What's in a Name? A Political Myth?

Michael Mernagh
mernaghm@hotmail.com

Abstract

This paper first reviews three methodologies for deriving a data frame that represents all of the electorate, to permit a comparison of the alphabetic distribution of TDs ("Teachtaí Dála") who are members of the "Dáil" (the lower house of the Irish parliament) with the alphabetic distribution of the population. The paper then uses statistical graphs to assess how the Irish electorate have voted since 1922. (In Ireland, candidates' names are listed alphabetically on ballot papers.) The paper concludes with an opinion on whether Ireland ought to retain alphabetical listing.

Keywords: ballot randomisation; donkey voter; Electoral Register; Benford's law; Markov chain.

1   Definition

A *donkey voter* [11] or *top-to-bottom voter* [13] is someone who votes for candidates based solely upon the sequential order in which they appear on a ballot sheet, rather than taking the time to number the candidates in the voter's own thought-out order of preference.

In countries where voting is compulsory, apathetic voters sometimes cast donkey votes just to avoid a fine.

2   Finding an Adequate Comparison Frame

The Central Statistics Office (CSO) in Ireland conducts total enumerations of the population in the quinquennial census. Nevertheless, the CSO neither compiles nor maintains databases of named persons. The CSO abides by the principle of using information for statistical purposes only. This reassures the public's confidence in the CSO as an independent body.

3   The Electoral Register

The national Electoral Register records electors in 'Polling Books' for 3,400 District Electoral Divisions.

Random samples are over-represented in larger households, which contain more electors, and so have a higher probability of selection than addresses that contain a smaller number of electors. A sampling bias also arises from the non-listing of households in which no member appears on the Electoral Register.

In their 1973 paper on Alphabetical Voting, Robson and Walsh [10] used as a benchmark frame the alphabetical distribution of a random sample of 2,100 people from the national Electoral Register. They grouped those names into five sets: A–C, D–G, H–L, M–O and P–Z, with an average of 420 names in each set.

In a written answer to Parliamentary Question Number 484 on 17 Feb 2004, the Data Protection Commission [2] made it clear that it is no longer legal (since 2001) to use the "full" Electoral Register for anything other than electoral or statutory use, even if it were possible to get hold of it.

The "edited" register lists only persons who have indicated that they have no objection to their details being used for purposes other than electoral or other statutory uses.

Persons contacted for interview who had opted out of the edited register might well raise objections and seek to find out how their names had been obtained. If illegal uses of the register became widespread, then it could be brought into disrepute—perhaps to the point where some people might choose not to register to vote at all.

Apart from that, a bias would result from using the edited version, because it would be improbable that those who opt out are a random sample of the full version.

4   Matheson's Methodology

In 1894, Sir Robert E. Matheson [7] issued his *Special Report on Surnames in Ireland* as an Appendix to the *Twenty-ninth Report of the Registrar-General of Marriages, Births and Deaths in Ireland*. It was the first detailed official work on surnames in Ireland

Matheson's methodology was to list all surnames (including their variations) that accounted for five or more births in the year 1890. He tabulated some 2600 names with the total number of births for each of those names.

He then multiplied the numbers of births in 1890 by 44.8 (which was the overall Birth Rate per 1,000 at that time) and rounded the result to the nearest hundred to estimate the population size of each surname stratum.

He listed the 100 most numerous names in the country at large with an estimate of the numerical frequency of each surname. These estimates are available at Freepages [3].

Matheson's printed opus is in two parts. The first part is a long table of statistics based on the registration of births in all of Ireland in 1890. For each name, there are six columns: the total registered, number registered in each of the four provinces, and notes on the counties in which each name was principally found.

The second part presents a list and index of names with variant forms and includes four fascinating, and sometimes amusing, chapters on spelling, contractions, interchangeable names, and English and Gaelic forms.

Matheson's ingenious method gave a rational approximation to the frequencies of surnames in Ireland in 1890. Incidentally, the roundings in the arithmetic generated relatively small margins of error that are best estimated using Chebyshev's inequality.

Madison's frequency table of surnames covered the island of Ireland. This is a disadvantage if used to compare the distribution of surnames of candidates successful in general elections in the Republic of Ireland with the corresponding distribution among the electorate.

5   Telephone Directories

I adopted a different approach to derive similar data. I comprehensively sampled all six residential telephone directories that covered the Republic of Ireland for the year 2009. I counted the numbers of pages devoted to each surname categorised by initial letter. I multiplied those numbers by the average number of entries per page. Then, I aggregated and rounded the results to the nearest 100 to estimate the total number of residential phone users. Lastly, I multiplied those data by a grossing factor to reflect the total population according the census of 2006.

Figure 1
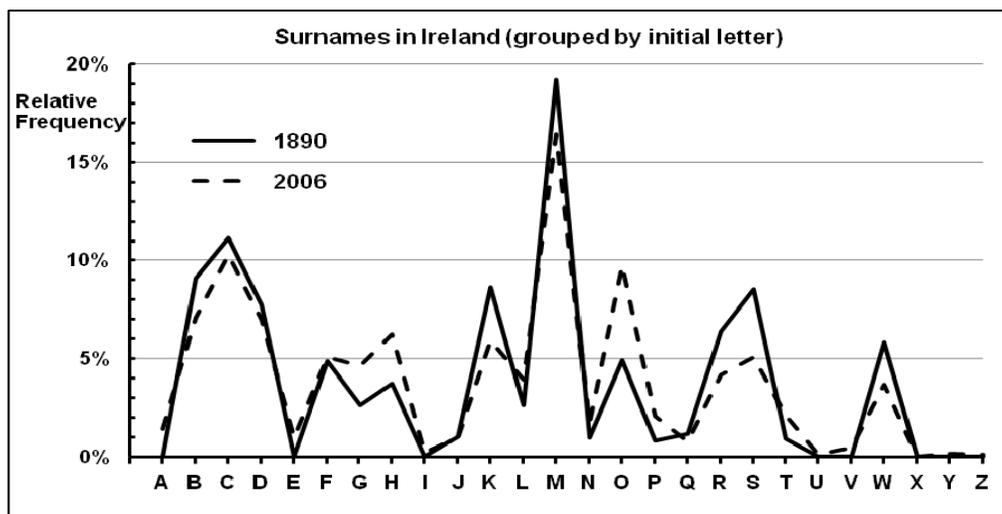
6 Comparison of Methodologies

Although the compilation methodologies differ, the sets of data for 1890 and 2006 are highly correlated.

The most noticeable change is the decrease in the relative frequency of surnames that begin with the letter "M" and the partial switch of frequencies from the reduction of names beginning with "S" to increase the number of names beginning with "O".

For his analysis, Matheson used the most commonly found forms of surnames. Thus, he listed Shea rather than O'Shea and Sullivan rather than O'Sullivan. After 1890, there was a tendency to resume the "O" in names that had previously dropped it. Surnames beginning with "O" are found to cluster in southwestern Ireland.

Analysis of each regional telephone directory shows that surnames beginning with "M" are relatively more numerous in the northern half of the island of Ireland. The decrease with respect to "M" is explained by the fact that the compilation for 1890 included Northern Ireland (where names beginning with "Mac" or "Mc" predominate), whereas the compilation for 2006 covered only the Republic of Ireland. The 2006 data correspond to the area represented by TDs in the Dáil.

