Statistical Science
2002, Vol. 17, No. 4, 420–435

The Mathematics and Statistics of Voting Power

Andrew Gelman, Jonathan N. Katz and Francis Tuerlinckx

Abstract. In an election, *voting power*—the probability that a single vote is decisive—is affected by the rule for aggregating votes into a single outcome. Voting power is important for studying political representation, fairness and strategy, and has been much discussed in political science. Although power indexes are often considered as mathematical definitions, they ultimately depend on statistical models of voting. Mathematical calculations of voting power usually have been performed under the model that votes are decided by coin flips. This simple model has interesting implications for weighted elections, two-stage elections (such as the U.S. Electoral College) and coalition structures. We discuss empirical failings of the coin-flip model of voting and consider, first, the implications for voting power and, second, ways in which votes could be modeled more realistically. Under the random voting model, the standard deviation of the average of n votes is proportional to $1/\sqrt{n}$, but under more general models, this variance can have the form $cn^{-\alpha}$ or $\sqrt{a - b\log n}$. Voting power calculations under more realistic models present research challenges in modeling and computation.

Key words and phrases: Banzhaf index, decisive vote, elections, electoral college, Ising model, political science, random walk, trees.

1. INTRODUCTION

To decide the outcome of a U.S. Presidential election, your vote must be decisive in your own state, and then your state must be decisive in the Electoral College. The different states have different populations, different numbers of electoral votes and different voting patterns, so the probability of casting a decisive vote will vary between states. Even in much simpler settings, weighted voting can lead to unexpected results. For example, consider an election with four voters representing constituencies of unequal size, who are given weights of 12, 9, 6 and 2, respectively, for a weighted majority vote. The voter with 2 votes has zero voting power in that these 2 votes are irrelevant to the election outcome, no matter what the other three voters do, and the other three voters have equal power in that any two of them can determine the outcome. The voter with 12 votes has no more power than the voter with 6. In fact, this system is equivalent to assigning the voters 1, 1, 1 and 0 votes.

Examples such as these have motivated the mathematical theory of *voting power*, which is generally defined in terms of the possibilities that a given voter or set of voters can affect the outcome of an election. Ideas of voting power have been applied in a range of settings, including committee voting in legislatures, weighted voting (as in corporations) and hierarchical voting systems, such as the U.S. Electoral College and the European Council of Ministers. We restrict ourselves in this paper to discussion of weighted majority

Andrew Gelman is Professor of Statistics and Political Science at Columbia University, New York, New York 10027 (e-mail: gelman@stat.columbia.edu; http://www.stat.columbia.edu/~gelman/). Jonathan N. Katz is an Associate Professor of Political Science in the Division of the Humanities and Social Sciences of the California Institute of Technology, Pasadena, California 91125 (e-mail: jkatz@caltech.edu; http://jkatz.caltech.edu). Francis Tuerlinckx is in the Faculty of Psychology and Educational Sciences, Katholieke Universiteit Leuven, Belgium (e-mail: francis.tuerlinckx@psy.kuleuven.ac.be).

voting in two-party systems, with the Electoral College as the main running example.

The major goals of voting power analysis are (1) to assess the relative power of individual voters or blocs of voters in an electoral system, (2) to evaluate the system itself in terms of fairness and maximizing average voting power, (3) to assign weights so as to achieve a desired distribution of voting power and (4) to understand the benefits of coalitions and bloc voting.

This paper reviews some of the mathematical research on voting power and presents some new findings, placing them in a political context. Many of the best-known results in this field are based on a convenient but unrealistic model of random voting, which we describe in Section 3. Section 4 compares to empirical election results and in Section 5, we move to more complicated models of public opinion and evaluate their effects on voting power calculations. One of the challenges here is in understanding which of the results on voting power are robust to reasonable choices of model. Section 6 concludes with an assessment of the connections between voting power and political representation.

The main results presented in this article are the following:

1. For the random voting model, we review well-known findings on voting power in weighted and two-stage elections.
2. We critically evaluate the random voting model, both in theory and in relation to data from elections, and discuss the general relevance of voting power in politics and the mistaken recommendations that have arisen from simplistic voting power calculations.
3. We review recent work and present new results on probability models for voters on trees. These results are mathematically interesting and point the way to further work connecting multilevel probability models of voters, empirical data on the variability of elections and political science models of public opinion.

2. BACKGROUND

2.1 Areas of Application of Voting Power

There are a variety of voting situations in which votes are not simply tallied, and so voting power is a nontrivial concept. We consider weighted majority voting and the closely related setting of two-level voting, where the representatives casting the weighted vote are themselves elected by majority vote in separate districts. This is the structure of the U.S. Electoral College and, implicitly, the European Council of Ministers (where each minister represents an individual country whose government is itself democratically elected).

Voting power measures have been developed in statistics, mathematics, political science and legal literatures for over 50 years. Key references are Penrose (1946), Shapley and Shubik (1954) and Banzhaf (1965). Felsenthal and Machover (1998) gave a recent review. This article focuses on weighted and two-stage voting systems such as the Electoral College (which has been analyzed by Mann and Shapley, 1960; Banzhaf, 1968; Brams and Davis, 1974; and Gelman, King and Boscardin, 1998, among others). Power indexes have strong and often controversial implications (see, e.g., Garrett and Tsebelis, 1999), and so it is important to understand their mathematical and statistical foundations.

Voting power has also been applied to more complicated voting schemes, which we do not discuss here, such as legislative elections with committees, multiple chambers, vetoes and filibusters (Shapley and Shubik, 1954). Luce and Raiffa (1957) discussed power indexes as an application of game theory, following Shapley (1953). The most interesting problems of legislative voting arise when considering coalitions that can change, so that individual voters are motivated to form coalition structures that increase their power. This game-theoretic structure differs from the examples in this article, where the system of coalitions is fixed and votes are simply added at each level.

2.2 Definitions of Voting Power

Voting power can and has been defined in a variety of ways (see Straffin, 1978; Saari and Sieberg, 1999; Heard and Swartz, 1999), and fairness of an electoral system can be defined even more generally (see, e.g., Beitz, 1989; Gelman, 2002). In this article we shall use the definition based on the probability that a vote affects the outcome of the election. Consider an electoral system with n voters. We use the notation i for an individual voter, $v_i = \pm 1$ for his or her vote, $v = (v_1, \dots, v_n)$ for the entire vector of votes and $R = R(v) = \pm 1$ for the rule that aggregates the n individual votes into a single outcome. It is possible for R to be stochastic (because of possible ties or, even more generally, because of possible errors in vote counting). We shall assume various distributions on v and rules $R(v)$ which then together induce distributions on R.

Voting power can also be defined without any probability distributions, simply as the number of combinations of the other $n-1$ voters for which vote i will be decisive, but, as we note at the beginning of Section 3, this is equivalent to the probability definition under the model that all vote combinations are equally likely.

The probability that the change of an individual vote v_i will change the outcome of the election is

$$\begin{aligned} \text{power}_i &= \text{power of voter } i \\ &= \Pr(R = +1 \text{ if } v_i = +1) \\ &\quad - \Pr(R = +1 \text{ if } v_i = -1). \end{aligned} \tag{1}$$

If your voting power is zero, then changing your vote from -1 to $+1$ has no effect on the probability of either candidate winning.

The probabilities in (1) refer to the distribution of the *other* $n-1$ voters, unconditional on v_i. Voting power represents the causal effect of changing your vote with all the other votes held constant—the direct effect on R of changing v_i, not the information that v_i might convey about the other $n-1$ votes.

We shall consider the following situations:

1. For simple *weighted voting*, we assume n voters with weights $w_i, i = 1, \dots, n$, and an aggregation rule $R(v) = \operatorname{sign}(\sum_{i=1}^n w_i v_i)$; that is, the winner is determined by weighted majority. Weighted voting is important for its own sake (e.g., in public corporations) and also as the second level of two-stage voting.
2. For *two-stage voting*, we assume the voters are divided into J jurisdictions, with the winner within each jurisdiction decided by majority vote. Each jurisdiction j has a weighted vote of w_j at the second level, and the overall winner is decided by the weighted majority of the winners in the jurisdictions.

We assume that ties at all levels are decided by coin flips.

