Chapter X: Modeling the outcomes of vote-casting in actual elections

Nicolaus Tideman and Florenz Plassmann

X.1 Introduction

How often do events of interest to voting theorists occur in actual elections? For example, what is the probability of observing a voting cycle—an outcome in which no candidate beats all other candidates in pairwise comparison by majority rule? When there is a candidate who beats all others in such pairwise comparisons—a Condorcet winner—what is the probability that a voting method chooses this candidate? What is the probability that voters have an incentive to vote strategically—that is, cast their votes in ways that do not reflect their true preferences? Voting theorists have analyzed these questions in great detail, using a variety of statistical models that describe different distributions of candidate rankings. But there has been no systematic effort to determine which statistical model comes closest to describing the distribution of rankings of candidates in actual elections. Thus we know how often various voting events occur under different statistical models, but not how often voting events occur in actual elections. This chapter provides a framework for answering this question.

We consider elections in which each voter is asked to submit a strict ranking of m candidates. We interpret the rankings submitted by all voters as the outcome of a statistical model of vote-casting that yields a vector with $m!$ components, representing the possible strict rankings of the m candidates. To assess the probabilities of voting events, we need to know the likelihoods of these vectors—are all vectors of rankings equally likely or are some more likely to occur than others?

We identify a statistical model of vote-casting that comes very close to describing the distribution of vectors of rankings in actual three-candidate elections.

We find it intuitive to view the statistical model of vote-casting as a two-part process. One part describes the distribution of the expected shares of the voters who will report each of the $m!$ rankings. The other part describes the distribution of the observed shares of the voters who report each of the $m!$ rankings, given a vector that describes the expected shares of these rankings

Voting theorists have proposed various models to describe possible distributions of the expected shares. However, these models were generally proposed for purposes other than describing rankings in actual elections, and we are not aware of a systematic investigation of whether any of these models is at all likely to describe rankings in actual vote-casting processes.

We consider, for the case of three candidates, ten models that have been proposed by others as well as two new models. We evaluate these twelve models with two sets of voting data. Our first data set was assembled by Nicolaus Tideman in 1987 and 1988, and it consists of individual ballot information for 87 elections that we transform into 883 three-candidate elections.[1] Our second data set consists of 913 three-candidate "elections" that we construct from the "thermometer scores" that are part of the surveys conducted by the American National Election Studies (ANES).[2] The results that we obtain from these two rather different

[1] These data have been analyzed previously by Feld and Grofman (1990 and 1992), Felsenthal et al. (1993), Felsenthal and Machover (1995), Tideman and Richardson (2000), Regenwetter et al. (2002), and Tideman (2006).

[2] ANES survey data have been used in several previous analyses of voting. For example, Chamberlin and Featherston (1986) use scores from ANES surveys administered in 1972, 1974, 1976, and 1978 to construct combinations of four candidates. Regenwetter et al. (2002 and 2003) analyze the thermometer scores of the three major candidates in the four ANES surveys administered in 1968, 1980, 1992, and 1996 to construct combinations of three candidates. Our method of constructing three-candidate "elections" is the same as that in these earlier analyses.

data sets are very consistent; they indicate that the combination of a spatial model of voting to describe the distribution of the vector of expected shares and a multinomial model to describe the distribution of the vector of actual shares fits the observed results of three-candidate elections much better than any of the models that have so far been used in theoretical analyses of voting events, and well enough that it may be difficult to devise an alternative model that would fit actual election data significantly better.

While the spatial model fits the observed data very well, it is too complex to permit the type of calculation of the probabilities of voting events that theorists have undertaken so far (as, for example, in Gehrlein, 2002). Thus we envisage using this model instead for Monte Carlo simulations, to generate data that have the same characteristics as data from actual elections. Simulating several million such elections will permit us to estimate the probabilities of various voting events. Such simulations would be unnecessary if there were enough data from actual elections in which voters rank the candidates to determine the frequencies of rare voting events with a reasonable degree of accuracy. But there are not nearly enough ranking data for this task. For example, Brams and Fishburn (2001) and Saari (2001) use data from a single election—the 1999 election for president of the Social Choice and Welfare Society—to illustrate and analyze the properties of different voting methods. Chamberlin and Featherston (1986) analyze data from five presidential elections of the American Psychological Association, while Regenwetter et al. (2002) use data from 12 elections for positions in professional organizations. Tideman's data set of 87 elections is one of the largest data sets of elections with multiple candidates that voting analysts have used. Even our two data sets with 883 and 913 three-candidate elections cannot provide reliable information about the frequencies of rare voting events. But if we are able to infer the underlying model of vote-casting from these elections, then data simulated under this model can reveal the frequencies of voting events of interest.

The remainder of this chapter is organized as follows: in section X.2 we formalize the two-part model of vote casting and introduce the twelve statistical models of the distribution of the expected shares. In section X.3 we explain our strategy for assessing the accuracy of these models, and we describe our data and report the results of our statistical analysis in section X.4. In section X.5 we illustrate the practical relevance of our analysis: we use the different models to predict the frequencies of voting cycles ("Condorcet's paradoxes") in our two data sets and show that several popular models come nowhere close to predicting the observed frequencies. We conclude in section X.6.

X.2 A statistical model of vote-casting

Consider an election with m candidates, in which each of n voters submits a strict ranking of the candidates. There are $m!$ possible strict rankings, and n_r voters submit ranking r, $r = 1, \ldots, m!$. Let the discrete random variable N_r describe the frequencies with which $n_r = 0, \ldots, n$ voters submit ranking r, with $\Sigma n_r = n$, and let $N = \{N_1, \ldots, N_{m!}\}$ be a random vector defined on the N_r. Define p_r as the expected share of ranking r among the n submitted ballots, with $\Sigma p_r = 1$ so that $p = \{p_1, \ldots, p_m!\}$ is a vector of length $m!$ of expected shares. Let P be a random vector of length $m!$ that is defined on the collection of feasible p. A statistical model of vote-casting consists of specifications of both N and P.

Previous analyses of the frequencies of voting events have focused predominantly on specifying P. The requirement that $\Sigma p_r = 1$ implies that the support of all permissible models of P is contained in the unit $(m! - 1)$-simplex. Some models assume that P is defined on the entire $m!$-simplex, while others assume that P is defined on a strict subset of the simplex. In the following subsection, we introduce twelve models of P and discuss and compare their properties. In subsection

X.2.2, we introduce two intuitive contenders for the distribution of N for a given realization of P.

X.2.1 A statistical model of P

Statistical models of P are mappings from the unit $(m! - 1)$-simplex to $[0, \infty]$. They differ in the probability density or probability mass that they assign to different p as well as in the subsets of the unit $(m! - 1)$-simplex that form the support of the mapping. It is straightforward to describe the differences among the probability structures but less straightforward to illustrate the differences in support for general m. However, it is customary in the theoretical literature of voting events to restrict the number of candidates, and many theoretical results are available only for elections with three candidates. We continue this tradition and restrict our analysis in this chapter to elections with $m = 3$; this permits us to represent the corresponding 5-simplex (or hexateron) as a three-dimensional octahedron, which we use to derive intuitive graphical illustrations of the differences in the support of different models. In this octahedron, each of the six vertices represents a vector p with one probability equal to 1 and the remaining five probabilities equal to zero, each of the 15 edges (including the three virtual edges connecting pairs of opposite vertices) represents a vector p with four probabilities equal to zero while the remaining two p_r sum to 1, and each of the 20 (real and virtual) faces represents a vector p with three probabilities equal to zero while the remaining three p_r sum to 1. The point at the center of the octahedron represents the vector of equal probabilities with each $p_r = 1/6$. Although such a three-dimensional representation of a five-dimensional space cannot distinguish among all vectors of probabilities permitted by the 5-dimensional space, it is nevertheless sufficient to illustrate the differences in the support of all but one of the models that we analyze. To simplify the exposition, we label the three candidates A, B, and C, and order the six rankings {ABC, ACB, CAB, CBA, BCA, BAC} so that ABC is ranking $r = 1$, ACB is ranking $r = 2$, and so on.

We classify the statistical models that describe the distribution of P as being in one of four categories. Models in the first category have support of the entire 5-simplex and accept every feasible vector p as a potential source of elections. Models in categories 2 through 4 assign zero probability as the source of an election to all points in the simplex except for subsets of Lebesgue measure zero that contain the subspaces of permitted probability vectors. Models in category 2 are of zero dimensionality and consist of either a single point or a single set of symmetric points within the simplex. Models in this category are described by discrete probability distributions rather than continuous probability distributions. Models in categories 3 and 4 are of higher dimensionality; those in category 3 are specified by linear restrictions on the unit simplex, while those in category 4 impose non-linear restrictions on the unit simplex. We describe each model below and summarize the properties of all twelve models in table X.1.

X.2.2.1 Models whose support is the entire unit simplex

Our first model, the Impartial Anonymous Culture (IAC), was proposed in Kuga and Nagatani (1974) and Gehrlein and Fishburn (1976a). This model assumes that all points within the 5-simplex are equally likely. Figure X.1a shows the support of IAC—the entire octahedron. Several voting theorists have used IAC to calculate the probability that specific voting events will occur,[3] but they generally emphasize that they do not necessarily believe that equally likely probabilities of

---

[3] For example, Saari (1990) uses this assumption to analyze the probability of strategic voting under different voting methods, Gehrlein (2002) uses it to analyze the probability of observing Condorcet's paradox, and Cervone et al. (2005) use it to analyze the probability that a Condorcet candidate, if it exists, will win the election.

strict rankings describe vote-casting.[4] We are not aware of any formal test of the statistical adequacy of the IAC assumption.

