to 2181.0, without any loss of observations. Furthermore, it has consequences for the coefficients in the model. The coefficients of the squared issue distances are now much smaller, failing to attain statistical significance. Further, the coefficients of the interaction terms of the Conservative dummy and the class dummies are reduced in size and fail to attain statistical significance, and even the coefficient of party identification is considerably reduced. This may indicate that feelings towards the parties and their leaders are highly correlated with these variables, however, since the deviance is considerably reduced by including the feeling variables, they are not redundant. Instead, the most plausible interpretation is that respondents' feelings toward parties and their leaders acts as mediators between the other other variables and voting choice, because it is the most proximal causal factor. Certainly, the apparent disappearance of the effects of many other variables does not indicate that they are irrelevant, but that their effect on the dependent variable is "overshadowed" by the feelings towards parties and their leaders. That being said, the fact that some of the coefficients are now small and statistically insignificant in model V_5 suggests that they can be dropped from the model without much loss in predictive power, thus achieving more parsimonious predictions. This is borne out by the coefficient estimates of model V_6, all predictors without statistically significant estimates are dropped. Estimates of coefficients of those variables that are kept in the model change little and there is hardly an increase in the deviance between the model and the data. The differences between the deviance of models V_5 and V_6 is 7.38. The likelihood ratio test of model V_5 against V_6 is therefore, with 5 degrees of freedom, statistically insignificant ($p = 0.194$).

Model V_6 now forms the point of departure for the model that is used to recover tactical and non-tactical voting types. The predictor variables for this model are now used throughout for the mixture components – that correspond to equation (5) or (9) – of the finite mixture model.

Tactical voting to avoid wasting one's vote means, as discussed previously, to choose from a restricted set of electorally viable parties or candidates, instead of the full set of available parties and candidates. Yet this still leaves open the question about what criteria for viability of a party or candidate are used by the voters. If voters only take into account their experience from the past, they will use the parties' district vote shares from the *previous* election as criterion. If they are able to form "rational expectations", that at least on average predict the outcome of the election at district level more or less accurately, the district vote shares of the parties in the *current* will be the criterion for their viability. If voters take into account both information from the past and information available in the form of opinion polls published during the campaign to the election, then a combination of district vote shares form the previous election and parties' current popularity in the polls will be the criterion.

For the empirical analysis of tactical voting during the 2010 election to the UK House of Commons, obtaining district level results from the previous election and the current election is relatively straightforward. Per-constituency electoral results for the 2005 and 2010 elections to the House of Commons can be obtained from The Electoral Commission of the United Kingdom (The Electoral Commission, 2014). The application of the 2005 data for predicting the viability of parties within districts in the 2010 election is less straightforward however, because between 2005 and 2010 many voting district boundaries were changed. Fortunately, notional 2005 results for the 2010 districts are available from Norris (2010) (for a discussion of the construction of notional results, see e.g. Borisyuk et al., 2010).

The combination of district results and poll results to obtain dynamically adjusted party viability is of course confronted with the problem that poll results usually refer to the parties' popularity on a national level, which is not the level at which judgments about the electoral viability of parties in a district are made. In this paper this problem is addressed by following a common practice of projecting electoral results in Britain, which rests on the so-called uniform national swing (UNS) approach (e.g. BBC, 2010): The national swing here refers the difference between parties' popularity in the current polls and their national level vote share in the previous election. These differences are then uniformly added to the parties' vote shares in each district: Suppose $n_{2005,j}$ is the percentage vote share of party j in the 2005 election, p_{tj} is the percentage support of party j at time t in the polls (where time is measured in days) and $c_{2005,ij}$ the percentage vote share of party j in district i in 2005. Then the projected vote share of this party in district i for time t is $\hat{c}_{tij} = c_{2005,ij} + (p_{t-1,j} - n_{2005,j})$. This works however only for the three major parties, Conservatives, Labour, and Liberal Democrats, for whom polls report percentage shares of support at the national level, but not for the other parties, including SNP and Plaid Cymru, which do not have much weight at the national level, but are highly competitive in Scottish and Welsh districts, in some of them even winning pluralities. In order include these parties into the projection of district-level vote shares a modification of the uniform national swing approach is used: Polls usually report support percentages for all parties other than the three major ones at least in combination, hence it is possible to compute a national swing with respect to all other parties as a group. This component of the swing is then added for each district to the party of the group of other parties that has achieved the relatively largest vote share in the previous election.

The incentive to vote tactically may vary across districts according to the respective vote shares of the parties. First, voters may be more likely to avoid wasted votes, the more "hopeless" third candidates or parties are. Second, voters may be more likely to avoid wasted votes the close the race is in a district. This suggests two variables for the prediction of levels of tactical voting, first, the difference in support between the second-placed and third-placed parties or candidates in a district – as a measure of the "hopelessness" of third parties – and, second, the difference in support between the first-placed and the second-placed party in the district, as a measure of the closeness of the race.

Estimation results for the corresponding variants of the finite mixture model are presented in tables 2 and 3. Table 2 shows the estimates of the coefficients of the logit equation for probability to vote tactically. $F_{2005,1}$ and $F_{2005,2}$ are fit to the data under the assumption that the viability of parties or candidates is determined from their results in their previous elections. F_{dyn1} and F_{dyn2} are fit to the data under the assumption that the viability of parties/candidates is determined form previous results updated by information from polls and applying the Uniform National Swing technique described earlier. $F_{2010,1}$ and $F_{2010,2}$ are fit to the data under the assumption that the viability is determined from the district results of the current election. Models $F_{2005,1}$, F_{dyn1}, and $F_{2010,1}$ contain only the variable that represents the "hopelessness" of third parties, as described in the previous paragraph (where the variable is constructed from districts results of the previous election, dynamically updated results and district results of the current election, respectively). Models $F_{2005,2}$, F_{dyn2}, and $F_{2010,2}$ add to these a variable representing the closeness of the race in the district, as described in the previous paragraph.

