Exploring Voting Blocs within the Irish Electorate: A Mixture Modeling Approach *

Isobel Claire Gormley
University College Dublin, Ireland

Thomas Brendan Murphy
Trinity College Dublin, Ireland

April 10, 2007

*Isobel Claire Gormley is a lecturer, School of Mathematical Sciences (Statistics), University College Dublin, Dublin 4, Ireland. E-mail: claire.gormley@ucd.ie; Web: http://mathsci.ucd.ie/. Thomas Brendan Murphy is a lecturer, Department of Statistics, School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland. E-mail: murphybt@tcd.ie; Web: http://www.tcd.ie/Statistics/staff/brendanmurphy.shtml.

Isobel Claire Gormley was supported by a Government of Ireland Research Scholarship in Science, Engineering and Technology provided by the Irish Research Council for Science, Engineering and Technology, funded by the National Development Plan. A substantial part of this work was completed during a visit to the Center for Statistics in the Social Sciences, University of Washington which was supported by NIH grant (8 R01 EB002137-02).

We would like to thank Prof. Michael Marsh, Pat Carroll, and Joan Burton for supplying the various data sets used in this study. We would also like to thank Prof. Adrian Raftery, the members of the Center for Statistics and the Social Sciences and the members of the Working Group on Model-based Clustering at the University of Washington for suggestions that contributed enormously to this work. The detailed suggestions made by the editor, associate editor and anonymous referees have greatly improved the paper.

Abstract

Irish elections use a voting system called proportional representation by means of a single transferable vote (PR-STV). Under this system, a voter expresses their vote by ranking some (or all) of the candidates in order of preference. Which candidates are elected is determined through a series of counts where candidates are eliminated and surplus votes are distributed.

The electorate in any election forms a heterogeneous population, that is, voters with different political and ideological persuasions would be expected to have different preferences for the candidates. The purpose of this article is to establish the presence of voting blocs in the Irish electorate, to characterize these blocs and to estimate their size.

A mixture modeling approach is used to explore the heterogeneity of the Irish electorate and to establish the existence of clearly defined voting blocs. The voting blocs are characterized by their voting preferences which are described using a ranking data model. In addition, the care with which voters choose lower tier preferences is estimated in the model.

The methodology is used to explore data from two Irish elections. Data from eight opinion polls taken during the six weeks prior to the 1997 Irish presidential election are analyzed. These data reveal the evolution of the structure of the electorate during the election campaign. In addition, data that record the votes from the Dublin West constituency of the 2002 Irish general election are analyzed to reveal distinct voting blocs within the electorate; these blocs are characterized by party politics, candidate profile and political ideology.

Keywords: Elections, Ranking Data, Mixture Models

1 Introduction

In elections, members of the electorate exhibit different voting behaviors by choosing to vote for different candidates. The difference in voting behavior may be due to allegiance to a political party, choosing familiar candidates, choosing geographically local candidates or one of many other reasons. The different voting behaviors lead to a collection of votes from a heterogenous population.

While several high profile governments employ plurality voting, a large proportion of the world's democracies feature some form of proportional representation (PR) voting system (Quinn and Martin, 2002). In particular, the proportional representation by means of a single transferable vote (PR-STV) system is employed in elections in Ireland, Malta, Australia and New Zealand as well as in certain city elections in the USA and by many professional bodies. Regenwetter et al. (2006) highlight the exceptional richness of PR-STV data and declare a need for probabilistic models of the voting process. We concentrate on studying Irish elections where PR-STV is employed. A description of the Irish electoral system is given in Section 2.

The purpose of this article is to establish the presence of voting blocs in the Irish electorate, to characterize these blocs and to estimate their size. Tam (1995) studied Asian voting behavior within the American political arena via a multinomial logistic regression model and concluded that Asians should not be treated as a monolithic group. Holloway (1995) examined the differences between voting blocs when analyzing United Nations roll call data using a multidimensional scaling technique. We propose using mixture models to explore the heterogeneity in the Irish electorate and to establish the existence of clearly defined voting blocs. In addition, we determine the voting preferences that characterize these blocs. This greatly adds to the understanding of how the Irish electoral system (PR-STV) works in practice.

The mixture modeling approach assumes that the electorate consists of a collection of homogeneous sub-populations (voting blocs). Two different models are proposed for modeling the voting patterns within the sub-populations: the Plackett-Luce model (Section 3.1) and Benter's model (Section 3.2). The suitability of these models as modeling tools in the context of the Irish voting system is discussed in Section 3.3. These mixture models capture voting behaviors that we would expect to find, for example voters with strong political party allegiance. In addition, this modeling approach allows for the estimation of the care with which voters choose lower tier preferences; this is an issue of great interest to political scientists and electoral candidates alike. The proposed mixture models are fitted by maximum likelihood using the EM algorithm (Section 4.2).

We concentrate on two elections, the 1997 Irish presidential election and the 2002 Irish general election. These elections are quite different in character; the general election elects the government and party politics are believed to play an important role (Marsh, 2000), whereas party politics are believed to only play a minor role in the presidential election (van der Brug et al., 2000). Eight opinion polls were completed during the campaign of the 1997 presidential election. The data from these polls show the development of the candidate support over the campaign and they also allow us to study the development of voting blocs during the campaign.

We notice when analyzing the presidential election data, in Section 5.1, that the number of components (voting blocs) in the fitted mixtures tends to increase throughout the campaign. Towards the end of the campaign the characteristics of the mixture components stabilize, but the mixture weights continue to vary. This suggests that clearly defined voting blocs exist within the electorate, but that the proportion of voters in each group varies later in the campaign.

Analyzing the general election data allows us explore the role of political parties in Irish governmental elections. The decomposition of the electorate into voting blocs allows us to explore the heterogeneity in voting behavior and to then carefully study the support for various parties. In Section 5.2, we use mixture models to analyze data from the Dublin West constituency from the 2002 Irish general election. The mixture analysis shows that there is strong political party support in this election, because voters tend to give their high preferences to candidates from the same political party or to parties of a particular persuasion. Additionally, the analysis suggests voters rank lower tier preferences with reduced care when the electorate heterogeneity is accounted for.

We conclude, in Section 6, by discussing the models used in this study. We suggest extensions of the model and more general applications within political science and elsewhere.

2 Irish elections

Irish elections use a voting system called proportional representation by means of a single transferable vote (PR-STV). Under this system voters rank some (or all) of their favorite candidates in order of preference. The votes are totalled through a series of counts, where candidates are eliminated, their votes are distributed, and surplus votes are transferred between candidates. An in depth description of the electoral system, including the method of counting votes is given in Sinnott (1999) and good introductions to the Irish political system are given in Coakley and Gallagher (1999) and Sinnott (1995).

A brief overview of the vote counting process is given here. For illustration purposes, the transfer of votes in the Dublin West constituency, where there were three seats available for election, is shown in Table 1. Under the PR-STV electoral system a constituency specific 'quota' of votes is calculated which depends on the number of seats available and the number of valid votes cast. The quota for the Dublin West constituency was calculated to be 7498. Thus once any candidate at any counting stage obtained or exceeded 7498 votes this candidate was elected.

At the first stage of the counting process the number of first preference votes obtained by each candidate is totalled. Candidates Bonnie and Smyth got the lowest number of first preferences and, as neither would ever be able to exceed the quota of votes required, were eliminated from the race. Thus at the second stage of counting their 748 and 134 votes respectively were transferred to the candidates given the second place preference on those ballot forms. Seventy five of these votes were non-transferrable i.e. no second place preferences were expressed. Also, as candidate Lenihan got 8086 first preference votes he was the first candidate to be elected. His 588 votes in excess of the quota were transferred at the third stage of counting to those candidates given second place preferences on those ballot forms. The 588 of Lenihan's votes that were transferred were randomly selected from his 8086 first preference votes. At the fourth stage of the counting process, after the previous transfers, candidate McDonald was not be able to reach the quota and was thus eliminated from the race. Her 2524 votes were then transferred to the next most preferred remaining candidates detailed on each of the ballots — 487 of these were non-transferrable votes. Subsequent to the transfer of McDonald's votes, Higgins' 7853 votes exceeded the quota and thus he was elected. At the fifth stage of the counting process Morrissey was eliminated and his 2662 votes were transferred to those remaining candidates ranked next on the ballot forms — 359 of Morrissey's votes were non-transferrable. At the sixth and final stage of counting Doherty-Ryan had the least number of votes and as her elimination left only one candidate, Burton was elected. Thus Lenihan was elected outright on first preference votes, but Higgins and Burton were subsequently elected during the counting process.

While the PR-STV system has many proponents, it also has many opponents. Sinnott (1995) describes some of the potential problems with the PR-STV system in an Irish context. Other potential flaws are explained in Katz (1984) and Brams and Fishburn (1984).

It has been argued that the PR-STV voting system puts too little emphasis on the political parties and too much emphasis on the candidates (Katz, 1984; Blais, 1991) and thus can lead to fracticious governments; this potential problem is examined in Sinnott (1995) where it is concluded that this problem does not manifest itself to a great degree in Irish elections. Our analysis in Section 5.2 supports the idea that this problem is not very serious in the 2002 general election.

Table 1: The transfer of votes in the Dublin West constituency. The numbers marked in boldface indicate that the candidate was elected. Three seats were available. The - symbol indicates that a candidate has been eliminated from the election. The quota required for guaranteed election for this constituency is 7498.

Candidate	**Party**	**Count**					
		1	2	3	4	5	6
Bonnie, R.	Green Party	748	—	—	—	—	—
	(GP)		-748				
Burton, J.	Labour	3810	4020	4079	4375	5125	**6300**
	(Lab)		+210	+59	+296	+750	+1175
Doherty-Ryan, D.	Fianna Fáil	2300	2386	2698	3056	3728	—
	(FF)		+86	+312	+358	+672	-3728
Higgins, J.	Socialist Party	6442	6660	6731	**7853**	**7853**	**7853**
	(SP)		+218	+71	+1122		
Lenihan, B.	Fianna Fáil	**8086**	**8086**	**7498**	**7498**	**7498**	**7498**
	(FF)			-588			
McDonald, M.	Sinn Féin	2404	2498	2524	—	—	—
	(SF)		+94	+26	-2524		
Morrissey, T.	Progressive Democrats	2370	2480	2554	2662	—	—
	(PD)		+110	+74	+108	-2662	
Smyth, J.	Christian Socialist Party	134	—	—	—	—	—
	(CSP)		-134				
Terry, S.	Fine Gael	3694	3783	3829	3982	4863	5669
	(FG)		+89	+46	+153	+881	+806
Non-transferable			75	75	562	921	2668
			+75	+0	+487	+359	+1747

2.1 The 1997 presidential election

The eighth (and current) President of Ireland, Mary McAleese, was originally elected in 1997. The presidential election had five candidates: Mary Banotti, Mary McAleese, Derek Nally, Adi Roche, and Rosemary Scallon.

