Evaluating Voting Methods by their Probability of Success:

An Empirical Analysis

T. Nicolaus Tideman
Department of Economics, Virginia Polytechnic Institute and State University
Blacksburg, VA 24061
ntideman@vt.edu

Florenz Plassmann*
Department of Economics, State University of New York at Binghamton
Binghamton, NY 13902-6000
fplass@binghamton.edu

This version: February 10, 2008

Abstract:

Voting methods for elections with more than two candidates have traditionally been evaluated in terms of their logical properties. We propose that voting methods should be evaluated instead by identifying the statistical properties of the process by which election outcomes are generated and then finding the method with the best success in identifying the proper winner. We consider four intuitive models of voter behavior that might govern this statistical process. From an empirical analysis of "election-like" data we conclude that the spatial model of voting provides the best description of the process by which election outcomes are generated. The "Estimated Centrality" voting method that was proposed by Good and Tideman (1976) and which is derived from the spatial model of voting is most likely to identify the best candidate, at least when the electorate is reasonably large and voters do not behave strategically.

Journal of Economic Literature Classification Codes: C4, D72

Keywords: Borda, Condorcet, Estimated Centrality, voter behavior, Pólya distribution

---

* Corresponding author. Phone: (607) 777-4304, Fax: (607) 777-2572. Some results are preliminary and subject to refinement.

1. INTRODUCTION

What method should we use to determine the winner in elections with three or more candidates? Voting theorists have proposed a large number of competing methods, which sometimes select different candidates as the winner. The customary way that these methods are evaluated is by comparing their logical properties. Because no method possesses all properties that are desired and there is no consensus on which properties are most important, there is also no consensus on which voting method is best. In this paper, we propose a different strategy for comparing voting methods. The existing literature on voting consists almost exclusively of theoretical analyses of voting methods, and references to actual elections are made only to illustrate voting paradoxes. We suggest that it would be more fruitful to begin instead with an analysis of actual elections, and to draw inferences about the relative attractiveness of different voting methods on the basis of a tested model of the statistical process that generates election outcomes.

To determine what such a statistical analysis has to offer, we need to define what makes a voting method attractive. While there is no agreement on which properties are most important, a property that would seem highly desirable is the ability to identify the "correct" or "best" outcome."[1] Beginning with the Marquis de Condorcet, various authors have interpreted voting methods as maximum likelihood estimators (MLEs) of the correct outcome.[2] They discovered that, because different voting methods yield the MLE depending on circumstances, no single method is most likely to identify the correct outcome in all circumstances. However, instead of analyzing the theoretical circumstances under which voting methods fail to identify the correct outcome, it will be more useful, we argue, to estimate how frequently different methods fail to

[1] See, for example, Young (1988, 1995), Risse (2001, 2005), Saari (2003, 2006).

[2] See, for example, Young (1986), Conitzer and Sandholm (2005), Drissi-Bakhkhat and Truchon (2004), and Truchon (2006). Young (1995) motivates the maximum likelihood method as a form of compromise that does not require the existence of a correct outcome.

identify the correct outcome in *actual* elections. The goal of such an analysis is to find a voting method that is *on average* more likely than other methods to identify the correct outcome. A group of voters who used such a method in all elections could thus expect to make more correct choices over all elections than if they used other voting methods, even if their method was not necessarily the best in every single election. Such a voting method would arguably be more attractive than the others.

We start our analysis with the premise that there is a correct or best outcome, and that the purpose of voting is to identify this outcome. We acknowledge that this premise may be controversial. Voting can be meaningful even in the absence of a correct outcome if it helps voters make collective decisions in ways that they find satisfactory. It is also not obvious what "the correct outcome" means. If consensus on the correct outcome could be reached, then there would be no need to vote. Still, voters might agree that the outcome is best that leads to the lowest collective utility loss among the voters, but not agree on which outcome this is.

The notion that there exists a correct outcome has been part of the analysis of voting for a very long time, for example, in the literature on Condorcet's Jury Theorem, and we view our analysis as a part of this tradition. The outcome does not need to be objectively correct; it suffices that the voters agree on the criterion that defines the best outcome. We therefore treat the terms "correct outcome" and "best outcome" as synonyms.

Determining which voting method is on average most likely to identify the correct outcome requires that we discover the statistical process that generates election outcomes. As a simple example of how this task can be approached, consider a three-candidate election in which voters are asked to rank the candidates. Tallying the ballots yields a number of votes for each of the six possible rankings. These six numbers can be viewed as one realization of a statistical model that specifies the probabilities of observing all possible combinations of frequencies for the six rankings.

Viewed from this perspective, the task of finding the best voting method is analogous to the task of determining the best estimator of an unknown statistical parameter in the analysis of observational data. In our case, the unknown parameter specifies the correct winner. Previous theoretical analyses of voting are thus comparable to the derivation of the theoretical properties of different statistical estimators. In the same way in which different estimators may each be optimal for data from a particular statistical model, different voting methods may each be optimal for voting data that follow different statistical models of the vote-casting process. Thus it is not all that surprising that theoretical analyses of voting have been unable to identify an objectively best voting method, since these analyses have not been linked to empirical inquiries into statistical properties of the vote-casting process.

The first step in the analysis of observational data is to determine the statistical model that is most likely to have generated the available data. The search for the best estimator can begin only after this model has been identified. Our first task is therefore to determine a statistical model of voter behavior that can be regarded as generating the outcomes of the vote-casting process in elections. Once we have specified the statistical model that describes the frequencies of the rankings as a function of the circumstances in a specific election and the frequencies of different electoral circumstances, we can then devise a method that is most likely to identify the correct winner as a function of the observed frequencies of the rankings. Our surprisingly strong results indicate that this line of inquiry is worth pursuing.

We call the model that governs the structure of the probabilities in such a statistical model a model of voter behavior. While there are many theoretically plausible models of voter behavior, we consider four models to be interesting candidates for our initial inquiry. First, it may simply be incorrect to assert that in any given election, some rankings of candidates are systematically more likely to occur than others. The corresponding model of voter behavior assumes that in every election, all rankings are equally likely, which implies that voting results contain no information that is useful for identifying the correct winner. Analyzing election data with this model therefore provides information about whether our proposed strategy of inquiry has any merit. If there is evidence that voters discriminate among rankings in systematic ways, then it is meaningful to ask whether their voting behavior makes any of the known voting methods more likely than others to identify the correct winner. The Borda method and the Condorcet-Kemeny-Young method (henceforth "Condorcet method") are arguably the most widely discussed voting methods, and we consider two models of voter behavior for which these methods are MLEs. Alternatively, the widely-used spatial model of voting (see Enelow and Hinich, 1984 and 1990) may describe voter behavior best. This model provides the motivation for Good and Tideman's (1976) voting method, which Tideman (2006) calls the "Estimated Centrality" method. Currently we are only able to evaluate the Estimated Centrality method for elections with three candidates, so we restrict our empirical analysis to such elections and leave the extension to elections with more than three candidates for future research.

In the remainder of this paper, we assess the validity of these four models of voter behavior, using data from 913 "elections" that we construct from the thermometer scores that are part of the surveys conducted by the American National Election Studies (ANES) at the University of Michigan. Although it is possible that such survey data differ from election data in systematic ways, an important advantage of using survey rather than actual election data is that survey respondents are less likely to have any reason to report rankings that differ in systematic ways from their true rankings. Thus our data set is likely to provide us with information about voter behavior that is little affected by strategic considerations.

We emphasize that we do not evaluate any of these voting methods according to their resistance to strategizing. The incorporation of estimates of resistance to strategizing into the analysis may lead to different conclusions about the relative attractiveness of voting methods. Saari (1990) proposes ways of assessing a voting method's theoretical susceptibility to strategizing, but he undertakes his analysis under the assumption that all voting outcomes have the same probability of occurrence. Our analysis indicates that assessing a voting method's resistance to strategizing on the basis of a model of the vote-casting process developed from actual election data is likely to provide further insights.

In Sections 2 and 3, we develop the statistical framework, the four models of voter behavior, and our strategy for model assessment. We describe our data and report the results of our statistical analysis in Section 4. We find strong evidence against the hypothesis that all rankings are equally likely in every election. We find strong evidence in favor of the hypothesis that the probabilities of rankings are generated by the combination of the spatial model of voting and a specified error structure. In Section 5, we use the three models of the vote casting process that correspond to the Estimated Centrality, the Borda, and the Condorcet methods to simulate artificial elections, and we assess the frequencies with which different voting methods determine the correct winner. In addition to the three voting methods already mentioned, we evaluate the Instant-Runoff method, the Maximin method, and the Plurality method. We find that the accuracy of each method depends strongly on the number of voters, far more so than it depends on the generating mechanism. However, for any given number of voters there are significant differences in the accuracy of different voting methods, so it is valuable to take account of accuracy when choosing a voting method. In general, the Plurality method performs worst of all and the Instant-Runoff method performs notably worse than the other four methods. While the differences among the other four methods are small, many differences are statistically significant.