Figure 1 compares the data derived from the telephone directories with the result of the classic research by Matheson.

In passing, it may be observed that an inverse square root transformation of the rank order of surnames (grouped by initial letter), linearises the relationship of rank order with surname initial letter frequencies.

For the 1890 data, the relationship between the rank order of a name and the frequency of the name can be linearised, as shown in Figure 2.

A similar linearised relationship exists between the rank order of letters and their relative frequencies in the 2006 data, as shown in Figure 3.

7 An Adaptation of Benford's "Law"

It might be expected, prima facie, that roughly the same number of surnames would begin with each letter of the alphabet and that the proportion of surnames beginning with any given letter would be roughly uniformly 1/26.

However, for many kinds of alphabetic data, the distribution of initials is highly skewed. A precise mathematical relationship, (known as *Benford's law* for numeric data) seems to hold (when adapted to model alphabetic data). See Plus maths [9].
This law does not work for truly random sets of data. It works best for data that are neither completely random nor overly constrained, but rather lie somewhere in between. These data can be wide ranging, and are typically the result of several processes, with many influences.

Figure 2

Linearised model of the 100 most frequent surnames in Ireland in 1890

$y = 88790x$
$R^2 = 0.9879$

$X = \dfrac{1}{Sqrt(n+1)}$ where n = rank and 0 < n < 101

Figure 3

Figure 3: Linearised model of the estimated frequencies of surnames in Ireland in 2006
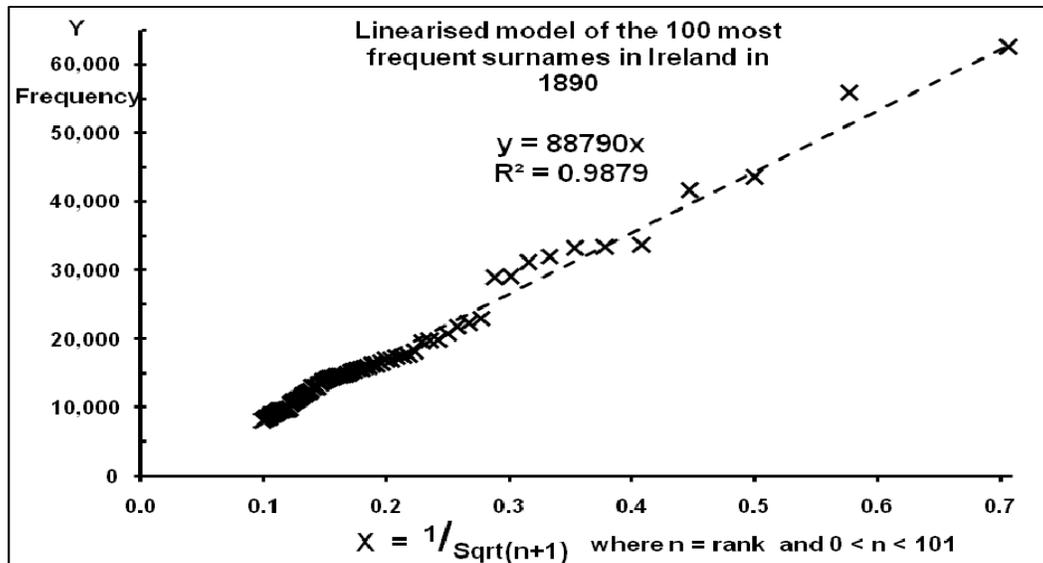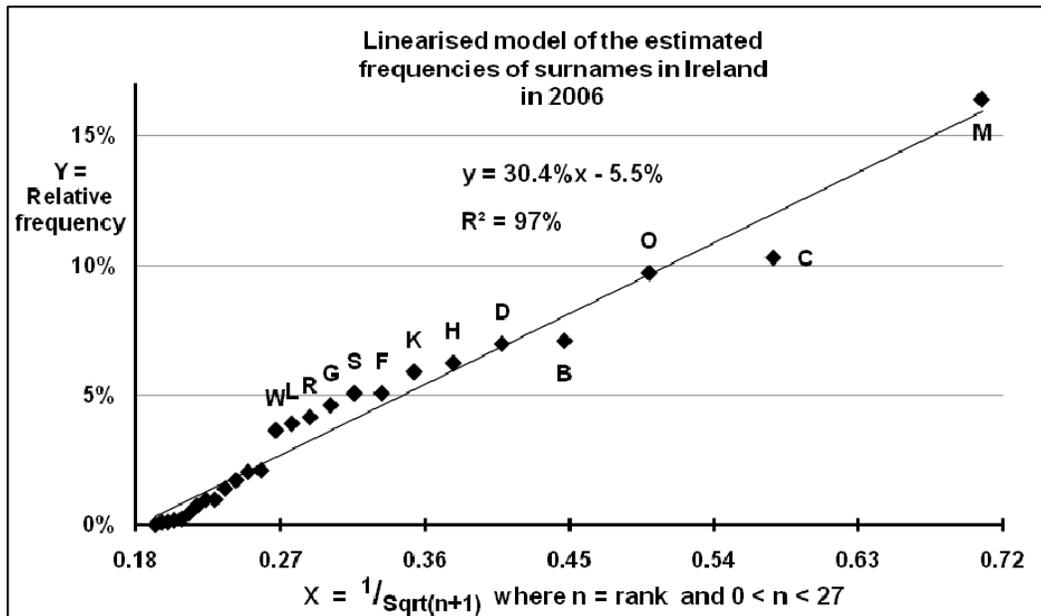
The expected proportion of surnames beginning with any letter is

$$\log_{27}[(n+1)/n],$$

where $0 < n < 27$ is the frequency rank of the letter. The cumulative function of

$$y = \log_{27}\{(n+1)/n\}$$

is

$$\Sigma y = \log_{27}(n+1).$$

8 Review of All Results in the 28 General Elections from 1922 to 2007

The analysis in this section is based on a historical list of TDs provided by Wikipedia [12].

The R^2 statistics measure how well the alphabetic distributions of TDs' surnames correspond with the comparison population.

A pattern can be observed in the graph of the R^2 values over time.

Figure 4 shows that from 1922 to 1965, the Irish electorate tended more to favour candidates who had a higher alphabetic ranking on the ballot papers.

The lowest R^2 value was in 1969.

The 1960s saw a surge of economic growth in modern Ireland. The establishment of a national television station in 1961 broadened political debate among the electorate. Free secondary education was introduced in 1966 for all social classes in both urban and rural areas. In 1972, the voting age was reduced from 21 to 18. Ireland joined the European Economic Community in 1973.

From 1973 to 2007, the trend has been away from alphabetical voting. In 2002, the R^2 value had regained the high level recorded in 1922.

Figure 5 shows the randomness of the spreads of the relative frequencies of surnames from 1922 to 2007.

Figure 6 compares the frequency of surname initials among TDs with the frequency of surname initials in the population. There are two outliers in Figure 6. They indicate that on average, historically, candidates whose surnames began with *B* had an advantage over candidates whose surnames began with *M*.