The estimates for models $F_{2005,1}$, F_{dyn1}, and $F_{2010,1}$ reported in table 2 suggest that context indeed matters for tactical voting: The positive coefficient of the difference in support between the second- and third-placed party in a district indicates that the tendency to vote tactically increases with the "hopelessness" of third parties. The notion that closeness matters for tactical voting seems however to find little support by the estimates models $F_{2005,2}$, F_{dyn2},

Table 2: Coefficients of the mixture component of tactical voting models

	$F_{2005,1}$	$F_{2005,2}$	F_{dyn1}	F_{dyn2}	$F_{2010,1}$	$F_{2010,2}$
(Intercept)	-3.028^{***}	-3.197^{***}	-3.058^{***}	-3.418^{***}	-3.162^{***}	-4.230^{***}
	(0.430)	(0.749)	(0.362)	(0.578)	(0.392)	(0.731)
Δ Support 2nd and 3rd	7.647^{***}	8.126^{**}	7.787^{***}	8.274^{***}	8.966^{***}	11.406^{***}
	(2.191)	(2.812)	(1.439)	(1.564)	(1.812)	(2.344)
Δ Support 1st and 2nd		0.665		1.830		4.051^{*}
		(2.301)		(2.103)		(1.834)
Log-likelihood	−1048.1	−1048.1	−1032.8	−1032.3	−1028.8	−1025.7
Deviance	2096.3	2096.2	2065.6	2064.7	2057.5	2051.5
N	2177	2177	2177	2177	2177	2177

Notes: Maximum likelihood estimates with standard errors in parentheses. p-value symbols: $***$: $p < .001$, $**$: $p < .01$, $*$: $p < .05$.

and $F_{2010,2}$: Throughout the coefficient of the difference in support between the first-placed and second-placed party is positive, contradicting the notion that closeness of the race is an incentive for tactical voting (this notion implies a negative coefficient). Furthermore, in $F_{2005,1}$ and F_{dyn1} this coefficient fails to attain statistical significance.

Table 3 shows the coefficients of the mixture components, which describe how voters choose from among the alternatives they take into consideration (that is, from all alternatives if they vote sincerely and form the viable alternatives if they vote tactically). The coefficient estimates do not lead to new conclusions relative to model V_6 discussed earlier. By and large the coefficient estimates in table 3 show the same direction as the coefficient estimates for model V_6 reported in table 1, but tend to be moderately larger in absolute size.

A comparison of the goodness-of-fit of models $F_{2005,1}$, F_{dyn1}, and $F_{2010,1}$ may give some hints on whether voters use only information about past performance of the parties to come to judgments about their viability or whether they update these judgments they receive during the campaign. Model F_{dyn1} fits considerable better to the data than model $F_{2005,1}$, which suggests that voters do update their judgments. Perhaps surprisingly, the goodness of fit of model $F_{2010,1}$, which in effect implies that voters use information available to them only *after* the election fits even better to the data, although the improvement of fit relative to model F_{dyn1} is rather modest. This finding allows two interpretations: either voters indeed form "rational expectations" or the Uniform National Swing technique does not give a perfect account of how

Table 3: Coefficients of the choice component of tactical voting models

	$F_{2005,1}$	$F_{2005,2}$	F_{dyn1}	F_{dyn2}	$F_{2010,1}$	$F_{2010,2}$
Party identification	2.025^{***}	2.026^{***}	2.038^{***}	2.037^{***}	2.044^{***}	2.054^{***}
	(0.104)	(0.104)	(0.103)	(0.103)	(0.105)	(0.105)
Feeling party	5.835^{***}	5.833^{***}	6.076^{***}	6.078^{***}	6.079^{***}	6.078^{***}
	(0.432)	(0.432)	(0.445)	(0.444)	(0.445)	(0.445)
Feeling party leader	2.450^{***}	2.450^{***}	2.304^{***}	2.290^{***}	2.479^{***}	2.429^{***}
	(0.384)	(0.384)	(0.389)	(0.389)	(0.392)	(0.393)
Minor parties	-1.281^{***}	-1.280^{***}	-1.288^{***}	-1.292^{***}	-1.248^{***}	-1.246^{***}
	(0.142)	(0.142)	(0.142)	(0.142)	(0.143)	(0.144)
LibDem × Professional	0.299	0.302	0.173	0.163	0.246	0.247
	(0.184)	(0.184)	(0.183)	(0.184)	(0.187)	(0.187)
LibDem × Business	0.812^{***}	0.810^{***}	0.670^{***}	0.653^{***}	0.753^{***}	0.733^{***}
	(0.194)	(0.194)	(0.195)	(0.196)	(0.199)	(0.199)
Log-likelihood	−1048.1	−1048.1	−1032.8	−1032.3	−1028.8	−1025.7
Deviance	2096.3	2096.2	2065.6	2064.7	2057.5	2051.5
N	2177	2177	2177	2177	2177	2177

Notes: Maximum likelihood estimates with standard errors in parentheses. p-value symbols: $***$: $p < .001$, $**$: $p < .01$, $*$: $p < .05$.

voters update their viability judgments in the light of information obtained from published poll results. Which criterion for viability is preferable – that based on dynamically updated past results or that based on the constituency results of the actual 2010 election? Judging from the goodness of fit of the various models, the 2010 results criterion seems preferable. Yet intuitively it seems more plausible not to rely on a criterion that seems to rest on the assumption that voters base their judgments on information available only after the election. Given that the difference in goodness of fit between the models using these two criteria is not very large (whether it is statistically significant cannot be tested using a likelihood-ratio test, because models based on the two different viability criteria are not nested and consume the same degrees of freedom) in the following model F_{dyn1} is retained for further examination.

4.2 Convergent Validity: The Finite Mixture Approach and a "Traditional" Method of Identifying Tactical Voters

In the post-election survey of the British Election Study, respondents were not only asked about their vote decision, but also about the reasons for their decisions, in a semi-open format. Among the response categories respondents could choose from were "I really prefer another party but it stands no chance of win" and it was further recorded if respondents volunteered a response like "I vote tactically". If the finite mixture model is to deliver alternative, if not improved estimates of tactical voting it should at least lead to similar results as traditional methods of the discovery of tactical voting. The degree to which this is the case will be examined in the following.

Figure 2 shows the combined percentages of respondents who either stated to really have preferred another party or to have voted tactically and the percentage of tactical voting obtained from the sample average of respondents' posterior probabilities to have voted tactically obtained from the finite mixture model discussed above (in particular from model F_{dyn1}). The estimated percentage of tactical voting obtained from the finite mixture model is, as the figure shows, slightly higher than the percentage of stated voting reasons that indicate tactical voting. Does that indicate that the finite mixture model leads to an overestimate of tactical voting? Not necessarily – stating to have voted tactically is not a necessary condition to have done so. It is not implausible that some voters state e.g. that the party has the best policies if this is the party with the best policies among only those they consider to be viable.