Some of these candidates were endorsed by political parties and some were independent candidates. Specifically the candidates and their endorsing parties were: Mary Banotti (Fine Gael), Mary McAleese (Fianna Fáil and Progressive Democrats), Derek Nally (Independent), Adi Roche (Labour), and Rosemary Scallon (Independent). Mary Banotti, Derek Nally and Adi Roche were considered to be liberal candidates where Mary McAleese and Rosemary Scallon were deemed the more conservative candidates.

Seven opinion polls and an exit poll, taken on polling day, were completed during the election campaign. Four of the opinion polls were conducted by Irish Marketing Surveys (IMS) during the two months prior to the election. Approximately 1100 respondents, drawn from 100 sampling areas, were interviewed for each poll. Interviews took place at randomly located homes with individuals selected according to a socioeconomic quota. A range of sociological questions were asked of each respondent as was the respondent's voting preference, if any, for each of the candidates. These preferences were in effect utilized as each respondent's vote. Everyone included in the poll data expressed at least one preference — in fact each poll has slightly over the required 1100 respondents.

The other three opinion polls were conducted by the Market Research Bureau of Ireland (MRBI), again during the two month electoral campaign. A similar sampling methodology as used in the IMS polls was employed — 100 Primary Sampling Units (PSU's) were selected from census data, and from each PSU 10 interviews were conducted using a random route procedure. The sample was quota controlled by age, gender, and socioeconomic class. In each of the three MRBI polls there was what could be considered as missing data — an average of 150 respondents in each poll either replied don't know, won't vote or refused to give their preferences. While such missing data is important, examining such voters and their covariates is a further research area. In this article only those who expressed at least one preference in the opinion polls are examined.

On the day of the presidential election, October 30th 1997, Lansdowne Market Research conducted an exit poll where 2498 voters were interviewed at 150 polling stations in all 41 Irish constituencies.

As stated, the aim of this article is to explore heterogeneity in the Irish electorate and to determine what voting characteristics define the partition of the electorate into voting blocs with respect to the 1997 presidential election. As party politics are meant to play a lesser role in presidential elections, our interest lies in determining if the partition of the electorate is along party lines or if voters are more candidate orientated. Further, as the polls detailed here were taken chronologically, a focus of the analysis of these data sets is to trace the development of the heterogeneity of the electorate. As the polls were conducted using similar methodology we deem the comparison of respondents of different polls to be justified.

A detailed description of the entire presidential election campaign, including the nomination and selection of candidates, is given by Marsh (1999). The sources of the poll data are given in Appendix A.

2.2 The 2002 general election

Ireland had its most recent general election on May 17th, 2002. One hundred and sixty six politicians were elected to be members of the Dáil (Irish government) from forty two constituencies. This election saw the introduction of electronic voting, for the first time, in three constituencies (Dublin North, Dublin West, and Meath). The remaining thirty nine constituencies had paper ballots.

In Dublin West there were three seats to be filled with nine candidates running for election. The nine candidates represented eight political parties, with Fianna Fáil having two candidates. The electorate was 53780 and there were a total of 29988 valid votes cast. The actual votes from this constituency are analyzed in this work. These data were previously analyzed using exploratory data analysis techniques by Laver (2004).

The voting data from the Dublin North, Dublin West and Meath constituencies are publicly available and the sources are given in Appendix A.

Again the substantive issue of interest is the size and characteristic voting patterns of the voting blocs within the Irish electorate. For example, two candidates in the Dublin West constituency represent the same political party — does this have an influence on the make up of the partition of voters?

3 Models for ranking data

Suppose that N candidates are running in an election and that the electorate consists of M voters. Each vote consists of a ranking $\underline{x}_i = (c(i,1), c(i,2), \ldots, c(i,n_i))$ where $c(i,j)$ records the candidate that was ranked in the jth position by voter i. Marden (1995) refers to this as the order representation of a ranking. Let $n_i \leq N$ be the number of preferences expressed by voter i, where N is the maximum number of preferences. Let $\mathbf{x} = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_M)$ denote the data set that contains all of the votes. For notational convenience in later sections, we define $c(i, n_i + 1), c(i, n_i + 2), \ldots, c(i, N)$ to be any ordering of the candidates not selected by voter i; the choice of ordering does not affect the results.

Several models are implemented in the literature to model ranking data. The Bradley-Terry model (Bradley and Terry, 1952) examines competition between a set of individuals as a set of pairwise comparisons from which an 'ability parameter' can be inferred and thus a ranking of the competitors can be formed. We take a different approach where all competitors are simultaneously compared with each other, rather than in a pairwise manner. Bradlow and Fader (2001) model the simultaneous movement of multiple items up and down a ranking over time within a Bayesian framework with an exploding multinomial-logit likelihood. Johnson et al. (2002) take a Bayesian latent variable approach to modeling ranking data from multiple evaluators who may use different ranking criteria. They include parameters in their hierarchical model to accommodate ties within the rankings. Graves et al. (2003) model car racing results by using a combination of the Bradley-Terry model with the Luce model and Stern's model to form their 'attrition model' which estimates driver ability. A step-wise approach is taken where the probability of a driver finishing in last place is examined and from this the final permutation of drivers is built.

We look at two models for ranking data: the Plackett-Luce model (Plackett, 1975) and Benter's model (Benter, 1994). Both models were originally developed within the context of analyzing the permutation of horses at the end of a race. Each horse has an associated parameter, interpreted as the probability of the horse winning the race, and the final permutation of horses is a combination of these parameters. A PR-STV ballot form can be thought of in a similar manner to that of the final permutation of horses; for example, once a candidate has been chosen they cannot be selected again and not all candidates must be ranked by a voter. Thus we employ both the Plackett-Luce and Benter models to model a PR-STV ranking as a sequential process where the next most preferred candidate is selected; Fligner and Verducci (1988) refer to models of this form as *multi stage models*. The manner in which both models are applied in the current context are detailed in the following sections.

3.1 Plackett-Luce model

The Plackett-Luce model has a single "support" parameter $\underline{p} = (p_1, p_2, \ldots, p_N)$, where $\sum_{j=1}^{N} p_j = 1$ and under this model the probability of a partial ranking $\underline{x}_i$ is

$$\mathbf{P}\{\underline{x}_i|\underline{p}\} = \prod_{t=1}^{n_i} \frac{p_{c(i,t)}}{\sum_{s=t}^{N} p_{c(i,s)}}. \tag{1}$$

An interpretation of this model is that the parameter value p_j records the probability of candidate j being given a first preference. The probability of a candidate being given a lower preference is proportional to the support parameter value for that candidate. At each preference level, the parameter values are normalized to produce valid probability values. Thus at each preference level a voter's selection is modeled as choosing the candidate most likely to be chosen first, conditional on which candidates, if any, have already been ranked.

3.2 Benter's model

Benter (1994) proposed a variant of the Plackett-Luce model for modeling horse races. The Plackett-Luce model suffers from the property that the probability of a horse with a low winning probability placing highly in a race is too small.

Benter's model has two parameters $\underline{p} = (p_1, p_2, \ldots, p_N)$ (the support parameter) where $\sum_{j=1}^{N} p_j = 1$ and $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ (the dampening parameter). Under Benter's model, the probability of a partial ranking $\underline{x}_i$ is

$$\mathbf{P}\{\underline{x}_i | \underline{p}\} = \prod_{t=1}^{n_i} \frac{p_{c(i,t)}^{\alpha_t}}{\sum_{s=t}^{N} p_{c(i,s)}^{\alpha_t}}. \tag{2}$$

It is reasonable to assume that $0 \leq \alpha_t \leq 1$, which makes lower preference choices at least as random as higher preference ones (see Proposition 1, Appendix B). In any case, $\alpha_1 \equiv 1$ and $\alpha_N \equiv 0$ for all models; this avoids over parameterization of the model.

Benter's model has a similar interpretation to the Plackett-Luce model, but when lower place probabilities are being computed the probabilities are dampened down to model the way in which lower preferences may be chosen less carefully than higher preferences by a voter. Under Benter's model the log odds of selecting candidate j over candidate l at choice level t is $\alpha_t \log(p_j/p_l)$. Thus the tth level dampening parameter α_t can be interpreted as how the log odds of selecting candidate j over candidate l is affected by the selection being made at choice level t. Since α_1 is constrained to be 1 for identifiability reasons, at the first choice level the log odds is unaffected. An α_2 value of 0.8, for example, indicates that the log odds is 'dampened' by a fifth to model the manner in which the second selection was made with less certainty than the first. Thus the model has greater entropy than the Plackett-Luce model which is itself a special case of Benter's model with $\underline{\alpha} = \underline{1} \equiv (1, 1, \ldots, 1)$. The estimation of dampening parameters is of interest as the care with which voters express their lower preferences is an attribute of the voting process in which political scientists and electoral candidates have an interest.

3.3 Model validity and interpretation

One concern associated with voting (and choice) models is the issue of independence of irrelevant alternatives (IIA) (see Train, 2003, for example). The Plackett-Luce and Benter models exhibit IIA within choice levels as the ratio of the probability of choosing one alternative over another is independent of the other available alternatives. In the Plackett-Luce model, the ratio of the probabilities is independent of choice level whereas in Benter's model the ratio varies with choice level due to the dampening parameter in the model. While it can be argued that IIA is an unsatisfactory property in some situations, in this application the models appear to give a realistic representation of the choice process. A detailed description of the relationship between IIA and ranking models is given in Marden (1995, Section 5.13.1).

Plackett-Luce models and Benter models were fitted, using maximum likelihood, to each of the eight opinion polls from the 1997 presidential election campaign and to the 2002 general election data from the Dublin West constituency. The estimated model parameters and their associated standard errors are reported in Figure 1, Table 2 and Table 3. The approximate standard errors reported throughout the article are derived within the EM algorithm as proposed by McLachlan and Peel (2000, Section 2.15) and McLachlan and Krishnan (1997, Chapter 4). The covariance matrix of the estimated model parameters is approximated by the empirical observed information matrix which is computed using the score function of the complete data log-likelihood. Computation of second order partial derivatives (the difficulty in obtaining which necessitated the use of the EM algorithm originally) is therefore avoided.