2.3 Claims in the Literature and Connections to Empirical Data

Social scientists have studied both the theoretical and the empirical implications of voting power, mostly in political applications, but also in areas such as corporate governance (Leech, 2002). Researchers have looked at fairness to individuals and also at the effects of unequal voting power on campaigning and the allocation of resources (see Snyder, Ting and Ansolabehere, 2001). The probability that a single vote is decisive in an election is also relevant in studying the responsiveness of an electoral system to voter preferences and the utility of voting (see Riker and Ordeshook, 1968; Ferejohn and Fiorina, 1974; Aldrich, 1993; Edlin, Gelman and Kaplan, 2002).

The probability of a vote being decisive is important directly—it represents your influence on the electoral outcome, and this influence is crucial in a democracy—and also indirectly, because it could influence campaigning. For example, one might expect campaign efforts to be proportional to the probability of a vote being decisive, multiplied by the expected number of votes changed per unit of campaign expense, although there are likely strategic complications since both sides are making campaign decisions. Thus, campaigning strategies have been studied in the political science literature as evidence of voting power (Brams and Davis, 1974, 1975; Colantoni, Levesque and Ordeshook, 1975; Stromberg, 2002).

Perhaps the most widely publicized normative political claim from the voting power literature is that, in two-stage voting systems with proportional weighting (that is, $w_j \propto n_j$), voters in larger jurisdictions have disproportionate power (Penrose, 1946; Banzhaf, 1965). Under a simple (and, in our judgment, inappropriate) model, the voting power in such systems is approximately proportional to $\sqrt{n_j}$ (see Section 3.2). This has led scholars to claim that the U.S. Electoral College favors large states (Banzhaf, 1968), a claim that we and many others have disputed (see Section 4).

In political science these theories have been checked with empirical data in various ways. Most directly, voting power depends on the probability that an election is a tie. This probability is typically so low that it is difficult to estimate directly; for example, in the past 100 years, there have been about 20,000 contested elections to the U.S. House of Representatives, and none of them has been tied. However, the probability of a tie can be estimated by extrapolating from the empirical frequency of close elections (see Mulligan and Hunter, 2001); for example, about 500 of the aforementioned House of Representatives elections were decided within 1,000 votes. If we define $f_{\overline{V}}$ to be the distribution of the difference $\overline{V}$ in vote proportions between the two leading candidates, then

$$\Pr(\text{tie election}) \approx \frac{f_{\overline{V}}(0)}{n} \tag{2}$$

in an election with n voters. Regression-type forecasting models for $\overline{V}$ have been used to estimate voting power for specific elections (see Section 4). In elections with disputed votes and possible recounts, so that no single vote can be certain to be decisive, the probability of affecting the outcome of the election can still be identified with the probability of a tie to a very close approximation (see the Appendix of Gelman, Katz and Bafumi, 2002).

3. THE RANDOM VOTING MODEL

We begin with the assumption that votes are determined by independent coin flips, which we call *random voting*. As we discuss in Section 4, the random voting model is empirically inappropriate for election data. We devote some space to this model, however, because it is standard in the voting power literature.

Under random voting, all 2^n vote configurations are equally likely, and so the power of voter i is simply $2^{-(n-1)}$ times the number of configurations of the other $n-1$ voters for which voter i is decisive (and counting semidecisive configurations, in which votes are exactly tied, as 1/2). Voting power calculations can thus be seen as combinatorical. However, we see the probabilistic, rather than the counting, derivation as fundamental.

3.1 Weighted Voting

To calculate the power of voter i with weight w_i in a simple weighted majority voting system, we define the total weighted vote of all the others as $V_{-i} = \sum_{k \neq i} w_k v_k$ and we define the sum of the squares of all the weights as $W^2 = \sum_{k=1}^{n} w_k^2$. Then, for weighted voting,

$$\text{power}_i = \Pr(|V_{-i}| < w_i) + \tfrac{1}{2}\Pr(|V_{-i}| = w_i). \tag{3}$$

From the random voting model, we can immediately derive that $E(V_{-i}) = 0$ and $\text{sd}(V_{-i}) = \sqrt{\sum_{k \neq i} w_k^2} = \sqrt{W^2 - w_i^2}$. If certain regularity conditions hold—if the number of voters is large enough, no single voter or small set of voters is dominant and there are no discrete features in the weights (as in the introductory example, where all but one of the weights is divisible by 3)—then we can think of the distribution of V_{-i} as approximately normally distributed and we can approximate (3) by $\Phi(w_i/\sqrt{W^2 - w_i^2}) - \Phi(-w_i/\sqrt{W^2 - w_i^2})$, where Φ is the cumulative normal distribution function.

If $w_i^2 \ll W^2$, voting power is approximately close to linear in w_i. For example, in the Electoral College, the values of w_i for the 50 states and the District of Columbia range from 3 to 54, with a total of 538. We can calculate power_i for each state (assuming random voting) and compare to the linear approximation, $\text{power}_i \approx \sqrt{2/\pi}(w_i/W)$. The linear fit has a relative error of less than 10% for all states.

As this calculation for the Electoral College illustrates, voting power paradoxes such as illustrated in the first paragraph of this article are unlikely to occur except in the context of the discreteness of very small voting systems. Such situations have occurred (see Felsenthal and Machover, 2000), but in our opinion they are fundamentally less interesting than the results on two-stage voting and coalitions that we review below. In weighted voting settings where the w_i's display central limit theorem-type behavior, voting power (given random voting) is approximately proportional to weight, as one would intuitively expect.

3.2 Two-Stage Voting

In two-stage voting, one must first compute the voting power of each jurisdiction j and then the power of each of the n_j voters within a jurisdiction. As described above, if the number of jurisdictions is large and none is dominant, it is reasonable to approximate the voting power of jurisdiction j as proportional to w_j under random voting.

For an individual voter i in jurisdiction j, let V_{-i} be the sum of the other $n_j - 1$ votes in the jurisdiction. Then the probability that vote i is decisive within jurisdiction j is

$$\Pr(V_{-i} = 0) + \tfrac{1}{2}\Pr(|V_{-i}| = 1)$$
$$= \begin{cases} \dbinom{n_j - 1}{(n_j - 1)/2}, & \text{if } n_j \text{ is odd,} \\[2ex] \dbinom{n_j - 1}{n_j/2}, & \text{if } n_j \text{ is even,} \end{cases}$$

under the random voting model. Unless n is very small, this can be approximated as $\sqrt{2/(\pi n_j)}$, which is a special case of (2).

With random voting, the votes inside and outside jurisdiction j are independent, and so the power of voter i in jurisdiction j, for two-stage random voting, is approximately proportional to

$$\text{power}_i \approx \frac{w_j}{\sqrt{n_j}}, \tag{4}$$

given the conditions stated at the end of the previous section. This result has led commentators to suggest that a fair allocation of weights in a two-stage voting system is proportional to the square root of the number of voters in the jurisdiction (Penrose, 1946), with perhaps some minor modifications due to the combinatorics of a discrete number of jurisdictions (Felsenthal and Machover, 2000). However, we disagree with these recommendations because of systematic flaws in the random voting model, as we discuss in Section 4.

3.3 Individual and Average Voting Power

The structure of a voting system can affect the power of individual voters and the average power of all n voters in interesting ways. Figure 1 illustrates an example with simple majority voting and various two-stage electoral systems with nine voters. These trees (and the accompanying calculations) illustrate the benefits under random voting of being in a large jurisdiction; they also illustrate the negative side: voters who are left tend to do worse than under majority rule.

One way to study the total effect of two-stage voting is to compute the average probability of decisiveness for all of the n voters. It has been proved (and we sketch a proof in the next paragraph) that, under the random voting model, this average voting power is maximized under simple popular vote (majority rule) and is lower under any other system. Figure 1 illustrates this point: the coalitions benefit their members but lower the average probability of decisiveness.