2. A Statistical Model of the Outcome of the Vote-Casting Process

Consider an election with M candidates in which N voters are asked to each submit a ranking of the candidates. There are $M!$ possible strict rankings. If p_r is the probability that a voter submits ranking r, $r = 1, \ldots, M!$, and $\Sigma p_r = 1$, then we can view the number of votes for ranking r, N_r, as a random variable whose statistical model describes the probabilities of the different possible outcomes of the vote-casting process. This statistical model can be specified through the joint distribution of the N_rs. Assume first that the p_rs are deterministic, that they are the same for all voters, and that all voters submit their rankings independently.[3] Then the N_rs follow a multinomial distribution with density function[4]

$$f(N_1,\ldots,N_{M!};N,p_1,\ldots,p_{M!}) = \prod_{r=1}^{M!} p_r^{\,N_r}\,\frac{N!}{\prod_{r=1}^{M!} N_r!}, \tag{1}$$

whose first two moments are $E[N_r] = Np_r$, $\mathrm{Var}[N_r] = Np_r(1 - p_r)$, and $\mathrm{Cov}[N_r, N_s] = -Np_rp_s$.

Note that we model the outcome of the vote-casting process in terms of the probabilities with which a voter submits any of the possible rankings, and not in terms of the probabilities that a voter ranks a particular pair of candidates one way or the other. The latter has been the standard framework for evaluating voting methods as maximum likelihood estimators since

---

[3] We relax the assumption that the ps are deterministic in Section 3.5.

[4] Generally, the multinomial distribution describes the probabilistic structure of any series of independent and identical Bernoulli trials with constant probabilities and multiple possible outcomes.

Condorcet's *Essai*. A frequently made assumption in this framework is that these probabilities are independent across pairs of candidates, which permits voters to have non-transitive preferences.[5] In contrast, modeling the probabilities of submitting different rankings of all of the candidates imposes transitivity on the voters' preferences naturally, and is therefore more likely to reflect the actual statistical model that governs the outcome of the vote-casting process.

Now assume that there is a correct ranking r^* and that the highest ranked candidate in r^*, candidate m^*, is the correct winner. We incorporate the notions of a correct ranking and correct winner into the statistical model of the N_rs through a model of voter behavior that describes the probabilities with which voters chose any of the rankings as functions of r^* and m^*. We consider four models of voter behavior. The first model describes the possibility that voters do not systematically discriminate among the candidates, so that all rankings of candidates are equally likely. The next two models are inspired by the Borda method and the Condorcet method respectively, and each model is specified by a single parameter whose value can vary across elections. Our fourth model of voter behavior is a spatial model of voting, which provides the motivation for Good and Tideman's (1976) "Estimated Centrality" voting method, and which can be specified by four parameters for any given election. We motivate and describe the four models as well as our strategies for their assessment in the following section.

3. Four models of Voter behavior

3.1. *Equal probabilities*

The simplest model of voter behavior assumes that all rankings are equally likely, or $p_r = 1/(M!)$. The voting literature refers to this assumption as the "impartial culture condition," and voting

---

[5] See Young (1995, p. 55) and Drissi-Bakhkhat and Truchon (2004, p. 167).

theorists frequently use it to calculate the likelihood that certain voting events will occur. For example, Saari (1990) uses this assumption to analyze the likelihood of strategic voting under different voting methods, Gherlein (2002) uses it to analyze the likelihood of observing Condorcet's paradox, and Cervone *et al*. (2005) use it to analyze the likelihood that a Condorcet candidate, if it exists, will win the election. Although voting theorists generally emphasize that they do not necessarily believe that equal probabilities describe voter behavior better than any other model does, the frequent use of this assumption suggests that it is informative to analyze whether this model is at all likely to describe actual voter behavior.

If the probabilities that voters cast their ballots for either the correct ranking or the correct winner do not differ from the probabilities that voters choose any of the other rankings and options, then voting is at best a useful mechanism for making a collective decision but not a useful mechanism for identifying the correct outcome. If the model of equal probabilities describes voter behavior best, then any attempt to determine the voting method most likely to identify either the correct ranking or the correct winner will be futile. Thus the model of equal probabilities represents a benchmark against which we measure the other three models.

3.2. *A model of voter behavior inspired by the Borda method*

When a voter ranks a candidate in position k, $k = 1, \ldots, M$, the Borda method assigns $M - k$ points to this candidate. Candidate m's Borda score is the sum of the points that m obtains from all voters, and the candidate with the highest Borda score is the Borda winner. An alternative way of calculating m's Borda score is to consider the $M - 1$ pairs of candidates that involve m, and for each pair give m one point for each ballot that ranks m ahead of the other candidate.

To derive a model of voter behavior that justifies the Borda method, we define the probabilities of the rankings, p_r, $r = 1, \ldots, M!$, in such a way that the Borda winner is a maximum likelihood estimate of the correct candidate. Consider any two rankings r and s in which the correct candidate m^* is ranked in positions k_r and k_s respectively. If a voter submits ranking r rather than ranking s, then the difference in the points that the Borda method assigns to candidate m^* is $k_r - k_s$. Recall our assumptions that the p_rs are the same for all voters and that the N voters submit their rankings independently. For the Borda method to be a maximum likelihood estimator of m^*, the difference in the log of the likelihood function from submitting ranking r rather than s must therefore be proportional to the difference in points that the Borda method assigns to m^* under the two rankings, or

$$\ln p_r - \ln p_s = \alpha (k_r - k_s) \tag{2}$$

for some constant α. This implies that

$$p_r = c_1 e^{\alpha k_r} \tag{3}$$

with

$$c_1 = 1 / \left((M-1)! \sum_{m=0}^{M-1} e^{m\alpha} \right) \tag{4}$$

to ensure $\Sigma p_r = 1$. The assumption that the p_r's are the same for all voters implies that α is the same for all voters. To confirm that such probabilities imply that the candidate with the greatest Borda score is most likely to be the correct winner, denote the ranking supplied by voter n by $r(n)$, and compute the likelihood function LH_B of the N rankings as

$$LH_B \propto \prod_{n=1}^{N} p_{r(n)} = \left(c_1^N \right) e^{\alpha \sum_{n=1}^{N} k_{r(n)}} \tag{5}$$

where $\Sigma k_{r(n)}$ is candidate m^*'s Borda score. The candidate with the highest Borda score is therefore the maximum likelihood estimate of the correct candidate m^*.

Table 1 shows the p_rs required by the Borda model of voter behavior, for an election with three candidates A, B, and C and a given value of α. The entry in each row of the final column is the probability of the rankings in the three preceding columns, conditional on the correct candidate being the one listed in the respective column heading. Because the Borda method assigns scores to candidates according to their position, it assigns the same probability to all rankings that rank the correct candidate in the same position.

Table 1. Probabilities of rankings for the statistical model inspired by the Borda method

Correct candidate			
A	B	C	
Ranking	Ranking	Ranking	Probability
ABC, ACB	BAC, BCA	CAB, CBA	$p_r = c\ e^{2\alpha}$
BAC, CAB	ABC, CBA	ACB, BCA	$p_r = c\ e^{\alpha}$
BCA, CBA	CAB, ACB	BAC, ABC	$p_r = c$

We calibrate the Borda model to observed election data by applying the Borda voting method. The evaluation of a single election provides one estimate for α as the value that minimizes the sum of squared differences between the observed and the predicted shares of the M! rankings. Application of the Borda method to multiple elections yields multiple estimates of α, which we use to calibrate the distribution of α. For our Monte Carlo analysis in Section 5, we assume that α follows a gamma distribution with the same mean and variance as the estimates of the αs.