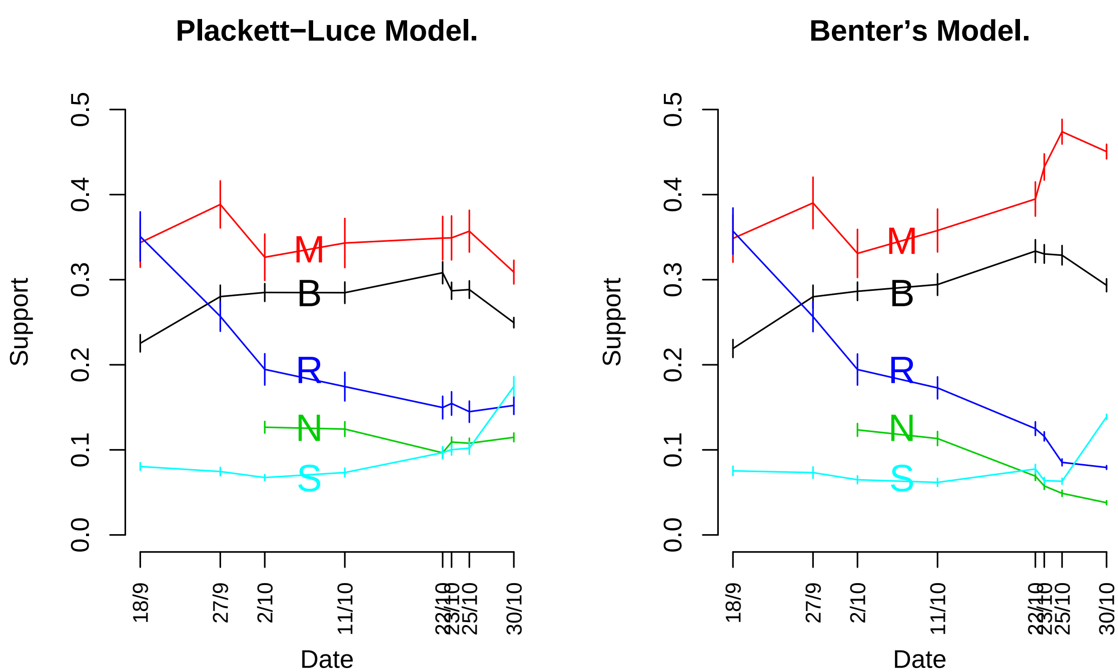

Figure 1: A graphical representation of the maximum likelihood estimates of the Plackett-Luce support parameter and the Benter support parameter for each of the eight polls from the 1997 presidential election campaign. Note that Nally was not a candidate when the first two polls were taken. Two standard errors either side of each estimate are also illustrated.

Figure 1 demonstrates the support parameter associated with each presidential

Table 2: The values of the Benter dampening parameter for each of the polls from the 1997 presidential election campaign. The fourth value of the dampening parameter is not computed for the first two polls as there were only four candidates when the poll was taken. Standard errors of the estimates are given in parentheses.

Date	**Poll**	**Dampening Parameter**			
18/9	IMS	1.00	0.80 (0.07)	1.00 (0.08)	—
27/9	MRBI	1.00	0.94 (0.00)	1.00 (0.09)	—
2/10	IMS	1.00	1.00 (0.08)	1.00 (0.08)	1.00 (0.09)
11/10	MRBI	1.00	1.00 (0.08)	1.00 (0.08)	1.00 (0.10)
22/10	MRBI	1.00	1.00 (0.05)	0.95 (0.07)	1.00 (0.10)
23/10	IMS	1.00	0.98 (0.04)	0.80 (0.05)	0.99 (0.07)
25/10	IMS	1.00	0.92 (0.04)	0.73 (0.05)	0.63 (0.07)
30/10	Lansdowne	1.00	0.73 (0.03)	0.18 (0.03)	0.00 (0.04)

Table 3: Maximum likelihood estimates of the Plackett-Luce and Benter support parameters for the 2002 Dublin West constituency election data; the proportion of the total first preference votes for each candidate is included for comparison purposes. The Benter dampening parameter estimate is $\underline{\hat{\alpha}} = (1.00, 0.92, 0.66, 0.44, 0.89, 0.94, 0.94, 0.00, 0.00)$ with the associated standard errors less than 1×10^{-4}. The largest standard error of the support parameters under either model was 2×10^{-6}.

Candidate	**Abbrev.**	**Party**	**First Preference**	**Plackett-Luce Estimate**	**Benter Estimate**
Bonnie, R.	Bon	GP	0.03	0.07	0.06
Burton, J.	Bur	Lab	0.13	0.16	0.16
Doherty-Ryan, D.	Do-Ry	FF	0.08	0.11	0.11
Higgins, J.	Hig	SP	0.22	0.16	0.17
Lenihan, B.	Len	FF	0.27	0.18	0.20
McDonald, M.	McD	SF	0.08	0.06	0.06
Morrissey, T.	Mor	PD	0.08	0.12	0.11
Smyth, J.	Smy	CSP	0.00	0.02	0.01
Terry, S.	Ter	FG	0.12	0.12	0.12

candidate under both the Plackett-Luce and Benter models. A general popularity ordering of McAleese, Banotti, Scallon, Roche and then Nally emerges under both models. The most striking feature of the plots is perhaps the rapid decline in support for Adi Roche. Roche began as favorite for the presidential seat but, after criticism from co-workers in a Chernobyl project about her style of work and claims that she was an unsuitable person to be president, her campaign never recovered. McAleese and Banotti maintained first and second place across the polls but Rosemary Scallon's position also improved. Early criticisms of the conservative candidate fizzled out as the campaign developed and as her professional presentation skills became more evident she finished in a respectable third place.

The model parameter estimates differ in the final polls where the Plackett-Luce estimates seem to shrink together but the Benter estimates become more dispersed. This can, in part, be explained by the fact that in the 30/10 poll people were encouraged to express all preferences. Interestingly, the support parameter estimates under Benter's model in the 30/10 poll (the exit poll) are very similar to the first preference proportions for each candidate. The dampening parameters for this poll (Table 2) complement these estimates as it is clear that lower place preferences are strongly dampened in this poll, thus giving a lot of priority to higher preferences.

The dampening parameters associated with the exit poll data give a good demonstration of the value of estimating such parameters — here they give an illustration of how many preference levels have an effect on estimating the support parameters of the model. The third level dampening parameter of 0.18 suggests that the third place preferences are only made with around one fifth of the certainty that the first place preferences are. Also $\alpha_4 = 0$ suggests that voters select the remaining candidates with equal probability.

Similar types of effects are apparent when examining the estimated support parameters for the voting data in the Dublin West constituency (Table 3). Again the Plackett-Luce parameters seem to shrink towards the mean — lower support parameters are pulled up (e.g. Bonnie from 0.03 to 0.07) and larger support parameters are pulled down (e.g. Higgins from 0.22 to 0.16). The shrinkage of the Benter estimates towards the first preference proportions is less extreme but again the dampening parameter values go some way in explaining this.

The dampening parameters for the Dublin West data suggest that lower preferences should be taken into account when modeling such data: $\alpha_6 = \alpha_7 = 0.94$ shows that the sixth and seventh level votes are nearly as influential as first place preferences.

One exception to the pattern of shrinkage of the support parameters is McDonald — she got 8% of the first preference votes and estimated support parameters of 0.06 under both the Plackett-Luce and Benter models. McDonald was a Sinn Féin candidate in the Dublin West election. Sinn Féin have a well defined body of support in Ireland in that they are "the only party committed to achieving a democratic socialist republic and the end of British rule in Ireland". Thus voters would tend to rank Sinn Féin candidates first or else not at all thus explaining why McDonald's support parameters are close to the first preference proportions. This in turn suggests there is a group of such voters present in the Irish electorate among potentially many others.

While the above analyses provide evidence to suggest the types of ranking models introduced are both applicable and necessary, the grouping structure within the electorate is not exposed. The next section offers a method for exploring the electorate using mixtures of the Plackett-Luce and Benter models. This approach provides an easily interpretable model for the heterogeneity in the electorate.

4 Mixture models

Suppose that the population of voters consists of K sub-populations (voting blocs) where voters belong to sub-population k with probability π_k and given that a voter is in sub-population k their vote follows an $f_k(\underline{x}_i)$ density. Then the probability of each vote is

$$\mathbf{P}\{\underline{x}_i\} = \sum_{k=1}^{K} \pi_k f_k(\underline{x}_i)$$

which is a finite mixture model. We assume that $f_k(\underline{x}_i) = f(\underline{x}_i|\underline{\theta}_k)$ where $\{f(\underline{x}_i|\underline{\theta}_k) : \underline{\theta}_k \in \Theta\}$ is a parametric family of densities (Plackett-Luce with $\underline{\theta}_k = \underline{p}_k$ or Benter with $\underline{\theta}_k = (\underline{p}_k, \underline{\alpha})$).

Many applications of mixtures have been reported in the statistics literature and these are extensively reviewed in Titterington et al. (1985), McLachlan and Basford (1988), and McLachlan and Peel (2000). The use of mixture models for analyzing ranking data is discussed in Marden (1995) (and references therein), Stern (1993) and Murphy and Martin (2003).

4.1 Mixture constraints

The proposed mixture models allow us to constrain the parameters in the different components; this offers modeling flexibility.

The Plackett-Luce model is a special case of the Benter model with dampening parameter $\underline{\alpha} = \underline{1} = (1, 1, \ldots, 1)$. Therefore, we investigate the option of constraining the $\underline{\alpha}$ value to be identically $\underline{1}$ or leaving it unconstrained.

We also have the option of forcing one component in the mixture to be a uniform component; that is a component with $\underline{p}_k = (1/N, 1/N, \ldots, 1/N)$. This uniform component can "soak up" any outlying data values and allows for better modeling of the remaining data. This use of a noise component is analogous to including a

Poisson noise term in model-based clustering (Fraley and Raftery, 1998, 2002). The usefulness of including a noise component in mixture models for ranking data has been demonstrated in Murphy and Martin (2003) and Gormley and Murphy (2006). Additionally, D'Elia and Piccolo (2005) use a uniform component in a mixture model for ranking data but in a different context.

Thus four different types of model were fitted to the data:

1. a mixture of Plackett-Luce models,
2. a mixture of Plackett-Luce models constrained such that one component is fixed to be a noise component,
3. a mixture of Benter models and
4. a mixture of Benter models constrained such that one component is fixed to be noise component.

In both of the Plackett-Luce mixture models each α_t is, by definition, constrained to be 1 whereas in both the Benter mixture models $\underline{\alpha}$ is to be estimated.

4.2 Fitting mixture models

The mixture models were fitted using maximum likelihood methods; that is, the likelihood

$$L(\underline{\pi}, \mathbf{p}, \underline{\alpha}|\mathbf{x}) = f(\mathbf{x}|\underline{\pi}, \mathbf{p}, \underline{\alpha}) = \prod_{i=1}^{M} \left[\sum_{k=1}^{K} \pi_k f(\underline{x}_i|\underline{p}_k, \underline{\alpha}) \right], \tag{3}$$

is maximized with respect to $\mathbf{p} = (\underline{p}_1, \underline{p}_2, \cdots, \underline{p}_K)$, $\underline{\pi} = (\pi_1, \pi_2, \cdots, \pi_K)$ and $\underline{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_N)$, in the case where $\underline{\alpha}$ is not constrained to be $(1, 1, \cdots, 1)$.