That two measures of tactical voting lead to the same average levels is not a proof of convergent validity, however. A further requirement is that both measures are actually realated. In the present case, this means that with an increased posterior probabilitiy of a tactical vote obtained from the finite mixture model the proportion of respondents who state to have voted tactically or to rather prefer another party should also increase. This is borne out by figure 3 which visualizes, in the form of a conditional density plot, the relation between the posterior probability of a tactical vote and the proportions of stated reasons for the vote. However, the relation between the posterior probability of a tactical vote and the stated reasons for voting is not perfect: If the relation would be perfect, the gray areas should form a triangle from the lower left corner of the diagram to its top right corner. Instead, the diagram shows that whatever the posterior probability of a tactical vote, the proportion of stated vote reasons indicating a tactical vote are lower than the model implied posterior probability, a finding that mirrors that of figure 2, where the average of stated tactical votes is lower than the average

Figure 2: Percentages of tactical voting – based on stated reasons and posterior probabilities obtained from the finite mixture model

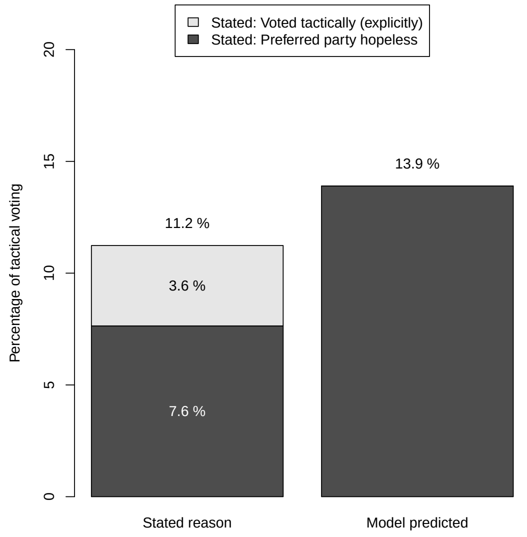
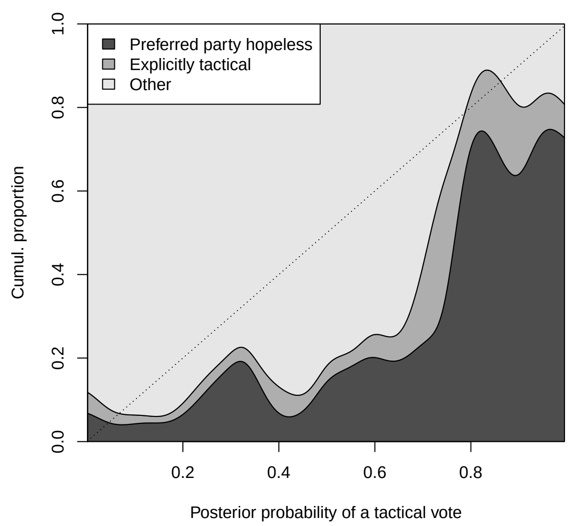


Figure 3: The relation between model predicted tactical voting and stated voting reasons: A conditional density plot



of posterior probabilities of tactical voting.

Since in the post-election survey of the British Election Study respondents were not only asked about the reasons for their vote, but also – if they stated to have preferred another party without a chance to win the seat or to have voted tactically – what party they "really" preferred, it is possible to probe even further into the convergent validity of the traditional measure of tactical voting and tactical voting reconstructed by the finite mixture model. This is done by constructing a "party-preference" variable in the following way: For those respondents who have stated to have preferred another party that did not have a chance to win or explicitly stated to have voted tactically, the party-preference variable contains the responses of the follow-up question about which party they "really" referred. For those respondents who gave another reason, the party preference variable contains the responses to the question about their voting decision.

This party-preference variable then is used to conduct two further checks for convergent validity: First, a conditional logit discrete choice model with the same chose predictors as in V_6 and F_{dyn1} is fitted to these party preferences, the resulting model being denoted by P_6. From this party-preference model, predicted probabilities are obtained and compared with the predicted choice probabilities of V_6 and the predicted choice probabilities of the non-tactical voting mixture component of the finite mixture model F_{dyn1}. If the choice probabilities of P_6 and of the non-tactical component of F_{dyn1} agree closely, or at least more closely than those of V_6 do with P_6, then this will point to a convergent validity of the two ways of reconstructing sincere party preferences. Second, one can compare model deviances between the party preference variable and these choice probabilities as a measure for how well they explain sincere party preferences. If there is a good convergent validity, then the deviance of the non-tactical component of F_{dyn1} should not be much worse than P_6 and perhaps much better than that of V_6.

Table 4 shows the coefficient estimates of the basic discrete choice model V_6 of reported votes in the post-election survey of the BES developed at the beginning of the previous subsection, the estimates of the coefficients of the conditional choice component of the finite mixture model F_{dyn1}, and the discrete choice model fitted to the "real" party preferences P_6.

Table 4: Comparing the performance of the tactical voting model – coefficients and goodness of fit

	V_6	F_{dyn1}	P_6
Party identification	1.806^{***}	2.038^{***}	2.037^{***}
	(0.089)	(0.103)	(0.096)
Feeling party	5.445^{***}	6.076^{***}	6.759^{***}
	(0.389)	(0.445)	(0.436)
Feeling party leader	2.044^{***}	2.304^{***}	2.179^{***}
	(0.337)	(0.389)	(0.373)
Minor parties	-1.569^{***}	-1.288^{***}	-1.268^{***}
	(0.129)	(0.142)	(0.129)
LibDem × Professional	0.265	0.173	0.315
	(0.160)	(0.183)	(0.177)
LibDem × Business	0.565^{***}	0.670^{***}	0.420^{*}
	(0.169)	(0.195)	(0.193)
Log-likelihood	−1094.2	−1032.8	−931.2
Deviance	2188.4	2065.6	1862.3
N	2177	2177	2170

Notes: Maximum likelihood estimates with standard errors in parentheses. p-value symbols: $***$: $p < .001$, $**$: $p < .01$, $*$: $p < .05$.

There are two things of note here: First, a comparison between the model deviances of V_6 and F_{dyn1} indicates that the latter fits considerably better to the actual voting choices than the former. Second, by and large the coefficients of F_{dyn1} are closer than the coefficients of V_6 to those of P_6. (The model deviance of model P_6 should however not compared with those of V_6 or F_{dyn1} because it is fitted to a different response variable.)