3.3. *A model of voter behavior inspired by the Condorcet method*

While the Borda method assigns a score to each candidate, the Condorcet method assigns a score to each of the $M!$ possible rankings. To assign a score to ranking r, the Condorcet method counts, for each ballot, the number of pairs of candidates that are ranked on the ballot in the same way as in ranking r. The sum over all N ballots of the number of pairs ranked the same on the ballot as in r is ranking r's Condorcet score. The ranking with the highest Condorcet score is the winning ranking by the Condorcet method, and the winning candidate is the candidate at the top of the winning ranking.[6]

To derive a model of voter behavior that justifies the Condorcet method, we define the probabilities of the rankings p_r, $r = 1, \ldots, M!$, in such a way that the Condorcet method is a maximum likelihood estimator of the correct candidate. Define n_{sr} as the number of pairs of candidates that are ranked the same in ranking s as in ranking r. Consider any two ballots whose rankings, r and s, agree with the correct ranking, r^*, on n_{rr*} and n_{sr*} pairs respectively. If a voter submits a ballot with ranking r rather than s, then this voter's contribution to ranking r^*'s Condorcet score changes by $n_{rr*} - n_{sr*}$. For the Condorcet method to be a maximum likelihood estimator of r^*, the difference in the log of the likelihood function from submitting ranking r rather than s must be proportional to the difference between n_{rr*} and n_{sr*}, or

$$\ln p_r - \ln p_s = \alpha\,(n_{rr*} - n_{sr*}) \tag{6}$$

for some constant α. This implies that

[6] See Kemeny (1959) and Young (1988).

$$p_r = c_2\, e^{\alpha n_{rr^*}} \tag{7}$$

with

$$c_2 = 1 \Big/ \left(\sum_{m=0}^{\frac{(M-1)M}{2}} f(m,M) e^{m\alpha} \right) \tag{8}$$

to ensure $\Sigma p_r = 1$, where $f(m, M)$ is the frequency distribution of Kendall's τ.[7] Thus if we denote the ranking on voter n's ballot by $r(n)$, then the likelihood function LH_C of the N rankings is

$$LH_C \propto \prod_{n=1}^{N} p_{r(n)} = \left(c_2^N\right) e^{\left(\alpha \sum_{n=1}^{N} n_{r(n)r^*}\right)} \tag{9}$$

where $\Sigma n_{r(n)r^*}$ is ranking r^*'s Condorcet score. The candidate with the highest Condorcet score is therefore the maximum likelihood estimate of the correct ranking r^*.

Table 2 shows the p_rs of the Condorcet model of voter behavior for an election with three candidates A, B, and C. As before, the entry in each row of the final column is the probability of

Table 2. Probabilities of rankings for the statistical model inspired by the Condorcet method

Correct ranking						
ABC	ACB	CAB	CBA	BCA	BAC	
Ranking	Ranking	Ranking	Ranking	Ranking	Ranking	Probability
ABC	ACB	CAB	CBA	BCA	BAC	$p_r = c\ e^{3\alpha}$
BAC, ACB	ABC, CAB	ACB, CBA	CAB, BCA	CBA, BAC	BAC, ABC	$p_r = c\ e^{2\alpha}$
BCA, CAB	BAC, CBA	ABC, BCA	ACB, BAC	CAB, ABC	CBA. ACB	$p_r = c\ e^{\alpha}$
CBA	BCA	BAC	ABC	ACB	CAB	$p_r = c$

[7] See Kendall and Gibbons (1990, pp. 91-92)

the rankings in the preceding columns, conditional on the ranking with the greatest probability of being selected by voters being the one listed in the respective column heading.

3.4. *A spatial model of voter behavior*

Assume that voters care about the "attributes" of candidates. These attributes form a multi-dimensional "attribute space." Every voter has an indifference 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 voters agree on the locations of the actual candidates. If attribute space has at least $M – 1$ dimensions and the candidates are in "general position," where a slight change in the position of any one candidate does not change the dimensionality of the space that they span, then the positions of the M candidates in attribute space span an $M – 1$ dimensional "candidate space" that is a subspace of attribute space. Voters' indifference maps are defined in candidate space through their definition in attribute space.

To complete the model, we need to specify the distribution of the voters' ideal points and the shapes of the voters' indifference maps. We follow Good and Tideman (1976) and assume that the positions of voters' ideal points in attribute space follow a spherical multivariate normal distribution and that these positions are independent of each other, which implies that the "relative" ideal points in the candidate space are spherical multivariate normal as well.[8] 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

[8] A copy of Good and Tideman (1976) is available at http://bingweb.binghamton.edu/~fplass/GoodTideman.pdf.

concentric spheres centered on the voter's ideal point. None of these assumptions is conceptually necessary and each could be replaced—at a cost of more complex calculations—if there is evidence that is does not represent observed voting data sufficiently well.[9]

We define the correct winner m^* as the candidate whose election leads to the lowest collective utility loss among the voters, and we assume that the collective utility loss is the sum of the losses of the individual voters. In conjunction with the assumptions about the voters' ideal points and preferences, this implies that the mode of the distribution of ideal points in candidate space represents the socially most attractive location for a candidate and that the best candidate among those who are available is the one who is closest to the mode.

Now assume that 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. The expected share of votes that each ranking receives is the integral of the density of voter ideal points over the region of the candidate space in which voters choose this ranking. Because it is straightforward to illustrate the computation of vote shares graphically for the case of three candidates and because we will restrict our empirical analysis to elections with three candidates, we describe this computation for $M = 3$, where the candidates' positions in attribute space span a two-dimensional "candidate plane."[10] In this case, our assumptions about the distribution of voters' ideal points and the loss of utility in terms of the distance from a voter's ideal point to a candidate's location imply that the relative ideal points in the candidate plane have a circularly

[9] See Good and Tideman (1976) for a discussion.

[10] The case when all candidates' attributes lie in a single line requires special treatment because not all of the 6 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 $M > 3$.

symmetric bivariate normal distribution, and that every voter's indifference contours are concentric circles around his ideal point.[11]

To determine how voters rank the three candidates, consider the triangle in the candidate plane that is formed by the locations of the three candidates, A, B, and C. 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, *P*. For the voters' ideal points in each sector, the distances to the locations of the three candidates have a unique rank order.

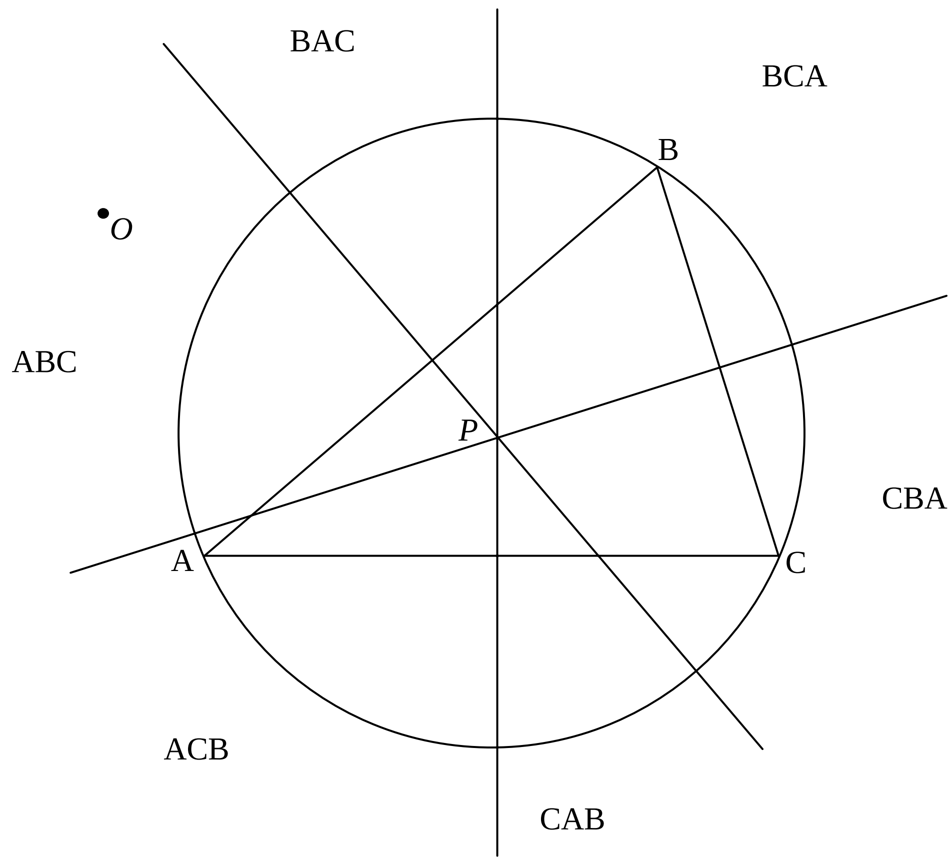

Figure 1. Division of the candidate plane into six sectors by drawing the perpendicular bisectors of the three sides of the triangle formed by the candidates' locations, and the associated rank orders of the sectors. (The figure is taken from Good and Tideman, 1977, p. 372.)

[11] See Good and Tideman (1976, p. 374).

These rank orders are indicated in Figure 1, together with the mode of the circular bivariate normal distribution at *O*. 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. These six integrals determine the probabilities p_r, $r = 1, \ldots, 6$ of the six rankings in the statistical model of the outcome of the vote casting process with 3 candidates. 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 sector and the two lines that form the sectors come equally close to *O*.