9 Analysis of All TDs Elected in 2007

The analysis of all TDs elected in 2007 is based on a report of the Government of Ireland [5]. Figure 7 illustrates the divergences of the frequencies of TDs' surnames from the corresponding population frequencies. The divergences are randomly scattered.

Figure 8 shows that the distribution of the surnames of TDs elected in 2007 does not conform to "Benford's Law". *The inference to be*

Figure 4

Figure 5

Figure 6
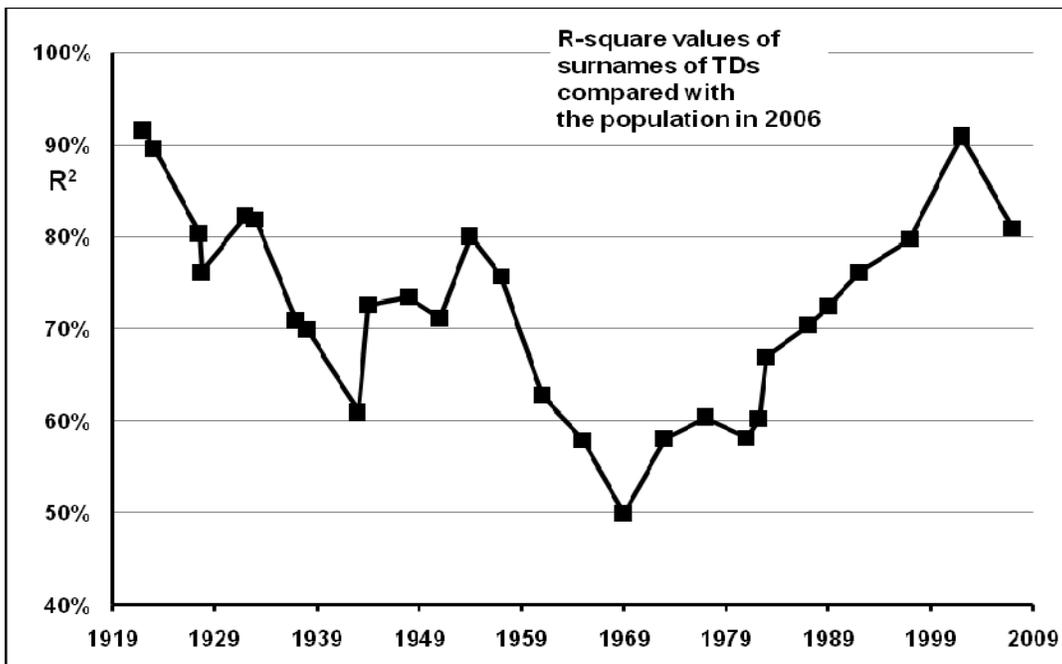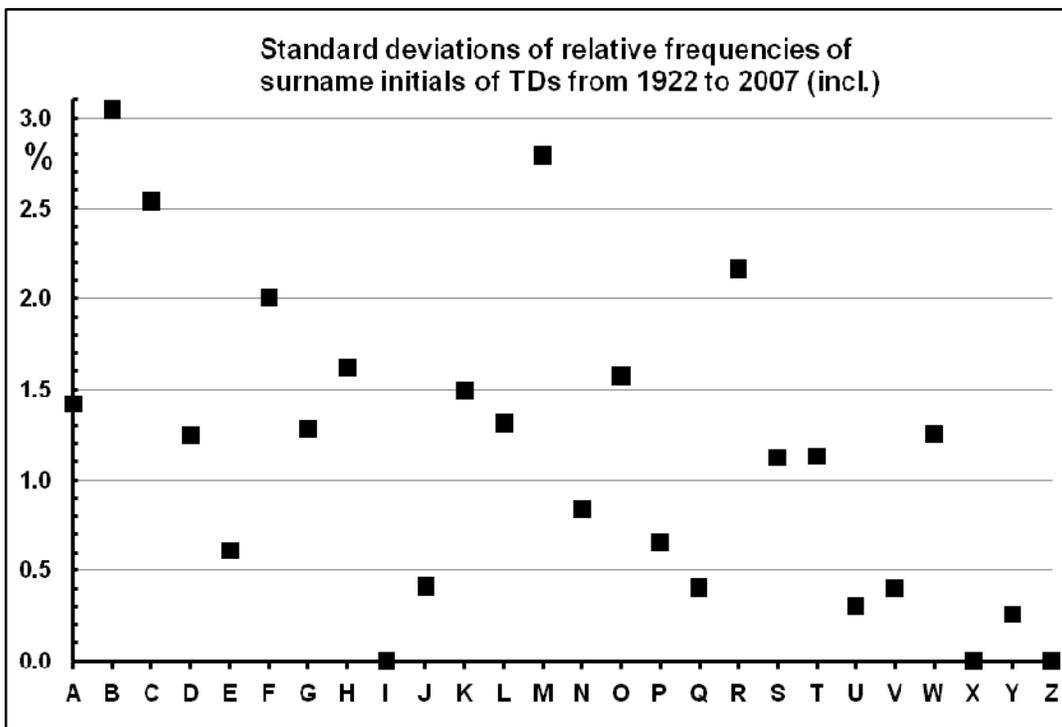
Scatterplot comparing TDs with the population in 2006
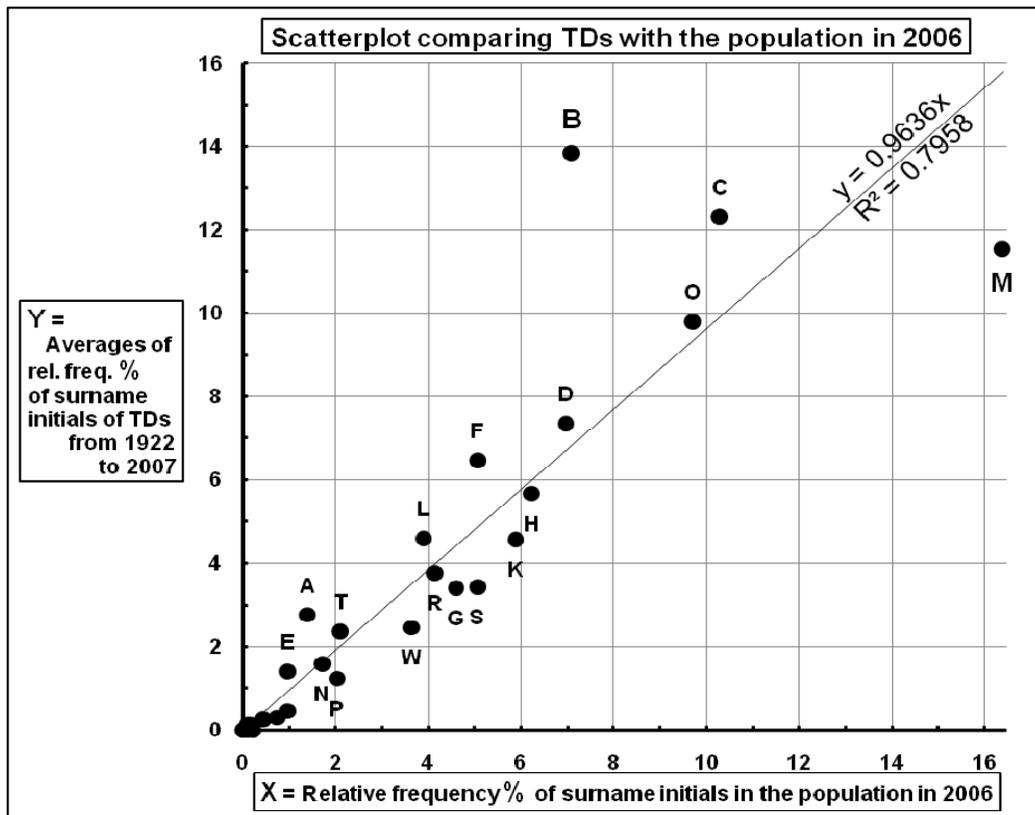
made is that the electorate did not tend to vote in accordance with the order of appearance of candidates' names on a ballot paper.