To prove the general result, we start by re-expressing voting power in terms of the probability of *satisfaction*—that is, of a voter's preferred candidate winning the election (see Straffin, 1978). Under the random voting model, we can rewrite (1) as

$$\begin{aligned}\text{power}_i &= \Pr(R=+1 \text{ if } v_i=+1)\\ &\quad + \Pr(R=-1 \text{ if } v_i=-1) - 1\\ &= 2\Pr(R=+1 \text{ and } v_i=+1)\\ &\quad + 2\Pr(R=-1 \text{ and } v_i=-1) - 1\\ &\qquad \text{(assuming random voting)}\\ &= 2\Pr(\text{voter } i \text{ is satisfied}) - 1,\end{aligned}$$

and so, under random voting, maximizing average voting power is equivalent to maximizing the average probability of satisfaction. Only voters on the winning side will be satisfied and so, conditional on the total vote, average satisfaction is maximized by assigning the winner to the side supported by more voters, which is simply majority rule. This theorem can also be seen as a corollary of more general results in graph theory (see Lemma 6.1 of Friedgut and Kalai, 1996).

One way to understand the result is that the winner-take-all rule within coalitions magnifies small differences (e.g., a vote of 10–8 within a coalition is transformed to 18–0 at the next stage in the tree), which has the effect of amplifying noise (if the election is thought of as a system of communicating individual preferences up to the top of the tree). The least noisy system is majority rule, with no coalitions at all. An analogy is to scoring in a single game of ping pong (the first player to get 21 points wins) versus a tennis match (a three-level system where a player must win a majority of sets, which in turn comprise games and points). If points are scored independently, then scoring is fairer in ping pong than in tennis.

4. EMPIRICAL RESULTS ON VOTE DISTRIBUTIONS AND VOTING POWER

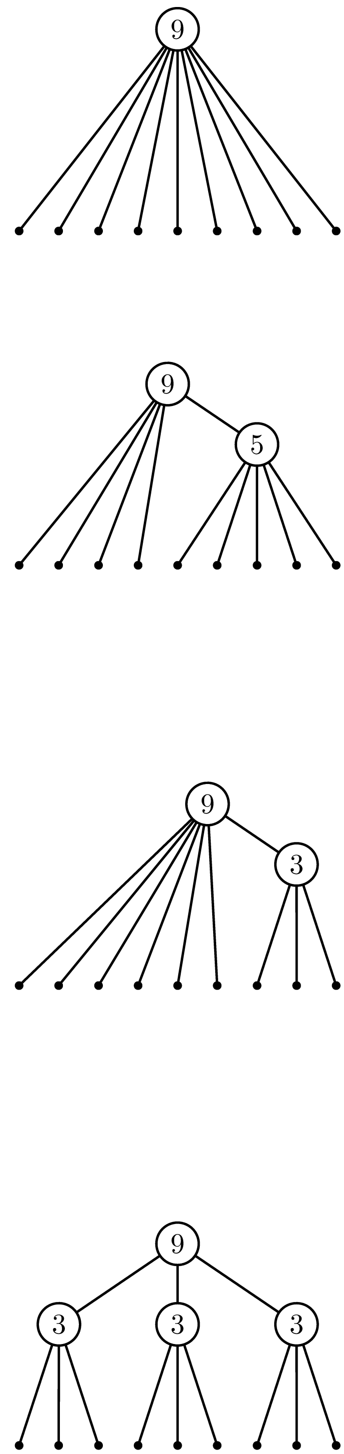
The random voting model is a natural starting point for studying voting power, but it is obviously unrealistic. From a probabilistic perspective, one can imagine developing more complex stochastic processes for voting that allow for correlations and unequal probabilities. From the tradition of analysis of voting data in political science, it would make sense to set up regression-type models and perform inference using available data from elections and committee votes. Yet another direction would be to estimate the probability of casting a decisive vote directly from empirical data, as has been done for judicial data by Heard and Swartz (1999). This would be possible with voting within legislatures using estimates such as those of Poole and Rosenthal (1997) of legislators' ideal points.

In this section we discuss the empirical failings of the random voting model and why this has major implications for voting power. Section 5 discusses more complex probability models for votes that move in the direction of realism.

4.1 Closeness of Elections, the Number of Voters and Voting Power

The random voting model makes predictions that are not valid for real elections. For example, in an election with 1 million voters, the random voting model implies that the proportional vote margin should have a mean of 0 and a standard deviation of 0.001. In reality, large elections are typically decided by much more than 0.1 of a percentage point.

Before discussing potential model improvements, we consider here the voting power implications of empirical problems with the random voting model.



A. No Coalitions

A voter is decisive if the others are split 4-4:

$$\Pr(\text{Voter is decisive}) = \binom{8}{4} 2^{-8} = 0.273$$

Average $\Pr(\text{Voter is decisive}) = 0.273$

B. A Single Coalition of 5 Voters

A voter in the coalition is decisive if others in the coalition are split 2-2:

$$\Pr(\text{Voter is decisive}) = \binom{4}{2} 2^{-4} = 0.375$$

A voter not in the coalition can never be decisive:

$$\Pr(\text{Voter is decisive}) = 0$$

Average $\Pr(\text{Voter is decisive}) = \frac{5}{9}(0.375) + \frac{4}{9}(0) = 0.208$

C. A Single Coalition of 3 Voters

A voter in the coalition is decisive if others in the coalition are split 1-1 and the coalition is decisive:

$$\Pr(\text{Voter is decisive}) = \frac{1}{2} \cdot \frac{50}{64} = 0.391$$

A voter not in the coalition is decisive with probability:

$$\Pr(\text{Voter is decisive}) = \binom{5}{1} 2^{-5} = 0.156$$

Average $\Pr(\text{Voter is decisive}) = \frac{3}{9}(0.391) + \frac{6}{9}(0.156) = 0.234$

D. Three Coalitions of 3 Voters Each

A voter is decisive if others in the coalition are split 1-1 and the other two coalitions are split 1-1:

$$\Pr(\text{Voter is decisive}) = \frac{1}{2} \cdot \frac{1}{2} = 0.250$$

Average $\Pr(\text{Voter is decisive}) = 0.250$

FIG. 1. *An example of four different systems of coalitions with nine voters, with the probability of decisiveness of each voter computed under the random voting model. Each is a "one person, one vote" system, but they have different implications for probabilities of casting a decisive vote. Joining a coalition is generally beneficial to those inside the coalition but hurts those outside. The average voting power is maximized under A, the popular-vote rule with no coalitions. From Katz, Gelman and King* (2002).

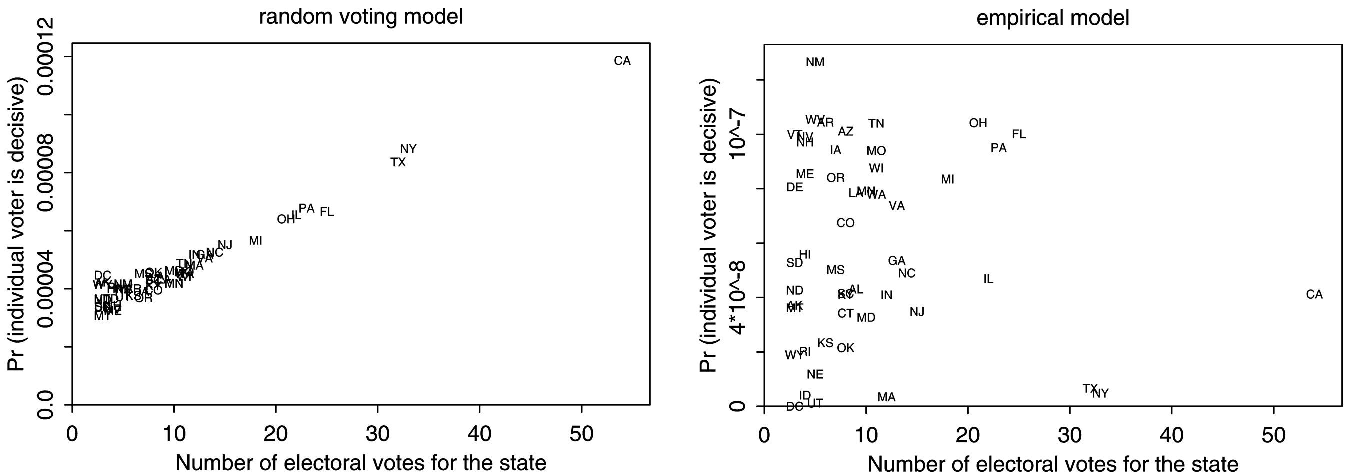


FIG. 2. *Voting power for individuals in the 2000 U.S. Presidential election, by state under the random voting model and under an empirical model based on randomly perturbing the actual election outcomes with variation at the national, regional, state and local levels. Voters in large states have disproportionate power under the random voting model, but not under the empirical model.*

Clearly, real elections are less close, and thus individual voters have less power, than predicted under random voting, but how are the results of *relative* voting power affected?