Several variations on IAC were developed specifically to examine the probability of observing Condorcet's paradox, and we consider three of them. Our second model, $IAC_b(k_b)$ was introduced in Gehrlein (2004). As with IAC, the support of $IAC_b(k_b)$ is the entire simplex, but the probabilities of $IAC_b(k_b)$ differ in a subtle way from those of IAC. Assume that one candidate, say C, is ranked last no more often than either A or B, so that the frequency with which C is ranked last is a real number k_b between 0 and 1/3. $IAC_b(k_b)$ assumes that the probability of ABC is distributed uniformly on the interval $[0, k_b]$, and that the probability of BAC equals k_b minus the probability of ABC. The probabilities of the four other rankings ACB, BCA, CBA, and CAB are distributed uniformly on the tetrahedron whose components sum to $1 - k_b$. Thus while the components of P are determined simultaneously under IAC, they are determined sequentially in two steps under $IAC_b(k_b)$. Lepelley (1995) considers the case of this model when $k_b = 0$, so that the support of the model is limited to a subset of the unit simplex. We consider this limiting case $IAC_b(0)$ below as model seven.

Gehrlein (2006) describes two variations on $IAC_b(k_b)$ that he calls $IAC_t(k_t)$ and $IAC_c(k_c)$. $IAC_t(k_t)$ assumes that one candidate is ranked first no more often than the other two candidates, while $IAC_c(k_c)$ assumes that one candidate is ranked in the middle no more often than the other two candidates.[5] In both models, the probabilities of the six rankings are determined analogously to those in $IAC_b(k_b)$. We analyze $IAC_t(k_t)$ and $IAC_c(k_c)$ as models three and four. The exact values of p in models 2 – 4 depend on the unknown parameters k_b, k_t, and k_c. In appendix X.1 we describe how we use our two data sets to estimate the parameters of these three models as well as the models described below.

---

[4] See, for example, Gehrlein (2002, p. 169) and Cervone et al. (2005, p.182).

[5] $IAC_b(k_b)$ measures the proximity of voter preferences to the case of single-peakedness (which occurs at $k_b = 0$), while the other two representations measure the proximities to "single-troughness" and "single-centeredness."

Models 1 – 4 have each five degrees of freedom per election, which implies that they are able to describe perfectly every observable vector *p*. Thus it is not possible to evaluate their accuracy through a likelihood of the realization of *N* given the most likely *p* because these models assign the same (or quite similar likelihoods for models 2 – 4) to all possible vectors *p*. In section X.3.1 we describe our strategy for assessing the accuracy of such models.

X.2.2.2 Models whose support is composed of 0-dimensional subsets of the unit simplex

The most restrictive of the models that restrict the support to a proper subset of the unit simplex is our fifth model, the impartial culture (IC). This model assumes that the support of *P* is a single point at the center of the simplex where all rankings are equally likely, or p_r = 1/6 for *r* = 1, …, 6, so that IC has no degrees of freedom. Figure X.1b shows IC's support as the octahedron's center. Its computational simplicity made IC popular in early Monte Carlo studies (see, for example, Campbell and Tullock, 1965), but there is now considerable empirical evidence that IC does not describe actual elections (see Regenwetter et al., 2006, for a summary).

Our sixth model, which was proposed in Chamberlin and Featherston (1986) and which we call Unique Unequal Probabilities (UUP), assumes that in every election, each candidate occupies a specifiable ranking niche (first, second, etc.), and that for each possible ranking of the candidates described by these niches, there is a constant probability that this ranking will be used by a voter. UUP restricts the support of *P* to a set of six points, one for each permutation of the rankings. Figure X.1c shows the support of UUP—six points in symmetric locations in the octahedron. Unlike IC, UUP does not specify the values of the six probabilities, so it has five degrees of freedom that determine the six probabilities p_r. These five degrees of freedom, constant across elections, can be estimated. Note that UUP has five degrees of freedom to fit all elections, while models 1 – 4 have five degrees of freedom for each election. Because all six probabilities can be equal, UUP includes IC as a limiting case.

X.2.2.3 Models whose support is more than 0-dimensional and is specified by linear restrictions on the unit simplex

Our seventh model is the limiting case of $\mathrm{IAC}_b(k_b)$ with $k_b = 0$ that was proposed in Lepelley (1995). We refer to this model as SPP because it ensures single-peaked (group) preferences, making a voting cycle impossible. The support of SPP consists of all vectors p with either $p_4 = p_5 = 0$ (A is never ranked last), $p_2 = p_3 = 0$ (B is never ranked last), or $p_1 = p_6 = 0$ (C is never ranked last). Each of these restrictions eliminates two rankings, implying that the remaining possibilities can be viewed as a tetrahedron; if we could see in five dimensions, then the support of SPP would be three tetrahedrons, of which any two share an edge where four rankings are assigned zero probability. Figure X.1d depicts this support in three dimensions.

Our eighth model, the Dual Culture (DC) proposed in Gehrlein (1978), assumes that the probabilities of opposite rankings are equal, that is, $p_1 = p_4$, $p_2 = p_5$, and $p_3 = p_6$. In contrast, our ninth model, the Uniform Culture (UC) proposed in Gehrlein (1982), assumes that the probabilities of neighboring rankings are equal, that is, $p_1 = p_2$, $p_3 = p_4$, and $p_5 = p_6$. Because neither model specifies the probabilities of any of the three pairs, each model has three degrees of freedom per election. Each set of three equalities specifies a plane in the 5-simplex. Both the support of DC and the support of UC can be represented by a triangle determined by the midpoints of three edges of the octahedron. These triangles appear to be coplanar in the octahedron (see figure X.1e), but in five dimensions their only common point is their centers where all probabilities equal 1/6. Thus both models include IC as special case.

X.2.2.4 Models whose support is a curved subset of the unit simplex

As tenth and eleventh models, we investigate two models for which the Borda voting method and the Condorcet voting method, respectively, are maximum likelihood estimators. The “Borda model” supposes that there is a “best candidate” and, for every candidate, evaluates the evidence that this candidate is best. Such evidence is measured by the number of times a candidate is ranked second plus twice the number of times the candidate is ranked first. Conitzer and Sandholm (2005) show that this measure of the evidence is optimal only if the probabilities for both rankings with the best candidate first are equal, the probabilities for both rankings with the best candidate second are equal, and the probabilities for both rankings with the best candidate third are equal. That is, $p_1 = p_2$, $p_3 = p_6$, and $p_4 = p_5$ if A is best, $p_1 = p_4$, $p_2 = p_3$, and $p_5 = p_6$ if B is best, and $p_1 = p_6$, $p_2 = p_5$, and $p_3 = p_4$ if C is best. In addition, the probabilities for the pairs of rankings must follow a geometric sequence, so that

$$p_r = c_1 e^{\beta w_r} \qquad \text{(X.1)}$$

where $w_r = 0, 1, 2$ denotes 1 minus the position that the “best candidate” occupies in ranking r (the ranking’s contribution to the candidate’s Borda score), β is a constant, and

$$c_1 = 1/\left(2\sum_{i=0}^{2} e^{i\beta}\right) \qquad \text{(X.2)}$$

to ensure $\Sigma p_r = 1$. Thus the Borda model has one degree of freedom (β) per election. Because the Borda model does not distinguish between rankings that rank a given candidate in the same position but differ in the positions they assign to the other candidates, its support includes the midpoints of the three edges of the octahedron joining rankings that list a given candidate first (where $\beta = \infty$), but the support includes none of the vertices. The three curved lines in figure X.1f that start at the center of the octahedron (where $\beta = 0$; here the Borda model nests IC) and end midway between the pairs of vertices where a given candidate is first ($\beta = \infty$) depict the support of the Borda model for $0 < \beta < \infty$. The differences among the three equality restrictions imply that these three curved lines lie in three different planes.

Analogous to Conitzer and Sandholm's (2005) derivation of the Borda model, we define our eleventh model so that the Condorcet voting method is a maximum likelihood estimator of the ranking that is most favorable in terms of the statistical model that we assume has generated the election data (the "correct ranking"). While the Borda voting method assigns a score to each of the m candidates, the Condorcet voting method assigns a score to each of the $m!$ possible rankings.[6] Let r^* be the correct ranking and let $n_{rr^*} = 0, \ldots, 3$ be the number of pairs of candidates that are ranked the same in ranking r and ranking r^*. The Condorcet model specifies the components of p as

$$p_r = c_2 e^{\gamma n_{rr^*}} \qquad \text{(X.3)}$$

[6] Condorcet's explanation of his method (Condorcet, 1785) was opaque and contained errors; Kemeny (1959) proposed the same voting method in the twentieth century, and Young (1988) explained how Condorcet's intention could be understood despite his errors.

where γ is a constant and

$$c_2 = 1 / \left(\sum_{i=0}^{3} f(i,3) e^{i\gamma} \right) \tag{X.4}$$

to ensure $\Sigma p_r = 1$, where $f(i, 3)$ is the frequency distribution of Kendall's τ.[7] Like the Borda model, the Condorcet model has one degree of freedom (γ) per election.

The six corkscrew-shaped lines in figure X.1g that start at the center of the octahedron (where the Condorcet model nests IC at $\gamma = 0$) and end at the six vertices ($\gamma = \infty$) depict the support of the Condorcet model for $0 < \gamma < \infty$. The actual support in the 5-simplex is six elongated corkscrews that each span a three-space specified by two opposite vertices (for example, ABC and CBA) and the two points that are midway between pairs of vertices whose orderings both differ from one of the opposite vertices by a permutation of one pair of adjacent candidates (for example, BAC and ACB both differ from ABC by such permutations).

None of the 11 models discussed so far are based on a stated belief that the associated distributions of P might actually describe rankings in actual elections. IAC, IC, UUP, DC, and UP assume that various components of p are equally likely, for the sake of algebraic tractability. $\text{IAC}_b(k_b)$, $\text{IAC}_t(k_t)$, $\text{IAC}_c(k_c)$ and SSP seek to describe rankings that have meaningful interpretations for the problem of defining probabilities of observing Condorcet's paradox. The Borda and Condorcet models are rationalizations of claims about how one ought to determine the winner in an election.