The fitting of mixtures using maximum likelihood can be implemented using the EM algorithm (Dempster et al., 1977). To use the EM algorithm, a membership label is introduced for each voter such that $z_{ik} = 1$ if voter i belongs to component k and $z_{ik} = 0$ otherwise. The likelihood of the observed data and the unobserved labels is called the complete-data likelihood,

$$L_C(\underline{\pi}, \mathbf{p}, \underline{\alpha}|\mathbf{x}, \mathbf{z}) = f(\mathbf{x}, \mathbf{z}|\underline{\pi}, \mathbf{p}, \underline{\alpha}) = \prod_{i=1}^{M} \prod_{k=1}^{K} \left[\pi_k f(\underline{x}_i|\underline{p}_k, \underline{\alpha}) \right]^{z_{ik}}. \tag{4}$$

The EM Algorithm (Algorithm 1, Appendix C.1) involves an E-step that replaces the missing data $\mathbf{z}$ with its expected value given the current parameter estimates and an M-step that maximizes the complete-data log-likelihood (4) computed with the estimates of $\mathbf{z}$.

The EM algorithm for fitting Benter mixtures is straightforward in principle, but the M-step is difficult in practice. The ECM algorithm (Meng and Rubin, 1993) proved to be useful when fitting mixtures of Benter models. This algorithm replaces maximization in the M-step with a series of easier conditional maximization steps.

In this case, the conditional maximizations were with respect to $\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_K$ and $\underline{\alpha}$.

The conditional maximizations in the M-step are implemented using the MM algorithm (Algorithm 2, Appendix C.3) (Lange et al., 2000; Hunter and Lange, 2004). This algorithm works by constructing a function that minorizes the objective function and then maximizing the minorizing function. This process is iterated leading to a sequence of parameter estimates giving increasing objective function values.

Convergence of the EM algorithm was assessed using the Aitken acceleration estimate of the final maximized likelihood. The algorithm is considered to be converged when the current likelihood value is within a tolerance of the Aitken estimate (Böhning et al., 1994; Lindsay, 1995).

4.3 Model comparison

Many different mixture models were fitted to the election data sets by varying the component models and the number of components. We require a criterion for comparing the fitted models. In any case, the chosen model should coincide with the political theory on voting behavior in Irish elections.

The Bayesian Information Criterion (BIC) is widely used to compare models. The BIC is defined to be

$$BIC = 2(\text{maximized likelihood}) - 2(\text{number of parameters})\log(M).$$

The BIC can be viewed as a criterion which rewards model fit, but penalizes model complexity. Table 4 details the parameters to be estimated within each type of model considered.

Table 4: The number of parameters in the various types of mixtures proposed for modeling the Irish election data.

Model	**Proportions**	**Support**	**Dampening**
Plackett-Luce	$K-1$	$K(N-1)$	0
Plackett-Luce (with Noise)	$K-1$	$(K-1)(N-1)$	0
Benter	$K-1$	$K(N-1)$	$N-2$
Benter (with Noise)	$K-1$	$(K-1)(N-1)$	$N-2$

The usual justification for the use of BIC is that, for regular problems, it provides an approximation to the Bayes factor for comparing models under certain prior assumptions (Kass and Raftery, 1995). Finite mixture models do not satisfy the regularity conditions for this approximation to be valid, but there is much in the literature to support its use in a mixture modeling context. Leroux (1992) showed that the number of components in the mixture, as estimated by the BIC, is at least as large as the true number of components, for large sample sizes. Keribin (1998, 2000) proved that the BIC is a consistent indicator, almost surely, of the number of components due to its appropriate penalizing term. In addition, the literature details many successful applications of the use of BIC as a model selection tool within the context of mixture models (see for example Fraley and Raftery (1998) and Dasgupta and Raftery (1998)).

Cross-validated likelihood was also proposed as a tool for determining the appropriate number of components in a mixture by Smyth (2000). Models are judged on their performance in out-of-sample prediction, as estimated in a cross validation manner. It is proposed as a practical alternative to the more Bayesian BIC approach but is more computationally expensive.

The BIC was used as the main model comparison tool in this analysis. In practice, the cross-validated likelihood method invariably suggested the maximum number of groups fitted as the best model. This was deemed to be a case of over-fitting. The BIC consistently returned the most parsimonious and interpretable models; the model selected using BIC is compared to the political theory of Irish voting behavior in Section 5.

5 Analysis

The proposed mixture model approach for exploring heterogeneity within the Irish electorate is demonstrated on Irish presidential and general election data. The analysis of the electorates of these elections using this approach establishes that there are homogeneous sub-populations of voters in the electorate and the form of these voting blocs is revealed. Laver and Marsh (1999) discuss the factors that influence voting behavior in Irish elections, in particular they highlight the effect of social class, the importance of candidate personality and the influence of party policy. As a result of these influences we expect to establish the existence of multiple voting blocs in the Irish electorate.

5.1 The 1997 presidential election

Mixtures of Plackett-Luce models and mixtures of Benter models, with up to 10 components, were fitted to the 1997 presidential election data sets. The BIC was used as the model selection criterion. For all polls (with the exception of two) the Plackett-Luce model with varying numbers of components was selected. In some of these polls a mixture of Plackett-Luce models which included a noise component was deemed the best model. For the two polls on 23/10 and 30/10 mixtures of Benter models were selected but the difference in BIC between the Benter and Plackett-

Luce mixtures was small. Thus mixtures of Plackett-Luce models are reported for all polls (Table 5) for ease of comparison.

Table 5: Parameter estimates when mixtures of Plackett-Luce models were fitted to each of the eight presidential election poll data sets are reported. Standard errors associated with these estimates are given in parentheses.

Date	**Banotti**	**McAleese**	**Nally**	**Roche**	**Scallon**	$\hat{\pi}_k$
18/9	0.23 (0.005)	0.34 (0.014)	-	0.35 (0.014)	0.08 (0.002)	*1.00*
27/9	0.28 (0.007)	0.39 (0.014)	-	0.26 (0.009)	0.07 (0.002)	*1.00*
2/10	0.32 (0.010)	0.42 (0.029)	0.07 (0.004)	0.16 (0.011)	0.02 (0.002)	*0.60* (0.090)
	0.20	0.20	0.20	0.20	0.20	*0.40* (0.036)
11/10	0.56 (0.059)	0.09 (0.010)	0.11 (0.011)	0.20 (0.027)	0.05 (0.005)	*0.27* (0.062)
	0.20 (0.004)	0.50 (0.010)	0.10 (0.004)	0.14 (0.006)	0.07 (0.003)	*0.73* (0.073)
22/10	0.75 (0.123)	0.03 (0.007)	0.10 (0.011)	0.09 (0.019)	0.03 (0.005)	*0.14* (0.059)
	0.28 (0.003)	0.53 (0.034)	0.05 (0.034)	0.10 (0.006)	0.04 (0.004)	*0.55* (0.102)
	0.18 (0.007)	0.32 (0.015)	0.10 (0.009)	0.16 (0.016)	0.24 (0.037)	*0.31* (0.137)
23/10	0.92 (0.001)	0.02 (0.012)	0.03 (< 0.001)	0.02 (0.002)	0.02 (0.005)	*0.16* (0.047)
	0.02 (0.025)	0.92 (0.002)	0.01 (0.003)	0.02 (0.002)	0.04 (0.002)	*0.20* (0.047)
	0.33 (0.002)	0.47 (0.008)	0.05 (0.004)	0.12 (0.005)	0.03 (0.003)	*0.44* (0.089)
	0.20	0.20	0.20	0.20	0.20	*0.20* (0.030)
25/10	0.96 (0.164)	0.01 (< 0.001)	0.01 (< 0.001)	0.02 (0.002)	0.01 (0.001)	*0.16* (0.042)
	0.00 (< 0.001)	1.00 (0.010)	0.00 (< 0.001)	0.00 (< 0.001)	0.00 (< 0.001)	*0.14* (0.037)
	0.25 (< 0.001)	0.61 (< 0.001)	0.04 (< 0.001)	0.07 (< 0.001)	0.03 (0.002)	*0.49* (0.057)
	0.20	0.20	0.20	0.20	0.20	*0.21* (0.026)
30/10	0.81 (0.054)	0.01 (0.001)	0.06 (0.001)	0.07 (0.004)	0.05 (0.003)	*0.23* (0.028)
	0.01 (0.001)	0.83 (0.004)	0.02 (< 0.001)	0.03 (0.001)	0.11 (0.003)	*0.27* (0.028)
	0.25 (0.011)	0.58 (0.011)	0.04 (0.017)	0.08 (0.025)	0.04 (0.042)	*0.36* (0.036)
	0.12 (0.017)	0.09 (0.041)	0.15 (0.002)	0.17 (0.004)	0.47 (0.003)	*0.14* (0.078)

Examination of Table 5 shows that the Irish electorate began the 1997 presidential campaign as a single voting bloc which then partitioned over the course of the campaign.

At the beginning of the campaign, as demonstrated by the first two polls, the electorate appeared to be composed of a single component which had larger levels of support for the three most high profile candidates — Banotti, McAleese and Roche. However Roche's support dropped by almost 10% between the polls taken on 18/9 and 27/9. As mentioned in Section 3.3, shortly after the initial nominations of candidates Adi Roche, who up until then had been the bookies favorite, was publicly criticized by fellow workers and her popularity dropped off significantly. This drop in support for Roche continued throughout all the polls detailed.

A month before polling day, demonstrated by the 2/10 poll, 40% of the electorate were best modeled as noise. The electorate appears to have become partitioned into a noise group and the original group who supported the high profile candidates of Banotti, McAleese and Roche. Perhaps Roche's drop off in popularity left some undecided voters.

By 11/10, the future pattern of the presidential race became clear. The Banotti and McAleese camps emerged strongly with the voting bloc weighted towards McAleese making up the larger 73% of the electorate. Notably, the group who favored McAleese also appear to have a good level of support for Banotti.

Between the polls conducted on 11/10 and the 22/10 a great deal of controversy arose in the presidential campaign. It was reported that Mary McAleese had sympathies with the republican party Sinn Féin which would have had a detrimental effect on her support. Further fuel was added to these allegations when the president of the Sinn Féin party gave McAleese the party's backing. Throughout this period McAleese consistently denied the claims and after defending her position well in nationally broadcast current affairs program on October 20th she re-established herself. In fact, the false allegations had a larger detrimental effect on her presidential competitors, some of whom had publicly castigated her about the allegations.

These events are mirrored by the results of the polls taken on 22/10 and 23/10. On 22/10 the electorate is composed of three voting blocs. Again the strongly Banotti group was present, the strongly McAleese group (with some Banotti support) was the largest group making up 55% of the electorate and 31% of the electorate seemed to be nearly a noise component with a conservative flavor. The larger support in this third group was for the two conservative candidates McAleese and Scallon. Scallon's performance in the campaign was beginning to win her votes.

