Developing the Empirical Side of Computational Social Choice

T. Nicolaus Tideman and Florenz Plassmann

Department of Economics, Virginia Polytechnic Institute and State University (ntideman@vt.edu)
and Department of Economics, State University of New York at Binghamton (fplass@binghamton.edu)

Abstract*

The empirical side of computational social choice has received relatively little attention. This paper discusses some interesting questions in public choice that can be addressed through empirical inquiries. The possibilities for such inquiries are illustrated with the example of how an empirical inquiry can be used to estimate which voting rule is most likely to identify the best option.

Introduction

One way of classifying work in computational social choice is to divide it between deductive work and inferential work. Deductive work is concerned with theorems, that is, with what can be proved mathematically. Inferential work is concerned instead with what can be regarded as plausible, based on evidence. Most of the work that has been done in computational social choice has been in the deductive tradition. This paper advances the argument that there is interesting and valuable work to be done in the inferential, or empirical, mode. We discuss the possibilities for interesting work in the empirical mode of computational social choice first in terms of the problem of developing models of election outcomes that can be used to estimate the frequencies with which various phenomena will be observed, and then in terms of estimating the probabilities with which different voting rules will select the best option.

Statistical Models of Vote Casting

Consider an election with m candidates and n voters. There are $m!$ strict rankings (rankings without ties) of the candidates. Assume that every voter submits a ballot with one of these strict rankings, and that n_r, $r = 1, \ldots, m!$ voters submit ranking r, with $\Sigma n_r = n$. Let N_r be a random variable that describes the distribution of n_r and let $N = (N_1, \ldots, N_{m!})$ be a random vector with joint density function

$$f_N(n_1, \ldots, n_{m!}; \pi_N) \quad (1)$$

where π_N is a vector of parameters. We call a specification of f_N a "model of N."

For the models of N that we consider below, π_N contains the expected shares of the $m!$ rankings. Define p_r as the expected share of the ballots with ranking r, with $\Sigma p_r = 1$. Let P_r be a random variable that describes the distribution of p_r, and let $P = (P_1, \ldots, P_{m!})$ be a random vector of vote share probabilities with joint density function

$$f_P(p_1, \ldots, p_{m!}; \pi_P) \quad (2)$$

where π_P is a vector of parameters. We call a specification of f_P a "model of P."

We define a statistical model of vote casting as a combination of a model of N and a model of P. The model of N describes the distribution of and the dependence among the ballots in any given election (given a specific realization of P), while the model of P specifies the probabilities of the vote shares (thus p_r is the expected vote share $E[N_r/n]$), and describes how they vary across elections, if they do.

Models of N

If voters submit their ballots independently, then f_N is the pdf of a binomial distribution if $m = 2$ and the pdf of a multinomial distribution if $m > 2$. For both models, $\pi_N = \{p_1, \ldots, p_{m!}, n\}$ with $\Sigma p_r = 1$, $E[N_r] = np_r$ (and thus $E[N_r/n] = p_r$), $Var[N_r] = np_r(1 - p_r)$, and $Cov[N_r, N_s] = -np_rp_s$.

An intuitive way of modeling dependent ballots is by conceiving of voting in terms of an urn model, with different possibilities for the replacement rule. While the binomial and multinomial distributions describe drawing with replacement, drawing without replacement leads to the hypergeometric distribution if $m = 2$ or the multivariate hypergeometric distribution if $m > 2$, with parameter vector

---

* Parts of this paper are taken verbatim from our three joint papers cited in the references; this paper collects material that is spread over these sources. Anyone wishing to cite this paper should look for corresponding material in one of these papers.

$\pi_N = \{p_1, \ldots, p_{m!}, n, \phi\}$, with $\phi > n$.[1] The first two moments are $E[N_r] = np_r$, $Var[N_r] = np_r(1 - p_r)\Phi$, and $Cov[N_r, N_s] = -np_rp_s\Phi$, with $\Phi = (\phi - n) / (\phi - 1)$. As ϕ approaches infinity, the distribution converges to the binomial distribution (if $m = 2$) or multinomial distribution (if $m > 2$).