In contrast, our final model, the spatial model of voting, is based on plausible models of distributions of voter and candidate characteristics. As in other spatial models of voting, our spatial model assumes that voters care about the "attributes"

---

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

of candidates; these attributes form a multi-dimensional "attribute space."[8] 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 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.[9] Voters' indifference maps are defined in candidate space through their definitions in attribute space.

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 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.[10]

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

---

[8] See Davis et al. (1970), and Enelow and Hinich (1984 and 1990).

[9] 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 $m > 3$.

[10] 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.

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, T. For the voters' ideal points in each sector, the distances to the locations of the three candidates have a unique rank order. These rank orders are indicated in figure X.2, 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.[11] These six integrals determine the 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. If O is exactly at the triangle's circumcenter T, then the spatial model coincides with our eighth model, DC, that assumes $p_1 = p_4$, $p_2 = p_5$, and $p_3 = p_6$. The spatial model has four degrees of freedom per election, and in appendix X.1 we explain how we parameterize the model.

The support of the spatial model in the 5-simplex is sufficiently complex to make it difficult to represent it in two or three dimensions, but we can offer some insights. Every vertex and every edge of the 5-simplex is included in the support of the spatial model. Of the 20 faces of the 5-simplex, each specified by three vertices, 18 are included in the support of the spatial model. The two faces that are not included in the support of the spatial model are the ones that are defined by rankings that form a voting cycle: ABC, BCA, CAB and CBA, BAC, ACB. There are 15 (three-dimensional) cells in the 5-simplex, each specified by four vertices and spanning a three-space on the "surface" of the 5-simplex. Of these 15 cells, nine are included in the support of the model and six are not. The nine that are included in the support are the ones for which the two vertices that are not among

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

the four in the defining set differ by a permutation of the positions of two candidates.

Within the 5-simplex, the support of the spatial model consists of six curved hyper-cells (four-spaces), that is, one curved hyper-cell for each ordering of the candidates. Figure X.3 offers some insights into the shape of these hyper-cells. Each triangle in the figure represents the subset of the 5-simplex that is consistent with fixed shares for three rankings. The largest triangle imposes the restrictions $p_{ABC} = p_{ACB} = p_{CAB} = 1/6$, and shows the intersection of the support of the spatial model with the 2-simplex consistent with these restrictions (formed by $p_{CBA} + p_{BCA} + p_{BAC} = 1/2$). For the middle triangle p_{CAB} is increased to 1/3 and the intersection of the support of the spatial model with the plane described by $p_{CBA} + p_{BCA} + p_{BAC} = 1/3$ is plotted as the curve in this triangle. Finally, for the smallest triangle, p_{CAB} is increased to 1/2 and intersection of the support of the spatial model with the plane described by $p_{CBA} + p_{BCA} + p_{BAC} = 1/6$ is plotted as the curve in this triangle. The combination of the three curves suggests that, if we increased p_{CAB} continuously from 0 to 2/3, we would be looking at a saddle-shaped figure, from the front of the saddle. The saddle we show is symmetric because $p_{CBA} = p_{BCA}$, and it tilts to one side if $p_{CBA} \neq p_{BCA}$.

The spatial model nests DC (when the locations of T and O coincide) and thus IC, and we confirmed empirically that it also nests the Borda model but not the Condorcet model.

X.2.2 A statistical model of N

Two intuitive contenders to describe the distribution of N for given n and p are the multinomial distribution and the multivariate Pólya distribution. The density function of the multinomial distribution is

$$f(n_1,\ldots,n_{m!};n,p_1,\ldots,p_{m!}) = \frac{n!}{\prod_{r=1}^{m!} n_r!}\prod_{r=1}^{m!} p_r^{\,n_r}\,, \qquad \text{(X.5)}$$

with first two moments $E[N_r] = np_r$, $Var[N_r] = np_r(1 - p_r)$, and $Cov[N_r, N_s] = -np_rp_s$. An intuitive way of motivating the multinomial distribution is to assume that the vector p describes the probabilities with which a voter submits any of the six rankings, that these probabilities are the same for all voters, and that voters submit their votes independently, in which case the N_rs follow a multinomial distribution. Note that the multinomial distribution has no unknown parameters besides n and the p_rs, which implies that the model of P must provide a complete explanation of the expected vote shares as well as their variation.

The multivariate Pólya distribution can be motivated by relaxing the assumption that voters submit their votes independently. Consider the possibility that, if one voter submits a particular ranking r, the probability that the next voter will submit the same ranking increases while the probabilities that this voter will submit any of the other rankings decrease. If it is true that the probabilities of observing first ranking r and then ranking s are identical to observing first ranking s and then ranking r, then the probabilities behave like those associated with urns to which balls are being added. In such a case, the distribution of N is described by the multivariate Pólya distribution with density function

$$f(n_1,\ldots,n_{m!};n,p_r,\ldots,p_{m!},\delta) = \frac{n!}{\prod_{r=1}^{m!} n_r!}\frac{\Gamma(\delta)}{\Gamma(n+\delta)}\prod_{r=1}^{m!}\frac{\Gamma(n_r+\delta p_r)}{\Gamma(\delta p_r)} \qquad \text{(X.6)}$$

whose first two moments are $E[N_r] = np_r$, $Var[N_r] = np_r(1 - p_r)\Psi$, and $Cov[N_r, N_s] = -np_rp_s\Psi$, where $\Psi = (n + \delta) / (1 + \delta)$.[12] As δ approaches infinity, Ψ approaches 1, and the multivariate Pólya distribution converges to the multinomial distribution. This genesis of the multivariate Pólya distribution suggests that $1/\delta$ is a measure of the dependence between voters.

The multivariate Pólya distribution can be derived alternatively by relaxing the assumption that the vote-share vector of the multinomial distribution is deterministic. The assumption of a deterministic vote-share vector might not be acceptable if the random vector P is specified on a strict subset of the $m!$-simplex. Models 5 – 12 that we describe in section X.2.1 assume that P is defined on either spots or lines of zero width and thus assign zero probability to all points of the simplex except for the subsets of Lebesgue measure zero that contain the spots and lines of zero width. For example, our fifth model, IC, assumes that P describes a single spot at the center of the simplex, and IC therefore does not provide good descriptions of observed vote-casting processes with variances of the vote shares around the center of the simplex that differ from $5n/36$ in the three-candidate case. However, if the vote-share vector of the multinomial distribution is not deterministic but rather a draw from a random vector Q, then the resulting model of vote-casting can accommodate probability vectors that are outside the support of a particular model of P. A natural assumption is that Q follows an $m!$-variate Dirichlet distribution with parameter vector δp, where δ is inversely proportional to the variances of Q. Compounding the multinomial distribution with the Dirichlet distribution yields the multivariate Pólya distribution with density function X.6. This genesis suggests that δ indicates how well a model of P that is defined on a strict subset of the $m!$-simplex describes observed vote-casting processes: the larger the value of δ, the smaller are the variances of Q and the less additional permitted but unexplained variation in the p_rs is necessary to fit the resulting model of vote-casting to actual elections.

---

[12] See Mosimann (1962, pp. 67 – 68).

X.3 Model evaluation

X.3.1 Evaluation of models with fewer than five degrees of freedom per election

For models with fewer than five degrees of freedom per election (models 5 – 12, whose support is a proper subset of the 5-simplex), we use likelihood calculations to assess how well the model can explain the vectors of vote shares that we observe in actual three-candidate elections, taking account of the fact that the models differ in their degrees of freedom. We derive the likelihoods by linking the vector p that is predicted by the model of P to the observed number of votes for each of the six rankings, assuming that these rankings follow either a multinomial or a multivariate Pólya distribution. For a set of elections whose outcomes are independent of each other, the likelihood function is proportional to the product over all elections of density function X.6 or X.7. For each model, we estimate the unknown parameters (if any) by maximizing the likelihood function over the observed election data. For nested models (for example, the Borda model nested in the spatial model or DC nested in the spatial model), a likelihood ratio test indicates which model yields a better fit of the data, given their different degrees of freedom. Nested as well as non-nested model can be compared by the Akaike and Bayesian information criteria (AIC and BIC) that account for differences in the degrees of freedom.[13] Models with lower values of AIC or BIC use degrees of

[13] These criteria are determined as AIC = – 2ln(L) + 2d and BIC = –2ln(L) + $d\,\overline{\ln(N_e)}$, where d is the total number of degrees of freedom, L is the maximum value of the likelihood function, and $\overline{\ln(N_e)}$ is the mean value of the log of the

freedom more efficiently in describing the data than models with higher AIC or BIC.

X.3.2 Evaluation of models with five degrees of freedom per election

The four models of *P* with five degrees of freedom per election (whose support is the entire 5-simplex) assign the same or quite similar likelihoods to all possible vectors *p*. This makes it impossible to evaluate their accuracy on the basis of the likelihood functions described above.

To assess these models, we use their assumptions to simulate multiple "elections" and determine whether these simulated data have the same statistical properties as observed election data. For each model of *P* whose support is a proper subset of the simplex, we calculate the mean (multi-dimensional Euclidean) distance from an observed vector of vote shares to the closest vector that this model permits as the source of the election. We measure the mean distance σ as the mean square root of the sum of the squared differences,

$$\sigma = \frac{1}{E}\sum_{e=1}^{E}\sqrt{\sum_{r=1}^{6}(p_{re}-s_{re})^2}\ , \qquad \text{(X.7)}$$

where E is the number of elections, s_{re} is the observed vote share of ranking r in election e, and p_{re} is the corresponding vote share predicted by the respective model. We place two subscripts on σ. The first indicates the source of the data, O for observed elections, and a number for a simulation with the corresponding model. The second subscript indicates which of models 5 – 12 was used to meas-

number of voters in an election. Thus BIC imposes a heavier penalty for the use of degrees of freedom than AIC.

ure distance to the nearest outcome permitted by a model. We compare σ_{ij}, i = 1, . . ., 6, 8, . . ., 12, j = 5, 6, 8, . . ., 12, computed from the simulated data, with the σ_{Oj}, j = 5, 6, 8, . . ., 12, computed from observed elections. If σ_{ij}, for any j, differs significantly from σ_{Oj}, then this is evidence that the simulated data differ in significant ways from the observed data and that model i is unlikely to be a good statistical description of the process that generated the observed data.