If the probability that voter i chooses party j as implied by model V_6 is denoted by $\pi_{ij}[V_6]$, the probability that voter i really prefers party j as implied by model P_6 is denoted by $\pi_{ij}[P_6]$, and the probability that a voter would choose party j if she voted sincerely as implied by model

Table 5: Symmetric Kullback-Leibler divergences between models V_6, F_{dyn1}, and P_6

	$\pi_{ij}[V_6]$	$\pi_{ij\|1}[F_{\text{dyn1}}]$	$\pi_{ij}[P_6]$
$\pi_{ij}[V_6]$	0.0	20.1	35.6
$\pi_{ij\|1}[F_{\text{dyn1}}]$	20.1	0.0	5.9
$\pi_{ij}[P_6]$	35.6	5.9	0.0

Notes: π_{ij} correspond to the probability that respondent j chooses alternative i; $\pi_{ij|1}$ correspond to the probability that respondent j chooses alternative i if she *votes sincerely.*

F_{dyn1} is denoted as $\pi_{ij|1}[F_{\text{dyn1}}]$, then the (dis-)similarity of the implications of V_6, P_6, and F_{dyn1} with respect to sincere votes can be compared by computing the (symmetric) Kullback-Leibler divergences between the vote distributions given by $\pi_{ij}[V_6]$, $\pi_{ij}[P_6]$, and $\pi_{ij|1}[F_{\text{dyn1}}]$ (since V_6 does not distinguish between sincere and tactical votes). The symmetric Kullback-Leibner divergence between $\pi_{ij}[V_6]$, $\pi_{ij}[P_6]$, and $\pi_{ij|1}[F_{\text{dyn1}}]$ is shown in table 5, from which it becomes obvious that P_6 and F_{dyn1} agree much more in terms of their predictions about sincere votes than V_6 and P_6.

The second above mentioned way to check how well the three models agree with respect to their implications for counterfactual sincere votes is to compare deviances of the stated party preferences from the three models. These deviances are shown in table 6, from which it becomes obvious that the finite mixture model, even though fitted to actual votes, predicts stated “real” party preferences almost as good as a discrete choice model fit directly to them, and clearly much better than a discrete choice model that does not distinguish between sincere and tactical voting.

To sum up the findings of this subsection, the traditional way to uncover tactical voting and the proposed finite-mixture approach do not agree perfectly, but they sufficiently do so to warrent the notion that they measure the same thing. Both the stated-reasons technique and the finite-mixture technique lead to comparable estimates of the level of tactical voting in the sample. The component of the finite mixture model pertaining to non-tactical votes leads to choice predictions very similar to those of a discrete choice model fitted to the stated party preferences. Finally, the non-tactical component of the finite mixture model explains these party preferences almost as good as a discrete choice model fitted directly to them.

Table 6: Comparison of V_6, F_{dyn1}, and P_6 in terms of predicting stated “real” party preferences (as opposed to actual votes) measured by the model deviance.

	Deviance
V_6	1895.4
F_{dyn1}	1868.0
P_6	1862.3

5 Discussion

5.1 The Relation of the Proposed Method to Some other Models

As mentioned in the introduction, the finite mixture model proposed in this paper is not the first attempt at recovering tactical voting without taking recourse to stated reasons of voting decisions. The first and in this regard still most notable approach is that taken by Alvarez, Nagler, and Boehmke (Alvarez and Nagler, 2000; Alvarez et al., 2006), who propose to include district-level electoral results into the set of predictor variables in discrete choice models and, in addition, also to include the interaction effects of these district-level results with individual-level predictors of voting choice.

The discrete choice model used by them is more complex than the one that forms the baseline of the finite mixture model proposed here – they use a multinomial probit model. While the multinomial probit model seems to be more attractive from the perspective of random-utility modeling, it not only poses considerable technical difficulties, but also restricts model specification to situations where the choice set is identical for all voters. This already limits the applicability of multinomial probit models to the case of voting in British elections, since not all British voters are faced with the same set of alternatives: In addition to the parties that field candidates in England, in Wales Plaid Cymru is another relevant contender that even gains plurality in some districts and the same can be said about Scotland and the SNP. Yet the main limitation, at least from the perspective of this paper, is that the approach taken by Alvarez et al. (2006) does not rest on an explicit specification of what it means by voting tactically to avoid wasting one's vote. The approach proposed in this paper rests on the notion that tactical voting means to vote for another alternative than the otherwise most preferred yet not electorally viable. The Alvarez-Nagler approach in contrast rests on the assumption that the viability is part of the set of independent variables that influence one's preferences.

That being said, the two main avenues – specifying voting calculi in terms of restricted (or non-restricted) choice sets and specifying preference functions that include the viability of alternatives – are not completely unrelated. Note that one can rewrite equation (5) with

$j \in \mathcal{S}_{hi}$ in the form

$$\pi_{ij|h} = \frac{w_{hij} \exp(\boldsymbol{x}'_{ij}\boldsymbol{\alpha})}{\sum_{k\in\mathcal{S}_i} w_{hik} \exp(\boldsymbol{x}'_{ik}\boldsymbol{\alpha})} \tag{9}$$

with "consideration dummies"

$$w_{hij} = \begin{cases} 1 & \text{for } j \in \mathcal{S}_{hi} \text{ and} \\ 0 & \text{for } j \notin \mathcal{S}_{hi}. \end{cases}$$

Suppose now one substitutes the binary "consideration dummies" w_{hij} by a function of some predictor variables, say

$$w_{hij} = \frac{\exp(\boldsymbol{u}'_{hij}\boldsymbol{\gamma})}{\sum_{k\in\mathcal{S}_i} \exp(\boldsymbol{u}'_{hik}\boldsymbol{\gamma})}$$

then $pi_{ij|h}$ becomes

$$\pi_{ij|h} = \frac{\exp(\boldsymbol{u}'_{hij}\boldsymbol{\gamma})[\sum_{l\in\mathcal{S}_i} \exp(\boldsymbol{u}'_{hil}\boldsymbol{\gamma})]^{-1} \exp(\boldsymbol{x}'_{ij}\boldsymbol{\alpha})}{\sum_{k\in\mathcal{S}_i} \exp(\boldsymbol{u}'_{hik}\boldsymbol{\gamma})[\sum_{l\in\mathcal{S}_i} \exp(\boldsymbol{u}'_{hil}\boldsymbol{\gamma})]^{-1} \exp(\boldsymbol{x}'_{ik}\boldsymbol{\alpha})} = \frac{\exp(\boldsymbol{u}'_{hij}\boldsymbol{\gamma} + \boldsymbol{x}'_{ij}\boldsymbol{\alpha})}{\sum_{k\in\mathcal{S}_i} \exp(\boldsymbol{u}'_{hik}\boldsymbol{\gamma} + \boldsymbol{x}'_{ik}\boldsymbol{\alpha})} \tag{10}$$