Table 1 summarises the average perpendicular distance of the cumulative datasets from both the cumulative Uniform and cumulative Benford distributions.

Table 1

Variable	Standard deviation of variable from	
	Uniform	Benford
1890 Population	6.3%	12.4%
2006 Population	7.0%	12.5%
TDs (2007)	6.2%	10.6%

On average, both population and TDs' datasets are closer to the uniform distribution than they are to Benford's.

10 Applying Markov Chain Theory to Derive Long-Run Results

What will be the distribution of TDs' surnames in the long-term? Construct a matrix of the relative frequencies of surnames beginning with the 23 letters of the alphabet in columns (excluding *I*, *X* and *Z*) by the 23 Dáils in rows from 1933 (the 8th Dáil) to 2007 (the 30th Dáil), inclusive.

The total of the elements in each row is unity (100%).

Calculations using Markov chains then indicate that in the future the distribution of surnames in the 34th Dáil, sometime in the future, will be as shown in Figures 9 and 10.

Since the lifetime of a Dáil can last for up to five years, this could be a 20-year prediction.

It is clear from Figure 9 that the predicted over-representation of surnames beginning with *B* will be matched by the under-representation of surnames beginning with *M*. Excluding *A* and *B*, there is a general oscillation between

Figure 7

Figure 8

Figure 9

Expected distribution of TDs' surnames in the long term
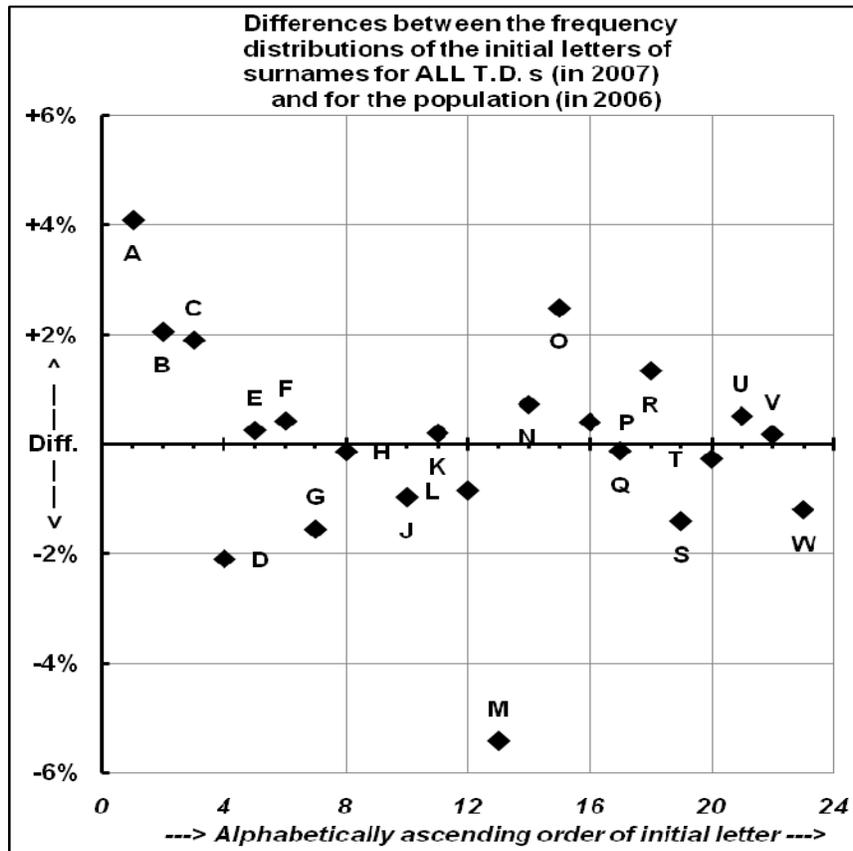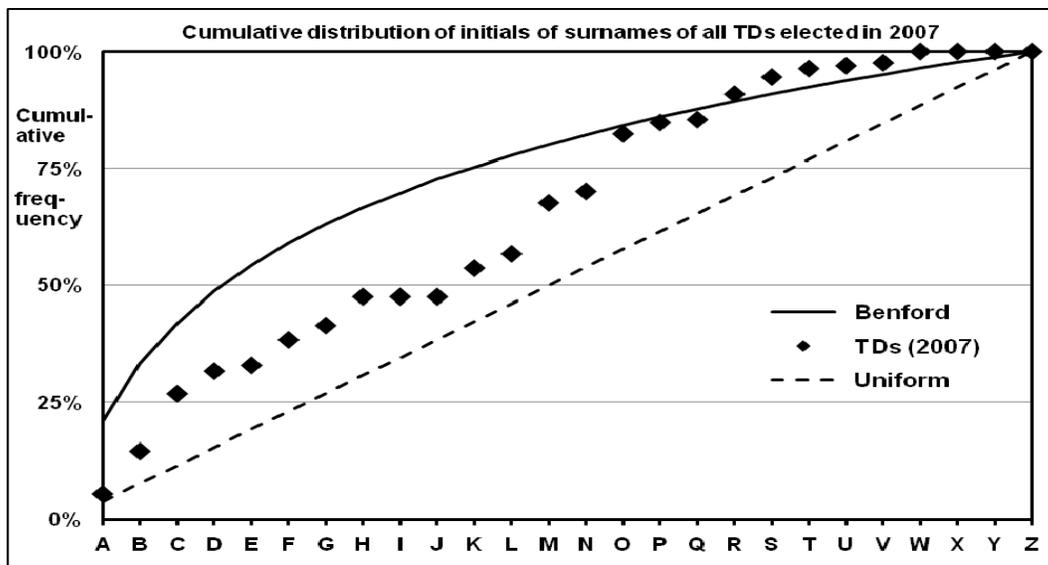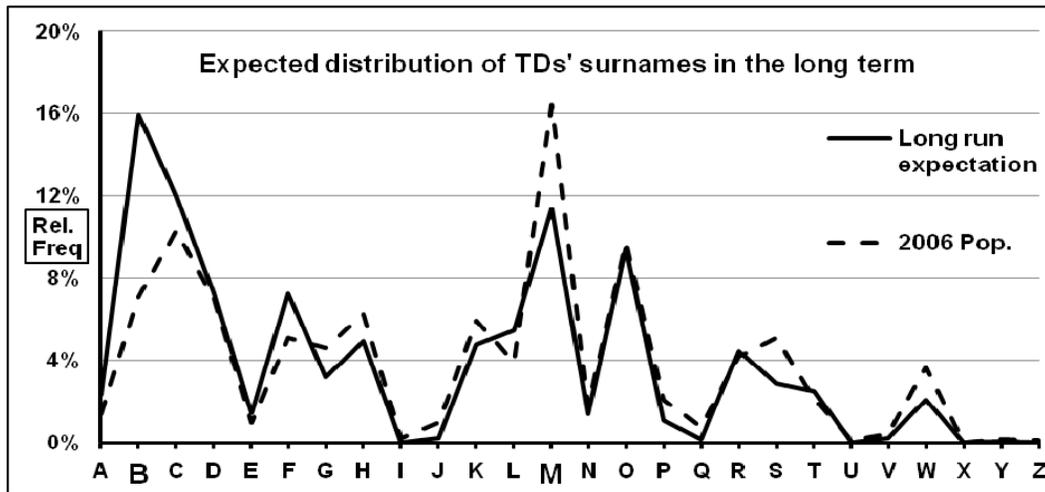
successive letters from *C* to *W* in the relative frequencies of surnames that begin with those letters.