X.4 Empirical evaluation of the twelve models

X.4.1 The data

Our first data set consists of 84 elections that were administered by the Electoral Reform Society (ERS) and tabulated by Nicolaus Tideman in 1987 and 1988 and three elections that he included from another source. For each election, we have individual ballot information about the strict ranking of candidates provided by each voter (the ballots did not permit ties). The number of voters in these elections ranges from 9 to 3,422, with a mean of 410.5, and the number of candidates ranges from 3 to 29, with a mean of 8.7. Most voters ranked only some of the candidates in these elections. We use these ballots to construct all possible combinations of three candidates within an election, treating each combination as one election with three candidates. We use a ballot in such a three-candidate election only if the voter ranks at all three candidates, which yields a total of 20,087 three-candidate elections with between 1 and 1,957 voters. We found that elections with too few voters contain mostly random noise and do not provide much information about the model of vote-casting. We therefore limit our analysis to elections with more than 350 voters, and our ERS data set contains 883 three-candidate elections with between 350 and 1,957 voters, with a mean of 716.4 voters.

The data from these 883 three-candidate elections are not independent, but this lack of independence is unlikely to affect our conclusions.[14] It is possible that such three-candidate combinations that are derived from rankings of more than three candidates are qualitatively different from rankings of elections with exactly three candidates—for example, because it is often simpler to rank three candidates than a larger number of candidates. However, we have individual ballot data for only eight genuine three-candidate elections, which is not sufficient to draw reliable statistical inference about the appropriate model of vote-casting.

Because no reasonable voting method is immune to strategic voting, it is possible that the rankings in our data set reflect voters' strategic considerations. We therefore examine a second ranking data set that is derived from survey data rather than election data, because the strategic considerations of survey respondents are likely to differ from those of voters. We assemble our second data set from the "thermometer" scores that are part of 18 surveys conducted by the American National Election Studies (ANES) between 1970 and 2004. These surveys are conducted every two years, and participants are asked to rate politicians on a scale from 0 to 100 (the thermometer). We refer to these persons as "candidates."

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. As before, we construct all possible combinations of three candidates within a year, for a total of 913 three-candidate combinations

---

[14] Dependence among the elections requires that the likelihood function be calculated from the conditional, rather than the marginal, distributions of the six vote-share vectors over all elections. However, because we use the likelihood functions to compare the accuracy of each model with that of the other models, ignoring the dependence is unlikely to affect these relative assessments.

from all 18 surveys, with between 759 and 2,521 responses and a mean of 1,566.7 responses. For simplicity, we will refer to the survey respondents as “voters” and to the three-candidate combinations as “elections.” For each response, we rank the three candidates according to their thermometer scores, thereby eliminating any information about the intensity of the voter’s preferences. If a response yields a strict ranking of candidates, then we count it as one vote for this ranking. Voters are allowed to assign equal scores to different candidates, and we adopt the following intuitive rule of accommodating ties: If all candidates are tied, then we count the response as 1/6 vote for each ranking, and if two candidates are tied, then we count the response as half a vote for each of the two possible strict rankings that break the tie. Thus our adjusted data set consists of the total number of votes for each of the six strict rankings in each of the 913 three-candidate elections.

Table X.2 shows, for both data sets, the number of elections with different numbers of candidates and the average number of voters in the original data sets for elections with 350 or more voters, the number of three-candidate elections we could have extracted from each of these elections, the number of three-candidate elections that had 350 or more voters that we did extract, and the average number of voters in each of these three-candidate elections with 350 or more voters.

It is notable that both data sets have voting cycles—the 913 ANES surveys have four cycles (0.44 percent), and the 20,087 ERS elections have 476 cycles (2.37 percent). However, there are only 101 voting cycles (1.45 percent) among the 6,794 ERS elections with 21 or more voters, and only six voting cycles (0.68 percent) among the 883 ERS elections with 350 and more voters. Thus the frequency of voting cycles falls fairly quickly as the number of voters increases.

X.4.2 Assessment of the eight models with fewer than five degrees of freedom per election

SPP predicts that, in every election, one of the candidates will never be ranked last. The fact that our two data sets contain predominantly elections in which every candidate is ranked last by some voters is conclusive evidence against the empirical relevance of this model, at least for our two data sets. We therefore do not consider SPP in our further analysis. Table X.3 reports the log-likelihood values and the AIC and BIC as well as our estimates of the Dirichlet δs for the remaining seven models whose support is a proper subset of the 5-simplex. For both data sets, the three measures of accuracy agree about the relative ranking of these models. IC, UUP, and DC have the smallest log-likelihoods, the largest values of AIC and BIC, the smallest values of δ, and thus the lowest accuracy. The Borda model consistently has the fourth highest accuracy, while UC and the Condorcet model are the third and second most accurate.

All our measures of accuracy indicate that the spatial model describes the observed data much better than any of the other seven models. Likelihood ratio tests of the spatial model and the nested IC, DC, and Borda models indicate that the improvement in the likelihood justifies the spatial model's additional degrees of freedom. For all models, AIC and BIC also suggest that, by a wide margin, the spatial model provides the best description of the ERS elections and the ANES surveys, despite its much larger use of degrees of freedom. The estimate of δ of the spatial model is close to being infinite, which indicates that the strict spatial model provides a very good explanation of almost all of the variation among the observed vote shares, and that perturbing the predicted vote-shares in the manner described by the Dirichlet process provides no significant improvement in the spatial model's fit. In contrast, the estimated δs of the other models are very small in comparison, meaning that adding variation through the Dirichlet process beyond the variation explained by the respective model improves the fit of these models considerably.

X.4.3 Assessment of the four models with five degrees of freedom per election

We analyze next whether the election data are consistent with any of the four models with five degrees of freedom per election whose support is the entire 5-simplex, IAC, IAC$_b$(k_b), IAC$_t$(k_t), and IAC$_c$(k_c). We report only the results from the ANES data because the results from the ERS data are qualitatively the same. We calibrated the three parameters—the shares of last, first, and middle ranks of the candidates with the fewest last, first, and middle ranks—as k_b = 0.218, k_t = 0.201, and k_c = 0.226. The fact that these values differ notably from the value of 1/3 predicted by IAC is some evidence against the hypothesis that IAC has generated the observed ANES data.

As discussed in section X.3.2, the accuracy of these four models cannot be evaluated by calculating a likelihood of observed outcomes given p, because a model with five degrees of freedom can match the six observed shares of any three-candidate election. We therefore drew, for models 1 – 6 and models 8 – 12, 1,000 samples of probability vectors from the distribution on the unit 5-simplex specified by the model, and evaluated these samples in terms of σ_{ij}, j = 5, 6, 8, …, 12. If any of the four models with five degrees of freedom per election has generated the observed data, then we would expect that, for data simulated with that model, σ_{ij}, will be similar to σ_{Oj}, j = 5, 6, 8, …, 12.

Columns 1 of table X.4 shows the σ_{Oj}s that we calculated from the observed ANES data, and columns 2 – 5 show the corresponding values of σ_{ij} that we calculated from our simulations. The differences are highly significant. We obtained the same result with other measures of distance.[15] Our results mean that the vectors of vote shares in the ANES data are much more clustered than what is permitted by any of the four models that assign equal probabilities to either all possible vectors (in case of IAC) or vectors in which two probabilities sum to no more than a value less than 1/3 (in case of IAC$_b$(k_b), IAC$_t$(k_t), and IAC$_c$(k_c)). Specifically, our estimate of an effectively infinite δ for the spatial model implies that the vote-share vectors that generate elections are clustered extremely closely around the vote share vectors predicted by the strict spatial model. We obtained the same results when we calibrated the parameters of models 2 – 4 from the ERS data and compared the σ_{ij}s with the ERS σ_{Oj}s. Thus the hypothesis that our two data sets were generated by IAC, IAC$_b$(k_b), IAC$_t$(k_t), or IAC$_c$(k_c) cannot be sustained.

In columns 2 – 8 of table X.5 we repeat this exercise with data that we simulated with the seven remaining models whose support is a proper subset of the 5-simplex. Again we only report the comparisons with the ANES σ_{Oj}s. With the exception of the data simulated under the spatial model, the σ_{ij}s are significantly different from the σ_{Oj}s, which is yet further evidence that neither of our two data sets is likely to have been generated by any of these models. In contrast, the σ_{12j}s that we calculated from the data simulated under the spatial model are not significantly different from the σ_{Oj}s. The one exception is $\sigma_{12,\,12}$, which is significantly smaller than $\sigma_{O,\,12}$. This suggests that, even though the generating mechanism of the ANES data is likely to be very close to the spatial model, it is not identical. One possible explanation is that our simulation framework assumes that voters know all candidates, even though not all voters in the ANES data ranked every candidate. We found that it is possible to improve upon the spatial model by relating candidate recognition to the δ of the multivariate Pólya distribution, but we leave a more thorough investigation of this issue for future research.