Drawing with replacement plus the addition of another ballot of the type drawn leads to the binomial-beta distribution if $m = 2$ and the multinomial-Dirichlet distribution (also known as multivariate Pólya distribution) if $m > 2$. One way of deriving these distributions is to let the $m! - 1$ independent share parameters of the binomial or multinomial distribution follow a beta or multivariate beta (= Dirichlet) distribution, respectively, with parameter vector $(p_1,\ldots,p_{m!}, \psi)$.[2] The first two moments of the resulting compound distributions are $E[N_r] = np_r$, $Var[N_r] = np_r(1 - p_r)\Psi$, and $Cov[N_r, N_s] = -np_rp_s\Psi$, where $\Psi = (\psi + n) / (\psi + 1)$. As ψ approaches infinity, the variance of the distribution imposed on the original share parameters approaches zero and the compound distribution converges to the original distribution of N. Further variations in the sampling and replacement procedures lead to distributions that can accommodate additional forms of dependence among the ballots (see, for example, Johnson *et al.*, 1997, pp. 200-231).[3]

In a model of vote casting that uses the multinomial distribution, the variance of each ballot share N_r/n converges to zero as the number of voters becomes large, or

$$\lim_{n\to\infty} Var\left[\frac{N_r}{n}\right] = \lim_{n\to\infty}\left(\frac{1}{n^2}Var[N_r]\right) = \lim_{n\to\infty}\left(\frac{n}{n^2}p_r(1-p_r)\right) = 0. \quad (3)$$

A comparable result holds in a model of vote casting that uses the multivariate hypergeometric distribution; the requirement that $\phi > n$ ensures that

$$\lim_{n\to\phi} Var\left[\frac{N_r}{n}\right] = \lim_{n\to\phi}\left(\frac{n}{n^2}p_r(1-p_r)\frac{\phi-n}{\phi-1}\right) = 0. \quad (4)$$

The fact that the hypergeometric distribution imposes an upper limit on n suggests that at best it can serve as an approximation of the model of N, which should not restrict the range of n.

In contrast, this variance does not converge to zero under the multinomial-Dirichlet distribution because

$$\lim_{n\to\infty} Var\left[\frac{N_r}{n}\right] = \lim_{n\to\infty}\left(\frac{n}{n^2}p_r(1-p_r)\frac{\psi+n}{\psi+1}\right) = \frac{1}{\psi+1}p_r(1-p_r). \quad (5)$$

Thus the multinomial-Dirichlet distribution increases the variance of the ballot shares beyond the variance described in the spatial model, by an amount that does not converge to zero as the number of voters becomes infinitely large.

The Spatial Model of P

The requirement that $\Sigma p_r = 1$ implies that all permissible models of P describe mappings from a unit $(m! - 1)$-simplex to the interval $[0, \infty]$. There are several competing models of vote casting for elections with more than two candidates in which voters are asked to rank the candidates. In Tideman and Plassmann (2012), we analyze how well a dozen different models of three-candidate elections describe ranking data from actual elections. We find the spatial model developed in Good and Tideman (1976) far out-performs the other models.

The spatial model assumes that voters care about the "attributes" of candidates; these attributes form a multi-dimensional "attribute space." Every voter has an in-difference map in attribute space, which contains an "ideal point" that describes the quantities of each attribute that the voter's ideal candidate would possess. Actual candidates also possess specifiable quantities of each attribute and therefore have locations in attribute space. We assume that attribute space has at least two dimensions and that the candidates are in "general position," where any slight change in the position of any one candidate does not change the dimensionality of the space that they span, so that the positions of the three candidates in attribute space span an two-dimensional "candidate plane" that is a subspace of attribute space.[4] Voters' indifference maps are defined in candidate space through their definitions in attribute space.