11 Representation of Women in the Dáil

Figure 11 shows that at least one woman has always been elected to the Dáil. From 1922 to 1969, the average number of women TDs was 3.65. Since 1969, that number has grown exponentially, reaching 23 in 2002.

From independence to the present day, 73 women have won 223 seats (an average of 3.05 election victories per woman). Figure 12 shows that frequency distribution.

Figure 13 shows the resemblance between the alphabetic distribution of surnames in the

population and the corresponding distribution for the 73 women who were elected down through the years.

12 Political Dynasties

For many years, a feature of Irish political life has been the tendency for sons and daughters to "inherit" the parliamentary seats previously held by their parents. This tends to happen most often in by-elections, where a bereaved candidate often attracts a significant "sympathy vote".

A sine wave curve accounts for nearly two-thirds of the variation in the dataset illustrated in Figure 4. It implies that the trend in the "inheritance" pattern in political history repeats itself every 125 years (which span five

Figure 10

Cumulative distributions of TDs' surnames

Figure 11

The chart "Women TDs" with equation $y = 7.08e^{0.03x}$, $R^2 = 0.75$, axes labeled "No." (vertical) and "X = Year minus 1969" (horizontal). Legend: Actual No.; Expon. (Actual No.)
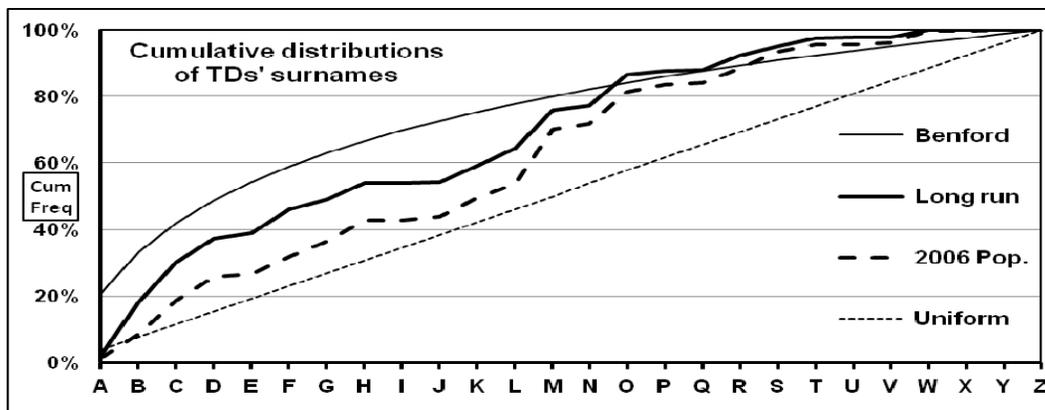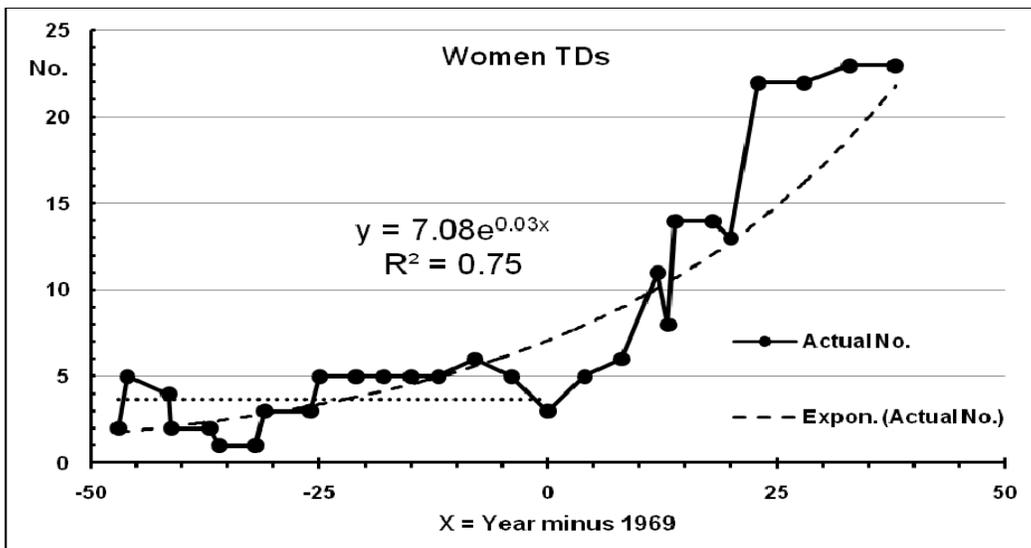
generations from great-great-grandparent to grandchild). See Figure 14.

13 Data Correlations

Table 2 shows the R^2 matrix of the datasets studied in this paper up to this point. Taking the variables pairwise, each R^2 value quantifies the amount of variation within one of those variables that can be accounted for by a linear model of the other variable.

The correlation of the TDs in 2007 with the population in 2006 is stronger than the correlation of the TDs in 2007 with the disaggregated cumulative Benford distribution.

14 Comparisons of "First-Preference" and "First-Past-The-Post" Voting Systems

The analysis in this section is based on a report of the Government of Ireland on the election of 2002 [4]. Figure 15 shows that the estimated frequencies of initial letters of surnames in the population of Ireland in 2006 explain 86% of the variation of first preference votes for successful candidates in the 2002 general election. This percentage rises to 92% for 2002 if all first preference votes for both successful and unsuccessful candidates are taken into account (see Figure 16).