---

[15] We weighted σ by its standard deviation and we also evaluated the ratios of

X.5 Cycles

As a final comparison of the contenders for the model of *P*, we examine how accurately each model predicts the observed occurrence of voting cycles in our two data sets. A common definition of a voting cycle is the absence of a strict pairwise majority rule winner (SPMRW), because there is a cycle (or at least a "semi-cycle"—with one or more ties) if no candidate beats all other candidates in pairwise comparisons. For IC and IAC, Gehrlein (2002) determined analytically the expected frequency of SPMRWs for elections with different numbers of voters. For IC he established that an SPMRW will exist in 58.34% of elections with 10 voters, 66.86% of elections with 20 voters, 73.46% of elections with 40 voters, and 91.23% of elections in the limit as the number of voters approaches infinity. For IAC he established that one can expect an SPMRW in 73.43% of elections with 10 voters, 81.99% of elections with 20 voters, 87.35% of elections with 40 voters, 91.05% of elections with 100 voters, and 93.75% of elections in the limit as the number of voters approaches infinity.

To estimate the frequency of SPMRWs for our models of *P*, we simulated one million elections each for IAC, IC, UUP, DC, UP, the Borda model, the Condorcet model, and the spatial model, and we recorded the number of SPMRWs in these simulated elections. To be able to compare our estimates with the observed frequency of SPMRWs in our two data sets, we had to account for the variations in the number of voters in the observed elections. To do so we recorded the numbers of voters in each of the 883 ERS elections and 913 ANES elections, and then simulated either 11,325 or 11,326 elections for each number of voters in the ERS data set (for a mean of 716.4 voters) and either 10,952 or 10,953 elections for each number of voters in the ANES data set (for a mean of 1566.7 voters). To assess

---

the likelihoods that we determined from the simulated and the observed data.

the differences among the models as the number of voters becomes either small or large, we also simulated ten sets of 1,000,000 elections each, with 10, 20, 40, 100, 500, 1,000, 10,000, 100,000, 500,000 and 1,000,000 voters per election.

Column 2 of table X.6 shows the share of elections with an SPMRW in the ERS and ANES data, and Columns 3 – 10 show the shares of elections with an SPMRW among the elections that we simulated with the eight models of *P*, for different numbers of voters. Consider first the two shaded rows with elections whose mean numbers of voters correspond to those of the ERS and ANES elections. We arranged the models of *P* so that models whose simulated shares are closer to the observed shares are further to the left. The predictions of the spatial model for the shares of elections with an SPMRW come closest to the observed shares, followed by the predictions of the Borda model, the Condorcet mode, UC, UUP, IAC, DC, and IC. (Although UUP predicts the share of SPMRWs for elections with a mean of 1566.7 voters slightly better than the spatial model, UUP's prediction for elections with a mean of 716.4 voters is much worse.) The shares predicted by the first five models are reasonably accurate, while the predictions of IAC, DC, and IC are much worse. This result is not surprising because the first five models all say that all elections will have SPMRWs as the number of voters approaches infinity, while IAC, DC, and IC all say that the limit of the share of elections with SPMRWs is less than one. That is, the first five models predict that voting cycles will occur very rarely, if at all, in elections with many voters, while IAC, DC, and IC predict that voting cycles will occur in more than 6% of all elections, even if the number of voters is very large. The fact that our simulations of IC and IAC for 10, 20, 40 and 100 voters yield frequencies of SPMRWs that correspond almost exactly to Gehrlein's analytic results suggests that our simulations are reliable. Thus the small number of voting cycles in our two data sets is further evidence that IAC, DC, and IC do not describe vote casting in actual elections.[16]

---

[16] Gehrlein (2002, p.189) reports analytic results for a model of *P* that we do not analyze here: the "maximum culture," or MC. He finds that only 90.83% of all elections generated under MC have an SPMRW as the number of voters ap-

Although the predictions of UC and UUP for the ERS and ANES elections are comparable to the predictions of the first three models, their predictions become notably different as the number of voters becomes either very small or very large. For elections with 10 voters, the predicted shares of UC and UUP differ by about ten percentage points from those of the spatial model. In addition, UUP predicts that every election has an SPMRW when there are 10,000 and more voters, while our simulations with the spatial model yielded about 500 elections with 10,000 voters without SPMRWs.

The shares predicted by the Borda and the Condorcet models are most similar to those of the spatial model, but even those two models underpredict the share of SPMRWs for elections with few voters and overpredict this share for elections with many voters; our simulations with 10,000 voters yielded fewer than 100 elections without SPMWR for the Borda model and fewer than 50 elections without SPMRW for the Condorcet model; much less than what is predicted by the spatial model.

Because the eight models make significantly different predictions about the number of voting cycles for elections with different numbers of voters, it is crucial to select the right model if one wants to make accurate predictions of the frequency of cycles in actual elections. The results of our analyses presented in this chapter suggest that the spatial model is likely to yield accurate predictions about the occurrence of voting events in actual elections, while all eleven other models are inadequate for this purpose.

X.6 Conclusion

---

proaches infinity. His result is strong evidence that the maximum culture also does not describe vote-casting in actual elections.

The starting point of our inquiry is the observation that no theoretical analysis of probability structures can tell us anything about the probability of observing vectors of rankings in actual elections. This is clearly not a new discovery. It parallels the mundane observation that no analysis in theoretical econometrics provides any information about the particular estimates that one obtains from analyzing specific data. The theoretical econometrics literature investigates the properties of different models that can describe data, but to analyze a specific data set, one needs to identify the model or the set of models that is most likely to have generated these data.

The relationship between our work and the theoretical literature on voting is like that between theoretical and empirical econometrics. The existing literature on voting is largely theoretical, in the sense that it does not seek to identify systematic patterns in ranking data from actual elections. The large literature on the probabilities of finding Condorcet cycles and the Condorcet efficiency of different voting rules has established that the results vary greatly across different models of P. But which of these models best describes the distribution of rankings in actual elections? The theoretical literature on voting cannot answer this question.

Our results suggest that a spatial model describes the statistical structure of P in actual elections much better than any other model that has been proposed so far, and so well that it may be difficult to find a model whose accuracy is significantly higher. We consider our result to be very encouraging, but more work needs to be done. For example, we draw our current conclusions on the basis of two data sets, one compiled from elections and the other from surveys. Our analyses suggest that the two data sets have somewhat different properties, but it is not clear whether these differences stem from their different sources or from the fact that the average ANES “election” has almost twice the number of voters than the average ERS election. Analyses of additional election data are necessary to answer this question and to determine the robustness of our results. We also focus exclusively on three-candidate elections, partly because this simplifies the exposition and makes it easier to relate our analysis to the previous literature, and partly because we are currently only able to evaluate the spatial model for three candidates. Extending our analysis to elections with more than three candidates will provide important insights about the general relative accuracy of the different models.

Our analysis applies to all inquiries into the frequency of rare voting events: the probability that strategic voting will alter the outcome of an election, the existence of dominant candidates, or the frequencies of voting paradoxes. Our framework makes it possible to develop realistic models of vote-casting for such analyses and thereby to improve significantly the accuracy of their predictions for actual elections. Such new inquiries into old questions are likely to yield interesting new insights.

X.7 References

Brams SJ, Fishburn P (2001) A nail-biting election. Soc Choice Welfare 18: 409-414

Campbell C, Tullock G (1965) A measure of the importance of cyclical majorities. Econ J 75: 300, 853-857

Cervone D, Gehrlein WV, Zwicker W (2005) Which scoring rule maximizes Condorcet efficiency under IAC. Theo Decis 58:2, 145-185

Chamberlin JR, Featherston F (1986) Selecting a voting system. J Polit 48:2, 347-369

Condorcet, MJAN, Marquis de (1785) Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix (Paris)

Conitzer V, Sandholm T (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, UK

Davis, OA, Hinich MJ, Ordeshook PC (1970) An expository development of a mathematical model of the electoral process. Am Polit Sci Rev 44: 426-448

DiDonato, AR, Hageman RK (1980) Computation of the integral of the bivariate normal distribution over arbitrary polygons. Naval Surface Weapons Center, Government Accession Number ADA102466

Enelow JM, Hinich MJ (1984) The spatial theory of voting: An introduction. Cambridge University Press

Enelow JM, Hinich MJ (1990) Advances in the spatial theory of voting. Cambridge University Press

Feld S, Grofman B (1990) Collectivities as actors. Ration Soc 2:4, 429-448

Feld S, Grofman B (1992) Who's afraid of the big bad cycle? Evidence from 36 elections. J Theor Polit 4:2, 231-237

Felsenthal, DS, Maoz Z, Rapoport A (1993) An empirical evaluation of six voting procedures: Do they really make any difference? Brit J Polit Sci 23:1, 1-27

Felsenthal DS, Machover M (1995) Who ought to be elected and who is actually elected? An Empirical Investigation of 92 Elections under Three Procedures. Elect Stud 14:2, 143-169

Gehrlein WV, Fishburn PC (1976) Condorcet's paradox and anonymous preference profiles. Public Choice 26: 1-18

Gehrlein WV (1978) Condorcet winners in dual cultures. Presented at the National Meeting of the Public Choice Society, New Orleans, LA

Gehrlein WV (1982) Condorcet efficiency and constant scoring rules. Math Soc Sci 2:123-130

Gehrlein WV (2002) Condorcet's paradox and the likelihood of its occurrence: Different perspectives on balanced preferences. Theo Decis 52:2, 171-199

Gehrlein WV (2004) Consistency in measures of social homogeneity: A connection with proximity to single-peaked preferences. Qual Quant 38: 147-171

Gehrlein WV (2006) Condorcet's paradox. Springer, Berlin and Heidelberg.

Good IJ, Tideman TN (1976) From individual to collective ordering through multidimensional attribute space. P R Soc Lon A 347: 371-385

Kemeny J (1959) Mathematics without numbers. Daedalus 88: 571-91

Kendall M, Gibbons JD (1990) Rank correlation methods. New York: Oxford University Press.