Now consider elections with $m = 3$ candidates. We follow Good and Tideman (1976) and assume that the positions of voters' ideal points in attribute space follow a spherical multivariate normal distribution, which implies

[1] The standard parameterization of hypergeometric distributions is $\pi_N = \{k_1, \ldots, k_{m!}, n\}$, with $\Sigma k_r > n$, so that $E[N_r] = nk_r/\Sigma k$. We reparameterize the k_rs as $k_r = p_r\phi$ with $\phi = \Sigma k_r$.

[2] The standard parameterization of beta and Dirichlet distributions is $\pi = \{k_1, \ldots, k_{m!}\}$ with $E[p_r] = k_r/\Sigma k$. We reparameterize the k_rs as $k_r = p_r\psi$ with $\psi = \Sigma k_r$ which yields $\pi_N = \{p_1, \ldots, p_{m!}, \psi\}$.

[3] Note that we model the distribution of the $m!$ rankings by specifying the expected shares of the rankings among the n ballots (the p_rs), rather than by specifying the probabilities with which individual voters submit these rankings. The expected shares equal these probabilities if indistinguishable voters submit their ballots independently, but a given vector of expected shares is consistent with a range of assumptions about the combination of probabilities and interdependence assigned to voters when there is dependence among the ballots. An alternative approach to model dependence among the ballots is to specify the probabilities with which individual voters submit the rankings, which then imply the expected vote shares. We do not follow this approach because we find it easier to calibrate our model to data from actual elections in terms of observable shares of rankings rather than in terms of unobservable individual probabilities.

[4] The case when all candidates' attributes lie in a single line requires special treatment because not all of the six possible rankings of the candidates occur, but it does not pose conceptual difficulties. See Good and Tideman (1976, pp. 380 – 381) for a description of the general case with $m > 3$.

that the distribution of "relative" ideal points in candidate space is bivariate normal. We further assume that every voter's utility loss from the choice of a particular candidate is the same increasing function of the distance between the candidate's location in candidate space and the voter's relative ideal point in candidate space, so that every voter's indifference surfaces are concentric spheres centered on the voter's ideal point.[5]

Suppose there is a set of candidates for which every voter submits a truthful ranking that reflects his ideal point, his indifference surfaces, and the positions of the candidates. To determine the vote share of each ranking, consider the triangle in the candidate plane that is formed by the locations of the three candidates. We divide the candidate plane into six sectors by drawing the perpendicular bisectors of the three sides of this triangle. These bisectors intersect at the triangle's circumcenter, *T*. For the voters' ideal points in each sector, the distances to the locations of the three candidates have a unique rank order. The integral of the density function of this distribution over each sector is the expected value of the fraction of the voters who rank the candidates in the order corresponding to the sector's rank order.[6] These six integrals determine the vote-share probabilities p_r of the six rankings. Note that even though sectors that are opposite each other have the same angle, they do not have the same integral of the density function (and therefore do not imply the same p_r), unless *O* is not inside either of the sectors and the two lines that form the sectors come equally close to *O*.

The vector of probabilities to be assigned to the six possible rankings of candidates has five degrees of freedom (since the probabilities sum to 1). However, the locations of three candidates in a plane have only four degrees of freedom that affect the probabilities of rankings; one degree of freedom is used because rotations of all candidate positions around the mode of the distribution of voter ideal points leave the probabilities unaffected. Equal movements of the positions of all candidates toward the circumcenter of the triangle formed by their locations also leave the probabilities unaffected, which uses another degree of freedom. As a consequence, the vector of probabilities that the spatial model assigns to the six possible rankings of candidates can take only some of its conceivable values. Thus unlike models with five degrees of freedom, the spatial model is directly falsifiable.[7]

We have found it convenient to use the four degrees of freedom as follows:

- Use one degree of freedom for the distance from the mode of the voters' ideal points to the circumcenter of the triangle formed by the candidates' locations. Call this line segment *L*.
- Use three degrees of freedom for the angles between *L* and the three perpendicular bisectors of the line segments connecting pairs of candidates.