The rank order of the sector that contains the mode of the bivariate normal distribution describes the correct ranking r^*, and the candidate whose election leads to the lowest collective utility loss among the voters (the correct winner, m^*) is the highest ranked candidate in r^*. Thus the Estimated Centrality method, using observable vote shares of the six rankings to identify the sector that contains the mode of the bivariate normal distribution, is a voting method inspired by the spatial model. The challenge is to identify locations for the borders between pairs of adjacent rankings that create sectors that match the six probabilities p_r as closely as possible to the six observed vote shares, $q_r = \hat{N}_r / N$, where $\hat{N}_r$ is the observed number of votes for ranking *r*. Because the construction of the relative locations from the observed q_rs has only four degrees of freedom while the derivation of the p_rs has five degrees of freedom, a perfect match is generally not possible.[12] Figure 2 shows one way of using the four degrees of freedom. The intersection

[12] The vote shares p_rs, which are defined relative to a given set of candidates' locations, have five rather than six degrees of freedom because they are independent of rotations around the mode *O*. The relative locations of the candidates that one can derive from the q_rs have only four degrees of freedom because they are not only independent of rotations around *O* but also independent of changes that move the locations of all candidates proportionately along rays emanating from the triangle's circumcenter.

of the perpendicular bisectors *P* is placed at the origin of a Cartesian coordinate system. The fact that the vote shares are independent of rotations around the mode of the distribution of voters' ideal points, *O*, permits us to rotate the coordinate system so that *O* is located on its horizontal axis. The first degree of freedom then specifies the distance between *P* and *O*. The remaining degrees of freedom specify the angles β_1, β_2, and β_3 formed by the line $\overline{PO}$ and the three perpendicular bisectors. Thus any set of values of the four degrees of freedom corresponds to a set of p_rs.[13]

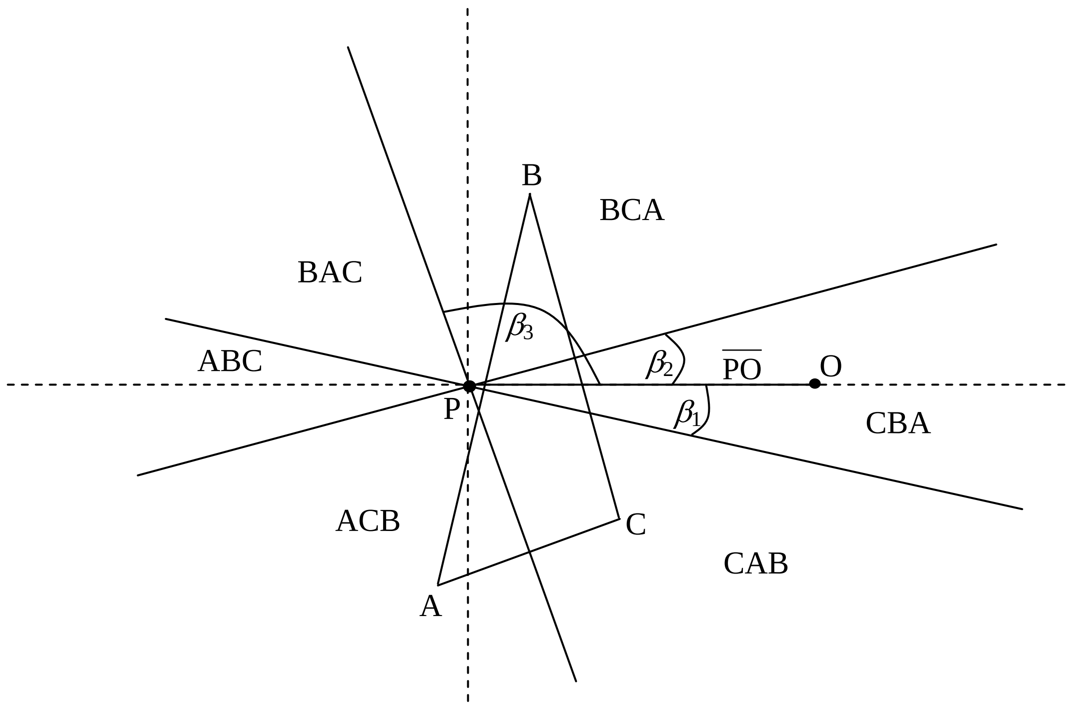

Figure 2. The four parameters $\overline{PO}$, β_1, β_2, and β_3 that define the spatial model.

We find the best match of the p_rs and q_rs by identifying the angles of the perpendicular bisectors and the location of the mode of the voters' ideal points that minimize the sum of the

[13] We use the algorithm described in DiDonato and Hageman (1980) to compute the integral of the bivariate normal distribution over each sector.

squared differences between the p_rs and q_rs, weighted by the inverses of their estimated variances,

$$WSSQ = \sum_{r=1}^{6} \frac{(p_r - q_r)^2}{Np_r(1 - p_r)}. \tag{10}$$

The computation of an election result under the Estimate Centrality method requires the numerical computation of $M!$ areas under an (M – 1)-variate normal distribution. Currently we are only able to compute such integrals for bivariate normal distributions, and we must therefore restrict our analysis to three-candidate elections. Because it is straightforward in theory to apply the Estimated Centrality method to elections with more than three candidates, it should be possible to extend our analysis.

We calibrate the spatial model to observed election data by applying the Estimated Centrality voting method. The evaluation of a single election provides one set of values for the angles of the bisectors and the location of the mode, and the evaluation of multiple elections provides information about the distributions of these parameters. In our Monte Carlo analysis in Section 4, we assume that the 3 angles follow a tri-variate Dirichlet distribution with four parameters (the three shares of a semi-circle and the variance parameter), and that the distance between P and O follows a Weibull distribution with two parameters.

3.5 *Model assessment*

To determine which model of voter behavior comes closest to describing the outcomes of actual elections, we assess each model's statistical adequacy using data from multiple elections. For each model we determine three different measures of fit. The first two measures assess how well each model explains the observed election data. The third measure assesses the degree to which each model can simulate election data that are reasonably similar to the observed data.

To assess how well the models explain each of the observed elections, we first determine each model's mean sum of weighted squared residuals, *MSWSR*, that is, the mean over the observed elections of the sum of squared differences between each of the $M!$ observed vote shares, q_r, and the corresponding vote share predicted by the best parameterization of the respective model, p_r, weighted by the inverse of the predicted variance in the number of voters:

$$MSWSR = \frac{1}{E}\sum_{e=1}^{E}\left(\sum_{r=1}^{M!}\frac{(p_{re}-q_{re})^2}{Np_{re}(1-p_{re})}\right), \qquad (11)$$

where E is the number of observed elections. The lower a model's *MSWSR*, the closer on average are the model's predicted vote shares to the actual vote shares.

As an alternative measure of fit, we determine each model's mean log-likelihood ratio,

$$MLLR = \frac{1}{E}\sum_{e=1}^{E}\ln\left(\prod_{r=1}^{M!}(p_{re})^{\hat{N}_{er}} \Big/ \prod_{r=1}^{M!}(q_{re})^{\hat{N}_{er}}\right) \qquad (12)$$

that is, the mean over the E elections of the logarithm of the likelihood of the predicted shares p_{re}, divided by the likelihood of the observed shares q_{re}. The algebraically greater a model's *MLLR*, the better is the model's fit.

An alternative way of assessing a model's adequacy is to ask whether the model is able to simulate artificial election data that have the same distribution as the observed election data. If the variation in the frequencies of rankings in actual elections differs from the sampling variation of the rankings predicted by the model of voter behavior, then this constitutes evidence against the model. A formal way of assessing the degree to which different models depart from actual elections is to introduce an "error function" in the form of an assumption that the p_rs for each election are random variables rather than being specified directly by a model of voter behavior. An intuitive assumption is that the p_rs follow a Dirichlet (multivariate beta) distribution with parameter vector $\alpha = (\delta p'_1, ..., \delta p'_{M!})$, where the parameter δ is a constant that is inversely proportional to the variances of the p_rs, and $(p'_1, ..., p'_{M!})$ is the vector of expected probabilities of the M! rankings, specified by the model of voter behavior. The use of the Dirichelet distribution to generate p_rs for a multinomial process that generates election outcomes leads to the multivariate Pólya (also known as the compound Dirichlet-multinomial) distribution with density function

$$f(N_1, ..., N_{M!}; N, \alpha_1, ..., \alpha_{M!}) = \frac{\Gamma(\delta)}{\Gamma(N+\delta)} \prod_{r=1}^{M!} \frac{\Gamma(N_r + \delta p_r)}{\Gamma(\delta p_r)}, \tag{13}$$

whose first two moments are $E[N_r] = Np_r$, $Var[N_r] = Np_r(1 - p_r)\psi$, and $Cov[N_r, N_s] = -Np_r p_s \psi$, where $\psi = (N + \delta) / (1 + \delta)$.[14] As the variance parameter δ approaches infinity, ψ approaches 1 and the multivariate Pólya distribution in (13) approaches the multinomial distribution in (1). Thus for each model of voter behavior we can estimate the value of δ for which the variations in frequencies of the simulated elections best match the variations in frequencies of the observed rankings. The greater the estimated value of δ, the better does the model describe the data.