Kotz S, Balakrishnan N, Johnson NL (2000) Continuous multivariate distributions. Wiley, New York

Kuga K, Nagatani H (1974) Voter antagonism and the paradox of voting. Econometrica 42: 1045-1067

Lepelley D (1995) Condorcet efficiency of positional voting rules with single-peaked preferences. Econ Design 1: 289-299

Merrill S (1984) A comparison of efficiency of multicandidate electoral systems. Am J Polit Sci 28: 23-48

Mosimann JE (1962) On the compound multinomial distribution, the multivariate beta distribution, and correlations among proportions. Biometrika 49: 65- 82

Regenwetter M, Grofman B, Marley AAJ (2002) On the model dependence of majority preference relations reconstructed from ballot or survey data. Math Soc Sci 43: 451-466

Regenwetter M, Marley AAJ, Grofman B (2003) General concepts of value restriction and preference majority. Soc Choice Welfare 21: 149-173

Regenwetter M, Grofman B, Marley AAJ, Tsetlin IM (2006) Behavioral social choice: Probabilistic models, statistical inference, and applications. Cambridge University Press

Saari DG (1990) Susceptibility to manipulation. Public Choice 64: 21-41

Saari DG (2001) Analyzing a nail-biting election. Soc Choice Welfare 18: 415-430

Tideman TN, Richardson D (2000) Better voting methods through technology: The refinement-manageability trade-off in the single transferable vote. Public Choice 103:1-2, 13-34

Tideman TN (2006) Collective decisions and voting. Ashgate, Burlington VT

Truchon M (2006) Borda and the maximum likelihood approach to vote aggregation. Mimeo

Young HP (1988) Condorcet's theory of voting. Am Polit Sci Rev 82:4, 1231-1244

Appendix X.1: Model parameterization, parameter estimates, and simulations:

Two of our twelve models, IC and IAC, do not have any parameters that need to be estimated to either fit the model to observed voting data or to undertake Monte Carlo simulations. We do not calibrate SPP because the model assigns zero probability to the possibility that every candidate will be ranked last, and we observed numerous elections in both data sets in which every candidate is ranked last. This appendix describes our strategy for calibrating the parameters of the remaining nine models and for simulating elections with each of these models. We report the calibrated values for the ERS and ANES data sets in Table X.A1.

(1) $\mathrm{IAC}_b(k_b)$, $\mathrm{IAC}_t(k_t)$, and $\mathrm{IAC}_c(k_c)$: The support of these models is the entire 5-simplex and each model can be calibrated to describe any set of observed vote shares by setting p_r equal to the observed vote share q_r. We calibrate the parameters k_b, k_t, and k_c as the mean over all elections of the smallest shares by which a candidate is ranked either last (k_b), first (k_t), or second (k_c). To simulate elections we assume that each parameter follows a beta distribution over the interval [0, 1/3], whose mean and variance coincide with those that we observe in the actual elections. We draw a share k from this beta distribution, determine p_1 as a draw from a uniform distribution on $[0, k]$, set $p_2 = k - p_1$, draw p_3, p_4, p_5, and p_6 from the unit 4-simplex, and rescale these 4 values so that they sum to $1 - k$. We then use these six shares to draw the number of votes for each of the six rankings from density function X.5 or X.6.

(2) UUP: The five parameters of this model are five of the six vote shares (that sum to 1) that describe the expected ranking. These parameters are the same for all elections, and we calibrate these parameters by identifying the share vector p that maximizes the likelihood function X.5 or X.6 over all elections. The maximum value of the likelihood function also determines the fit of UUP. Note that the five parameters of UUP are constant across elections, while the parameters of

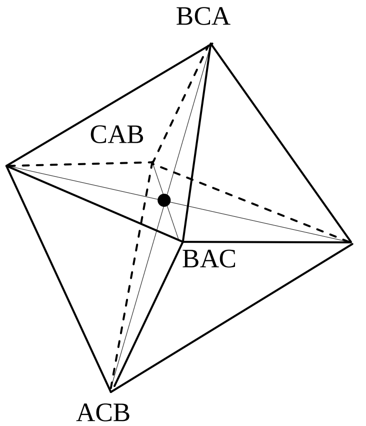
the other eight models are calibrated separately for each election. To simulate elections with UUP, we use the calibrated shares to draw six vote shares from the unit 6-simplex, and then use draws from density function X.5 or X.6 to determine the number of votes for the six rankings.

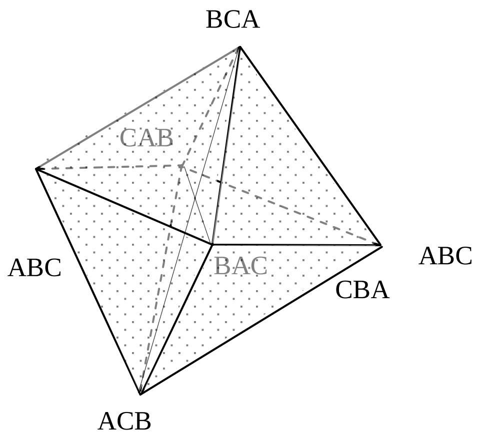
(3) DC and UC: Both models have two parameters each—two of the three pairs of probabilities of opposite (for DC) and neighboring (for UC) rankings. We fit these models to the data by using, for each election, the average value of the two observed vote shares of each pair of rankings as the predicted vote shares for that pair. To simulate elections with these models, we assume that the three pairs of shares follow a Dirichlet distribution whose mean and variance coincides with those of the shares that we observe in the actual elections. We draw three shares from the Dirichlet distribution that we use as input to draw six vote-shares from the unit 6-simplex, and then use the multinomial or multivariate Pólya distributions to draw the six N_rs.

(4) Borda and Condorcet: Each model has one parameter. For the Borda model the parameter β is the increase in log probability associated with an increase by one in the rank that a voter assigns to the "best" candidate. For the Condorcet model the parameter γ is the increase in log probability associated with a reduction of one adjacent-pair permutation in the difference between the "best" ranking and the one that a voter reports. We fit these models to the data by calibrating, for each election, the values of β and γ in equations X.1 and X.3, respectively, so that the resulting vector p maximizes the likelihood for the observed vote shares. To simulate elections with these models, we assume that β and γ follow a gamma distribution whose mean and variance coincides with those of the values of β and γ that we calibrated in the actual elections. We then draw a value of β or γ from the gamma distribution that we use to determine six shares from equations X.1 or X.2; we proceed as above to draw six vote shares from the unit 6-simplex and six N_rs from the multinomial or multivariate Pólya distribution.

(5) Spatial model: This model has four parameters, and figure X.A1 shows one way of using the four degrees of freedom (as in Good and Tideman, 1976). The intersection of the perpendicular bisectors T 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 T and O. The remaining degrees of freedom specify the angles β_1, β_2, and β_3 formed by the line $\overline{TO}$ and the three perpendicular bisectors. Thus any feasible set of values of the four degrees of freedom corresponds to a set of p_r. We calibrate the spatial model by placing, for each election, the borders between pairs of adjacent rankings and the distance in such a way as to create sectors that match the six probabilities p_r (the integrals over the triangular-shaped slices under the bivariate normal distribution) as closely as possible to the six observed vote shares, q_r.

To simulate elections with the spatial model, we assume that the three vote shares follow a Dirichlet distribution and that the distance between T and O follows a Weibull distribution; the means and variances of these distributions coincide with those of the parameters that we calibrated in each of the actual elections. We then use draws from the Dirichlet and the Weibull distributions to construct six shares, and proceed as above to draw six vote shares from the unit 6-simplex and six N_rs from the multinomial or multivariate Pólya distribution.