which is a regular conditional logit specification in which we have $\boldsymbol{u}_{hij}$ as a predictor for the viability of the alternatives. Yet to arrive at $w_{hij} = 0$ for some j we need $\exp(\boldsymbol{u}'_{hij}\boldsymbol{\gamma}) = 0$, hence $\boldsymbol{u}'_{hij}\boldsymbol{\gamma} = -\infty$. That is, a conditional logit model with predictors for the viability of alternatives can lead to a choice from a restricted set only in the limit, which one will certainly not reach if one tries to estimate $\boldsymbol{\gamma}$ from empirical data. Further, taking (4) and substituting in equations (6), (7), and (10) leads to

$$\pi_{ij} = \sum_{h=1}^{m} \frac{\exp(\boldsymbol{u}'_{hij}\boldsymbol{\gamma} + \boldsymbol{x}'_{ij}\boldsymbol{\alpha})}{\sum_{k\in\mathcal{S}_i} \exp(\boldsymbol{u}'_{hik}\boldsymbol{\gamma} + \boldsymbol{x}'_{ik}\boldsymbol{\alpha})} \frac{\exp(\boldsymbol{z}'_{hi}\boldsymbol{\beta})}{1 + \sum_{g=1}^{m} \exp(\boldsymbol{z}'_{gi}\boldsymbol{\beta})} \tag{11}$$

which cannot be transformed into a regular conditional logit format. On the other hand, the attempt to construct an analogous conditional logit model fails to lead to an identification of the parameters that specify the voting type probabilities φ_{hi}, because

$$
\begin{aligned}
\pi^*_{ij} &= \frac{\exp\left(\sum_{h=1}^m \boldsymbol{u}'_{hij}\boldsymbol{\gamma} + \boldsymbol{x}'_{ij}\boldsymbol{\alpha} + \sum_{h=1}^m \boldsymbol{z}'_{hi}\boldsymbol{\beta}\right)}{\sum_{k\in\mathcal{S}_i}\exp\left(\sum_{h=1}^m \boldsymbol{u}'_{hik}\boldsymbol{\gamma} + \boldsymbol{x}'_{ik}\boldsymbol{\alpha} + \sum_{h=1}^m \boldsymbol{z}'_{hi}\boldsymbol{\beta}\right)} \\
&= \frac{\exp\left(\sum_{h=1}^m \boldsymbol{u}'_{hij}\boldsymbol{\gamma} + \boldsymbol{x}'_{ij}\boldsymbol{\alpha}\right)\exp\left(\sum_{h=1}^m \boldsymbol{z}'_{hi}\boldsymbol{\beta}\right)}{\sum_{k\in\mathcal{S}_i}\exp\left(\sum_{h=1}^m \boldsymbol{u}'_{hik}\boldsymbol{\gamma} + \boldsymbol{x}'_{ik}\boldsymbol{\alpha}\right)\exp\left(\sum_{h=1}^m \boldsymbol{z}'_{hi}\boldsymbol{\beta}\right)} \\
&= \frac{\exp\left(\sum_{h=1}^m \boldsymbol{u}'_{hij}\boldsymbol{\gamma} + \boldsymbol{x}'_{ij}\boldsymbol{\alpha}\right)}{\sum_{k\in\mathcal{S}_i}\exp\left(\sum_{h=1}^m \boldsymbol{u}'_{hik}\boldsymbol{\gamma} + \boldsymbol{x}'_{ik}\boldsymbol{\alpha}\right)}.
\end{aligned}
\tag{12}
$$

To sum up, while the proposed finite mixture model arguably has similarities with a discrete choice model that involves the viability of alternatives as a predictor variable for preferences and choices, it cannot be reduced to it.

Another type of model to which the proposed finite mixture model bears resemblance is the random choice-set or random consideration-set model proposed by Ben-Akiva and Boccara (1995), the application of which to political science topics has been discussed by Steenbergen et al. (2011). It can even be argued that the finite mixture model is a modified or restricted version of a random consideration set model. The consideration-set approach envisages a two-stage process that ultimately leads to the choice of one element from a set of alternatives, e.g. the choice of a party or candidate to vote for. In the first stage, individuals decide for each alternative in the choice set $\mathcal{S}$ (which is identical for all individuals) whether to consider it or not, that is whether to keep alternative j in the individual consideration set $C_i \subseteq \mathcal{S}$ or not. In the second stage, individuals choose an alternative j from the consideration set with probability (note that the notation here and in the following differs a bit from Steenbergen et al. (2011), to keep it consistent with the earlier discussion)

$$\pi_{ij|C_i} = \frac{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\alpha})}{\sum_{k\in C_i}\exp(\boldsymbol{x}'_{ik}\boldsymbol{\alpha})}.$$

If it is assumed that the consideration set C_i is non-empty and the size of (number of elements in) the choice set is $|\mathcal{S}| = q$ then there are $m = q^2 - 1$ possible consideration sets. Suppose all these possible consideration sets are enumerated by numbers $h = 1, \dots, m$, so that any consideration set can be denoted by $\mathcal{S}_h$. Then the probability that individual i chooses alternative

j takes, with $\pi_{hij} = \pi_{ij|C_i=\mathcal{S}_h}$ and $\varphi_{hi} = \Pr(C_i = \mathcal{S}_h)$, the familiar form of equation (4). Thus the finite mixture model of this paper is quite similar to a random consideration set model. But is is also differs in terms of the number of possible consideration sets it admits. Suppose the finite mixture model is used to uncover whether British voters have voted sincerely ($h = 1$) or tactically ($h = 2$), as in the previous section, and an individual i has a total choice set of seven alternatives, then it admits $m = 2$ possible consideration sets, whereas a random consideration set model admits $m = 7^2 - 1 = 48$ consideration sets.