The results of the 23/10 poll indicate how well McAleese recovered and gained from the Sinn Féin controversy. The electorate really partitions at this stage into a group of Banotti supporters, a group of McAleese supporters, a group of voters who support the high profile candidates McAleese, Banotti and Roche and one fifth of the electorate are still modeled as noise. The results of the poll taken on 25/10 are very similar — the main theme of the four components remains the same, with the probability of belonging to each group altering slightly. The group with support for the candidates with the higher profiles (and supported by the larger parties) makes up almost half of the electorate.

The changes in the composition of the electorate between 25/10 and polling day pay tribute to Scallon's performance throughout her campaign — again the

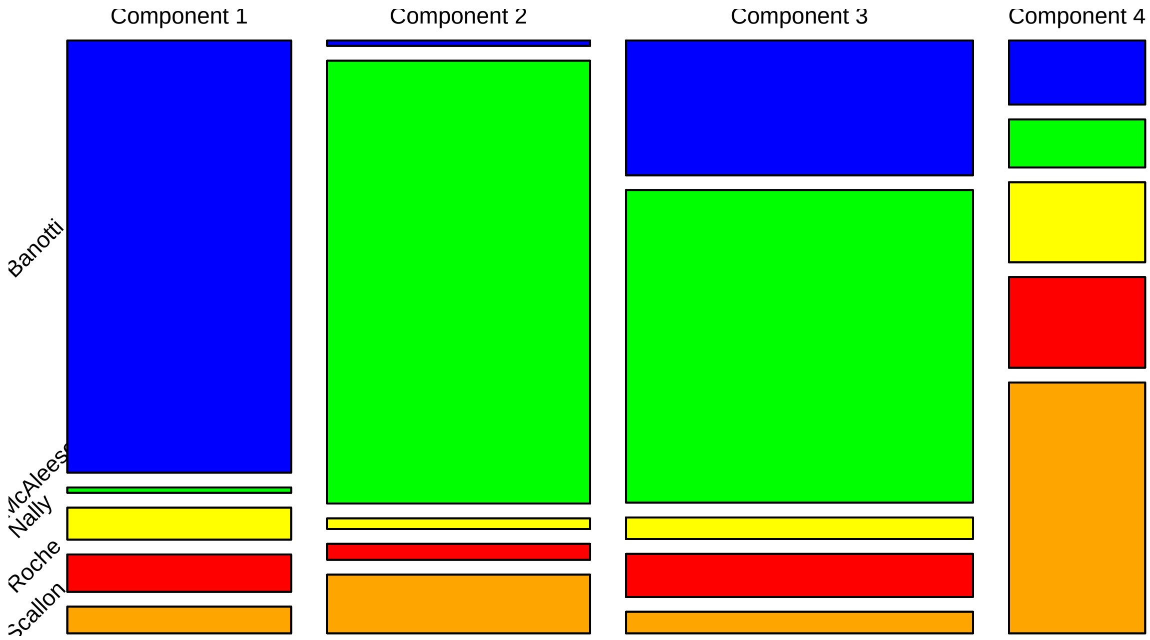

Figure 2: A mosaic plot representation of the mixture fitted to the Lansdowne exit poll data (conducted on 30/10) for the 1997 Presidential Election. The column widths represent the mixture proportions and the columns are divided into sections according to the support parameter within the component.

themes of each of the four voting blocks are similar but the estimated support parameters for each candidate drop in nearly every group, with the exception of Scallon. Her support parameters in each of the four groups are significantly higher than they were in the 25/10 poll. Figure 2 provides a graphical representation of the estimated models parameters of the exit poll.

In summary, the mixture model finds groups of voters which appear logical in the context of this presidential election; the voting blocs appear to be dominated by candidate personality and ideology. One possible explanation for the predominant choice of the Plackett-Luce model over Benter's model is that there were only five candidates in the election. Thus the electorate was very familiar with all of the candidates and lower preferences may have been made with as much certainty as higher preferences.

5.2 The 2002 general election

Mixtures of Plackett-Luce and Benter models were fitted to the data from the Dublin West constituency. The mixture with the highest BIC value was a fifteen component Benter mixture and is reported in Table 6 and Figures 3 and 4.

The mixture model reveals some interesting features in the electorate. The voting blocs can be summarized as follows:

Table 6: Fifteen component mixture of Benter models fitted to the Dublin West constituency data. The support parameter estimates all have standard errors less than 5×10^{-3} with the exception of two — Lenihan's support parameter in component 6 has a standard error of 8×10^{-3} and McDonald's support parameter in component 10 has an associated standard error of 1×10^{-2}. The final row of the table gives the mixture component probabilities whose standard errors were all less than 8×10^{-3}. Benter dampening parameter estimates were $\underline{\hat{\alpha}} = (1.00, 1.00, 0.95, 0.74, 0.57, 0.41, 0.28, 0.15, 0.00)$ with the associated standard errors all less than 1×10^{-2}.

Candidate	**Party**	**Components**														
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Bonnie	GP	0.01	0.02	0.00	0.03	0.01	0.01	0.13	0.04	0.13	0.06	0.03	0.01	0.00	0.04	0.01
Burton	Lab	0.01	0.24	0.00	0.17	0.13	0.09	0.17	0.25	0.22	0.39	0.01	0.03	0.05	0.04	0.01
Doherty-Ryan	FF	0.23	0.03	0.19	0.01	0.11	0.08	0.09	0.00	0.00	0.00	0.00	0.08	0.11	0.07	0.07
Higgins	SP	0.02	0.03	0.00	0.62	0.05	0.37	0.11	0.05	0.40	0.36	0.52	0.54	0.00	0.03	0.00
Lenihan	FF	0.70	0.11	0.80	0.03	0.45	0.36	0.13	0.00	0.00	0.02	0.00	0.20	0.55	0.07	0.22
McDonald	SF	0.02	0.00	0.00	0.07	0.05	0.00	0.06	0.01	0.09	0.00	0.42	0.13	0.00	0.72	0.00
Morrissey	PD	0.02	0.21	0.01	0.02	0.08	0.05	0.15	0.06	0.04	0.05	0.00	0.01	0.20	0.02	0.68
Smyth	CSP	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.02	0.00	0.00	0.00	0.00	0.01	0.00
Terry	FG	0.00	0.37	0.00	0.05	0.11	0.04	0.14	0.59	0.09	0.11	0.00	0.00	0.09	0.01	0.01
Probability		0.10	0.09	0.09	0.09	0.08	0.08	0.07	0.07	0.06	0.06	0.06	0.05	0.05	0.03	0.02

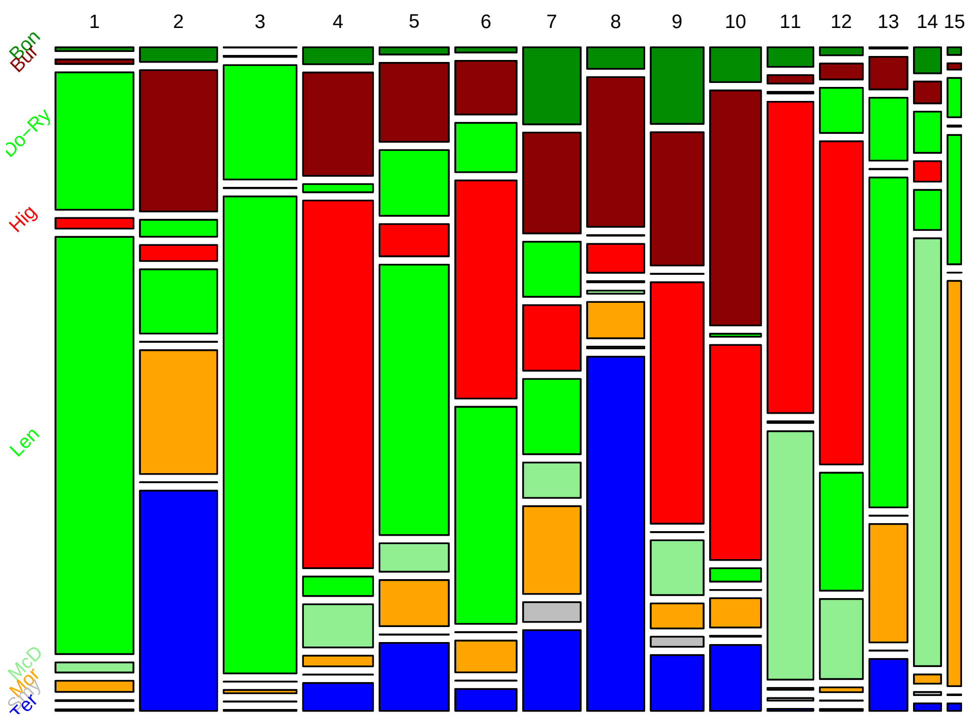

Figure 3: A mosaic plot representation of the 15 group Benter mixture model fitted to the Dublin West data. The width of each column illustrates the mixing proportion for each group and the sections within each column represent the support parameter for each candidate within that group.

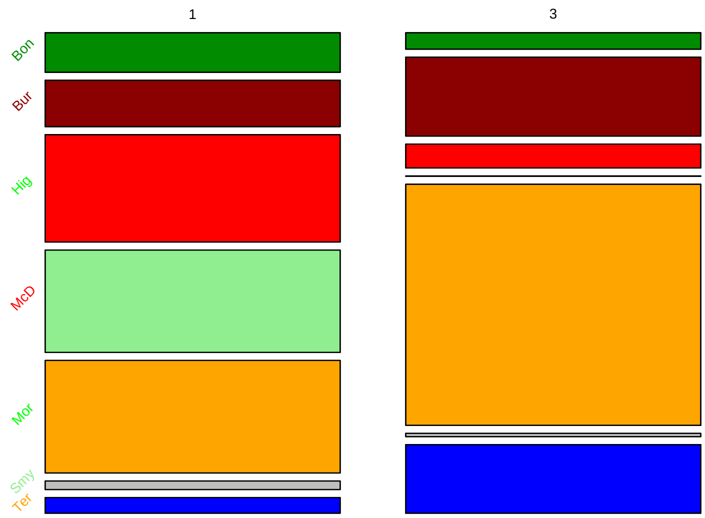

Figure 4: A mosaic plot representing the parameter estimates for groups 1 and 3 of the Benter mixture model fitted to the Dublin West data. The two Fianna Fáil candidates (Doherty-Ryan and Lenihan) have been removed to determine the subtle differences between the groups.

1. This component gives almost all of its support to Fianna Fáil. Lenihan gets more support than his running mate, Doherty-Ryan. This component is similar to Component 3, but there are subtle differences (see Figure 4 and Component 3).

2. The support is divided between Fine Gael, Labour and the Progressive Democrats. These are the three parties that are next largest after Fianna Fáil in terms of the number of seats held in government.

3. Almost all of the support is for Fianna Fáil. Again, Lenihan gets more support than his running mate, Doherty Ryan. This component differs from Component 1, in that the candidate support conditional on having ranked the Fianna Fáil candidates Lenihan and Doherty-Ryan in first and second place (in either order) is strongly for the Progressive Democrats candidate Morrissey (Figure 4). Interestingly, the Progressive Democrats were in coalition government with Fianna Fáil prior to the election. Thus this component would appear to be the voters in the electorate who were in favor of a Fianna Fáil/Progressive Democrats coalition government.