We model the distribution of the length of *L* as a Weibull distribution and the three angles as a multinomial Dirichlet distribution. We chose both distributions because they yield acceptable fits of the model to three different data sets. We calibrate the two parameters of the Weibull distribution so that the mean and standard deviation of the distribution correspond to the observed mean and standard deviation of *L*. We parameterize the Dirichlet distribution by $\lambda_i = \psi\alpha_i$, $i = 1, \ldots, 3$, where the three α_is are the shares of the three pairs of opposite angles, with $\Sigma\alpha_i = 1$, and ψ determines the variance of share *i* as $\alpha_i(1 - \alpha_i)/(\psi + 1)$. To calibrate α_i, we calculate the mean of angle *i* over all elections, and use its share among the three angles as α_i. We calibrate ψ as the mean of the three variance parameters implied by $\alpha_i(1-\alpha_i)/Var[\widehat{share_i}] - 1$, where $Var[\widehat{share_i}]$ is the observed variance of the share of angle *i* over all elections in the data set.

Our calibration of the spatial model to election data includes a finite value for the parameter ψ of the multinomial-Dirichlet model of *N* ($\psi = 330$), implying that the multinomial distribution of rankings is based on probabilities that differ slightly from those prescribed by the spatial model. This means that the probabilities associated with rankings are slightly "lumpy" compared to the smoothness of the spatial model. An alternative interpretation of this over-dispersion is that the rankings of different voters are not independent of one another, but rather are influenced by interactions among the voters. It may be interesting to see if methods can be developed to distinguish between these interpretations.

Calibrating the Spatial Model of P

Our data come from some private elections that Nic Tideman gathered and coded in the late 1980s, and from two sets of political surveys in the U.S. and in Germany. In Plassmann and Tideman (2011) we use these observed data to calibrate the unknown distributional parameters of

[5] None of these assumptions is conceptually necessary and each could be replaced—at a cost of more complex calculations—if there is evidence that it does not represent election data sufficiently well. See Good and Tideman (1976) for a discussion.

[6] In general, evaluating the spatial model for elections with *m* candidates requires numerical integration over *m*! non-central wedges of an (m – 1)-variate normal distribution. Currently we have such an algorithm only for the bivariate standard-normal distribution, which explains why we need to restrict our analysis, for the time being, to three-candidate elections

[7] The model known as the impartial anonymous culture, or IAC, is an example of a model with five degrees of freedom that can describe perfectly every conceivable vector of vote-shares. Such a model is falsifiable only indirectly. Because IAC has no unknown parameters, it is straightforward to use this model to simulate rankings. In Tideman and Plassmann (2012) we assess IAC by comparing ranking data simulated under IAC with observed ranking data from actual elections. We find that the distribution of rankings simulated under IAC is very different from the distribution of observed rankings. Thus despite its ability to describe perfectly every conceivable vector of vote-shares, IAC is not an accurate model of vote casting in actual elections.

the two distributions that describe the spatial model. In particular, we assume that L follows a Weibull distribution with scale parameter α = 0.6858 and shape parameter β = 2.4608, and that the three angles follow a Dirichlet distribution with parameters δ_1 = 26.47, δ_2 = 23.37, and δ_3 = 23.65.[8]

Assessing the ability of the Spatial Model to simulate ranking data with the same properties as observed ranking data

In Plassmann and Tideman (2011) we establish that rankings simulated with these distributions are extraordinarily similar to the rankings in our three data sets. To assess the degree of similarity between simulated and observed data, we begin by noting that for elections with three candidates, a ranking data set contains of a series of six-tuples whose components are the numbers of voters who submitted each of the six possible strict rankings in different elections. We normalize the components of the six-tuple to sum to 1 and refer to the result as a "normalized six-tuple," that is, an element of the unit 5-simplex. Our method of assessing the similarity of different ranking data sets begins with a subset of the unit 5-simplex that does not have full Lebesgue measure relative to the unit 5-simplex. Define such a subset as a "truly proper" subset. One convenient source of such truly proper subsets of the unit 5-simplex is those models of P whose support is less than the full unit 5-simplex, for example, the spatial model. Given a truly proper subset of the unit 5-simplex, we can measure the multi-dimensional Euclidean distance $\hat{\sigma}$ from any normalized six-tuple in a data set to the nearest element of the truly proper subset; this distance is the square root of the sum of the squared differences,