If the number of voters n_j in a district is moderate or large, we can use (2) to approximate the probability that a single voter is decisive within district j as

$$\text{(5)}\quad \Pr(\text{a vote is decisive within district } j) \approx \frac{f_j(0)}{n_j},$$

where f_j is the distribution of the *proportional* vote differential $\overline{V}_j = (1/n_j)\sum_{i=1}^{n_j} v_i$ within district j. If f_j has a mean of 0 and a fixed distributional form (e.g., normality), then $f_j(0) \propto 1/\mathrm{sd}(\overline{V}_j)$ and so

$$\text{(6)}\quad \begin{aligned} &\Pr(\text{a vote is decisive within district } j)\\ &\quad\approx \frac{1}{n_j \mathrm{sd}(\overline{V}_j)}. \end{aligned}$$

Thus, when comparing the probability of decisiveness for voters in districts of different sizes n_j (as in Section 3.2), the behavior of $\mathrm{sd}(\overline{V}_j)$ as a function of n_j is crucial. Under random voting, $\mathrm{sd}(\overline{V}_j) \propto 1/\sqrt{n_j}$, but for actual elections this is not generally true.

4.2 U.S. Presidential Elections

We focus on the most widely discussed example, the relative power of voters in different states in electing the President of the United States. In the Electoral College, each state gets two electoral votes plus a number approximately proportional to the state's population. Except for the smallest states, this means that w_j is approximately proportional to n_j (voter turnout varies slightly between states). The random voting model then implies [see (4)] that an individual's voting power should be approximately proportional to the square root of the population of his or her state; see the left panel of Figure 2.

Thus, the general conclusion in the voting power literature is that the Electoral College benefits voters in large states. For example, Banzhaf (1968) claimed to offer "a mathematical demonstration" that it "discriminates against voters in the small and middle-sized states by giving the citizens of the large states an excessive amount of voting power," and Brams and Davis (1974) claimed that the voter in a large state "has on balance greater potential voting power ... than a voter in a small state." Mann and Shapley (1960), Owen (1975) and Rabinowitz and Macdonald (1986) came to similar conclusions. This impression has also made its way into the popular press; for example, Noah (2000) stated, "the distortions of the Electoral College ... favor big states more than they do little ones." It has similarly been claimed that if countries in the European Union were to receive votes in its council of ministers proportional to their countries' populations, then voters in large countries would have disproportionate power (Felsenthal and Machover, 2000).

4.3 The $1/\sqrt{n}$ Rule

The above claims all depend on the intermediate result, under the random voting model, that the probability of decisiveness within a state is proportional to $1/\sqrt{n_j}$. The extra power of voters in large states derives from the assumption that elections in these states

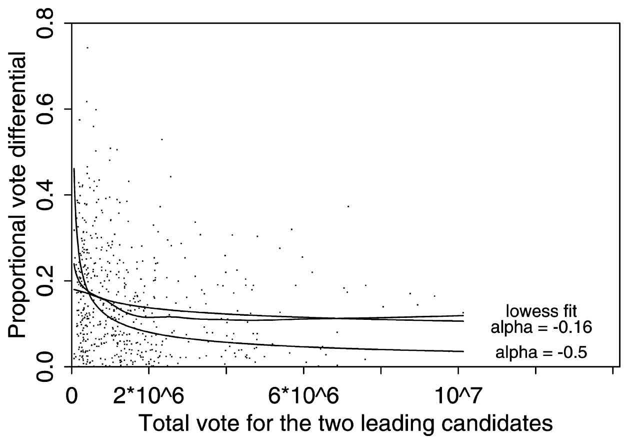


FIG. 3. *The margin in state votes for President as a function of the number of voters n_j in the state. Each dot represents a different state and election year from 1960 to 2000. The margins are proportional; for example, a state vote of* 400,000 *for the Democratic candidate and* 600,000 *for the Republican would be recorded as* 0.2. *Lines show the lowess (nonparametric regression) fit, the best-fit line proportional to* $1/\sqrt{n_j}$ *and the best-fit line of the form* cn_j^{α}. *As shown by the lowess line, the proportional vote differentials show only very weak dependence on* n_j. *The* $1/\sqrt{n_j}$ *line, implied by standard voting power measures, does not fit the data.*

are much more likely to be close. This assumption can be tested with data. For example, Figure 3 shows the absolute proportional vote differential $|\overline{V}_j|$ as a function of number of voters n_j for all states (excluding the District of Columbia) for all Presidential elections from 1960 to 2000.

We test the $1/\sqrt{n_j}$ hypothesis by fitting three different regression lines to $|\overline{V}_j|$ as a function of n_j. First, we use the lowess procedure (Cleveland, 1979) to fit a nonparametric regression line. Second, we fit a curve of the form $y = c/\sqrt{n_j}$, using least squares to find the best-fitting value of c. Third, we find the best-fitting curve of the form $y = cn_j^{\alpha}$. The best-fit α is -0.16 (with a standard error of 0.03), which suggests that elections tend to get slightly closer for larger n_j but with a relationship much weaker than $1/\sqrt{n_j}$. As the scale of the graph makes clear, it is impossible for the $1/\sqrt{n_j}$ rule to hold in practice, as this would mean extreme landslides for low n_j or extremely close elections for high n_j, neither of which, in general, will hold.

4.4 Empirical Estimates of Voting Power in Presidential Elections

Given that the $1/\sqrt{n_j}$ rule is not appropriate for real elections, how should voting power be calculated for two-stage elections? The right panel of Figure 2 shows a calculation for the 2000 Presidential election based on perturbing the empirical election results. Uncertainty in the election outcome is represented by adding normally distributed random errors ε at the national, regional, state and local levels: for Congressional districts i in state j within region k, the election outcomes are simulated 500 times by perturbing the observed outcome, $\overline{V}_i^{\text{obs}}$:

$$(7)\quad \overline{V}_i^{\text{sim}} = \overline{V}_i^{\text{obs}} + \varepsilon^{\text{nation}} + \varepsilon_k^{\text{region}} + \varepsilon_j^{\text{state}} + \varepsilon_i^{\text{district}}.$$

The $\overline{V}_i^{\text{sim}}$ values are then summed within each state j to get a simulation of the state-level vote differentials, $\overline{V}_j$. The error terms in the simulation (7) represent variation between elections, and the hierarchical structure of the errors represents observed correlations in national, regional and state election results (Gelman, King and Boscardin, 1998). For the simulation for Figure 2b, the error terms on the vote proportions have been assigned normal distributions with standard deviations 0.06, 0.02, 0.04 and 0.09, which were estimated from election-to-election variation of vote outcomes at the national, regional, state and Congressional district levels. (The pattern of results of the voting power simulation are not substantially altered by moderate changes to these variance parameters.)

Under the simulation, the probability that a voter is decisive within state j is given by (5), which we evaluate from the normal density function. We then compute in two steps the probability that the state's electoral votes are decisive for the nation. First, we update the distributions of the national and regional error terms $\varepsilon^{\text{nation}}$ and $\varepsilon_k^{\text{region}}$ using the multivariate normal distribution given the condition $\overline{V}_j = 0$. Second, we simulate the vector of election outcomes for all the other states under this condition and estimate the probability that the w_j electoral votes of state j are decisive in the national total. This last computation could be performed by counting simulations, but is made more computationally efficient by approximating the proportion of electoral votes received by either candidate as a beta distribution, as in Gelman, King and Boscardin (1998).

Comparing to the result under the random voting model, the empirical calculation shows much more variation between states (because some states, like New Mexico, were close, and others, like Massachusetts, were not) and no strong dependence on state size. In reality, but not in the random voting model, large states are *not* necessarily extremely close, and thus voters in large states do *not* have disproportionate voting power.