Figure 12

Histogram of numbers of times Women were elected to the Dáil. Vertical axis: Number of Women; Horizontal axis: Number of times elected.
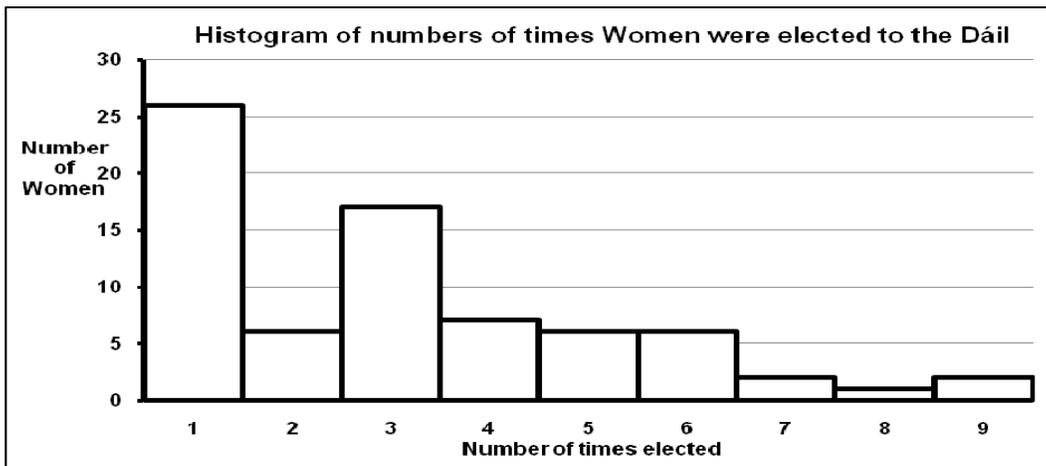
Table 2

R^2	I	II	III	IV	V	VI	VII	VIII
I	100%							
II	85%	100%						
III	1%	2%	100%					
IV	0%	0%	0%	100%				
V	62%	73%	14%	0%	100%			
VI	61%	81%	14%	0%	81%	100%		
VII	66%	80%	14%	0%	98%	87%	100%	
VIII	57%	77%	9%	0%	87%	88%	90%	100%

KEY to Table 2					
I	=	**1890 Population**	**V**	=	**Long run expectation**
II	=	**2006 Population**	**VI**	=	**ALL TDs (2007)**
III	=	**Benford Distribution**	**VII**	=	**ALL TDs (ALL years)**
IV	=	**Uniform Distribution**	**VIII**	=	**Women TDs (ALL years)**

The proportional representation system of voting does not necessarily always elect the candidates with the highest number of first preference votes.

One might wonder if the results under the "first-past-the-post" system would reflect the distribution of surnames in the population better, but the R^2 statistic for such a system is only 81%. Thus, in 2002 proportional representation reflected the general population better than the "first-past-the-post" system.

Similarly, based on a report of the Government of Ireland on the election of 2007 [5], Figure 17 shows that the estimated frequencies of surnames explain 79% of the variation of first preference votes for successful candidates in that election. This percentage rises to 89% for 2007 if all first preference votes for both

Figure 13

Frequency distribution of surnames of 73 Women TDs

Figure 14

R-square values of surnames of TDs compared with the population in 2006

$$EST = 77\% + 16\%[SIN\{2\pi(X+47)\div125\}]$$

X = Year minus 1922
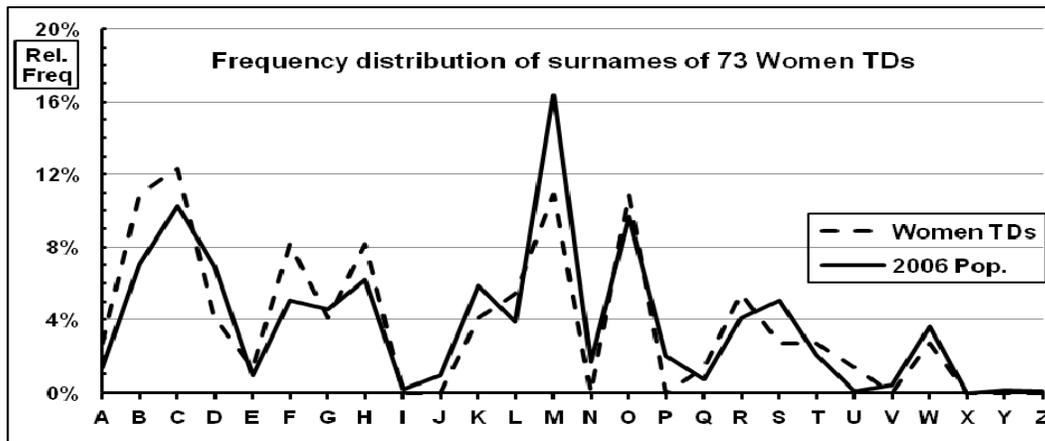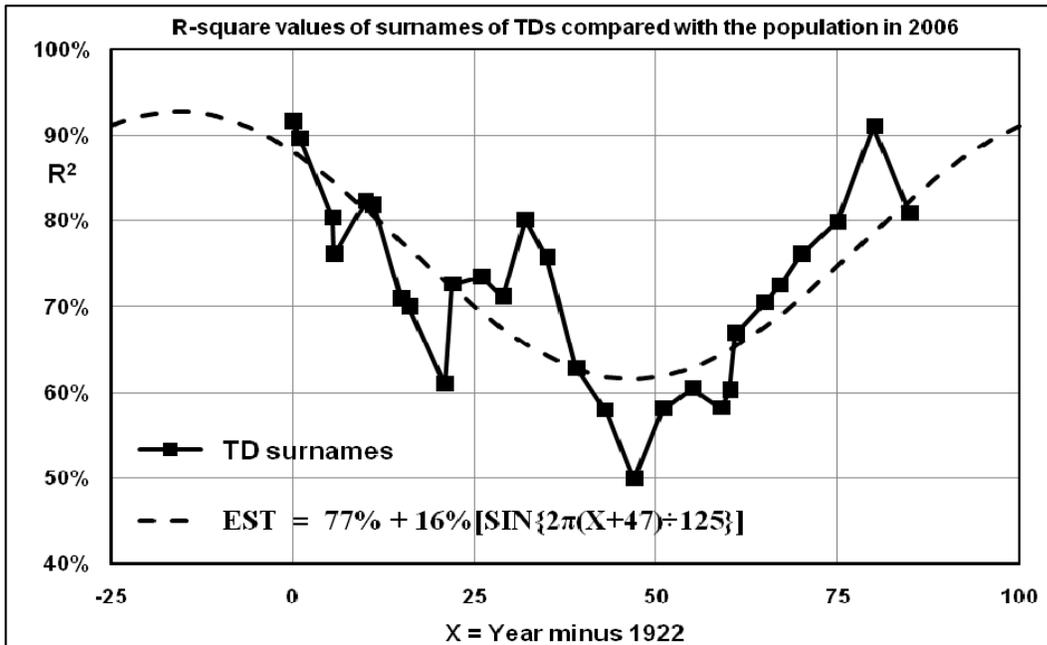
successful and unsuccessful candidates are taken into account (see Figure 18).

For 2007, the results under the "first-past-the-post" system do not reflect the distribution of surnames in the population any better (the R^2 statistic is also 79%).

The surplus votes from candidate with a very large number of first preference votes can, when transferred, benefit another candidate who did not poll so well. A system of "vote management" has developed, where the constituency is divided into canvassing areas based on such candidates' home bases.