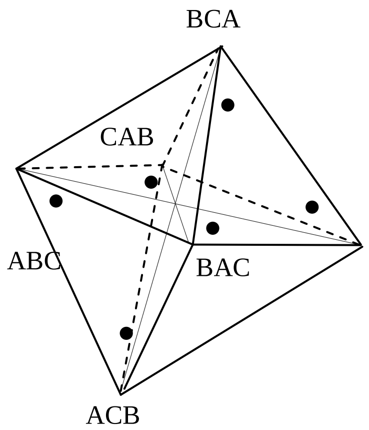


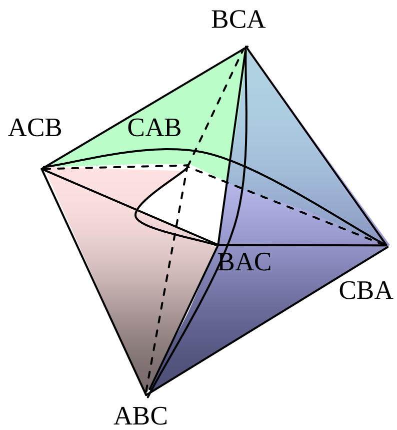


CBA

Fig. X.1a Three-dimensional representation of the support of IAC and variations

Fig. X.1b Three-dimensional representation of the support of IC





CBA

Fig. X.1c Three-dimensional representation of the support of UUP

Fig X.1d Three-dimensional representation of the support of SPP

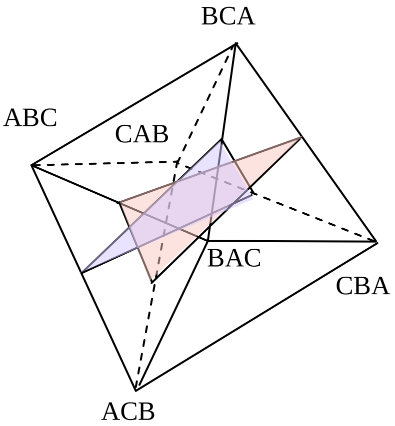


Fig. X.1e Three-dimensional representations of the support of DC and UC

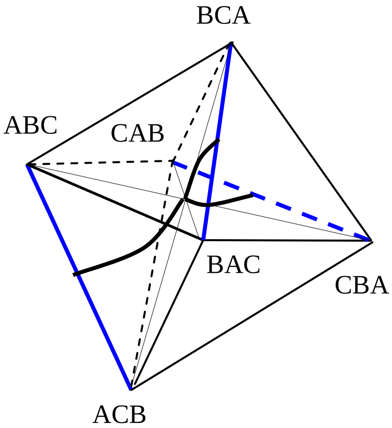


Fig.X.1f Three-dimensional representation of the support of the Borda model

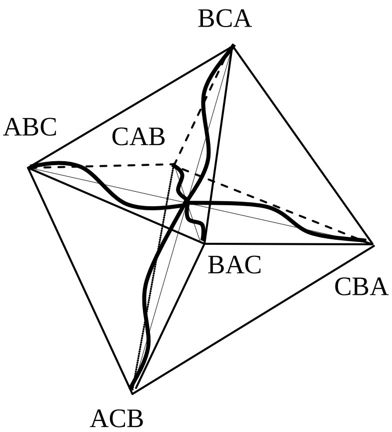


Fig. X.1g Three-dimensional

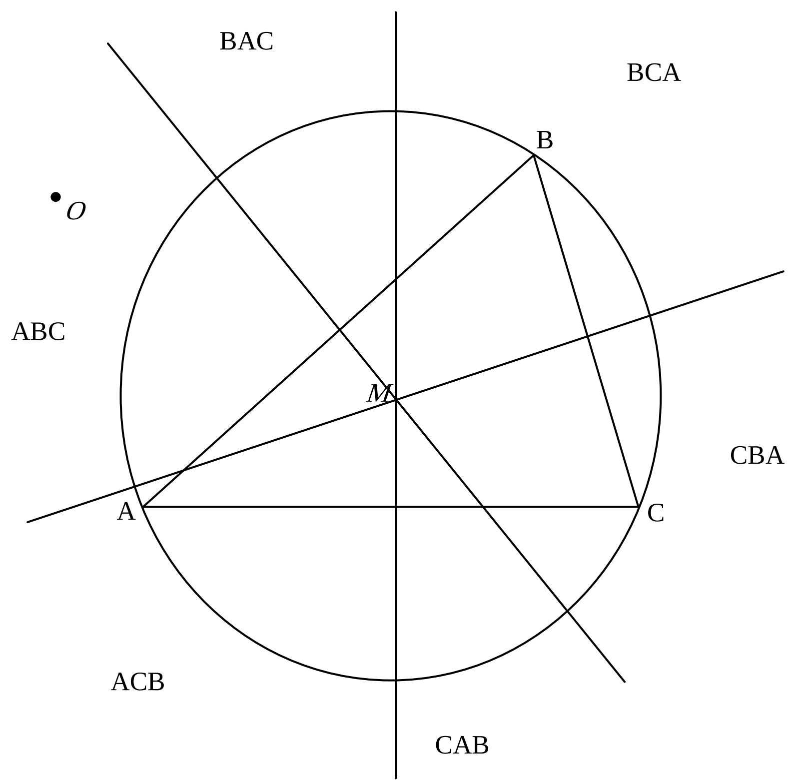

BAC
BCA
B
O
ABC
M
CBA
A
C
ACB
CAB

Fig. X.2 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, 1976, p. 372.)

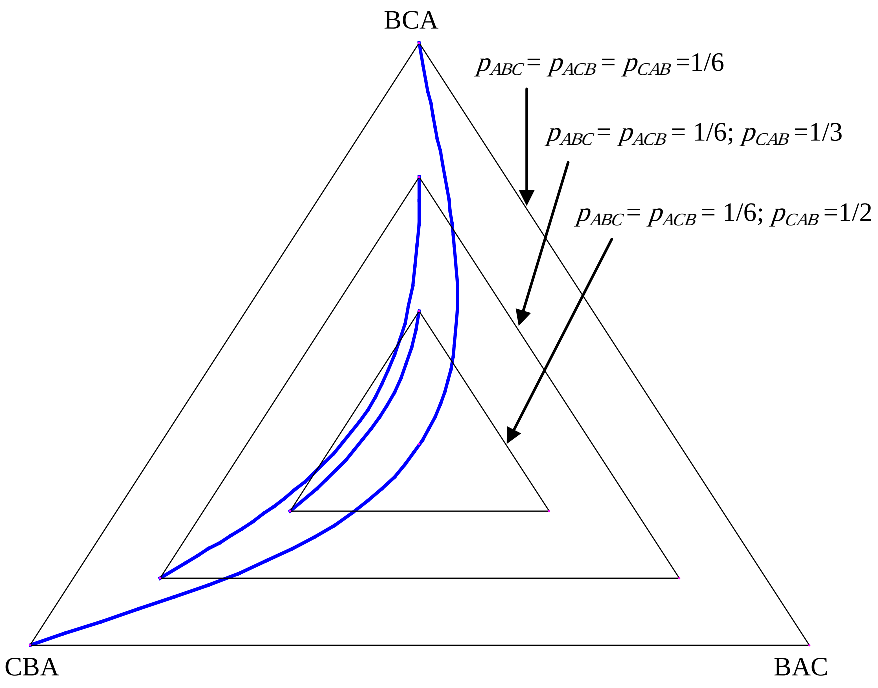

BCA
$p_{ABC} = p_{ACB} = p_{CAB} = 1/6$
$p_{ABC} = p_{ACB} = 1/6$; $p_{CAB} = 1/3$
$p_{ABC} = p_{ACB} = 1/6$; $p_{CAB} = 1/2$
CBA
BAC

Fig X.3 Two-dimensional representations of the parts of the support of the spatial model that are consistent with specified values for the shares of ABC, ACB, and CAB

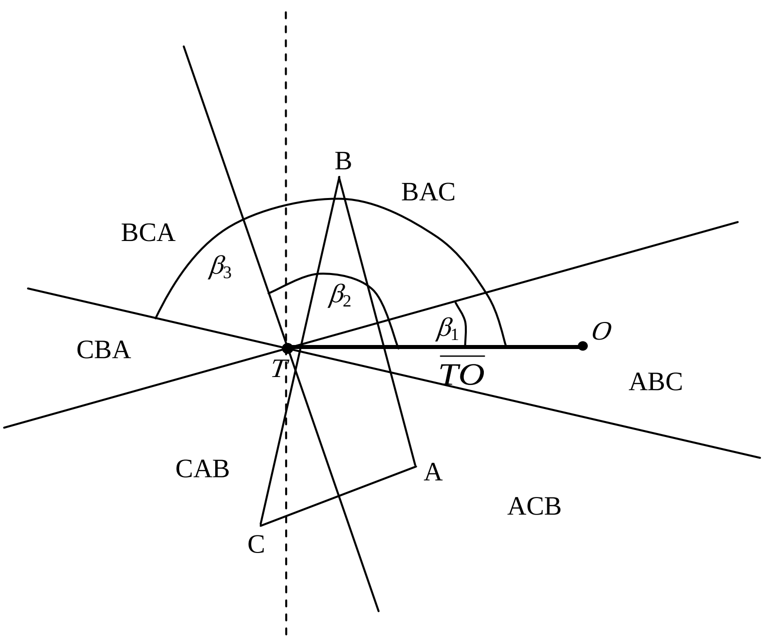


Fig. X.A1 The four parameters $\overline{TO}$, β_1, β_2, and β_3 that define a spatial model observation

Table X.A1: Parameter values calibrated from ERS and ANES data

Model →	IAC$_b$(k_b)	IAC$_t$(k_t)	IAC$_c$(k_c)	Borda	Condorcet
Data set ↓	k_b	k_t	k_c	β	γ
ERS	0.182	0.187	0.248	0.517	0.513
	(0.002)	(0.002)	(0.001)	(0.009)	(0.008)
ANES	0.218	0.201	0.226	0.358	0.360
	(0.002)	(0.002)	(0.002)	(0.006)	(0.006)

Model →	Spatial model			
		Angles of the perpendicular bisectors with the line $\overline{TO}$		
Data set ↓	$\overline{TO}$	β_1	β_2	β_3
ERS	0.597	0.548	1.549	2.560
	(0.009)	(0.011)	(0.011)	(0.011)

ANES	0.445	0.556	1.550	2.592		
	(0.008)	(0.012)	(0.015)	(0.012)		
		Corresponding angles between pairs of bisectors:				
		A_1	A_2	A_3		
ERS		1.130	1.000	1.011		
		(0.006)	(0.006)	(0.006)		
ANES		1.105	0.994	1.042		
		(0.010)	(0.011)	(0.010)		
Model →	UUP					
Data set ↓	p_1	p_2	p_3	p_4	p_5	p_6
ERS	0.294	0.167	0.127	0.108	0.120	0.184
ANES	0.233	0.150	0.140	0.148	0.160	0.169

Notes:

(1) The values in parentheses are standard errors.

(2) All values were calibrated from the multinomial model (the values calibrated from the multivariate Pólya model are very similar).

(3) The entries for UUP are the six shares, calibrated over all elections in the respective data set, which minimize the multinomial likelihood function.

The entries for the other six models are the means of the values that we calibrated for each election in the respective data set.

(4) For DC and UC, aggregation across elections is meaningless because of the arbitrariness of labeling a candidate A, B, or C, so we do not show any calibrated values for these two models.