A second difference between the finite mixture model and a random consideration set model is the specification of the consideration set probabilities φ_{hi}. In contrast to the finite mixture model, the random consideration set model considers the inclusion of alternatives into the consideration set as a random process, where it is assumed that the probability of individual i taking alternative j into consideration, i.e. $j \in C_i$, is given by

$$\Pr(j \in C_i) = \omega_{ij} = \frac{1}{1 + \exp(\gamma_j - \boldsymbol{u}'_{ij}\boldsymbol{\beta})} = \frac{\exp(\boldsymbol{u}'_{ij}\boldsymbol{\beta} - \gamma_j)}{1 + \exp(\boldsymbol{u}'_{ij}\boldsymbol{\beta} - \gamma_j)} \tag{13}$$

and that the inclusion probabilities are independent, i.e. $\Pr(j_1, j_2 \in C_i) = \Pr(j_1 \in C_i)\Pr(j_2 \in C_i)$. According to Ben-Akiva and Boccara (1995) and Steenbergen et al. (2011) this leads to

$$\varphi_{hi} = \Pr(C_i = \mathcal{S}_h) = \frac{\prod_{j\in\mathcal{S}_h}\omega_{ij}\prod_{j\notin\mathcal{S}_h}(1-\omega_{ij})}{1 - \Pr(C_i = \emptyset)}, \tag{14}$$

where $\Pr(C_i = \emptyset) = 1 - \prod_{j\in\mathcal{S}}(1 - \omega_{ij})$.

According to Steenbergen et al. (2011) a problem in estimating a random consideration set model is that the corresponding log-likelihood function is not globally weakly concave, for which reason they use a Bayes estimator with a weakly informative prior on the model parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. But a lack of concavity is only a problem for unmodified Newton-Raphson algorithms, that require the Hessian to be negative definite. Quasi-Newton algorithms that involve line-search may be robust in this regard. Yet the real problem is that Ben-Akiva and Boccara (1995)'s original specification in itself does not make sure that $\sum_{h=1}^{m}\varphi_{hi} = 1$, a requirement for φ_{hi} to be a probability, a problem that becomes apparent if the construction of

$\Pr(C_i = \emptyset)$ and equation (13) are substituted into equation (14):

$$\begin{aligned}\varphi_{hi} &= \frac{\prod_{j\in\mathcal{S}_h}\omega_{ij}\prod_{j\notin\mathcal{S}_h}(1-\omega_{ij})}{1-\Pr(C_i=\emptyset)} = \prod_{j\in\mathcal{S}_h}\omega_{ij}\times\frac{\prod_{j\notin\mathcal{S}_h}(1-\omega_{ij})}{\prod_{j\in\mathcal{S}}(1-\omega_{ij})}\\ &= \prod_{j\in\mathcal{S}_h}\frac{\omega_{ij}}{1-\omega_{ij}} = \prod_{j\in\mathcal{S}_h}\exp(\boldsymbol{u}'_{ij}\boldsymbol{\beta}-\gamma_j) = \exp\left(\sum_{j\in\mathcal{S}_h}\boldsymbol{u}'_{ij}\boldsymbol{\beta}-\sum_{j\in\mathcal{S}_h}\gamma_j\right).\end{aligned} \tag{15}$$

The last expression in the equation is an exponential and thus can become arbitrary high for large values of $\boldsymbol{u}'_{ij}\boldsymbol{\beta}$ or large negative values of γ_j. For the φ_{hi} to be probabilities they must however be restricted to be smaller than or equal to unity. Further they must sum to one. This means that a restriction that implements an upper limit is not sufficient, it means also that as one of these probabilities increases at least another one has to decrease.

One way to assure that $\sum_{h=1}^{m}\varphi_{hi} = 1$ is to put further constraints on the model parameters: If one writes $\boldsymbol{v}_{hi} := \sum_{j\in\mathcal{S}_h}\boldsymbol{u}_{ij}$ and requires that

$$\exp\left(\sum_{j\in\mathcal{S}_h}\gamma_j\right) = \sum_{h=1}^{m}\exp\left(\sum_{j\in\mathcal{S}_h}\boldsymbol{u}'_{ij}\boldsymbol{\beta}\right) = \sum_{h=1}^{m}\exp\left(\boldsymbol{v}'_{hi}\boldsymbol{\beta}\right)$$

then φ_{hi} of equation (15) becomes

$$\varphi_{hi} = \exp\left(\sum_{j\in\mathcal{S}_h}\boldsymbol{u}'_{ij}\boldsymbol{\beta}-\sum_{j\in\mathcal{S}_h}\gamma_j\right) = \frac{\exp\left(\sum_{j\in\mathcal{S}_h}\boldsymbol{u}'_{ij}\boldsymbol{\beta}\right)}{\exp\left(\sum_{j\in\mathcal{S}_h}\gamma_j\right)} = \frac{\exp\left(\boldsymbol{v}'_{hi}\boldsymbol{\beta}\right)}{\sum_{g=1}^{m}\exp\left(\boldsymbol{v}'_{gi}\boldsymbol{\beta}\right)}$$

so that the sum-to-one requirement of the probabilities φ_{hi} is satisfied for all parameter values $\boldsymbol{\beta}$. Further, if one writes $\boldsymbol{z}_{hi} = \boldsymbol{v}_{hi} - \boldsymbol{v}_{1i}$, so that $\boldsymbol{v}_{hi} = \boldsymbol{z}_{hi} + \boldsymbol{v}_{1i}$ and $\boldsymbol{z}_{1i} = 0$, one obtains, since $\boldsymbol{z}_{1i} = 0 \Rightarrow \exp\left(\boldsymbol{z}'_{hi}\boldsymbol{\beta}\right) = 1$,

$$\varphi_{hi} = \frac{\exp\left(\boldsymbol{v}'_{hi}\boldsymbol{\beta}\right)}{\sum_{g=1}^{m}\exp\left(\boldsymbol{v}'_{gi}\boldsymbol{\beta}\right)} = \frac{\exp\left(\boldsymbol{z}'_{hi}\boldsymbol{\beta}\right)/\exp\left(\boldsymbol{v}'_{1i}\boldsymbol{\beta}\right)}{\sum_{g=1}^{m}\exp\left(\boldsymbol{z}'_{gi}\boldsymbol{\beta}\right)/\exp\left(\boldsymbol{v}'_{1i}\boldsymbol{\beta}\right)} = \begin{cases}\dfrac{1}{1+\sum_{g=2}^{m}\exp\left(\boldsymbol{z}'_{gi}\boldsymbol{\beta}\right)} & \text{for } h=1\\[2ex] \dfrac{\exp\left(\boldsymbol{z}'_{hi}\boldsymbol{\beta}\right)}{1+\sum_{g=2}^{m}\exp\left(\boldsymbol{z}'_{gi}\boldsymbol{\beta}\right)} & \text{for } h>1\end{cases}$$

like in equations (6) and (7). In sum, with proper restrictions in place that assure its identification, the consideration set model becomes a finite mixture model similar to that proposed in this paper.