4. Empirical Analysis of the Outcomes of Vote-Casting Processes

4.1 The data

We assemble our ranking data from the thermometer scores that are part of the surveys conducted by ANES. These surveys are conducted every two years, and participants are asked to rate politicians on a scale from 0 to 100 (the "thermometer"). The list of persons includes the president and the vice president, the republican and democratic presidential candidates and vice presidential candidates (in election years), as well as past presidents and presidential candidates

[14] See Mosimann (1962, pp. 67-68).

who still play significant roles in the political arena. We refer to these persons as "candidates." In the surveys conducted before 1970, a candidate whom the survey respondent did not know received a score of 50 on the participant's answer sheet, while such a candidate was coded as "unknown" in the surveys from 1970 onward. To avoid ambiguities between unknown candidates and candidates evaluated at 50, we restrict our analysis to the 18 surveys conducted from 1970 to 2004.

The number of respondents in a survey ranges from 1,212 in 2004 to 2,705 in 1974, and the number of candidates included in the surveys ranges from 3 in 1986 and 1990 to 12 in 1976. We construct all possible combinations of 3 candidates within a year, so we have 1 combination in 1986 and 1990, $12!/(3!\cdot 9!) = 220$ combinations in 1976, and a total of 913 combinations from all 18 surveys. We treat each combination as one election with three candidates. (Recall that we restrict our analysis to three-candidate elections to be able to evaluate the spatial model with the Estimated Centrality method.) For every such election, we eliminate the survey responses that do not assign scores to all three candidates, yielding from 759 to 2,521 responses for the 913 three-candidate elections. For each response, we rank the three candidates according to their thermometer scores. If a response yields a strict ranking of candidates, then we count it as one vote for this ranking. Survey respondents are allowed to assign equal scores to different candidates; about 13 percent of all responses in our three-candidate elections are two-way ties and about 5 percent are three-way ties. While there are different ways of accommodating ties, we adopt the following intuitive rule: If all candidates are tied, then we count the response as 1/6 vote for each of the 6 possible rankings, and if two candidates are tied, then we count the response as half a vote for each of the two possible strict rankings.[15] Thus our adjusted data set

[15] We adopt this procedure only to calibrate our models of voter behavior, which requires the number of votes for each ranking for the calculation of the *MSWSR* and the *MLLR*. The standard methods for accommodating ties in

consists of the total number of votes for each of the six strict rankings in each of the 913 three-candidate elections.

We first count the frequency with which the three voting methods yield different outcomes in the 913 elections. The three voting methods chose the same winning ranking in 838 elections and the same winning candidate in 879 elections. Thus there is disagreement about the winning ranking in 8.2% of the elections and disagreement about the winning candidate in 3.7% of the elections. In most elections, the choice of voting method does not affect how the election is decided. However, the disagreement about the winning candidate in 34 elections suggests that it is worthwhile to inquire which voting method is most likely to identify the correct winner.

With respect to the winning ranking, the Borda method and the Condorcet method disagreed in 75 elections, the Borda method and Estimated Centrality disagreed in 74 elections, while the Condorcet method and Estimated Centrality disagreed in only 37 elections. In no election did the three voting methods choose three different rankings. With respect to the winning candidate, the Borda method and the Condorcet method disagreed in 34 elections, the Borda method and Estimated Centrality disagreed in 30 elections, while the Condorcet method and Estimated Centrality disagreed in only 14 elections. Thus the Condorcet method and Estimated Centrality agree more often with each other than either of them agrees with the outcome chosen by the Borda method.

4.2 Assessing the four models of voter behavior

We first report each model's *MSWSR* and its *MLLR*, together with their standard errors, in Columns 1 and 2 of Table 3. The two measures of fit suggest the same relative degrees of

---

more elegant ways (for example, Black's method—see Black, 1958, pp. 61 – 64) do not allocate shares of tied votes to the respective rankings.

Table 3. Assessment of models of voter behavior

	Analysis of observed election data[1]		Comparison of simulated with observed election data[2]					
	Mean sum of weighted squared differences[3]	Mean log-likelihood ratio	Multinomial distribution	Pólya distribution		Convex combination of Multinomial and Pólya distributions		
			Test statistic (*p* value for 10 df)	δ	Test statistic (*p* value for 9 df)	δ	Share of Multi-nomial	Test statistic (*p* value for 8 df)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Equal probabilities	207.28	-196.80	11,000,000[4]	18.5	28.405	18	0.0300	19.6140
	(4.42)	(4.26)	(0.0000)		(0.0000)			(0.0119)
Borda model	121.00	-116.79	2,388,599	32	73.4583	31	0.0525	21.2425
	(2.78)	(2.72)	(0.0000)		(0.0000)			(0.0065)
Condorcet model	88.36	-84.99	1,012,883	48	74.8475	45	0.0533	24.1243
	(2.25)	(2.00)	(0.0000)		(0.0000)			(0.0022)
Spatial model	0.97	-0.87	118.71	2,440	12.2153	675	0.6431	6.0730
	(0.06)	(0.05)	(0.0000)		(0.2014)			(0.6391)

Notes:

[1] The values in parentheses in Columns 1 and 2 are standard errors

[2] The values in parentheses in Columns 3, 5, and 8 are the chi-square tail-area probabilities under the hypothesis that the simulated and observed distributions of the sum of weighted squared differences between the predicted and observed vote shares are the same.

[3] To facilitate comparisons, we have multiplied the mean sums of weighted squared differences and their standard errors in Column (1) by 1,000,000.

[4] All residuals for the multinomial model were concentrated in the first bin.

accuracy for the four models. The fit of the equal-percentages model is far below that of the other three models, which constitutes strong evidence against this model and suggests that, in the typical election, voters report some rankings of candidates significantly more frequently than others. The fit of the Borda model is worse than that of the Condorcet model by about 50 percent. The fit of the spatial model is better than that of the other three models by about *two orders of magnitude*. Even though the spatial model uses three more degrees of freedom than the Borda and Condorcet models, the results indicate that, by a wide margin, the spatial model describes the observed election data best.

We next investigate the degree to which these models are able to simulate election data with the same distribution as the observed data. Simulating elections requires that we calibrate the models of voter behavior to the observed data to obtain estimates of the unknown model parameters and their distributions. The model of equal percentages does not have any unknown parameters because it simply predicts that each ranking has an equal chance of being chosen. Our estimates of the unknown parameters of the other three models are the values that minimize the weighted sum of squared differences between the observed and predicted vote shares in equation (10).[16] To facilitate replication of our analysis, we report these estimates in Table 4.[17] With respect to the distributions of these parameters, we assume that the unknown scale parameter α in the Borda and the Condorcet model follows a gamma distribution, whose parameters we calibrate to match the mean and variance of the estimates of the αs in Table 4. For the spatial model, we assume that the 3 angles follow a tri-variate Dirichlet distribution with

[16] Recall that each evaluation of the Estimated Centrality method yields one set of parameters estimates that minimize the weighted sum of squared differences.

[17] Recall that β_1, β_2, and β_3 specify the three angles formed by the line $\overline{PO}$ and the perpendicular bisectors, which implies that the average angle formed by the two perpendicular bisectors that form the sector that contains O is $\beta_1 + \pi - \beta_3 = 1.1060$, while the other two angles formed by the perpendicular bisectors are 0.9907 and 1.0449.

Table 4. Estimates of the parameters of the three models of voter behavior

Models of Voter Behavior	Angles of the three perpendicular bisectors with the line $\overline{PO}$			Distance from *O* to *P*	Scale parameter
	β_1	β_2	β_3	x	α
Spatial Model	0.5562 (0.0121)	1.5486 (0.0146)	2.5920 (0.0125)	0.4439 (0.0112)	
Borda Model					0.3445 (0.0062)
Condorcet Model					0.3438 (0.0060)

Note: Standard errors are shown in parentheses

four parameters (the three shares of a semi-circle and the variance parameter), and that the distance of *P* and *O* follows a Weibull distribution with two parameters. Again, we used the estimates in Table 4 to guide our calibration of these six parameters.