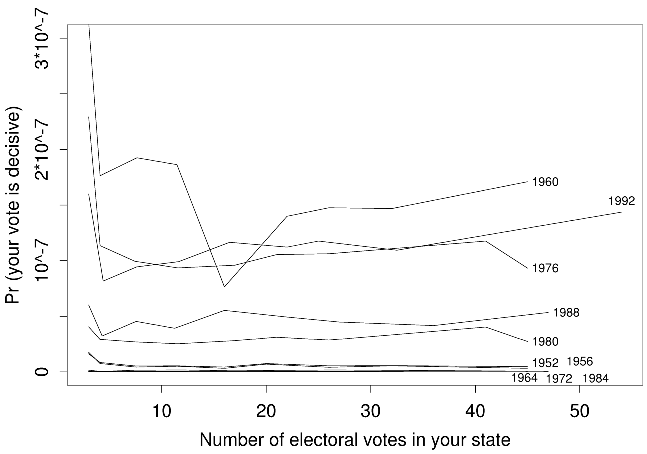


FIG. 4. *The average probability of a decisive vote as a function of the number of electoral votes in the voter's state, for each U.S. Presidential election from 1952 to 1992 (excluding 1968, when a third party won in some states). The probabilities are calculated based on a forecasting model that uses information available two months before the election. This figure is adapted from Gelman, King and Boscardin (1998). The probabilities vary little with state size, with the most notable pattern being that voters in the very smallest states are, on average, slightly more likely to be decisive.*

A slightly different empirically based method of computing voting power is described by Gelman, King and Boscardin (1998). A hierarchical linear regression model, based on standard election forecasting procedures used in political science, was used to obtain probabilistic forecasts for Presidential elections by state. The models were then used to compute the probability of decisive vote; Figure 4 shows the resulting average probability of decisiveness of voters in a state, as a function of the number of electoral votes in the state, for each election. The clearest pattern is that the smallest states have slightly higher voting power, on average; this is a result of the two "free" votes that each state receives in the Electoral College, so that the smallest states have disproportionate weights.

4.5 The Electoral College and Average Voting Power

As discussed in Section 3.3, under random voting, average voting power is maximized under a popular vote system. However, this result is highly sensitive to the assumption, under random voting, that all vote outcomes are equally likely (see Natapoff, 1996). Thus, if the question of average voting power is of practical interest, it is important to address it using actual electoral data.

Using the model (7) based on perturbing district-by-district outcomes (as was used to calculate the values displayed in Figure 2b), Katz, Gelman and King (2002) computed average voting power for Presidential elections under three electoral systems: popular vote, electoral vote and a hypothetical system of winner-take-all by Congressional district. They found average voting power to vary dramatically between elections (depending on the closeness of the national election); within any election, however, changing the voting rule had little effect on the average probability of decisiveness.

4.6 Empirical Evidence from Other Elections

Analyses of electoral data from U.S. Congress, U.S. state legislatures and European national elections also have found only very weak dependence of the closeness of elections on the number of voters (Mulligan and Hunter, 2001; Gelman, Katz and Bafumi, 2002). This persistent empirical finding, contradicting the $1/\sqrt{n}$ rule implied by random voting, casts strong doubt on the recommendations by mathematical analysts from Penrose (1946) to Felsenthal and Machover (2000) to apply standard power indexes to weighted voting.

The best approach for assigning weighted votes is still unclear, however. We have seen that assigning weights to equalize voting power under random voting is inappropriate in real-world electoral systems that do not follow the $1/\sqrt{n}$ rule. However, equalizing empirical voting power is also problematic, as the weights would then have to depend on local political conditions. For example, in the 2000 election, it would be necessary to lower the weight of New Mexico and raise the weight of Massachusetts (see Figure 2b), but then these weights might have to change in the future if New Mexico moved away or Massachusetts moved toward the national average. A reasonable default position is to assign weights in proportion to population size or perhaps population size to the 0.9 power (see Gelman, Katz and Bafumi, 2002), but these too are based on particular empirical analyses. In general, this fits in with the literature on evaluating voting methods and power indexes based on their performance in actual voting situations (see, e.g., Felsenthal, Maoz and Rapoport, 1993; Heard and Swartz, 1999; Leech, 2002).

5. STOCHASTIC PROCESSES FOR VOTERS

The simplest generalization of random voting is for votes to be independent but with probability p, rather than $1/2$, of voting $+1$. This is not a useful model; although it corrects the mean vote, it still predicts a standard deviation that is extremely small for large elections (Beck, 1975). It is necessary to go further and allow each voter to have a separate p_i and to model the distribution of these p_i's. If the probabilities p_i are given any fixed distribution not depending on n, then the distribution of the average vote, for large n, will converge to the distribution of the p_i's. The empirically falsified $1/\sqrt{n}$ rule then goes away, to be replaced by the more general expression (5), because the binomial variation from which it derives is minor compared to any realistic variation among the probabilities p_i. (This was noted by Good and Mayer, 1975; Margolis, 1977, 1983; and Chamberlain and Rothschild, 1981.)

The next step is give a dependence structure to the voters' probabilities p_i. It makes sense to build this dependence upon existing relationships among the voters. The most natural starting point is a tree structure based on geography; for example, the United States is divided into regions, each of which contains several states, each of which is divided into Congressional districts, counties, cities, neighborhoods and so forth. When modeling elections, it makes sense to use nested communities for which electoral data are available (e.g., states, legislative districts and precincts). It might also be appropriate to include structures based on nonnested predictors. For example, voters have similarities based on demographics as well as geography; nonnested models also have been used to capture "small-world" phenomena in social networks (Watts, Dodds and Newman, 2002).

Section 5.1 discusses an approach based on directly modeling the dependence of individual votes and Section 5.2 presents a model using correlated latent variables. It is interesting to see the different implications for observable outcomes that are obtained by these two families of probability models.

5.1 Discrete Modeling of Dependence of Votes

One idea for modeling votes, taken from the mathematical literature and based on the Ising model from statistical physics, is to model dependence of the votes v_i via a probability density proportional to $\exp(-\sum_{ij} c_{ij} v_i v_j)$, where c_{ij} represents the strength of the connection between any two voters i and j. Under this model, voters who are connected are more likely to vote similarly. If the voters form a tree structure, then the model implies correlation of the votes in the same local community. If the correlation is higher than a certain critical value, then the properties of average votes can be qualitatively different than in the random voting model, as we prove below (see Evans, Kenyon, Peres and Schulman, 2000).

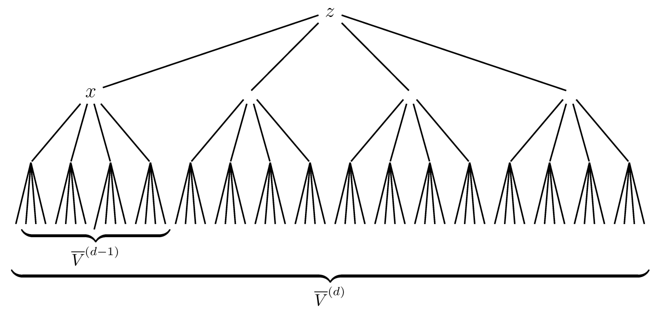


FIG. 5. *Notation for the derivation of the standard deviation of the average vote differential, $\overline{V}^{(d)}$, for the Ising models on k^d voters in a tree structure. This tree, unlike that in Figure 1, represents dependence between voters, not a coalition structure. Here, z represents the unobserved ± 1 variable at the root of the tree, x represents the unobserved ± 1 variable at one of the k branches at the next level and $\overline{V}^{(d-1)}$ is the average vote differential for the k^{d-1} voters in this branch. The derivation proceeds by determining the mean and variance of $\overline{V}^{(d)}$, conditional on z, in terms of the mean and variance of $\overline{V}^{(d-1)}$, conditional on x.*

5.1.1 *A stochastic model on a tree of voters.* To get a sense of how these models work, we derive some basic results for the Ising model on symmetric trees. Suppose we have $n = k^d$ voters arrayed in a tree of depth d with k branches at each node. At each node of the tree is a variable that equals ± 1 (see Figure 5). For the leaf nodes, this variable represents a vote; at the other nodes, the variable is unobserved and serves to induce a correlation among the leaves. The probability model states that, conditional on a parent node, the k children are independent, each with probability π of differing from the parent. We further assume that the marginal probabilities of $+1$ and -1 are equal.