Table X.1 Comparison of the twelve models of *P*

#	Model	Number of dimensions in subspace(s)	Parameters per election to be calibrated to fit the model to this election	Parameters to be calibrated to simulate data from this model	First proposal (of which we are aware) as a description of *P*
A. Models whose support is the entire 5-simplex:					
1.	IAC	5	5	0	Kuga & Nagatani (1974)
					Gehrlein & Fishburn (1976)
2.	$\text{IAC}_b(k_b)$	5	5	1	Gehrlein (2004)
3.	$\text{IAC}_t(k_t)$	5	5	1	Gehrlein (2006)
4.	$\text{IAC}_c(k_c)$	5	5	1	Gehrlein (2006)
B. Models whose support is one or more 0-dimensional subspaces of the 5-simplex:					

5.	IC	0	0	0	Campbell and Tullock (1965)
6.	UUP	0	5 (for all elections)	5	Chamberlin and Featherston (1986)
C. Models whose support is one or more more-than-0-dimensional subspaces, defined by linear restrictions:					
7.	SPP (IAC$_b$(0))	3	4	0	Lepelley (1995)
8.	DC	3	2	2	Gehrlein (1978)
9.	UC	3	2	2	Gehrlein (1982)
D. Models whose support is one or more more-than-0-dimensional subspaces, defined by nonlinear restrictions:					
10.	Borda	1	1	1	Conitzer and Sandholm (2005)
11.	Condorcet	1	1	1	This paper
12.	Spatial model	4	4	4	This paper

Notes:

(1) The four models in A can describe any set of observed vote-shares and the parameters equal the observed shares. We describe our strategy for simulating data under these models in the appendix.

(2) The 5 parameters of UUP are calibrated from all elections simultaneously, and the model's fit to any individual elections is assessed on the basis of these parameter values. The parameters of SPP, DC, UC, the Borda model, the Condorcet model, and the spatial model are calibrated for each election individually.

(3) The 5 parameters of UUP are constants in simulations from UUP. To simulate from any of the other models with unknown parameters, we assign distributions to all parameters and draw pseudo random numbers from these distributions that we use as inputs into the density functions X.6 and X.7.

Table X.2 Properties of the original ERS and ANES data sets and of the extracted three-candidate elections

	Number of candidates in the original elections (1)	Number of elections with 350 or more voters (2)	Average number of voters per election (3)	Potential number of three-candidate elections (with any number of voters) (4)	Number of three-candidate elections with 350 and more voters (5)	Average number of voters per three-candidate election (6)
ERS data:	3	0				
	4	6	614	24	4	1,005

	5	3	1,238	30	10	1,859	
	6	2	770	40	20	414	
	7	3	714	105	37	390	
	8	3	541	168			
	9	2	450	168			
	10	5	679	600	120	665	
	11	2	913	330	108	404	
	12 – 20	9	986	3,883	584	531	
	Total	35		5,348	883		
<u>ANES data:</u>	3	2	2,078	2	2	1,599	
	4	0					
	5	2	1,396	20	20	1,248	
	6	2	1,607	40	40	1,403	

7	5	1,858	175	175	1,540
8	3	2,184	168	168	1,699
9	2	1,561	168	168	1,228
10	1	2,257	120	120	1,826
11	0				
12	1	2,248	220	220	1,663
Total	18		913	913	

Table X.3 Assessment of seven models whose support is a subset of the 5-simplex

		883 ERS elections					913 ANES surveys				
Mean number of voters:		716.4					1566.7				
	DF for ERS; ANES	LLH Multinomial	LLH Pólya	δ	AIC	BIC	LLH Multinomial	LLH Pólya	δ	AIC	BIC
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
IC	0; 0	-113,551	-23,084	15.93	227,102	227,102	-195,310	-27,811	19.29	390,621	390,621
UUP	5;	-72,581	-22,133	25.51	145,172	145,204	-171,991	-27,586	21.42	343,992	344,019

	5										
DC	2,649; 2,739	-93,902	-23,028	17.22	193,102	210,036	-127,510	-26,915	29.70	260,498	278,099
UC	2,649; 2,739	-51,295	-21,091	43.92	107,888	124,822	-91,278	-26,000	45.23	188,034	205,635
Borda model	883; 913	-54,255	-21,215	40.16	110,276	115,921	-122,264	-26,679	32.53	246,354	251,218
Condorcet model	883; 913	-32,614	-19,590	93.49	66,994	72,639	-93,229	-25,986	44.96	188,284	193,148
Spatial model	3,532; 3,652	-14,267	-14,267	> 1.E08	35,598	58,177	-16,428	-16,428	> 1.E08	40,160	59,614

Notes:

(1) We calculated the AIC and BIC in columns 5, 6, 10, and 11 using the multinomial log-likelihood functions.

(2) For the spatial model, rounding errors prevented us from determining the exact values of the Pólya log-likelihood functions, but the differences between the multinomial and Pólya log-likelihood function values are less than 0.5.

Table X.4 Mean Euclidian distances σ calculated from observed and simulated data

	Observed data	Simulated data			
Data source:	ANES	IAC	$\mathrm{IAC}_b(k_b)$	$\mathrm{IAC}_t(k_t)$	$\mathrm{IAC}_c(k_c)$
			$k_b = 0.218$	$k_t = 0.201$	$k_c = 0.226$
Number of voters:	1566.7	1567	1567	1567	1567
Model:	(1)	(2)	(3)	(4)	(5)
IC	0.197	0.335	0.319	0.330	0.317
	(0.0021)	(0.0031)	(0.0030)	(0.0031)	(0.0030)

UUP	0.153	0.331	0.303	0.314	0.302
	(0.0017)	(0.0032)	(0.0032)	(0.0033)	(0.0032)
DC	0.148	0.253	0.237	0.249	0.248
	(0.0023)	(0.0030)	(0.0029)	(0.0029)	(0.0030)
UC	0.126	0.252	0.241	0.247	0.238
	(0.0020)	(0.0030)	(0.0028)	(0.0032)	(0.0029)

Borda model	0.149	0.272	0.265	0.260	0.255
	(0.0018)	(0.0029)	(0.0026)	(0.0031)	(0.0028)
Condorcet model	0.121	0.211	0.192	0.192	0.189
	(0.0015)	(0.0024)	(0.0020)	(0.0020)	(0.0021)
Spatial model	0.008	0.061	0.068	0.068	0.068
	(0.0002)	(0.0018)	(0.00110)	(0.0018)	(0.0018)

Note: Standard errors of estimate are in parentheses.

Table X.5 Mean Euclidian distances σ calculated from observed and simulated data

	Observed data	Simulated data						
Data source:	ANES	IC	UUP	DC	UC	Borda	Condorcet	Spatial model
Mean number of voters:	1566.7	1567	1567	1567	1567	1567	1567	1567
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	0.197	0.022	0.079	0.267	0.260	0.116	0.138	0.193
	(0.0021)	(0.0002)	(0.0004)	(0.0035)	(0.035)	(0.0019)	(0.0023)	(0.0022)

UUP	0.153	0.059	0.022	0.261	0.250	0.098	0.097	0.156
	(0.0017)	(0.0003)	(0.0002)	(0.0034)	(0.0034)	(0.0016)	(0.0019)	(0.0020)
DC	0.148	0.016	0.067	0.016	0.225	0.114	0.133	0.145
	(0.0023)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0019)	(0.0021)	(0.0023)
UC	0.126	0.016	0.061	0.231	0.016	0.047	0.076	0.127
	(0.0020)	(0.0002)	(0.0003)	(0.0030)	(0.0003)	(0.0006)	(0.0014)	(0.0020)

Borda model	0.149	0.018	0.068	0.267	0.111	0.019	0.077	0.142
	(0.0018)	(0.0002)	(0.0003)	(0.0035)	(0.0016)	(0.0002)	(0.0014)	(0.0019)
Condorcet model	0.121	0.017	0.048	0.265	0.158	0.058	0.018	0.120
	(0.0015)	(0.0002)	(0.0003)	(0.0034)	0.0022)	(0.0010)	(0.0002)	(0.0016)
Spatial model	0.008	0.008	0.020	0.006	0.006	0.007	0.018	0.006
	(0.0002)	(0.0002)	(0.0003)	(0.0008)	(0.0003)	(0.00021	(0.0003)	(0.0002)

Note: Standard errors of estimate are in parentheses.

Table X.6 Shares of strict pairwise majority rule winners (SPMRWs) in observed and simulated elections

	Observed elections	Simulated elections							
Mean number of voters (1)	(2)	Spatial model (3)	Borda model (4)	Condorcet model (5)	UC (6)	UUP (7)	IAC (8)	DC (9)	IC (10)
10		69.68%	67.39%	67.82%	81.44%	60.01%	73.46%	61.94%	58.44%
20		80.83%	79.07%	79.36%	89.67%	69.42%	81.98%	69.99%	66.85%
40		88.79%	88.01%	88.15%	94.52%	77.62%	87.31%	76.34%	73.46%
100		94.98%	95.27%	95.39%	97.73%	86.51%	91.07%	82.32%	79.69%
500		98.92%	99.41%	99.46%	99.54%	97.43%	93.20%	88.53%	85.90%

716.4	99.32%	99.36%	99.55%	99.62%	99.73%	98.40%	93.46%	91.18%	88.67%
1,000		99.47%	99.77%	99.82%	99.77%	99.14%	93.45%	90.00%	87.55%
1,566.7	99.56%	99.78%	99.89%	99.92%	99.91%	99.77%	93.65%	92.12%	89.59%
5,000		99.89%	99.98%	99.99%	99.95%	99.99%	93.69%	92.06%	89.51%
10,000		99.94%	99.99%	100.00%	99.98%	100.00%	93.75%	92.57%	90.06%
100,000		100.00%	100.00%	100.00%	100.00%	100.00%	93.73%	93.34%	90.82%
1,000,000		100.00%	100.00%	100.00%	100.00%	100.00%	93.71%	93.60%	91.10%

Note: All shares are determined from the results of one million simulated elections. To simulate a million elections with a mean number of 716.4 voters, we simulated either 11,325 or 11,326 elections with the number of voters from each of the 883 ERS elections. To simulate a million elections with a mean number of 1,566.7 voters, we simulated either 10,952 or 10,953 elections with the number of voters from each of the 913 ANES elections. For all other rows, we simulated a million elections, each with the number of voters shown in the first column.