$$\hat{\sigma} = \sqrt{\textstyle\sum_{r=1}^{6}(p_r - \hat{s}_r)^2}, \qquad (6)$$

where $\hat{s}_r$ is the r^{th} element of the observed normalized six-tuple and p_r is the r^{th} element of the nearest six-tuple that is an element of the truly proper subset of the unit 5-simplex. For a model of P, p_r is the vote-share probability of ranking r that the model predicts for this election. If the truly proper subset yields a distribution of distances of the normalized six-tuples of one data set that is different from the distribution for another data set, then we take this as evidence that the two data sets are different. If, on the other hand, the truly-proper subset yields very similar distributions of the distances of the normalized six-tuples in the two data sets, then this supports the hypothesis that these data originated from statistical processes that are similar enough so that analyses on one data set can substitute for analyses on the other. Such a comparison cannot be done with a subset of the 5-simplex that is not a truly proper subset (such as the support of the impartial anonymous culture, or IAC), because all distances in such a comparison would be zero.

Note that it is appropriate to evaluate the data with the same model of P that is used to simulate one of the data sets (that is, use the spatial model to evaluate data generated by the spatial model). Although any model tends to yield short distances to data that were simulated with the model itself, it is irrelevant how short the distances to simulated data are—what is relevant is whether the distances to simulated data and observed data are the same. Thus if a model yields shorter distances for data simulated with this model than for observed data, then this is evidence against the hypothesis that the observed data were generated with this model.

We view each distance $\hat{\sigma}_{ki}$ that we determine from the normalized six-tuple k in data set i as an independent draw from a random variable σ_i. A simple test for evaluating the similarity of the random variables σ_i and σ_j from which data sets i and j were drawn is to compare estimates of their moments, for example, their means. A more rigorous test is to determine the ecdfs of the observed draws from σ_i and σ_j and to use the Kolmogorov-Smirnov (KS) test to assess the likelihood that the two ecdfs represent the cdf of the same σ.[9] If the test suggests that data sets i and j have the same distributions of distances between observed and predicted normalized six-tuples, then this supports the hypothesis that the data sets were generated by mechanisms whose products cannot be distinguished.

An important question is how many elections we need to simulate to be reasonably certain that the KS test is reliable. Because the accuracy of the KS test increases with the number of observations used to derive the ecdfs, it would seem that we should simulate as many normalized six-tuples as possible. However, if the spatial model and the actual data generating mechanism are not identical but only very similar, then simulating too many six-tuples will cause us to ultimately reject the hypothesis that the simulated data and the observed data were generated by the same vote casting mechanism. On the other hand, the fewer six-tuples we simulate, the noisier are the simulated data and thus the p-value of the KS statistic.

Rather than trying to establish an optimal number of simulations that balances the two problems, we adopt the following three-step procedure:

1. Simulate the same number of six-tuples as there are in the observed election data set with which we want to compare the simulated data.

[8] To match the variance of the angles that we determine from observed election data, we parameterize the Dirichlet distribution so that each of the three shares is multiplied by the common constant 73.5008. Dividing the three values in the text by this number yields three shares that sum to 1.

[9] The KS test statistic is the greatest distance between the two ecdfs.

2. Repeat this exercise 999 times.
3. Evaluate the ecdf of the *p*-values that we compute at all iterations.

The more similar the spatial model is to the mechanism that has generated the observed data, the closer is the distribution of *p*-values to a uniform distribution and the closer is the ecdf to a straight line.[10] The advantage of using this procedure is that simulating a relatively small number of vote-share vectors prevents us from rejecting data that are "reasonably similar," while the repetitions ensure that we are not misled by an unusually small or large *p*-value. This procedure also solves the problem of which number of voters to assume for our simulations: if we simulate the same number of elections as there are in the data set, then we can use the observed numbers of voters for our simulated elections and thereby ensure identical distributions of voters in the observed and simulated elections.