5.2 Application of the Finite Mixture Model to Voting with Dual Ballots in Mixed Electoral Systems

First-past-the-post systems are not the only electoral systems that give opportunity or incentive to vote tactically. Voters may be lead to voting tactically even in proportional-representation (PR) systems (Cox, 1997). Firstly, as long as a parliament has less members than the size of the electorate, perfect representation is impossible and an effective threshold of representation exists. That is, even in PR systems there may be the possibility that a vote can be considered wasted by a voter if he decides to give it to a party that does not have a chance to surpass the threshold of representation. However, the lower this threshold is, the less likely this is to occur, because in order to be hopeless in the face of a low threshold of representation, a party must have a small number of supporters to begin with. Another, more often discussed motive to vote tactically lies in the fact that elections in PR systems typically lead to coalition governments. When faced with the possibility of multiple different government coalitions, voters may decide to vote, instead of their most preferred party, to a smaller party that is a likely partner in a coalition with the most preferred party. And if the most preferred party is likely to have more than one opportunities to form a coalition with other parties, a voter may choose to vote instead for that possible coalition partner, the inclusion of which into the coalition will lead to a composition of the government that as a whole most closely fits her preferences (Austen-Smith and Banks, 1988; Bargsted and Kedar, 2009). The incentive to vote tactically in this way may be the stronger, the more at risk a smaller coalition partner is not to pass the threshold of representation (Gschwend, 2007; Shikano et al., 2009).

The opportunities or incentives to vote tactically may further proliferate in mixed electoral systems with dual ballots, such as the German one. In elections to the German federal parliament, the Bundestag, voters are allowed to cast two ballots: The first ballot contributes to the votes for a party's candidate for the seat representing the seat in which the voter lives, and the second ballot contributes to the votes for a party list that determine the overall number of seats allocated to the party in the Bundestag. Thus a voter may be motivated to vote tactically in her first ballot, so not to waste it for a party candidate that has no chance of winning the district seat, and/or to to vote tactically in her second ballot, so as to support a likely coalition partner of the most preferred party, such that the preferred party is more likely to be included in a government coalition, and/or not to waste it for a party that has no chance to pass the threshold of representation.

The application of the finite mixture approach proposed in this paper to voting in mixed electoral systems with dual ballots is possible, if only a little more complicated. In this situation, the first-ballot votes (or vote intentions) are represented by one binary random variable $Y_{ij}^{(1)}$ where $Y_{ij}^{(1)} = 1$, if voter i casts her first ballot in favor of the candidate ran by party j, and $Y_{ij}^{(1)} = 0$ otherwise. Likewise, the second-ballot votes (or vote intentions) are represented by another binary random variable $Y_{ij}^{(2)}$, analogously constructed so that $Y_{ij}^{(2)} = 1$, if voter i casts her second ballot for the candidate list of party j, and $Y_{ij}^{(2)} = 0$ otherwise. Further one can construct two random variables $T_i^{(1)}$ and $T_i^{(2)}$ that represent the type of voting employed by voter i when casting her first and second ballot, respectively.

One could just apply the finite mixture model separately for the first- and second-ballot choices, but it may be more interesting and perhaps more efficient, to construct a comprehensive model for both ballots. Such a comprehensive model could involve a common set of predictor variables and coefficients for both choice probabilities in the mixing components of both the first and the second ballot, that is:

$$\begin{aligned}\Pr\left(Y_{ij}^{(1)} = 1|T_i^{(1)} = g\right) = \pi_{ij|g}^{(1)} &= \frac{\exp(\boldsymbol{x}_{ij}'\boldsymbol{\alpha})}{\sum_{k\in\mathcal{S}_{gi}^{(1)}}\exp(\boldsymbol{x}_{ik}'\boldsymbol{\alpha})}\\ \Pr\left(Y_{ij}^{(2)} = 1|T_i^{(2)} = h\right) = \pi_{ij|h}^{(2)} &= \frac{\exp(\boldsymbol{x}_{ij}'\boldsymbol{\alpha})}{\sum_{k\in\mathcal{S}_{hi}^{(2)}}\exp(\boldsymbol{x}_{ik}'\boldsymbol{\alpha})}\end{aligned} \tag{16}$$

where $\mathcal{S}_{gi}^{(1)}$ and $\mathcal{S}_{hi}^{(2)}$ are the consideration sets specific for the voting types employed in the first- and second-ballot vote choice, and $\boldsymbol{x}_{ik}$ is a vector of predictor variables of party preferences similar to those used in the application of the finite-mixture model to the 2010 BES data. Of course given that one would usually expect different strategems for first- and second-ballot tactical voting, it may be less sensible to employ a common specification for the probabilities $\varphi_{gi}^{(1)} = \Pr(T_i^{(1)} = g))$ and $\varphi_{hi}^{(2)} = \Pr(T_i^{(2)} = h))$.

If one assumes that probabilities (given the preference predictor variables and voting types) of the first- and second-ballot vote choices are (conditionally) indepedent, that is,

$$\Pr\left(Y_{ij}^{(1)} = 1, Y_{ij}^{(2)} = 1 | T_i^{(1)} = g, T_i^{(2)} = h\right) = \Pr\left(Y_{ij}^{(1)} = 1 | T_i^{(1)} = g\right) \Pr\left(Y_{ij}^{(2)} = 1 | T_i^{(2)} = h\right)$$

then the techniques developed for this paper do not need much modification to be applied to the analysis of tactical voting in dual-ballot mixed electoral systems. Such a conditional-independence assumption is not so artificial as it may seem: It means basically that the choice component of the model given by equation (16) is comprehensive enough to include all predictor variables relevant for electoral choice. On the other hand, exactly this substantive interpretation of the conditional-independence assumption may also motivate the desire to test it. Yet the questions about how such a test can be constructed and how the finite-mixture model can be modified to deal the violation of this independence assumption are not answered yet and require further research.

6 Conclusion

The present paper proposes a new method for empirically distinguishing tactical from non-tactical votes. At the core of this method lies the notion that, when voters abandon their most preferred alternative for another one that satisfies certain other criteria, they effectively choose from a reduced consideration set instead of the full choice set, of which the consideration set is a proper subset. In the case of Duvergerian tactical voting with the intent to avoid wasting one's vote, this reduced consideration set is circumscribed by the electoral viability of its members. This motive is made quite explicit by voters who in a survey state as reason for their vote that they rather preferred another party, but that it did stand a chance of winning. The basic idea here is that, by simple logic this means that such voters considered for choosing only those parties with such a chance.