5.1.2 *The distribution of the average of n votes.* One way to understand this model and to see how it differs from random voting is to study its implications for the probability that an individual voter is decisive in a coalition or district j of n_j voters. As discussed in Section 4, the probability of decisiveness is linked to the standard deviation of the proportional vote differential $\overline{V}_j$ in the district. The random voting model predicts $\mathrm{sd}(\overline{V}_j) \propto n_j^{-0.5}$, whereas empirical electoral data show much weaker declines on the order of $\mathrm{sd}(\overline{V}_j) \propto n_j^{-\alpha}$, with powers estimated at lower values such as $\alpha = 0.15$ (see Figure 3).

This model can easily be simulated starting at the top of the tree and working downward. However, this is computationally expensive for large values of n, so we use analytic methods to evaluate $\mathrm{sd}(\overline{V}^{(d)})$ as a function of $n = k^d$, as well as the parameters k and π that determine the stochastic process. The results we prove here also appear in Bleher, Ruiz and Zagrebnov (1995) and Evans et al. (2000).

The setup for our derivation is diagrammed in Figure 5. For an Ising-model tree of depth d, let $z = \pm 1$ be the equally likely values at the root and let $\overline{V}_d$ be the average value at the k^d leaves. The variance of $\overline{V}_d$ can be decomposed, conditional on the root value z, as

$$\begin{aligned}\operatorname{var}\left(\overline{V}^{(d)}\right) &= \operatorname{var}\left(\mathrm{E}\left(\overline{V}^{(d)}|z\right)\right) + \mathrm{E}\left(\operatorname{var}\left(\overline{V}^{(d)}|z\right)\right)\\ &= \tfrac{1}{2}\Big[\left(\mathrm{E}\left(\overline{V}^{(d)}|z=+1\right)-0\right)^2\\ &\qquad + \mathrm{E}\left(\left(\overline{V}^{(d)}|z=-1\right)-0\right)^2\Big]\\ &\quad + \tfrac{1}{2}\Big[\operatorname{var}\left(\overline{V}^{(d)}|z=+1\right)\\ &\qquad + \operatorname{var}\left(\overline{V}^{(d)}|z=-1\right)\Big]\\ &= \mu_d^2 + \sigma_d^2,\end{aligned} \tag{8}$$

where μ_d and σ_d are the conditional means and variances of $\overline{V}_d$ given $z = +1$:

$$\mu_d = \mathrm{E}\left(\overline{V}^{(d)}|z=+1\right), \quad \sigma_d^2 = \operatorname{var}\left(\overline{V}^{(d)}|z=+1\right).$$

This decomposition is useful because we can evaluate μ_d and σ_d^2 recursively.

At $d = 0$, the root of the tree is the same as the leaf, and $\overline{V}^{(0)} = z$. Thus,

$$\mu_0 = \mathrm{E}\left(\overline{V}^{(0)}|z=+1\right) = 1,$$

$$\sigma_0^2 = \operatorname{var}\left(\overline{V}^{(0)}|z=+1\right) = 0.$$

For $d \geq 1$ we note that $\overline{V}_d$ is the average of k identically distributed random variables $\overline{V}_{d-1}$ that are independent conditional on z (see Figure 5). Then we can use the symmetry of the underlying model to obtain

$$\begin{aligned}\mu_d &= \mathrm{E}\left(\mathrm{E}\left(\overline{V}^d|x\right)|z=+1\right)\\ &= (1-\pi)\mu_{d-1} + \pi(-\mu_d)\\ &= (1-2\pi)\mu_{d-1}\\ &= (1-2\pi)^d,\end{aligned} \tag{9}$$

with the last step being a recursive calculation starting with $\mu_0 = 1$. We evaluate σ_d^2 using the conditional variance decomposition and the fact that it is the mean of k independent components,

$$\begin{aligned}\sigma_d^2 &= \frac{1}{k}\left(\operatorname{var}\left(E\left(\overline{V}^d|x\right)|z=+1\right)\right.\\ &\qquad \left. + \mathrm{E}\left(\operatorname{var}\left(\overline{V}^d|x\right)|z=+1\right)\right)\\ &= \frac{1}{k}\left(4\pi(1-\pi)\mu_{d-1}^2 + \sigma_{d-1}^2\right),\end{aligned} \tag{10}$$

with the second term being the variance of a random variable that equals μ_{d-1} with probability $1-\pi$ or $-\mu_{d-1}$ with probability π. We can expand the recursion in (10) and insert (9) to obtain

$$\sigma_d^2 = 4\pi(1-\pi)k^{-d}\sum_{i=0}^{d}\left((1-2\pi)^2 k\right)^i,$$

and combining with (9) into (8) yields

$$\operatorname{var}\left(\overline{V}^{(d)}\right) = \xi^d + (1-\xi)k^{-d}\sum_{i=0}^{d}(\xi k)^i, \tag{11}$$

where, for convenience, we have defined

$$\xi = (1-2\pi)^2.$$

We evaluate (11) separately for three cases, depending on whether the factor ξk in the power series is less than 1, equal to 1 or greater than 1. For each, we focus on the limit of large d—that is, large n—since we are interested in modeling elections of thousands or millions of voters.

- If $\xi k < 1$, then we can expand (11) as

$$\begin{aligned}\operatorname{var}\left(\overline{V}^{(d)}\right) &= \xi^d + (1-\xi)k^{-d}\frac{1-(\xi k)^d}{1-\xi k}\\ &\approx \left(\frac{1-\xi}{1-\xi k}\right)\frac{1}{n} \quad \text{for large } n.\end{aligned} \tag{12}$$

  Thus, for $\xi < 1/k$, the standard deviation of the average of n votes is proportional to $1/\sqrt{n}$, just as in the random voting model but with a different proportionality constant. This will not be useful for us in modeling data with more gradual power-law behavior such as displayed in Figure 3.
- If $\xi k = 1$, then (11) becomes

$$\begin{aligned}\operatorname{var}\left(\overline{V}^{(d)}\right) &= \xi^d + (1-\xi)k^{-d}d\\ &= \frac{1}{n}\left(1 + \left(1-\frac{1}{k}\right)\log_k n\right)\\ &\approx \left(1-\frac{1}{k}\right)\frac{1}{n}\log_k n \quad \text{for large } n.\end{aligned} \tag{13}$$

  Thus, for $\xi = 1/k$, the standard deviation of the average vote margin is proportional to $\sqrt{(\log_k n)/n}$.
- If $\xi k > 1$, then (11) becomes

$$\begin{aligned}\operatorname{var}\left(\overline{V}^{(d)}\right) &= \xi^d + (1-\xi)k^{-d}\frac{(\xi k)^d - 1}{\xi k - 1}\\ &\approx \frac{k-1}{k-1/\xi}\xi^d \quad \text{for large } n.\end{aligned} \tag{14}$$

Since $d = \log_k n$, we can write $\xi^d = n^{-2\alpha}$, where α is a power less than 1/2; more specifically, $\alpha = -0.5 \log_k \xi$. We then can reexpress (14) as

$$\text{(15)}\quad \operatorname{var}\left(\overline{V}^{(d)}\right) \approx \left(\frac{k-1}{k-k^{2\alpha}}\right) n^{-2\alpha} \quad \text{for large } n.$$

Evans et al. (2000) generalized these power laws to nonregular trees.

5.1.3 *Fitting the model to electoral data.* In real elections, one can approximate the standard deviation of $\overline{V}_j$ for districts j as proportional to $n_j^{-\alpha}$, where α is some power less than 1/2 (see Section 4). As we just have shown, this corresponds to the Ising model with $\xi > k$.

Given α and k, we can solve for $\pi = \frac{1}{2}(1 - \sqrt{\xi}) = \frac{1}{2}(1 - k^{-\alpha})$. For example, the Presidential election data illustrated in Figure 3 can be fitted by $\alpha = 0.16$. For $k = 2, 3, 10, 100$, the best-fitting π's are $0.05, 0.08, 0.15, 0.26$.