4. The support is mainly for Joe Higgins of the Socialist party, but most of the remaining support is divided between the Labour and Sinn Féin candidates; historically, these would have been seen as left wing parties.

5. This component gives a lot of support to Lenihan, but divides its support quite evenly between Burton, Doherty-Ryan, Terry and Morrissey after that. This component appears to be predominantly candidate centered on Lenihan.

6. The support here is primarily for Higgins and Lenihan. These candidates are from very different parties, but both candidates have a very high profile within the constituency. There is reason to believe that this component may be geographically based.
7. This component shows almost uniform support for most of the major candidates in this constituency. Smyth and McDonald receive considerably less support than the other candidates.
8. The support is divided between the Fine Gael and Labour candidates. These two parties encourage voters to transfer their lower preferences between these parties. These parties are former coalition government parties (1994–1997).
9. Higgins and Burton get most of the support. Higgins and Burton are high profile candidates in the constituency. They are both from left-wing parties. The support for Bonnie, McDonald and Terry could be explained on similar party or idealistic grounds. This component has extremely low support for Fianna Fáil.
10. Burton and Higgins get most of the support with Terry receiving a moderate amount of support. Terry's party is closely linked to Burton's party (see Component 8).
11. This component shows support for the Socialist and Sinn Féin candidates. These are the two most left-wing candidates in the constituency.
12. Higgins, Lenihan and McDonald have the majority of the support. The candidates are from parties that are quite different. These candidates are all high profile.
13. This component has strong support for the candidates from the two government coalition parties (Fianna Fáil and Progressive Democrats).
14. McDonald of Sinn Féin receives the majority of the support in this component.
15. The support mainly goes to the Progressive Democrats candidate. The remaining support is for the two Fianna Fáil candidates. All most all of the support is for the previous government coalition parties.

The fifteen component mixture of Benter's model that was selected using BIC gives clear and meaningful voting blocs. The voting blocs confirm the idea that Irish elections are influenced by party politics, candidate profile and political ideology (Bowler and Farrell, 1991; Laver and Marsh, 1999; Marsh, 2000). In particular, party politics is a defining characteristic in many of the voting blocs. The mixture model found in this analysis provides strong support for the political theory of how Irish elections work.

The estimated dampening parameter $\underline{\hat{\alpha}}$ in the Benter mixture is also of interest (Table 6). The parameter estimate shows that the first two preferences are very carefully chosen ($\alpha_1 = \alpha_2 = 1$) and that later preferences become more random (higher entropy) as α_t decreases with t. The parameter estimates also suggest that the choice of candidate at the last two choice levels is essentially uniform (maximum entropy). This is interesting, because one could postulate that the high and low preferences are made very carefully and that the middle preferences are very random. However, the fitted estimate indicates that choices get more random as a ballot is completed. Laver (2004) noted that the median (and modal) number of preferences expressed by voters was three in this constituency. His findings may also support the idea that voters give a few top preferences carefully and after that they either don't select candidates or they select them in a less careful manner.

6 Conclusions

The proposed mixtures of Plackett-Luce and Benter models provide an interpretable model for PR-STV election data. The models can be used to discover and model any heterogeneity present in the voting behavior of the electorate while also modeling the care with which voters choose their preferences. Clearly defined and interpretable voting blocs within the Irish electorate are highlighted. The model fitting by maximum likelihood using the EM and MM algorithms provides an efficient method for fitting these models.

The use of a noise component in the mixture models was found to be advantageous. The component accounted for small groups of voters who didn't fit into the main groups in the mixture. As a result, extra mixture components were not required to model these small groups of voters. These results are in agreement with previous uses of noise components in mixtures.

The scope of the models proposed for PR-STV data in this article lies far beyond modeling Irish election data alone. While the larger global powers still favor a plurality electoral system most of the rest of the world employ some form of PR voting. In particular, an increasing number of European Union member states use a PR system (Regenwetter et al., 2006, Chapter 6.2). Moreover, PR is often involved in intricately balanced and therefore interesting political arenas such as that currently in place in Northern Ireland. Attempts have recently been made to establish an assembly with devolved powers from Britain through a PR-STV voting system — the methodology proposed here could be used to provide some insight into the make up of the polarized voting blocs in Northern Ireland.

Under the 2008 Democratic Party selection rules for the US Presidential election, delegates will be selected under a proportional representation system; delegates for the Democratic Party national convention are awarded to candidates in proportion to the proportion of the caucus vote the candidate receives. Variations on the models in this article could be applied to a time-series of polled voter preferences to explore the evolution of preferred presidential candidates and to examine the presence and structure of voting blocs.

Mixture models are also applicable in the study other electoral systems. In the approval voting system which is used for the election of the UN secretary general, voters select a subset of candidates and give an equal vote to each of them. This type of voting could be modeled using a mixture of independent Bernoulli models, that is a latent class model (Bartholomew and Knott, 1999). Under the cumulative voting system which is employed in some corporate governance elections, voters have N votes to distribute amongst the candidates; candidates can receive more than one vote from a voter. Data from these elections can be modeled using a mixture of multinomial distributions. Hence, the use of mixture models for analyzing voting data is more widely applicable than the situation considered here.

In addition, the mixture modeling methodology is applicable in the analysis of other choice data. Gormley and Murphy (2006) analyze Irish college application choices using a mixture of Plackett-Luce models and establish the existence of homogeneous groups of applicants. In particular, the proposed methodology could also be applied to the analysis of customer choice data in marketing applications, where customers express preferences for products.

The mixture of experts model offers an extension of the mixture model for situations where covariate information is available. In the case of opinion polls, where covariates are available, including this information could provide insight to the underlying factors which influence the form of the voting blocs in the electorate. The application of a mixture of experts model to the analysis of Irish presidential election data is given in Gormley (2006).

A Data sources

The various 1997 Irish presidential election opinion poll data sets were collected by the three companies: Lansdowne Market Research, Irish Marketing Surveys (IMS), and Market Research Bureau of Ireland (MRBI).

These data sets are available through the Irish Elections Data Archive (`http://www.tcd.ie/Political_Science/elections/elections.html`) and the Irish Opinion Poll Archive (`http://www.tcd.ie/Political_Science/cgi/`) which are maintained by Prof. Michael Marsh in the Department of Political Science, Trinity College Dublin, Ireland.

The voting data from the Dublin West constituency is available from the constituency returning officer's web page (`http://www.dublincountyreturningofficer.com`).

B Dampening increases entropy

Proposition 1 (Dampening Entropy) *Let $\underline{p} = (p_1, p_2, \ldots, p_N)$ be such that $p_j > 0$ for all j and $\sum_{j=1}^N p_j = 1$ and let $0 \le \alpha \le 1$. Let*

$$q(\alpha) = (q_1, q_2, \ldots, q_N) = \left(\frac{p_1^\alpha}{\sum_{j=1}^N p_j^\alpha}, \frac{p_2^\alpha}{\sum_{j=1}^N p_j^\alpha}, \ldots, \frac{p_N^\alpha}{\sum_{j=1}^N p_j^\alpha} \right).$$

Then, the Entropy$[q(\alpha)] = \mathbf{E}[q(\alpha)] = -\sum_{j=1}^N q_j(\alpha) \log q_j(\alpha)$ *is a decreasing function of α.*

Proof: We have:

$$\begin{aligned}
\frac{\partial \mathbf{E}}{\partial \alpha} &= -\sum_{j=1}^N q_j' [1 + \log q_j] \\
&= -\sum_{j=1}^N \left[\frac{p_j^\alpha}{\sum_{l=1}^N p_l^\alpha} \left\{ \log p_j - \frac{\sum_{l=1}^N p_l^\alpha \log p_l}{\sum_{l=1}^N p_l^\alpha} \right\} \right] \left[1 + \alpha \log p_j - \log \sum_{l=1}^N p_l^\alpha \right] \\
&= -\sum_{j=1}^N \Bigg[q_j \log p_j + \alpha q_j \{\log p_j\}^2 - q_j \log p_j \log \sum_{l=1}^N p_l^\alpha \\
&\qquad - q_j \sum_{l=1}^N q_l \log p_l - \alpha q_j \log p_j \sum_{l=1}^N q_l \log p_l + q_j \log \sum_{l=1}^N p_l^\alpha \sum_{l=1}^N q_l \log p_l \Bigg] \\
&= -\alpha \left[\sum_{j=1}^N q_j \{\log p_j\}^2 - \left\{ \sum_{j=1}^N q_j \log p_j \right\}^2 \right] \\
&\le 0
\end{aligned}$$

by the Cauchy-Schwarz inequality.

C Algorithms for fitting mixtures of Benter models

C.1 EM algorithm

Algorithm 1 (EM-Algorithm for Mixtures) *When fitting mixtures the EM algorithm reduces to the following steps:*

0. Let $h = 0$ and choose initial parameter estimates $\mathbf{p}^{(0)}$, $\underline{\alpha}^{(0)}$ and $\underline{\pi}^{(0)}$.

1. **E-Step:** *Compute the quantities*

$$z_{ik}^{(h+1)} = \frac{\pi_k^{(h)} f(\underline{x}_i | \underline{p}_k^{(h)}, \underline{\alpha}^{(h)})}{\sum_{k'=1}^K \pi_{k'}^{(h)} f(\underline{x}_i | \underline{p}_{k'}^{(h)}, \underline{\alpha}^{(h)})}$$

2. **M-Step:** *Let* $\pi_k^{(h+1)} = \sum_{i=1}^{M} z_{ik}^{(h+1)}/M$ *and maximize*

$$\sum_{k=1}^{K}\sum_{i=1}^{M} z_{ik}^{(h+1)} \log f(\underline{x}_i|\underline{p}_k, \underline{\alpha})$$

with respect to $\mathbf{p} = (\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_K)$ *and* $\underline{\alpha}$ *(if required).*

Call the maximizing values $\underline{\pi}^{(h+1)}$, $\mathbf{p}^{(h+1)}$ *and* $\underline{\alpha}^{(h+1)}$.

3. *If converged, then stop. Otherwise, increment* h *and return to Step 1.*

C.2 MM algorithm

The application of the M-step of the EM algorithm to Benter's model is difficult in practice. Below we detail the steps involved in the simplification of this problem via the MM algorithm where the creation of a surrogate function allows maximization of the objective function to be transferred. Lange et al. (2000) and Hunter and Lange (2004) provide similar methodology applied to the Plackett-Luce model.

To construct such a surrogate function the supporting hyperplane property of convex functions is exploited. Consider the convex function $f(x)$ with differential $df(x)$. Then

$$f(x) \geq f(y) + df(y)(x-y) \qquad x, y \geq 0 \tag{5}$$

provides a linear minorizing function that can be utilized as a surrogate function to which optimization can be transferred. By iteratively maximizing the surrogate function the objective function (in this case, the complete-data log-likelihood) is simultaneously driven uphill and maximum likelihood parameter estimates can be derived.