In Plassmann and Tideman (2011) we establish that the spatial model fits the data set derived from actual elections extraordinarily well, and that it fits the two data sets that we compiled from surveys reasonably well. It would be worthwhile to examine whether a different use of the degrees of freedom and/or the use other distributions to characterize data might yield an even better fit. In addition, it will also be valuable to locate additional data and check whether the spatial model and our calibration describe other data sets equally well.

Things to Do with a Statistical Model of Vote Casting in Three-Candidate Elections

One can use our statistically valid model of election outcomes to investigate any number of characteristics of elections, through simulations. For example, one can investigate the frequency of majority-rule cycles and other voting anomalies like a voting rule's failure to elect a Condorcet winner when one exists. There are not nearly enough ranking data from actual elections to investigate these phenomena empirically using only the observed data. However, there are enough ranking data from actual elections to infer the parameters of a statistical model of vote casting. Thus simulations permit analyses of the frequencies of low-frequency phenomena without the need for massive amounts of data from real elections. To be sure, using simulations for the analysis of such low-frequency phenomena is not a new idea, and simulations have been used for this purpose since at least the 1960s (see, for example, Campbell and Tullock, 1965). However, the results of such inquiries depend greatly on the model of vote casting that is used to simulate the ballots, and different models yield very different outcomes. What is new in our approach is to first identify a model that is capable of simulating ballots with the same statistical structure as actual ballots, so that the simulations provide information about what one should expect to happen in actual elections.

To demonstrate the power of statistical inquiries to contribute to computational social choice, we analyze the question of which voting rule is most likely to select the best candidate. It might seem that such an inquiry cannot possibly succeed, because if there were a way of establishing with certainty which candidate was best, then we would not have political disputes. However, we do not propose to determine which candidate is best in any particular election. Rather, by simulating the election process, we estimate the frequencies with which different voting rules select the best candidate. While there are different ways of defining who the best candidate is, we define the best candidate as the one closest to the mode of the voter ideal points. One way of justifying this definition is by assuming that voters seek to identify the candidate whom they consider best from a social perspective, and that the voters' perceptions are correct on average. Because we know the parameters that we used to simulate each election, we know which candidate is closest to the mode of the voters' ideal points and is therefore best. We then simulate a large number of elections, evaluate each election with multiple voting rules, and thus estimate, with acceptable degrees of accuracy and precision, the probability with which each voting rule identifies the best candidate.

The Probabilities with which Fourteen Voting Rules Identify the Best Candidate

Because random fluctuations are more prominent in small samples, voting rules tend to become more accurate as the number of voters increases. When the number of voters is small, the accuracy of a voting rule also depends on whether the number of voters is odd or even. Therefore we simulated elections with 10, 100, 1,000, 10,000, and 100,000 voters, and also with 11, 101, 1,001, 10,001, and 100,001 voters. For each election size, we simulated 1,000,000 elections.

We investigate 14 voting rules, of which seven are Condorcet consistent (electing the Condorcet winner—the candidate who beats all other candidates in pairwise comparisons—if there is one) and seven are not:

[10] If one examines two sets of data that are realizations of the same distribution, then one would expect that in 1,000 repetitions the *p*-value of the KS test statistic exceeds, say, 0.5 about 500 times and 0.95 about 50 times. Thus the cdf of the *p*-values is a straight line from [0, 0] to [1, 1]. Note that the uniform distribution is a limit result that holds only if both data sets accurately represent their underlying populations. If, for example, the observed data do not represent their underlying population perfectly, then the ecdf of *p*-values will deviate from a straight line even if the simulated data are drawn from the same population.