Next, one can imagine setting the parameter k so that the coefficient $\sqrt{(k-1)/(k-k^{2\alpha})}$ from (15) matches the coefficient c in the fitted curve, $\operatorname{sd}(\overline{V}_j) \approx c n_j^{-\alpha}$. However, this second fitting step is not so effective, because the coefficient in (15) has a narrow range of possibilities. For example, for $\alpha = 0.16$, $\sqrt{(k-1)/(k-k^{2\alpha})}$ ranges from a maximum of 1.15 (at $k = 2$) to a minimum of 1 (as $k \to \infty$), but the estimate of c from the data in Figure 3 is 1.74.

Even if the estimated c from this data set happened to be in the range (1, 1.15), we would not want to take the Ising model too seriously as a description of voters. Our purpose in developing such a stylized model here is primarily to show how simple conditions of connectedness can induce power laws that go beyond the random voting model.

5.2 Continuous Modeling of Latent Underlying Preferences

The other natural approach to modeling variation in opinion, deriving from preference models in social science, is to think of the votes v_i as independent but with structure on the probabilities p_i. A natural starting point is an additive model on the logit or probit scale: for example, in a hierarchical structure,

$$p_i = \Phi(\alpha_{\text{nation}} + \beta_{\text{region}_i} + \gamma_{\text{state}_i} + \gamma_{\text{district}_i} + \cdots).$$

More generally, a nonnested model has the form $p_i = \Phi((X\beta)_i)$, with geographic and demographic predictors X. This sort of model is consistent with understanding swings in votes (Gelman and King, 1994) and public opinion in the short term (Gelman and King, 1993) and long term (Page and Shapiro, 1992). Further work is needed to study how the votes in these models aggregate and their implications for voting power. Mathematically, this is related to models in spatial statistics and their implications for the sampling distribution of spatial averages (Whittle, 1956; Ripley, 1981), with the additional analytical difficulties that arise from the nonlinear probit transformation.

5.2.1 *A stochastic model on a tree of voters.* A starting point for theoretical exploration, by analogy to the Ising models discussed in Section 5.1, is to apply the additive model to a regular tree structure, with each node of the tree having a continuous value z. We consider a simple but nontrivial random walk model, in which independent error terms $\varepsilon \sim \mathrm{N}(0, \tau^2)$ are assigned to each node of the tree, and then, for each node, the value z is defined as the sum of the ε's for that node and all the nodes above it in the tree.

We work with a tree of depth D with k branches at each node, thus representing $N = k^D$ voters. The value z for a node at depth d of the tree is then the sum of $d + 1$ independent $\mathrm{N}(0, \tau^2)$ terms starting at the root and working down to the node.

For any of the k^D leaf nodes i (i.e., voters), the probability $p_i = \Pr(V_i = +1)$ is set to $\Phi(z_i)$. In our model, the values z_i at the leaves marginally have $\mathrm{N}(0, (D+1)\tau^2)$ distributions but with a correlation structure induced by the tree. Such a model has three parameters: D, k and τ, and we can explore the variation of average votes at different levels of aggregation, as a function of these parameters.

In the Ising model, we did not need to consider the depth D of the larger tree in evaluating the properties of average votes in subtrees of depth d. In contrast, the distribution of the votes in the random walk model depends on the higher branches of the tree. This makes sense from a political standpoint, because state-level votes, for example, are affected by national and regional as well as statewide and local swings.

5.2.2 *The distribution of the average of n votes.* As in Section 5.1, we shall determine the variance of $\overline{V}^{(d)}$, the proportional vote differential based on averaging $n = k^d$ voters. For this model, we compute the variance by counting the number of pairs of the n voters that are a distance $0, 1, 2, \ldots, d$ apart in the tree,

$$\mathrm{var}\Big(\overline{V}^{(d)}\Big) = \frac{1}{k^d}\left(1 + \sum_{\delta=1}^{d}(k-1)k^{\delta-1}\rho_\delta\right), \tag{16}$$

where ρ_δ is the correlation between the votes v_i at two leaves a distance δ apart in the tree. In deriving (16), we have used the fact that, from the symmetry of the model, each $v_i = \pm 1$ has a marginal mean of 0 and variance of 1.

For each δ, the correlation ρ_δ can be determined based on the bivariate normal distribution: if voters i and j are at a distance δ, then we can write

$$\begin{aligned}\rho_\delta &= \Pr(v_i = +1 \text{ and } v_j = +1)\\ &\quad + \Pr(v_i = -1 \text{ and } v_j = -1)\\ &\quad - \Pr(v_i = +1 \text{ and } v_j = -1)\\ &\quad - \Pr(v_i = -1 \text{ and } v_j = +1)\\ &= 2A - (1 - 2A)\\ &= 4A - 1,\end{aligned}$$

where A is the area in the positive quadrant of the bivariate normal distribution with mean 0 and variance matrix

$$\begin{pmatrix}(D+1)\tau^2 + 1 & (D-\delta+1)\tau^2\\ (D-\delta+1)\tau^2 & (D+1)\tau^2+1\end{pmatrix}.$$

The extra "+1" term in the variance here corresponds to the latent N(0, 1) error term in the probit model, $\Pr(v_i = +1) = \Phi(z_i)$. Evaluating the area of the normal distribution, we obtain

$$\rho_\delta = \frac{2}{\pi}\arcsin\left(\frac{(D+1-\delta)\tau^2}{(D+1)\tau^2+1}\right).$$

Thus,

$$\begin{aligned}\mathrm{var}\Big(\overline{V}^{(d)}\Big) = {}& \frac{1}{k^d}\left(1 + \frac{2}{\pi}(k-1)\sum_{\delta=1}^{d}k^{\delta-1}\right.\\ &\left.\times \arcsin\left(\frac{(D+1-\delta)\tau^2}{(D+1)\tau^2+1}\right)\right).\end{aligned} \tag{17}$$

Because of the power of k, the last terms (with higher values of δ) will dominate in the summation in (17). For these higher terms, the expression $((D+1-\delta)\tau^2)/((D+1)\tau^2+1)$ will be close to 0, and so the arcsine can be approximated by the identity function. We use this approximation to gain understanding of the behavior of $\mathrm{var}(\overline{V}^{(d)})$, knowing that we can compare to the exact formula (17) at any point. Approximating $\arcsin(x)$ by x yields

$$\begin{aligned}\mathrm{var}\Big(\overline{V}^{(d)}\Big) &\approx \frac{1}{k^d}\left(1 + \frac{2}{\pi}(k-1)\sum_{\delta=1}^{d}k^{\delta-1}\right.\\ &\qquad\left.\times\left(\frac{(D+1-\delta)\tau^2}{(D+1)\tau^2+1}\right)\right)\\ &\approx \frac{1}{k^d} + \frac{2}{\pi}\frac{\tau^2}{(D+1)\tau^2+1}\\ &\qquad \times\left(D(1-k^{-d}) + \frac{k}{k-1} - d\right).\end{aligned}$$

We further simplify by ignoring terms of order $k^{-d} = 1/n$, to obtain

$$\mathrm{var}\Big(\overline{V}^{(d)}\Big) \approx \frac{2}{\pi}\frac{\tau^2}{(D+1)\tau^2+1}\left(D + \frac{k}{k-1} - d\right). \tag{18}$$

This is simply a linear function in $\log n$ (recall that $n = k^d$, so $\log n = d\log k$),

$$\mathrm{var}\Big(\overline{V}^{(d)}\Big) \approx a - b\log n, \tag{19}$$

where the parameters a and b are determined by D, k and τ, the parameters of the underlying stochastic model. For numbers of voters n in the thousands to millions (as in our electoral data), we compared (19) to (17) and found the approximation to be essentially exact.

5.2.3 *Fitting the model to electoral data.* The model (19) is much different from a power law but it actually behaves similarly over a fairly wide dynamic range of n. For example, the Presidential votes by state displayed in Figure 3 have n ranging from 60,000 to 10 million. The best-fit line of the form (19) to these data is $\mathrm{var}(\overline{V}^{(d)}) = 0.20 - 0.0113\log n$. Assuming a normal distribution, this implies $\mathrm{E}(|\overline{V}^{(d)}|) = 0.8\sqrt{0.20 - 0.0113\log n}$, which we display in Figure 6 along with the previously fitted power-law curve. The two lines look almost identical, and it would be close to hopeless to try to distinguish between them from the data.