We then undertake the following Monte Carlo experiment for each model: draw a set of parameters from their respective distributions, use them to generate the p_rs for a given r^* and m^*, and use the p_rs to draw a set of N_rs from the multinomial distribution. Repeat these steps 1,000,000 times, evaluate the data at each step with the respective voting method, and compare the distribution of the squared residuals with the distribution of the squared residuals in the observed elections.

To compare these distributions, we generate the histogram of the weighted squared residuals of the observed data by first ordering the observed elections by the weighted squared residuals that we obtained in our analysis of the respective model of voter behavior, and then distributing the residuals into 11 bins of 83 residuals each. We then allocate the residuals from our evaluation of the simulated data into corresponding bins. Under the assumption that the differences between the numbers of observations in the bins from the simulated and the observed elections are asymptotically normally distributed, the sum of the squared differences divided by the respective variances is asymptotically chi-square distributed with 10 degrees of freedom.[18] Thus the right tail area of the chi-square distribution indicates the probability that the weighted squares of differences between observed and expected numbers of elections in each bin will be greater than or equal to the test statistic if the observed election data are generated by the respective model of voter behavior.

We first simulated 1,000,000 elections for each model under the assumption that the simulated rankings follow the multinomial distribution in (1). In Column 3 of Table 3, we report the chi-square test statistics, together with the corresponding probabilities (in parentheses) that the residuals of an analysis of the simulated data have the same distribution as the residuals of our analysis of the data from ANES. The large values of the test statistics indicate that none of the four models yields sufficient variation in the frequencies of rankings for the corresponding multinomial vote-generating process to be an adequate description of the data. For the model of Equal Probabilities, the Borda model, and the Condorcet model, the test statistics are especially high because the residuals from the simulated data are concentrated in the first few bins—in fact, *all* residuals for the model of Equal Probabilities are in the first bin. While it would have been possible to alter the test statistics by choosing different bin sizes, it is obvious that the distributions of the residuals from the simulated data are very different from those from the observed data. We thus decided that it was not worthwhile to analyze alternative sets of bins with different sizes. In comparison to the other three models, however, the test statistic of the

[18] This relationship holds only for a random (iid) sample. However, the fact that we demarcate the bins of the observed elections so that the probability of a weighted sum of squared differences being in any one bin is 1/11 made it straightforward to calculate the correlations between any two bins as a function of the number of voters. We found this correlation to be fairly small—the correlation is ρ = -0.1 if N = 1, ρ = 0.0043 if N = 75, ρ = 0.0071 if N = 150, ρ = 0.0085 if N = 300, and ρ = 0.0093 if N = 600. Thus the correlation increases at a decreasing rate and can be expected to be still below 0.01 if N = 5,000. Because the correlations are very small and our results are extremely strong, we decided that inference based on the chi-square distribution would be sufficiently accurate.

spatial model is fairly small, which indicates that the spatial model is a more promising candidate for the right model of voter behavior than any of the other three models.

To accommodate the extra variation in the frequencies of rankings, we repeated the analysis under the assumption that the probabilities of the six rankings follow the Pólya distribution in (13)—that is, they are randomized through a Dirichlet process before they become the input of the multinomial distribution. For each model, we estimated the variance parameter of the Dirichlet distribution, δ, as the value that minimizes the chi-square test statistic, so the corresponding chi-square test has 11 – 2 = 9 degrees of freedom. We report the estimates of the variance parameters and the test statistics in Columns 4 and 5 of Table 3. The large variance parameter of the spatial model implies that the additional variance contributed by the Dirichlet process is fairly small, and that the spatial model by itself already describes a large share of the variation in the data. In contrast, the variance parameters of the three other models are very small, indicating that the additional variance in the probabilities that results from the Dirichlet process is fairly large. These models of voter behavior can therefore explain only a small fraction of the variation of the observed election data, while the Dirichlet error process provides a sizeable share of the necessary variation.

All test statistics in Column 5 are much smaller than those in Column 3, which indicates that the Dirichlet process leads to better fits for all models. However, only the spatial model has a noticeably positive probability that the simulated data have the same distribution as the observed data. At a probability of 79.86%, we cannot reject the hypothesis, at all conventional levels of significance, that the two data sets were generated by the same process. We conclude that the spatial model provides a reasonably good description of the data generating process of the observed ANES election data.

5. SIMULATIONS

Although our empirical results suggest that the spatial model describes the statistical process of the outcomes of vote-casting much better than any of the other three models, they do not imply that the Estimated Centrality method is necessarily most likely to identify either the correct ranking or the correct winner. We next estimate the accuracy with which different voting methods identify either the correct ranking or the correct winner under different models of voter behavior.

The idea of a correct outcome that exists prior to the casting of votes, together with the notion that these votes are random variables, implies that voting can yield an incorrect outcome even when there are just two candidates. Thus when we report the frequency of incorrect outcomes under different voting methods for three candidates, it should be borne in mind that in our model of voting, the possibility of incorrect outcomes does not arise solely because there are three candidates but can happen with two candidates as well.

To improve the fit of the spatial model and thereby improve the predictive power of our simulations for actual elections, we experimented with additional error distributions and obtained an extremely close fit for the spatial model by assuming that $1/\delta$ follows a normal distribution, although the estimated mean value of δ tended to converge towards infinity. We were able to obtain an equally good fit for the spatial model by assuming that the N_rs follow a weighted combination of two independent distributions, one multinomial and the other Pólya. This combination requires the estimation of an additional weight parameter ω that describes the share of the multinomial distribution in the combined distribution. Because the multinomial and Pólya distributions are independent of each other in this combination, the first two moments of the combined distribution are $\mathrm{E}[N_r] = \omega Np_r + (1 - \omega)Np_r = Np_r$ and $\mathrm{Var}[N_r] = \omega^2 Np_r(1 - p_r) + (1 -$

$\omega)^2 N p_r(1 - p_r)\psi$, with $\psi = (N + \delta) / (1 + \delta)$. Thus the greater the estimated values of ω and δ, the closer this distribution is to the multinomial distribution and the better the corresponding model describes the data.

Columns 6 – 8 of Table 3 show the variance and weight parameters as well as the chi-square test statistics for the four models. The test statistic for the spatial model indicates an almost perfect fit, and the large share of the multinomial distribution as well as the large value of δ indicate that the Pólya distribution provides only the smaller part of the variation. We report the values for the other models for the sake of completeness—the new error distribution yields a better fit for all three models; the Equal Percentages model fails the hypothesis test for different distributions at 99% significance. However, the small weight on the multinomial distribution and as well as the small δ indicate that the error distribution contributes a much larger share to the variation of the frequencies than it does for the spatial model. Although it may be possible to construct error structures that would generate a good fit for the other three models, such an exercise would not alter the fact that these models of voter behavior do not generate the necessary variation in the data and are therefore unable to explain much of the observed data.

We use the parameter estimates in Columns 6 – 8 of Table 3 for the spatial model to simulate one million elections each for different numbers of voters—25, 100, 1,000, 10,000, 100,000, and 1,000,000 voters. We evaluate every election with the Borda method, the Condorcet method, and the Estimated Centrality method, as well as with the Instant Run-off method, the Maximin method, and the Plurality method. The Instant Run-off method selects the majority winner if there is one, and otherwise eliminates the candidate with the fewest votes and allocates the votes for that candidate to the candidates ranked second on those ballots. The Maximin method chooses the Condorcet winner if there is one, and it chooses the candidate with the smallest loss in pairwise comparisons in case of a voting cycle. The Plurality method chooses the candidate with the greatest number of first-place votes.

For each voting method, we count the number of times it does not identify the correct ranking and the correct winner. We report the results of our simulations in Table 5. The Plurality method performs by far the worst in all simulations, and the Instant Run-off method is second worst. The other four methods yield fairly similar results. The Estimated Centrality method identifies the correct ranking most often regardless of the number of voters, and it identifies the correct winner most often if the number of voters is large. For elections with 25 voters, the Borda method and the Estimated Centrality method perform nearly equally well in identifying the winner, and the Borda method has a small but statistically significant advantage for 100 voters. It is interesting to note that the number of incorrectly identified rankings and winners does not converge towards zero as the number of voters increases, which is a consequence of using the Dirichlet error structure in our generating mechanism.

For comparison, we also assess the accuracy of the six voting methods under the Borda and Condorcet models of voter behavior, reporting the results of our simulations in Tables 6 and 7. Because neither the Borda nor the Condorcet model offers a satisfying description of the observed ANES data, we evaluate strict Borda and Condorcet models employing only the multinomial distribution in (1) without any additional error structure. It is therefore worth emphasizing that neither of the two sets of simulations is likely to provide insights into the precision of voting methods in actual elections, but it is nevertheless informative to compare these simulations with those that we obtained under the spatial model. To generate the probabilities of the N_rs, we assume that the αs follow gamma distributions with the means and variances that are given in Table 4. We note the following three results.