Figure 15

Comparison of first preference votes for TDs elected in 2002 with the estimated population in 2006

$R^2 = 86\%$

First pref's (2002)

Est. Pop'n (2006)

Initial of Surname
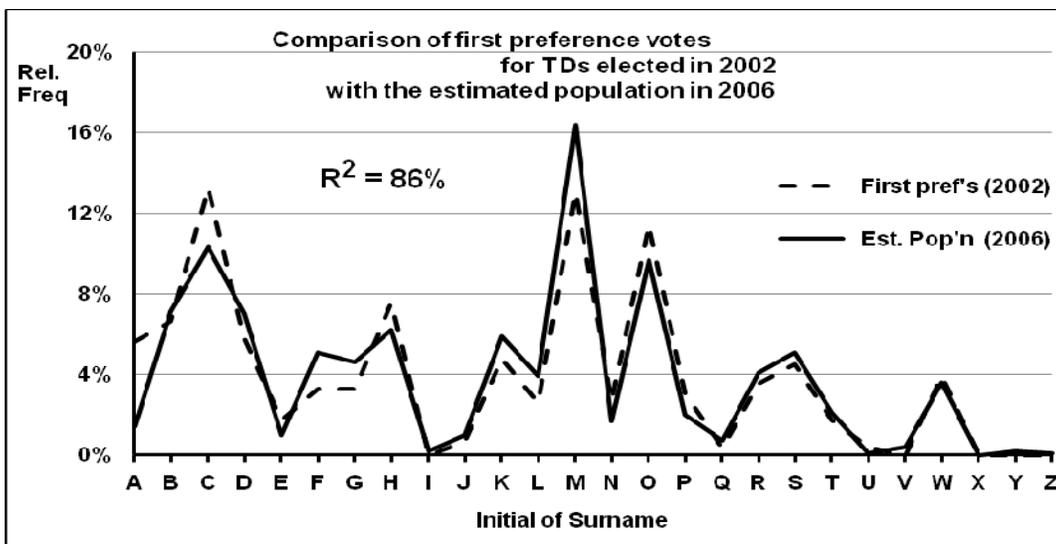
Figure 16
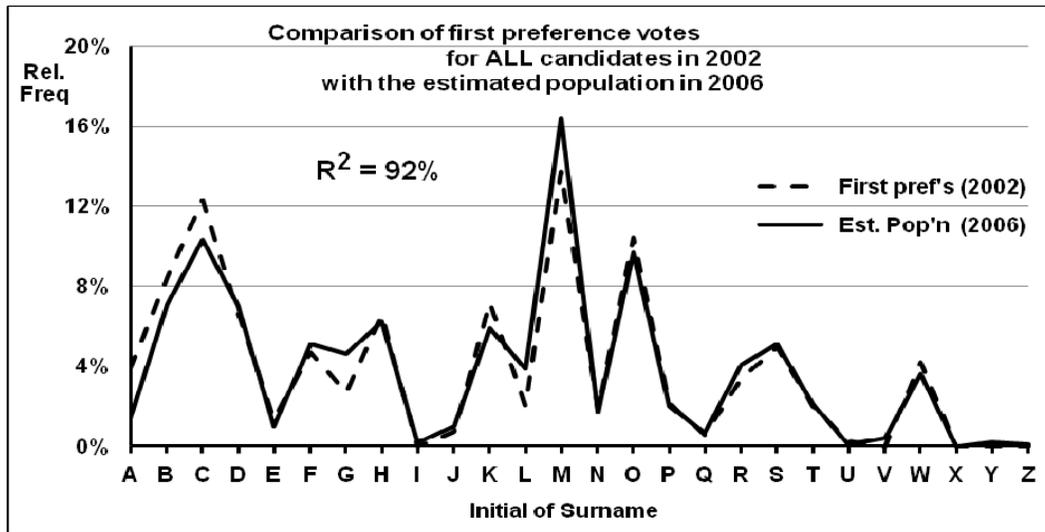
15 Summary

1. Matheson's data for 1890 are highly correlated with the set of data for 2006 extracted from the telephone directories.

2. The alphabetic distribution of TDs' surnames corresponds well with the general population.

3. From 1922 to 1969, the Irish electorate tended more and more to favour candidates who had a higher alphabetic ranking on the ballot papers, but this trend was reversed from 1969 onwards.

4. The divergences of the frequencies of TDs' surnames from the corresponding population frequencies are randomly scattered.

Figure 17
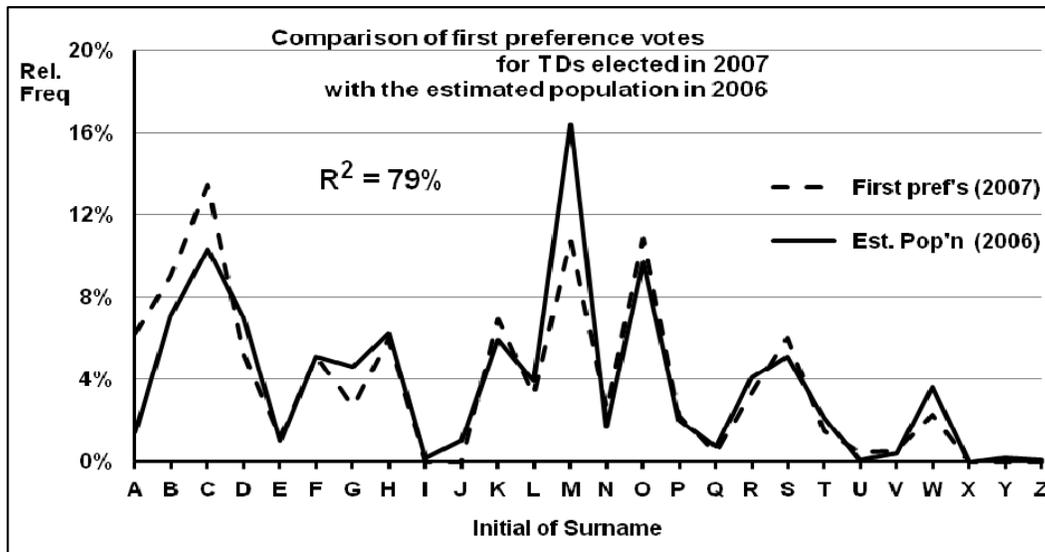
Figure 18
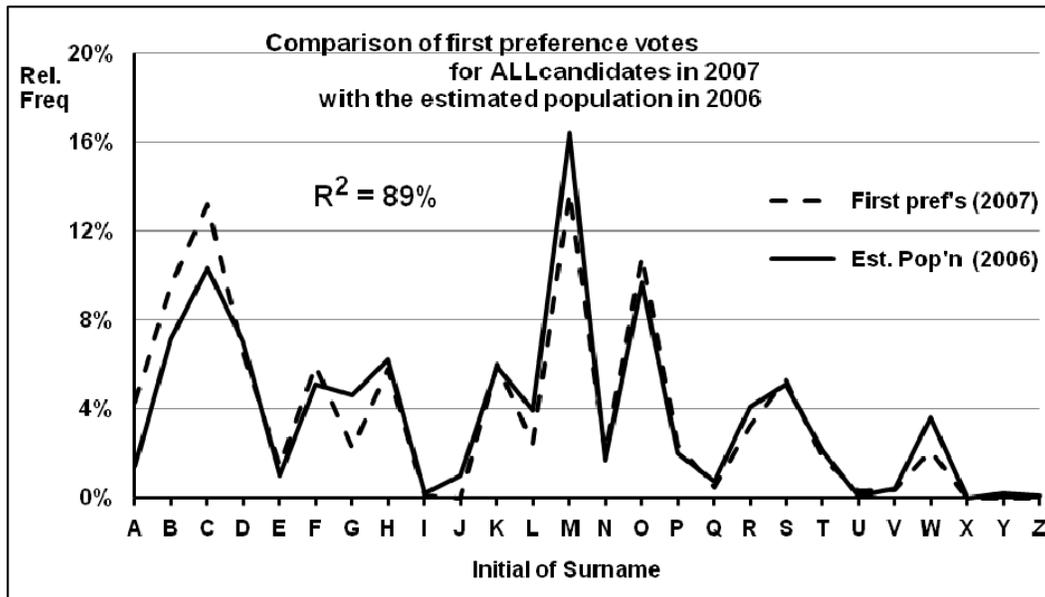
5. The divergences of the frequencies of TDs' surnames from the corresponding population frequencies are randomly scattered.