Consider the complete-data log-likelihood which we wish to maximize:

$$Q = \sum_{i=1}^{M}\sum_{k=1}^{K} \hat{z}_{ik}\{\log \pi_k + \sum_{t=1}^{n_i} \alpha_t \log p_{kc(i,t)} - \sum_{t=1}^{n_i} \log \sum_{s=t}^{N} p_{kc(i,s)}^{\alpha_t}\}. \tag{6}$$

Deriving the maximum likelihood estimates $\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_K$ and $\underline{\alpha}$ from this is not straightforward. The following sections demonstrate how parameter estimates were obtained.

C.2.1 Maximization with respect to $\underline{p}_k$.

In this case we treat α_t as a fixed constant $\bar{\alpha}_t$. In practice $\bar{\alpha}_t$ is the value of α_t at the previous iteration of the MM algorithm. By (5), the strict convexity of the $-\log(x)$ function implies that

$$-\log(x) \quad \geq \quad -\log(y) + 1 - \frac{x}{y}.$$

Thus,

$$-\log \sum_{s=t}^{N} p_{kc(i,s)}^{\bar{\alpha}_t} \;\; \geq \;\; -\log \sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\bar{\alpha}_t} \; +1 \; - \frac{\sum_{s=t}^{N} p_{kc(i,s)}^{\bar{\alpha}_t}}{\sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\bar{\alpha}_t}}.$$

where $\bar{p}_{kj}$ is a constant and in practice is the estimate of p_{kj} from the previous iteration of the MM algorithm.

It follows from (6) that, up to a constant,

$$Q(p_{kj}) \;\; \geq \;\; q \;\; = \;\; \sum_{i=1}^{M} \sum_{k=1}^{K} \hat{z}_{ik} \left\{ \sum_{t=1}^{n_i} \bar{\alpha}_t \log p_{kc(i,t)} - \sum_{t=1}^{n_i} \left(\frac{\sum_{s=t}^{N} p_{kc(i,s)}^{\bar{\alpha}_t}}{\sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\bar{\alpha}_t}} \right) \right\}$$

Again this modified Q function poses maximization problems with respect to p_{kj}. By the supporting hyperplane property of convex functions (5) the function $f(p) = -p^{\bar{\alpha}}$ becomes

$$-p^{\bar{\alpha}} \;\; \geq \;\; -\bar{p}^{\bar{\alpha}} \;\; - \;\; \bar{\alpha}\bar{p}^{\bar{\alpha}-1}(p - \bar{p})$$

which provides the surrogate function

$$q \geq \sum_{i=1}^{M} \sum_{k=1}^{K} \sum_{t=1}^{n_i} \hat{z}_{ik} \bar{\alpha}_t \log p_{kc(i,t)} - \sum_{i=1}^{M} \sum_{k=1}^{K} \sum_{t=1}^{n_i} \hat{z}_{ik} \left\{ \sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\bar{\alpha}_t} \right\}^{-1} \left\{ \sum_{s=t}^{N} \bar{\alpha}_t \bar{p}_{kc(i,s)}^{\bar{\alpha}_t - 1} p_{kc(i,s)} \right\}$$

up to a constant. By iterative maximization of this surrogate function we produce a sequence of p_{kj} values which converge to the maximum of Q with respect to p_{kj}. Differentiation of q with respect to p_{kj} gives

$$\frac{\partial q}{\partial p_{kj}} \;\; = \;\; \sum_{i=1}^{M} \sum_{t=1}^{n_i} \hat{z}_{ik} \left\{ \frac{\bar{\alpha}_t}{p_{kc(i,t)}} \mathbf{1}_{\{j=c(i,t)\}} \;\; - \;\; \left(\sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\bar{\alpha}_t} \right)^{-1} \left(\sum_{s=t}^{N} \bar{\alpha}_t \bar{p}_{kc(i,s)}^{\bar{\alpha}_t - 1} \mathbf{1}_{\{j=c(i,s)\}} \right) \right\} \quad (7)$$

where $\mathbf{1}_{\{j=c(i,s)\}}$ is an indicator function such that

$$\mathbf{1}_{\{j=c(i,s)\}} \;\; = \;\; \begin{cases} 1 & \text{if } j = c(i,s) \\ 0 & \text{otherwise.} \end{cases}$$

We denote

$$\omega_{kj} \;\; = \;\; \sum_{i=1}^{M} \sum_{t=1}^{n_i} \hat{z}_{ik} \bar{\alpha}_t \mathbf{1}_{\{j=c(i,t)\}}$$

and

$$\delta_{ijs} \;\; = \;\; \begin{cases} 1 & \text{if } j = c(i,s) \text{ and } 1 \leq s \leq n_i \\ 0 & \text{if } j \neq c(i,s) \text{ and } 1 \leq s \leq N \\ 1 & \text{if } s = N+1 \text{ and } j \neq c(i,l) \quad \forall l \text{ such that } 1 \leq l \leq N. \end{cases}$$

Therefore equating (7) to zero gives

$$\frac{\omega_{kj}}{p_{kj}} = \sum_{i=1}^{M}\sum_{t=1}^{n_i} \hat{z}_{ik} \left\{ \sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\bar{\alpha}_t} \right\}^{-1} \left\{ \sum_{s=t}^{(N+1)} \bar{\alpha}_t \bar{p}_{kj}^{\bar{\alpha}_t - 1} \delta_{ijs} \right\}$$

which implies that

$$\hat{p}_{kj} = \frac{\omega_{kj}}{\sum_{i=1}^{M}\sum_{t=1}^{n_i} \hat{z}_{ik} \left\{ \sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\bar{\alpha}_t} \right\}^{-1} \left\{ \sum_{s=t}^{(N+1)} \bar{\alpha}_t \bar{p}_{kj}^{\bar{\alpha}_t - 1} \delta_{ijs} \right\}}.$$

C.2.2 Maximization With Respect To $\underline{\alpha}$.

We return to the original complete data log likelihood function (6) and treat the problematic term as a function of α_t. p_{kj} is treated as a constant with $\bar{p}_{kj}$ denoting the estimate of p_{kj} from the previous iteration of the MM algorithm. Thus by (5)

$$-\log \sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\alpha_t} \geq -\log \sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\bar{\alpha}_t} + 1 - \frac{\sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\alpha_t}}{\sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\bar{\alpha}_t}}.$$

It therefore follows, up to a constant,

$$Q \geq q = \sum_{i=1}^{M}\sum_{k=1}^{K} \hat{z}_{ik} \left\{ \sum_{t=1}^{n_i} \alpha_t \log \bar{p}_{kc(i,t)} + \sum_{t=1}^{n_i} \left(\frac{-\sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\alpha_t}}{\sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\bar{\alpha}_t}} \right) \right\}.$$

Similar to the maximization with respect to p_{kj}, this surrogate function is still difficult to optimize. Also, the function $f(\alpha) = -\bar{p}^{\alpha}$ is a concave function and minorizing by use of a linear surrogate function is not possible. Bounding a convex function $f(x)$ around y in a using a quadratic gives

$$f(x) \leq f(y) + f'(y)(x-y) + \frac{1}{2}(x-y)^T B(x-y)$$

where $B - H(y) > 0$ and $H(y)$ is the Hessian $d^2 f(y)/dy^2$. Thus

$$\begin{aligned} \bar{p}^{\alpha} &\leq \bar{p}^{\bar{\alpha}} + (\log \bar{p})\bar{p}^{\bar{\alpha}}(\alpha - \bar{\alpha}) + 1/2(\alpha - \bar{\alpha})^2 (\log \bar{p})^2 \\ \Rightarrow -\bar{p}^{\alpha} &\geq -\bar{p}^{\bar{\alpha}} - (\log \bar{p})\bar{p}^{\bar{\alpha}}(\alpha - \bar{\alpha}) - 1/2(\alpha - \bar{\alpha})^2 (\log \bar{p})^2 \end{aligned}$$

because $(\log \bar{p})^2 > H(\bar{\alpha})$. Hence the surrogate function becomes

$$q \geq \sum_{i=1}^{M}\sum_{k=1}^{K} \hat{z}_{ik} \left[\sum_{t=1}^{n_i} \alpha_t \log \bar{p}_{kc(i,t)} + \sum_{t=1}^{n_i} \left(\sum_{s=t}^{N} \bar{p}_{kc(i,s)}^{\bar{\alpha}_t} \right)^{-1} \left\{ \sum_{s=t}^{N} \left(-\log \bar{p}_{kc(i,s)} \bar{p}_{kc(i,s)}^{\bar{\alpha}_t} (\alpha_t - \bar{\alpha}_t) \right. \right. \right.$$
$$\left. \left. \left. -1/2(\alpha_t - \bar{\alpha}_t)^2 (\log \bar{p}_{kc(i,s)})^2 \right) \right\} \right]$$

up to a constant. Iterative maximization of this surrogate function with respect to α_t leads to a sequence of $\hat{\alpha}_t$ values that converge to the maximum of Q. Thus

$$\frac{\partial q}{\partial \alpha_t} = \sum_{i=1}^{M}\left\{\sum_{k=1}^{K}\hat{z}_{ik}\left[\log\bar{p}_{kc(i,t)} + \left(\sum_{s=t}^{N}\bar{p}_{kc(i,s)}^{\bar{\alpha}_t}\right)^{-1}\left\{\sum_{s=t}^{N}\Big(-\log\bar{p}_{kc(i,s)}\bar{p}_{kc(i,s)}^{\bar{\alpha}_t} - (\alpha_t - \bar{\alpha}_t)(\log\bar{p}_{kc(i,s)})^2\Big)\right\}\right]\right\}.\mathbf{1}_{\{t\leq n_i\}}$$

which implies that

$$\hat{\alpha}_t = \frac{\sum_{i=1}^{M}\left\{\sum_{k=1}^{K}\hat{z}_{ik}\left[\log\bar{p}_{kc(i,t)} + \left(\sum_{s=t}^{N}\bar{p}_{kc(i,s)}^{\bar{\alpha}_t}\right)^{-1}\left\{\sum_{s=t}^{N} -\log\bar{p}_{kc(i,s)}\bar{p}_{kc(i,s)}^{\bar{\alpha}_t} + \bar{\alpha}_t(\log\bar{p}_{kc(i,s)})^2\right\}\right]\right\}.\mathbf{1}_{\{t\leq n_i\}}}{\sum_{i=1}^{M}\left\{\sum_{k=1}^{K}\hat{z}_{ik}\left(\sum_{s=t}^{N}\bar{p}_{kc(i,s)}^{\bar{\alpha}_t}\right)^{-1}\sum_{s=t}^{N}(\log\bar{p}_{kc(i,s)})^2\right\}.\mathbf{1}_{\{t\leq n_i\}}}.$$