Table 5. Accuracy of different voting methods under the Spatial model of voter behavior – incorporating model error*

Number of voters	Borda	Condorcet	Estimated Centrality	Instant Run-off	Maximin	Plurality	Borda	Condorcet	Estimated Centrality	Instant Run-off	Maximin	Plurality
	Number of times (out of 1,000,000) that the voting method fails to identify the correct *ranking*						Percentage of time that the voting method fails to identify the correct *ranking*					
25	456,104	461,629	448,492	524,008	456,075	550,711	0.4561	0.4616	**0.4485**	0.5240	0.4561	0.5507
	(498)	(499)	(497)	(499)	(498)	(497)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
100	284,790	291,792	281,253	418,659	290,313	441,100	0.2848	0.2918	**0.2813**	0.4187	0.2903	0.4411
	(451)	(455)	(450)	(493)	(454)	(497)	(0.0005)	(0.0005)	(0.0004)	(0.0005)	(0.0005)	(0.0005)
1,000	130,037	122,112	117,135	309,169	121,803	372,957	0.1300	0.1221	**0.1171**	0.3092	0.1218	0.3730
	(336)	(327)	(322)	(462)	(327)	(484)	(0.0003)	(0.0003)	(0.0003)	(0.0005)	(0.0003)	(0.0005)
10,000	94,144	75,583	72,289	285,608	75,535	362,903	0.0941	0.0756	**0.0723**	0.2856	0.0755	0.3629
	(292)	(264)	(259)	(452)	(264)	(481)	(0.0003)	(0.0003)	(0.0003)	(0.0005)	(0.0003)	(0.0005)
100,000	88,633	69,132	66,048	282,503	69,058	362,320	0.0886	0.0691	**0.0660**	0.2825	0.0691	0.3623
	(284)	(254)	(248)	(450)	(254)	(481)	(0.0003)	(0.0003)	(0.0002)	(0.0005)	(0.0003)	(0.0005)
1,000,000	88,596	68,224	65,054	282,201	68,160	361,889	0.0886	0.0682	**0.0651**	0.2822	0.0682	0.3619
	(284)	(252)	(247)	(450)	(252)	(481)	(0.0003)	(0.0003)	(0.0002)	(0.0005)	(0.0003)	(0.0005)
	Number of times (out of 1,000,000) that the voting method fails to identify the correct *winner*						Percentage of time that the voting method fails to identify the correct *winner*					
25	275,678	283,730	274,892	290,194	278,613	318,137	0.2757	0.2837	**0.2749**	0.2902	0.2786	0.3181
	(447)	(451)	(446)	(454)	(448)	(466)	(0.0004)	(0.0005)	(0.0004)	(0.0005)	(0.0004)	(0.0005)
100	156,539	163,837	157,888	191,992	162,360	222,868	**0.1565**	0.1638	0.1579	0.1920	0.1624	0.2229
	(363)	(370)	(365)	(394)	(369)	(416)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
1,000	65,185	63,547	61,038	87,071	63,352	167,353	0.0652	0.0635	**0.0610**	0.0871	0.0634	0.1674
	(247)	(244)	(239)	(282)	(244)	(373)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0004)
10,000	46,218	39,036	37,426	64,425	38,993	158,782	0.0462	0.0390	**0.0374**	0.0644	0.0390	0.1588
	(210)	(194)	(190)	(246)	(194)	(365)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0004)
100,000	43,235	35,469	33,828	61,454	35,421	158,206	0.0432	0.0355	**0.0338**	0.0615	0.0354	0.1582
	(203)	(185)	(181)	(240)	(185)	(365)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0004)
1,000,000	43,112	34,955	33,348	60,888	34,899	157,856	0.0431	0.0350	**0.0333**	0.0609	0.0349	0.1579
	(203)	(184)	(180)	(239)	(184)	(365)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0004)

Note: Standard errors are shown in parentheses

* As a result of a slight programming glitch, the results for Estimated Centrality are only estimates, to be refined later. (The results are reasonably precise, but it would have taken a few days to update them, which we plan to do at the next iteration of the paper).

Table 6. Accuracy of different voting methods under the Borda model of voter behavior – no model error

Number of voters	Borda	Condorcet	Estimated Centrality	Instant Run-off	Maximin	Plurality	Borda	Condorcet	Estimated Centrality	Instant Run-off	Maximin	Plurality
	Number of times (out of 1,000,000) that the voting method fails to identify the correct *ranking*						Percentage of time that the voting method fails to identify the correct *ranking*					
25	618,427	623,660	620,363	626,290	617,214	636,064	**0.6184**	0.6237	0.6204	0.6263	**0.6172**	0.6361
	(486)	(484)	(485)	(484)	(486)	(481)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
100	542,318	546,253	542,618	553,233	543,591	553,742	**0.5423**	0.5463	**0.5426**	0.5532	**0.5436**	0.5537
	(498)	(498)	(498)	(497)	(498)	(497)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
1,000	503,285	503,669	502,494	504,256	503,547	504,914	**0.5033**	0.5037	**0.5025**	0.5043	**0.5035**	0.5049
	(500)	(500)	(500)	(500)	(500)	(500)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
10,000	500,137	500,077	500,054	499,835	500,121	499,969	0.5001	0.5001	0.5001	0.4998	0.5001	0.5000
	(500)	(500)	(500)	(500)	(500)	(500)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
100,000	499,248	499,666	500,133	499,427	499,701	499,433	0.4992	0.4997	0.5001	0.4994	0.4997	0.4994
	(500)	(500)	(500)	(500)	(500)	(500)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
1,000,000	499,790	499,618	500,747	499,362	499,607	499,388	0.4998	0.4996	0.5007	0.4994	0.4996	0.4994
	(500)	(500)	(500)	(500)	(500)	(500)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
	Number of times (out of 1,000,000) that the voting method fails to identify the correct *winner*						Percentage of time that the voting method fails to identify the correct *winner*					
25	237,569	246,929	239,651	251,393	240,755	270,937	**0.2376**	0.2469	0.2397	0.2514	0.2408	0.2709
	(426)	(431)	(427)	(434)	(428)	(444)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
100	84,826	92,045	85,268	106,773	89,674	107,374	**0.0848**	0.0920	**0.0853**	0.1068	0.0897	0.1074
	(279)	(289)	(279)	(309)	(286)	(310)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
1,000	5,599	6,422	5,635	7,074	6,266	8,248	**0.0056**	0.0064	**0.0056**	0.0071	0.0063	0.0082
	(75)	(80)	(75)	(84)	(79)	(90)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
10,000	189	217	172	233	205	292	**0.0002**	0.0002	**0.0002**	0.0002	0.0002	0.0003
	(14)	(15)	(13)	(15)	(14)	(17)	(0.0000)	(0.0000)	0.0000	(0.0000)	(0.0000)	(0.0000)
100,000	6	8	2	7	8	7	0.0000	0.0000	**0.0000**	0.0000	0.0000	0.0000
	(2)	(3)	(1)	(3)	(3)	(3)	(0.0000)	(0.0000)	0.0000	(0.0000)	(0.0000)	(0.0000)
1,000,000	0	0	0	0	0	0	0	0	0	0	0	0
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)

Note: Standard errors are shown in parentheses

Table 7. Accuracy of different voting methods under the Condorcet model of voter behavior – no model error