6. Based on past results, in the long-term future, the over-representation of surnames beginning with *B* will be matched by an under-representation of surnames beginning with *M*. Surnames of TDs beginning with letters other than *B* and *M* will mirror their distribution in the general population.

7. Since 1969, the number of women elected to the Dáil continues to grow.

8. The distribution of the surnames of women TDs conforms closely to the population's distribution.

9. Political dynasties are a fact of Irish political history.

10. In general, results under the "first-past-the-post" system do not reflect the distribution of surnames in the population better than the proportional representation system.

11. In 2007, the electorate did not vote strictly in accordance with the order of appearance of candidates' names on the ballot papers.

12. In the 2002 and 2007 Irish general elections, candidates did not gain a significant advantage exclusively because of their alphabetical order of appearance on ballot papers.

16 Conclusion: Should Ireland Retain Alphabetical Listing?

In the 1986 case of *O'Reilly v Minister for Environment*, the Irish High Court upheld the constitutional validity of alphabetical listing against an equality-rights challenge. The court noted that despite its faults, A to Z does have the advantage of making it easy to find candidates on the ballot-paper.

Since 1965, the political party to which a candidate belongs is printed beside the candidate's name. (See Marsh [6].) Candidates' photographs and their party logos have appeared on ballot papers since 2002 (Office of the Attorney General [8]). These measures partially offset any perceived advantage arising from the listing of surnames alphabetically.

Furthermore, many Irish voters have strong political opinions and vote for the party for

which they have always voted. Intelligent voters go to the polls with their minds made up beforehand on how they choose to vote.

Usually, there are only between eight and 15 candidates on most ballot paper. Voters are not bewildered by a multitude of names and so are less likely to take the easy option to vote like donkeys, particularly as they are not obliged to list their preferences for all candidates.

Since 1979, Australia has used a system called Robson Rotation. Each ballot paper contains a different permutation of candidates. Each candidate's name appears a certain proportion of times at every position on the paper. This disperses the donkey votes equally and nullifies their impact on the result.

Computerised randomisation could be incorporated into the modern processes for printing ballot papers. The order of the candidates could also be rotated so that if there were ten candidates, each would head the ballot on one-tenth of the papers. Such a system was used in New York City from 1937 to 1947.

Electronic voting machines costing €52 million were tested in three constituencies in Ireland in the general election of 2002. Subsequently, much debate and serious doubt arose about the accuracy of the software (Coyle et al. [1]). The system was abandoned. The annual cost of insurance and storage for those idle machines is about €800,000. In the interests of openness, transparency and accountability, any randomisation software for ballot papers would have to pass stringent tests to satisfy public confidence.

We should never under-estimate the collective intelligence of the electorate. They are not such donkeys as is commonly supposed. When all of the votes have been cast and counted, Democracy is the only guarantor of the least evil. People elect governments whom they think will do the least harm, but do it very well. In the final analysis, we are governed by the best of a bad lot.

The only thing as bad as a 'donkey-vote' is a 'reverse donkey-vote' (or a 'bottom-to-top vote').

17 References

[1] Coyle, L., Cunningham, P. and Doyle, D. Validity of the Electronic Implementation of the Counting Rules — Dáil, European and Local Elections, Report for Commission on Electronic Voting, http://www.cev.ie/htm/report/first_report/pdf/Appendix%202E-Part1.pdf (accessed December 8, 2010).

[2] Data Protection Commission: Case Study 5/97: Use of Electoral Register to Prepare Mailing Lists or for Other Purposes Not Related to its Primary Function, http://www.dataprotection.ie/viewdoc.asp? DocID=160 (accessed December 8, 2010).

[3] Freepages: Table of the Hundred Most Numerous Surnames in Ireland in 1890, http://freepages.genealogy.rootsweb.ancestry.com/~irishancestors/Surnames/TableIrl.html (accessed December 8, 2010).

[4] Government of Ireland (2003) 29th Dáil General Election, May, 2002, Election Results and Transfer of Votes, http://www.oireachtas.ie/documents/a-misc/Elections2002Result.pdf (accessed December 8, 2010).

[5] Government of Ireland (2007) 30th Dáil General Election, May, 2007, Election Results and Transfer of Votes, http://www.oireachtas.ie/documents/publications/Electoral_Handbook1.pdf (accessed December 8, 2010).

[6] Marsh, M. Candidates or Parties? Objects of Electoral Choice in Ireland, www.tcd.ie/ines/files/personal_vote_PP_rev.pdf (accessed December 8, 2010).

[7] Matheson, R. E. (1894) Special Report on Surnames in Ireland, ISBN 184630052, http://www.whollygenes.com/Merchant2/merchant.mvc?Screen=PROD&Product_Code=IET0047 (accessed December 8, 2010).

[8] Office of the Attorney General, Statutory Instrument No. 16/2002 — Ballot Paper (Photographs and Emblems) Regulations, 2002, http://www.irishstatutebook.ie/2002/en/si/0016.html (accessed December 8, 2010).

[9] Plus Maths: Benford's Law, http://plusMaths.org/issue9/features/benford (accessed December 8, 2010).

[10] Robson, C. and Walsh, B. (1973) Alphabetical Voting: A Study of the 1973 General Election in the Republic of Ireland, Economic and Social Research Institute (ESRI), General Research Series No #71, Dublin.

[11] Wikipedia: Donkey Voter, http://en.wikipdia. org/wiki/Donkeyvote (accessed December 8, 2010).

[12] Wikipedia: Historical Lists of TDs, http://en.wikipedia.org/wiki/Category: Members _of_ the_3rd_D%C3%A1il (accessed December 8, 2010).

[13] Wikipedia: Top-to-bottom Voting, http://en.wikipedia.org/wiki/Donkey voting (accessed December 8, 2010).

About the Author

Michael Mernagh is a Chartered Statistician and a Fellow of the Royal Statistical Society. He was born in Enniscorthy, Ireland, in 1948. He graduated as a mature student in mathematical statistics at the Dublin Institute of Technology. He is now retired and lives near Cork city.