C.3 EM/MM algorithm

Algorithm 2 (EM-Algorithm for Mixtures via the MM-Algorithm) *When fitting mixtures of Benter's model using the EM algorithm (incorporating the MM algorithm) reduces to the following steps:*

0. *Let $h = 0$ and choose initial parameter estimates $\mathbf{p}^{(0)}$, $\underline{\alpha}^{(0)}$ and $\underline{\pi}^{(0)}$.*
1. **E-Step:** *Compute the quantities*

$$z_{ik}^{(h+1)} = \frac{\pi_k^{(h)} f(\underline{x}_i|\underline{p}_k^{(h)}, \underline{\alpha}^{(h)})}{\sum_{k'=1}^{K}\pi_{k'}^{(h)} f(\underline{x}_i|\underline{p}_{k'}^{(h)}, \underline{\alpha}^{(h)})}$$

2. **M-Step:** *Compute*

$$\pi_k^{(h+1)} = \frac{\sum_{i=1}^{M} z_{ik}^{(h+1)}}{M}$$

$$p_{kj}^{(h+1)} = \frac{\omega_{kj}}{\sum_{i=1}^{M}\sum_{t=1}^{n_i} z_{ik}^{(h+1)}\left\{\sum_{s=t}^{N}\bar{p}_{kc(i,s)}^{\bar{\alpha}_t}\right\}^{-1}\left\{\sum_{s=t}^{(N+1)}\bar{\alpha}_t\bar{p}_{kj}^{\bar{\alpha}_t-1}\delta_{ijs}\right\}}$$

(where $\bar{p}_{kj}$, $\bar{\alpha}_t$, ω_{kj} and δ_{ijs} are defined in Section C.2.1)

$$\alpha_t^{(h+1)} =$$

$$\frac{\sum_{i=1}^{M}\left\{\sum_{k=1}^{K} z_{ik}^{(h+1)}\left[\log \bar{p}_{kc(i,t)}+\left(\sum_{s=t}^{N}\bar{p}_{kc(i,s)}^{\bar{\alpha}_t}\right)^{-1}\left\{\sum_{s=t}^{N}-\log \bar{p}_{kc(i,s)}\bar{p}_{kc(i,s)}^{\bar{\alpha}_t}+\bar{\alpha}_t(\log \bar{p}_{kc(i,s)})^2\right\}\right]\right\}.\mathbf{1}_{\{t\leq n_i\}}}{\sum_{i=1}^{M}\left\{\sum_{k=1}^{K} z_{ik}^{(h+1)}\left(\sum_{s=t}^{N}\bar{p}_{kc(i,s)}^{\bar{\alpha}_t}\right)^{-1}\sum_{s=t}^{N}(\log \bar{p}_{kc(i,s)})^2\right\}.\mathbf{1}_{\{t\leq n_i\}}}.$$

3. If converged, then stop. Otherwise, increment h and return to Step 1.

References

Bartholomew, D. J. and Knott, M. (1999), *Latent variable models and factor analysis*, London: Edward Arnold, 2nd ed.

Benter, W. (1994), "Computer-based Horse Race Handicapping and Wagering Systems: A Report," in *Efficiency of Racetrack Betting Markets*, eds. Ziemba, W. T., Lo, V. S., and Haush, D. B., San Diego and London: Academic Press, pp. 183–198.

Blais, A. (1991), "The debate over electoral systems," *International Political Science Review*, 12, 239–260.

Böhning, D., Dietz, E., Schaub, R., Schlattmann, P., and Lindsay, B. (1994), "The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family," *Annals of the Institute of Statistical Mathematics*, 46, 373–388.

Bowler, S. and Farrell, D. M. (1991), "Voter Behavior Under STV-PR: Solving the Puzzle of the Irish Party System," *Political Behavior*, 13, 303–320.

Bradley, R. A. and Terry, M. E. (1952), "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons," *Biometrika*, 39, 324–345.

Bradlow, E. T. and Fader, P. S. (2001), "A Bayesian Lifetime Model for the "Hot 100" *Billboard* Songs," *J. Amer. Statist. Assoc.*, 96, 368–381.

Brams, S. J. and Fishburn, P. C. (1984), "Some Logical Defects of the Single Tranferable Vote," in *Choosing an Electoral System: Issues and Alternatives*, eds. Lijphart, A. and Grofman, B., New York: Praeger, pp. 147–151.

Coakley, J. and Gallagher, M. (1999), *Politics in the Republic of Ireland*, London: Routledge in association with PSAI Press, 3rd ed.

Dasgupta, A. and Raftery, A. (1998), "Detecting features in spatial point processes with clutter via model-based clustering." *J. Amer. Statist. Assoc.*, 93, 294—302.

D'Elia, A. and Piccolo, D. (2005), "A mixture model for preferences data analysis," *Computational Statistics and Data Analysis*, 49, 917–934.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. Ser. B*, 39, 1–38, with discussion.

Fligner, M. A. and Verducci, J. S. (1988), "Multistage Ranking Models," *J. Amer. Statist. Assoc.*, 83, 892–901.

Fraley, C. and Raftery, A. E. (1998), "How many clusters? Which clustering method? - Answers via Model-Based Cluster Analysis," *Computer Journal*, 41, 578–588.

— (2002), "Model-Based Clustering, Discriminant Analysis, and Density Estimation," *J. Amer. Statist. Assoc.*, 97, 611–612.

Gormley, I. C. (2006), "Statistical Models for Rank Data," Ph.D. thesis, University of Dublin, Trinity College.

Gormley, I. C. and Murphy, T. B. (2006), "Analysis of Irish third-level college applications data," *J. Roy. Statist. Soc. Ser. A*, 169, 361—379.

Graves, T., Reese, C. S., and Fitzgerald, M. (2003), "Hierarchical Models for Permutations: Analysis of Auto Racing Results," *J. Amer. Statist. Assoc.*, 98, 282–3291.

Holloway, S. (1995), "Forty Years of United Nations General Assembly Voting," *Canadian Journal of Political Science*, 17, 223–249.

Hunter, D. R. and Lange, K. (2004), "A tutorial on MM algorithms," *Amer. Statist.*, 58, 30–37.

Johnson, V. E., Deaner, R. O., and van Schaik, C. P. (2002), "Bayesian Analysis of Rank Data With Application to Primate Intelligence Experiments," *J. Amer. Statist. Assoc.*, 97, 8–17.

Kass, R. E. and Raftery, A. E. (1995), "Bayes factors," *Journal of the American Statistical Association*, 90, 773–795.

Katz, R. S. (1984), "The single transferable vote and proportional representation," in *Choosing an Electoral System: Issues and Alternatives*, eds. Lijphart, A. and Grofman, B., New York: Praeger, pp. 135–145.

Keribin, C. (1998), "Estimation consistante de l'ordre de modèles de mélange," *C. R. Acad. Sci. Paris Sér. I Math.*, 326, 243–248.

— (2000), "Consistent estimation of the order of mixture models," *Sankhyā Ser. A*, 62, 49–66.

Lange, K., Hunter, D. R., and Yang, I. (2000), "Optimization transfer using surrogate objective functions," *J. Comput. Graph. Statist.*, 9, 1–59, with discussion, and a rejoinder by Hunter and Lange.

Laver, M. (2004), "Analysing structure of party preference in electronic voting data," *Party Politics*, 10, 521–541.

Laver, M. and Marsh, M. (1999), "Parties and Votes," in *Politics in the Republic of Ireland*, eds. Coakley, J. and Gallagher, M., London: Routledge & PSAI Press, 3rd ed., pp. 152–176.

Leroux, B. G. (1992), "Consistent estimation of a mixing distribution," *Ann. Statist.*, 20, 1350–1360.

Lindsay, B. (1995), *Mixture Models: Theory, Geometry and Applications*, Hayward, CA: Institute of Mathematical Statistics.

Marden, J. I. (1995), *Analyzing and modeling rank data*, London: Chapman & Hall.

Marsh, M. (1999), "The Making of the Eight President," in *How Ireland Voted 1997*, eds. Marsh, M. and Mitchell, P., Boulder, CO: Westview and PSAI Press, pp. 215–242.

— (2000), "Candidate Centered but Party Wrapped: Campaigning in Ireland under STV," in *Elections in Australia, Ireland, and Malta under the Single Transferable Vote*, eds. Bowler, S. and Grofman, B., Ann Arbor, MI: The University of Michigan Press, pp. 114–130.

McLachlan, G. J. and Basford, K. E. (1988), *Mixture models: Inference and applications to clustering*, New York: Marcel Dekker Inc.

McLachlan, G. J. and Krishnan, T. (1997), *The EM algorithm and extensions*, New York: John Wiley & Sons Inc.

McLachlan, G. J. and Peel, D. (2000), *Finite Mixture models*, New York: John Wiley & Sons.

Meng, X.-L. and Rubin, D. B. (1993), "Maximum likelihood estimation via the ECM algorithm: a general framework," *Biometrika*, 80, 267–278.

Murphy, T. B. and Martin, D. (2003), "Mixtures of Distance-Based Models for Ranking Data," *Computational Statistics and Data Analysis*, 41, 645–655.

Plackett, R. L. (1975), "The analysis of permutations," *Applied Statistics*, 24, 193–202.

Quinn, K. M. and Martin, A. D. (2002), "An Integrated Computational Model of Multiparty Electoral Competition," *Statistical Science: Voting and Elections*, 17, 405–419.

Regenwetter, M., Grofman, B., Marley, A. A. J., and Tsetlin, I. M. (2006), *Behavioral Social Choice. Probabilistic Models, Statistical Inference and Applications.*, New York: Cambridge University Press.

Sinnott, R. (1995), *Irish voters decide: Voting behaviour in elections and referendums since 1918*, Manchester: Manchester University Press.

— (1999), "The Electoral System," in *Politics in the Republic of Ireland*, eds. Coakley, J. and Gallagher, M., London: Routledge & PSAI Press, 3rd ed., pp. 99–126.

Smyth, P. (2000), "Model selection for probabilisitc clustering using cross-validated likelihood," *Statistics and Computing*, 9, 63—72.

Stern, H. S. (1993), "Probability Models on Rankings and the Electoral Process," in *Probability Models and Statistical Analyses For Ranking Data*, eds. Fligner, M. A. and Verducci, J. S., New York: Springer-Verlag, pp. 173–195.

Tam, W. K. (1995), "Asians — A monolithic voting bloc?" *Political behaviour*, 17, 223–249.

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985), *Statistical analysis of finite mixture distributions*, Chichester: Wiley.

Train, K. E. (2003), *Discrete Choice Methods with Simulation*, Cambridge: Cambridge University Press.

van der Brug, W., van der Eijk, C., and Marsh, M. (2000), "Exploring Uncharted Territory: The Irish Presidential Election 1997," *British Journal of Political Science*, 30, 631–650.