Number of voters	Borda	Condorcet	Estimated Centrality	Instant Run-off	Maximin	Plurality	Borda	Condorcet	Estimated Centrality	Instant Run-off	Maximin	Plurality
	Number of times (out of 1,000,000) that the voting method fails to identify the correct *ranking*						Percentage of time that the voting method fails to identify the correct *ranking*					
25	432,522	412,053	428,079	455,208	407,937	490,475	0.4325	0.4121	0.4281	0.4552	**0.4079**	0.4905
	(495)	(492)	(495)	(498)	(491)	(500)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
100	197,018	181,291	193,744	233,906	180,076	248,129	0.1970	0.1813	0.1937	0.2339	**0.1801**	0.2481
	(398)	(385)	(395)	(423)	(384)	(432)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
1,000	16,717	14,232	16,716	20,654	14,236	24,746	0.0167	**0.0142**	0.0167	0.0207	**0.0142**	0.0247
	(128)	(118)	(128)	(142)	(118)	(155)	-0.0001	-0.0001	(0.0001)	-0.0001	-0.0001	-0.0002
10,000	480	409	515	646	409	811	0.0005	**0.0004**	0.0005	0.0006	**0.0004**	0.0008
	(22)	(20)	(23)	(25)	(20)	(28)	(0.0000)	(0.0000)	0.0000	(0.0000)	(0.0000)	(0.0000)
100,000	10	9	13	13	9	15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(3)	(3)	(4)	(4)	(3)	(4)	(0.0000)	(0.0000)	0.0000	(0.0000)	(0.0000)	(0.0000)
1,000,000	0	0	0	0	0	0	0	0	0.0000	0	0	0
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
	Number of times (out of 1,000,000) that the voting method fails to identify the correct *winner*						Percentage of time that the voting method fails to identify the correct *winner*					
25	267,310	262,860	266,376	265,767	256,956	288,965	0.2673	0.2629	0.2664	0.2658	**0.2570**	0.2890
	(443)	(440)	(442)	(442)	(437)	(453)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0005)
100	111,877	106,487	110,426	122,723	104,545	129,441	0.1119	0.1065	0.1104	0.1227	**0.1045**	0.1294
	(315)	(308)	(313)	(328)	(306)	(336)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
1,000	8,896	7,783	8,985	8,323	7,702	12,111	0.0089	**0.0078**	0.0090	0.0083	**0.0077**	0.0121
	(94)	(88)	(94)	(91)	(87)	(109)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
10,000	277	243	290	253	240	426	0.0003	**0.0002**	0.0003	0.0003	**0.0002**	0.0004
	(17)	(16)	(17)	(16)	(15)	(21)	(0.0000)	(0.0000)	0.0000	(0.0000)	(0.0000)	(0.0000)
100,000	2	4	8	4	4	6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(1)	(2)	(3)	(2)	(2)	(2)	(0.0000)	(0.0000)	0.0000	(0.0000)	(0.0000)	(0.0000)
1,000,000	0	0	0	0	0	0	0	0	0.0000	0	0	0
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)

Note: Standard errors are shown in parentheses

First, the accuracy with which different voting methods predict the correct ranking and winner depends on and varies greatly with the underlying model of voter behavior: for different models, different voting methods are most likely to identify the correct ranking and winner. This confirms the theoretical result in the voting literature that no single voting method is the best estimator of the correct winner under all possible circumstances. For the Borda model, the Borda method identifies the winner statistically significantly more accurately than the other four voting methods only for 25 voters. As the number of voters increases, the Estimated Centrality method does better than the Borda method. None of the voting methods identify the correct ranking well under the Borda model. For the Condorcet model, the Maximin method is generally slightly more accurate than the Condorcet method, and both methods are significantly more accurate than the other four methods for 25 and 100 voters.

Second, all six voting methods are less likely to identify the correct ranking and winner under the spatial model plus its assumed error structure than under the Borda or Condorcet models without any error structure. As the number of voters increases, the accuracy of all six voting methods increases at a much faster rate for the strict Borda and Condorcet models than for the spatial model with its assumed error structure: with these two strict models, all voting methods are able to identify the correct winner in every single one of the one million simulated elections when the number of voters is one million.

Third, under the Borda model, the degree of accuracy of all voting methods with respect to the correct ranking does not improve beyond 50%, even if the number of voters is large. This is explained by the fact that the Borda model of voter behavior assigns the same probability to all rankings that rank the winning candidate at the same position. Thus if the Borda model were the correct model of voter behavior, then it would be futile to try and estimate correct rankings beyond the top candidate.

6. Conclusion

The main purpose of our paper is to suggest a new direction for the evaluation of voting methods. Rather than analyzing all possible cases under which different voting methods may generated counter-intuitive outcomes, we suggest that it is more promising to analyze how different voting methods perform in actual elections. Our initial analysis indicates that the spatial model of voting may be a more suitable description of voter behavior than several other plausible models, and that the Estimated Centrality voting method may be somewhat more likely to identify the correct winner in actual elections than the popular Borda and Condorcet methods.

We emphasize that our current analysis is far from being conclusive; it constitutes only the beginning of a larger research program with at least three different lines of inquiry. First, we draw our conclusions in this paper on the basis of a single data set that is derived from surveys. We need to test whether our results will continue to hold for data derived from real elections, and whether it is even possible to describe all elections with a single model of voter behavior. Still, the fact that the spatial model fits the data so much better than the three competitors that we examine suggests that the spatial model may very well continue to be a reasonable description of voter behavior in other elections.

Second, we examine only four models of voter behavior, and it is probably possible to improve upon the spatial model that we use. The fact that we did not obtain a satisfying fit of our spatial model until we introduced some additional variation through the Dirichlet model indicates that our spatial model does not yet describe voter behavior perfectly. We have only examined a very simple version of the spatial model, and it is likely that a more general spatial model will fit the observed data even better. Third, our analysis has implications for all inquiries into the frequency of certain voting events, be that the probability that strategic voting will alter the outcome of an election, the existence of dominant candidates, or the likelihood of different voting paradoxes. Rather than assuming that all rankings are equally likely, our framework makes it possible to incorporate more realistic models of voter behavior into such analyses and thereby to improve the accuracy of their predictions. In this context it is worth reiterating that our results regarding the attractiveness of the Estimated Centrality method are based on the assumption that voters do not vote strategically. Any recommendation about which voting method one ought to use in actual elections needs to take the method's resistance to strategizing into account. Fortunately, our framework quite naturally lends itself to these kinds of inquiries.

References

Black, Duncan. 1958. *The Theory of Committees and Elections*. Cambridge University Press: Cambridge.

Cervone, Davide; William Gehrlein; William Zwicker. 2005. "Which Scoring Rule Maximizes Condorcet Efficiency Under Iac." *Theory and Decision* **58**:2, 145-185.

Condorcet, Marie Jean Antoine Nicolas de Caritat, Marquis de, Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix (Paris, 1785).

Conitzer, Vincent and Tuomas Sandholm. 2005. "Common Voting Rules as Maximum Likelihood Estimators." In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI-05)*, pp. 145-152, Edinburgh, Scotland, UK.

DiDonato, A.R. and R.K. Hageman. 1980. "Computation of the Integral of the Bivariate Normal Distribution Over Arbitrary Polygons." Naval Surface Weapons Center, Government Accession Number ADA102466.

Drissi-Bakhkhat, Mohamed and Michael Truchon. 2004 "Maximum Likelihood Approach To Vote Aggregation With Variable Probabilities" *Social Choice and Welfare* **23**: 161-185.

Enelow, James M. and Melvin J. Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. Cambridge University Press.

Enelow, James M. and Melvin J. Hinich. 1990. *Advances in the Spatial Theory of Voting*. Cambridge University Press.

Gehrlein, William V. 2002. "Condorcet's Paradox and the Likelihood of Its Occurrence: Different Perspectives on Balanced Preferences." *Theory and Decision* **52**:2, 171-199.

Good, I. Jack and T. Nicolaus Tideman. 1976. "From Individual to Collective Ordering through Multidimensional Attribute Space." *Proceedings of the Royal Society of London* (Series A) **347**: 371-385.

Kemeny, John. 1959. "Mathematics without Numbers." *Daedalus* **88**: 571-91.

Kendall, Maurice and Jean D. Gibbons. 1990. Rank Correlation Methods. New York: Oxford University Press.

Mosimann, James E. 1962. "On the Compound Multinomial Distribution, the Multivariate Beta Distribution, and Correlations Among Proportions." *Biometrika* **49:** 65-82.

Risse, Mathias. 2001. "Arrow's Theorem, Indeterminacy, and Multiplicity Reconsidered." *Ethics* **111**: 706-734.

Risse, Mathias. 2005. “Why the Count de Borda Cannot Beat the Marquis de Condorcet” *Social Choice and Welfare* **26**: 95-113.

Saari, Donald G. 1990. “Susceptibility to Manipulation.” *Public Choice* **64**: 21-41.

Saari, Donald G. 2003. “Capturing the ‘Will of the People.’” *Ethics* **113**: 333-349.

Saari, Donald G. 2006. “Which is Better: The Condorcet or Borda Winner?” *Social Choice and Welfare* **26**: 107-129.

Tideman, T. Nicolaus. 2006. *Collective Decisions and Voting*. Burlington, VT: Ashgate.

Truchon, Michel. 2006. “Borda and the Maximum Likelihood Approach to Vote Aggregation.” *Mimeo*.

Young, H. Peyton. 1986. “Optimal Ranking and Choice from Pairwise Comparisons.” In Grofman B. and G. Owen (eds.) Information Pooling and Group Decision Making. Greenwich, Conn: JAI Press, pp. 113-122.

Young, H. Peyton. 1988. “Condorcet’s Theory of Voting.” *The American Political Science Review* **82**(4): 1231-1244.

Young, H. Peyton. 1995. “Optimal Voting Rules.” *Journal of Economic Perspectives* **9**(1): 51-64.