A. Seven voting rules that are not Condorcet consistent:

1. The Alternative Vote
Score each candidate by the number of voters who rank the candidate first. A candidate wins if he receives a majority of the votes. If not, eliminate the candidate with the lowest score and restart the count on the reduced set of candidates.

2. The Anti-plurality rule
Score each candidate by number of times the candidate is ranked last. The candidate with the lowest score wins.

3. The Borda rule
Score each candidate by the sum, over all other candidates, of the number of voters who favor the candidate in a pairwise comparison. The candidate with the greatest score wins.

4. The Bucklin rule
Score each candidate by the number of voters who rank the candidate first. If a candidate has a majority, that candidate wins. If not, then add to each candidate's score the number of voters who rank the candidate second. Continue in this way until at least one candidate has a majority. At that point, the candidate with the greatest score wins.

5. The Coombs rule
If a candidate wins is ranked first by a majority of the votes, that candidate wins. If not, score each candidate by the number of voters who rank the candidate last and eliminate the candidate with the greatest score. Restart the count on the reduced set of candidates.

6. The Estimated Centrality rule
Use the spatial model to locate the boundaries between voters with different rankings of the candidates. The winner is the candidate at the top of the ranking representing the preferences of the voters at the mode of the distribution of voters' ideal points.

7. The Plurality rule
Score each candidate by the number of times the candidate is ranked first. The candidate with the greatest score wins.

B. Seven voting rules that are Condorcet consistent

1. The Alternative Schwarz rule
Start with the Schwartz set, that is, the union of smallest sets of candidate, none of which are beaten in pairwise comparisons by any candidate outside the set. Eliminate all candidates not in the Schwartz set. If the Schwartz set contains more than one candidate, eliminate the candidate with the fewest first-place votes, considering only the candidates in the Schwartz set. If there is still more than one candidate, restart the count on the reduced set of candidates. When all candidates but one have been eliminated, the winner is the candidate who remains.

2. The Alternative Smith rule
Like Alternative Schwartz, but using the Smith set, that is, the smallest set of candidates that beat every candidate outside the set in pairwise comparisons.

3. The Black rule
If a Condorcet winner exists, that candidate wins. If there is no Condorcet winner, then the winner is the candidate with the greatest Borda score.

4. The Copeland rule
Score each candidate by the difference between the number of candidates that the candidate beats in pairwise comparisons and the number of candidates that beat the candidate. The candidate with the greatest score wins.

5. The Dodgson rule
Score each candidate by the number of pairs of adjacent candidates on individual ballots that must be reversed to make the candidate a Condorcet winner. The candidate with the lowest score wins.

6. The Kemeny rule
Score each ranking of the candidates according to the total number of pairs of candidates on all ballots that are ranked in the same order as in the ranking that is being scored. The candidate at the top of the ranking with the greatest score wins.

7. The Nanson rule
Score all candidates by their Borda scores. Eliminate all candidates with scores less than or equal to the average. If there is more than one candidate left, restart the count on the reduced set of candidates.

Tables 1 and 2 show the results. Some conclusions that can be drawn from the tables are:

- The number of voters has a very large effect on the probability that any voting rule will select the best candidate.
- Whether the number of voters is even or odd has a negligible effect on the probability that any voting rule will select the best candidate, except in elections with very few voters. All voting rules are somewhat more likely to elect the best candidate in elections with 11 voters than they are in elections with 10 voters.
- The anti-plurality rule is by far the worst of the rules examined. Unless the number of voters is small, the Bucklin rule is nearly as bad.
- The other 12 rules are fairly similar, although there are notable systematic differences.