From the a statistical perspective, having several types of voters with different patterns of voting, e.g. sincere or tactical voting, comes down to assuming that the probability distribution that describes votes is a mixture of several distinctive distributions – the mixture components – that characterize these voting types. From this finite mixture model, a method was devised to estimate the distribution of these components – the mixing distribution – and from there the posterior probabilities of individual voters to exhibit one of these components in their voting behavior. The method turned out to be a typical application of an EM algorithm to estimate parameters of a probability distribution in the present of missing information by maximizing a likelihood function that marginalizes over the missing data.

The proposed method then was applied to estimating the amount of tactical voting during the 2010 General Election to the House of Commons of the UK. In addition to illustrating the method, it also served as a validity check. As a measure of convergent validity, results obtained with the new method were compared with those obtained by a more traditional method to identify tactical voting: recording the reasons that voters give for their voting choice in an electoral survey. The result of this comparison appears reassuring in so far as the new method lead to an estimate of the prevalence of tactical voting roughly at the same level as the old method, and as on an individual level, posterior probabilities of tactical voting established with the new method showed a positive association with respondents' statements to have voted tactically.

The paper also compared the new method with some other models of electoral choice. This comparison showed that the new method cannot be reduced to discrete choice modeling with the viability of alternatives as a component of voters' utility function, but it showed that it can be characterized as a restricted version of an improved random consideration-set approach. In discussing the relation with the latter models, the paper also points to some identification problems in the way they were so far employed to political science problems and suggested a possible solution to these problems. Finally the technical aspects of expending the proposed model to the case of dual-ballot voting in mixed electoral systems was briefly discussed.

Whether potential technical difficulties in applying the finite mixture approach to dual-ballot voting in mixed electoral systems can be overcome is one question, and it appears that its answer is positive. Whether the method can be made empirically fruitful for an understanding of phenomena such as ticket-splitting in such electoral systems is another question, the answer of which depends on whether explanations of ticket-splitting and tactical voting in such systems can be integrated into the reduced-consideration-set framework. Clearly tactical voting in proportional representation systems that avoids voting for parties that are unlikely to pass the threshold of representation has the same formal structure as tactical voting in first-past-the-post systems (Cox, 1997). The difference is merely the size of the consideration set of viable alternatives.

The analysis of the relevance of policy-balancing (Bargsted and Kedar, 2009) will perhaps not require an approach such as the one put forward here, since it can be well addressed in the framework of a simple probabilistic choice model: If the probability to vote for a party is related to policy distance, then voters located in terms of their policy preferences between two parties will vote for either of these parties with roughly equal probability. And if they have the opportunity two cast two ballots, they will the more likely split these ballots the closer to equality their choice probabilities in favor of either of these parties are.

Threshold insurance for a coalition partner (Gschwend, 2007; Shikano et al., 2009) of a preferred party may however be a potential application of the method proposed here. Suppose a voter prefers a party who is likely to be one of the largest parties in parliament, but to need a coalition partner to become a governing party, suppose further that this party also has a viable candidate in a voters' district. Then she will probably cast her first ballot for this party. However, if the preferred coalition partner is in danger of falling below the threshold of representation with respect to its share of list votes, she may depart form her preferred major party in her second ballot and instead vote for this potential coalition partner. This kind of strategic voting with the second ballot can also be conceived of a restriction of the consideration set. Yet the criterion for inclusion into this restricted consideration set then is not viability, but availability as a coalition partner for the preferred party. And if threshold insurance is prevalent in an electorate of a dual-ballot system, one should find that such restriction to coalition partners in second-ballot votes increases, the more voters perceive the representation of coalition partners in peril. Whether this can be borne out empirically with the help of the method proposed in this paper is one of the potential avenues of further research.

References

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(01): 57–75.

Austen-Smith, David and Jeffrey Banks, 1988. "Elections, Coalitions, and Legislative Outcomes". *The American Political Science Review* 82(2): 405–422.

Bargsted, Matias A. and Orit Kedar, 2009. "Coalition-Targeted Duvergerian Voting: How Expectations Affect Voter Choice under Proportional Representation". *American Journal of Political Science* 53(2): 307–323.

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

Ben-Akiva, Moshe and Bruno Boccara, 1995. "Discrete Choice Models with Latent Choice Sets". *International Journal of Research in Marketing* 12(1): 9–24.

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.

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

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.

Downs, Anthony, 1957. *An Economic Theory of Democracy.* New York: Harber & Row.

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

Evans, Geoffrey and Anthony Heath, 1993. "A Tactical Error in the Analysis of Tactical Voting: A Response to Niemi, Whitten and Franklin". *British Journal of Political Science* 23(1): 131–137.

Franklin, Mark, Richard Niemi and Guy Whitten, 1994. "The Two Faces of Tactical Voting". *British Journal of Political Science* 24(4): 549–557.

Goldthorpe, John, 1999. "Modelling the Pattern of Class Voting in British Elections, 1964-1992". In Evans, Geoffrey (Ed.), "The End of Class Politics? Class Voting in Comparative Context", Oxford: Oxford University Press, pp. 59–82.

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., John Curtice, Roger Jowell, Geoffrey Evans, Julia Field and Sharon Witherspoon, 1991. *Understanding Political Change: The British Voter, 1964-1987.* Oxford: Pergamon Press.

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

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.

McLachlan, Geoffrey and Thriyambakam Krishnan, 2007. *The EM Algorithm and Extensions.* John Wiley & Sons.

Niemi, Richard G., Guy Written and Mark N. Franklin, 1992. "Constituency Characteristics, Individual Characteristics and Tactical Voting in the 1987 British General Election". *British Journal of Political Science* 22(02): 229–240.

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/`

Shikano, Susumu, Michael Herrmann and Paul W. Thurner, 2009. "Strategic Voting under Proportional Representation: Threshold Insurance in German Elections". *West European Politics* 32(3): 634–656.

Steenbergen, Marco R., Dominik Hangartner and Catherine E. de Vries, 2011. "Choice under Complexity: A Heuristic-Systematic Model of Electoral Behavior". Annual Meeting of the Midwest Political Science Association, March 31-April 3, Chicago, IL.

The Electoral Commission, 2014. "Election and Referendum Reports".
`http://www.electoralcommission.org.uk/our-work/publications/election-and-referendum-reports`