Table 1. Frequencies with which 14 voting rules fail to select the best option when the number of voters is even

	10	100	1,000	10,000	100,000
Alternative Vote	30.52%	11.64%	6.42%	5.69%	5.56%
Anti-plurality	37.21%	20.37%	16.21%	15.65%	15.62%
Borda	29.17%	11.65%	6.84%	6.20%	6.07%
Bucklin	31.62%	16.59%	13.19%	12.76%	12.76%
Coombs	30.29%	11.77%	6.53%	5.78%	5.67%
Estimated Centrality	29.44%	11.06%	6.20%	5.37%	5.30%
Plurality	30.14%	11.79%	6.82%	6.20%	6.09%
Alternative Schwartz	30.22%	11.61%	6.38%	5.65%	5.52%
Alternative Smith	30.51%	11.62%	6.38%	5.65%	5.52%
Black	29.06%	11.41%	6.37%	5.64%	5.51%
Copeland	29.67%	11.60%	6.39%	5.65%	5.52%
Dodgson	29.66%	11.57%	6.37%	5.64%	5.51%
Kemeny	30.23%	11.58%	6.37%	5.64%	5.51%
Nanson	29.65%	11.56%	6.37%	5.64%	5.51%

Table 2. Frequencies with which 14 voting rules fail to select the best option when the number of voters is odd

	11	101	1,001	10,001	100,001
Alternative Vote	28.43%	11.59%	6.42%	5.68%	5.59%
Anti-plurality	36.21%	20.31%	16.26%	15.67%	15.57%
Borda	28.15%	11.69%	6.84%	6.20%	6.08%
Bucklin	29.26%	16.28%	13.18%	12.81%	12.69%
Coombs	28.40%	11.73%	6.53%	5.78%	5.70%
Estimated Centrality	27.93%	11.09%	6.11%	5.39%	5.32%
Plurality	29.13%	11.80%	6.82%	6.18%	6.13%
Alternative Schwartz	28.38%	11.59%	6.39%	5.64%	5.55%
Alternative Smith	28.38%	11.59%	6.39%	5.64%	5.55%
Black	28.14%	11.53%	6.38%	5.63%	5.54%
Copeland	28.39%	11.60%	6.40%	5.64%	5.55%
Dodgson	28.25%	11.54%	6.38%	5.63%	5.54%
Kemeny	28.25%	11.54%	6.38%	5.63%	5.54%
Nanson	28.13%	11.54%	6.38%	5.63%	5.54%

- Except when the number of voters is small, the Estimated Centrality rule has the greatest probability of selecting the best candidate. However, its suitability as a voting rule is limited both by the fact a computer is required to calculate the winner and by the fact that at present there is no algorithm for applying the rule to elections with more than three candidates.
- For elections with 1,000 or more voters, all Condorcet consistent voting rules have higher probabilities of electing the best candidate than any of the non-Condorcet consistent rules other than Estimated Centrality.

We emphasize that the estimated probabilities of selecting the best candidate are not the sole criterion by which to select a voting rule. It is also necessary to consider such factors as the understandability of the rule for voters and the susceptibility of the rule to strategic manipulation.

We hope that this illustration of the potential of the empirical side of computational public choice will motivate others both to find data sets with which to determine whether our results hold more generally and to expand the range of public choice questions that are explored with empirical methods.

References

Campbell, C., and Tullock, G. 1965. “A measure of the importance of cyclical majorities.” *Economic Journal* **75,** 853 – 857.

Good, I.J. and Tideman, T.N. 1976. “From individual to collective ordering through multidimensional attribute space.” *Proceedings of the Royal Society of London* (Series A) **347**, 371 – 385.

Johnson, N.L, Kotz, S. and Balakrishnan, N. 1997. *Discrete Multivariate Distributions*. New York: Wiley.

Plassmann, F. and Tideman, T.N. 2011. “How to assess the frequency of voting paradoxes and strategic voting opportunities in actual elections.” Mimeo. A copy is available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1911286.

Tideman, T.N. and F. Plassmann. 2011. “Which voting rule is most likely to choose the best option?” available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1939105

Tideman, T.N. and Plassmann, F. 2012. “Modeling the outcomes of vote-casting in actual elections.” In: Dan Felsenthal and Moshé Machover (eds.) *Electoral Systems: Paradoxes, Assumptions, and Procedures*, Springer (forthcoming). A copy is available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1627787.