We can map the fitted values $a = 0.020$ and $b = 0.0113$ to D, k and τ in (18). Since we are fitting three parameters to two, we can set one of them arbitrarily; for simplicity, we set $k = 2$ for a binary tree. Then the fitted values are $N = 12$ million and $\tau = 0.175$. As with the Ising model in the previous section, we do not want to take these parameters too seriously; these estimates are merely intended to give insight into the sort of model that could predict patterns of vote margins that occur in real electoral data.

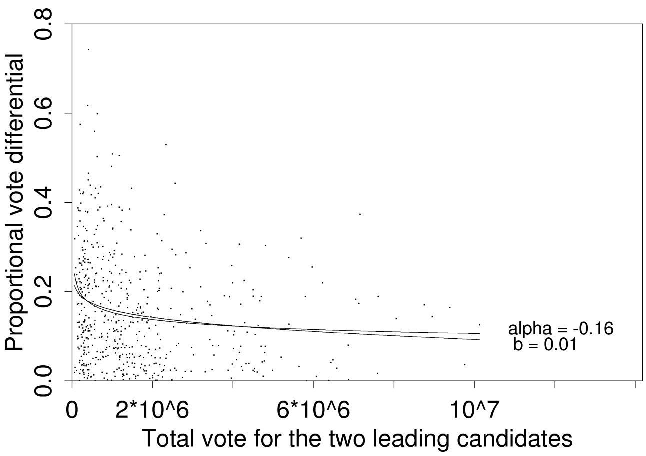


FIG. 6. *The proportional margin in state votes for President as a function of the number of voters n_j in the state, repeated from Figure* 3. *The best-fit lines of the form cn_j^{α} and $\sqrt{a - b\log n_j}$ are displayed. The power law is consistent with the Ising model described in Section* 5.1 *and the logarithmic form is consistent with the random walk model described in Section* 5.2. *Both fit the data much better than the $1/\sqrt{n_j}$ curve predicted by the random voter model* (*see Figure* 3).

6. DISCUSSION

Voting power is important for studying political representation, fairness and strategy, and has been much discussed in political science. Although power indexes are often considered as mathematical definitions, they ultimately depend on statistical models of voting. As we have seen in Section 3, even the simplest default of random voting is full of subtleties in its implications for voting power. However, as seen in Sections 4 and 5, more realistic data-based models lead to drastically different substantive conclusions about fairness and voting power in important electoral systems such as the U.S. Electoral College. Further work is needed to develop models of individual voters in a way consistent with available data on elections and voting, and to understand the implications of these models for voting power.

We conclude with a discussion of the fundamental connections between individual voting power and political representation.

6.1 Fundamental Conflict between Decisiveness of Votes and Legitimacy of Election Outcomes

Our mathematical and empirical findings do not directly address normative questions such as, "Which electoral system should be used?" or, in a legislature, "How should committees or subcommittees be assigned?" Let alone more fundamental questions such as, "Is it desirable for the average voting power to be increased?" After the 2000 election, some commentators suggested that it would be better if close elections were *less* likely, even though close elections are associated with decisiveness of individual votes, which seems like a good thing.

The issue of the desirability of close elections raises a conflict between two political principles: on one hand, *democratic process* would seem to require that every person's vote has a nonzero chance (and, ideally, an equal chance) of determining the election outcome. On the other hand, very close elections such as Florida's damage the *legitimacy* of the process, and so it might seem desirable to reduce the probability of ties or extremely close votes.

No amount of theorizing will resolve this difficulty, which also occurs in committees and leads to legitimacy-protecting moves such as voting with an informal straw poll. The official vote that follows is then often close to unanimous as the voters on the losing side switch to mask internal dissent. This article's theoretical findings on the benefits of coalitions imply that such behavior is understandable but in a larger context can reduce the average voting power of individuals.

6.2 Limitations of Individualistic Measures of Group Power

We must also realize that individual measures of political choice, even if aggregated, cannot capture the structure of group power. For one thing, groups that can mobilize effectively are solving the coordination problem of voting and can thus express more power through the ballot box (Uhlaner, 1989). For an extreme example, consider the case of Australia, where at one time Aboriginal citizens were allowed, but not required, to vote in national elections, while non-Aboriginal citizens were required to vote. Unsurprisingly, turnout was lower among Aboriginals. Who was benefiting here? From an individual-rights standpoint, the Aboriginals had the better deal, since they had the freedom to choose whether to vote, but as a group, the Aboriginals' lower turnout would be expected to hurt their representation in the government and thus, probably, hurt them individually as well. Having voting power is most effective when you and the people who share your opinions actually vote.

ACKNOWLEDGMENTS

We thank Yuval Peres for alerting us to the Ising model on trees described in Section 5.1 and suggesting its application to votes. We also thank Peter Dodds, Amit Gandhi, Hal Stern, Jan Vecer, Tian Zheng and two referees for helpful discussions and comments. This work was supported in part by NSF Grants SES-9987748 and SES-0084368.

REFERENCES

ALDRICH, J. H. (1993). Rational choice and turnout. *American Journal of Political Science* **37** 246–278.

BANZHAF, J. F. (1965). Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review* **19** 317–343.

BANZHAF, J. F. (1968). One man, 3.312 votes: A mathematical analysis of the Electoral College. *Villanova Law Review* **13** 304–332.

BECK, N. (1975). A note on the probability of a tied election. *Public Choice* **23** 75–79.

BEITZ, C. R. (1989). *Political Equality: An Essay in Democratic Theory*. Princeton Univ. Press.

BLEHER, P. M., RUIZ, J. and ZAGREBNOV, V. A. (1995). On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice. *J. Statist. Phys.* **79** 473–482.

BRAMS, S. J. and DAVIS, M. D. (1974). The 3/2's rule in presidential campaigning. *American Political Science Review* **68** 113–134.

BRAMS, S. J. and DAVIS, M. D. (1975). Comment on "Campaign resource allocations under the electoral college," by C. S. Colantoni, T. J. Levesque and P. C. Ordeshook. *American Political Science Review* **69** 155–156.

CHAMBERLAIN, G. and ROTHSCHILD, M. (1981). A note on the probability of casting a decisive vote. *J. Econom. Theory* **25** 152–162.

CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74** 829–836.

COLANTONI, C. S., LEVESQUE, T. J. and ORDESHOOK, P. C. (1975). Campaign resource allocations under the electoral college (with discussion). *American Political Science Review* **69** 141–161.

EDLIN, A., GELMAN, A. and KAPLAN, N. (2002). Rational voters vote for Social good. Technical report, Dept. Statistics, Columbia Univ.

EVANS, W., KENYON, C., PERES, Y. and SCHULMAN, L. J. (2000). Broadcasting on trees and the Ising model. *Ann. Appl. Probab.* **10** 410–433.

FELSENTHAL, D. S. and MACHOVER, M. (1998). *The Measurement of Voting Power*. Elgar, Cheltenham, UK.

FELSENTHAL, D. S. and MACHOVER, M. (2000). Enlargement of the EU and weighted voting in its council of ministers. Report VPP 01/00, Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science.

FELSENTHAL, D. S., MAOZ, Z. and RAPOPORT, A. (1993). An empirical evaluation of six voting procedures: Do they really make any difference? *British Journal of Political Science* **23** 1–27.

FEREJOHN, J. and FIORINA, M. (1974). The paradox of not voting: A decision theoretic analysis. *American Political Science Review* **68** 525–536.

FRIEDGUT, E. and KALAI, G. (1996). Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.* **124** 2993–3002.

GARRETT, G. and TSEBELIS, G. (1999). Why resist the temptation to apply power indices to the European Union? *Journal of Theoretical Politics* **11** 291–308.

GELMAN, A. (2002). Voting, fairness, and political representation (with discussion). *Chance*. **15**(3) 22–26.

GELMAN, A., KATZ, J. N. and BAFUMI, J. (2002). Why standard voting power indexes don't work. An empirical analysis. Technical report, Dept. Statistics, Columbia Univ.

GELMAN, A. and KING, G. (1993). Why are American presidential election campaign polls so variable when votes are so predictable? *British Journal of Political Science* **23** 409–451.

GELMAN, A. and KING, G. (1994). A unified method of evaluating electoral systems and redistricting plans. *American Journal of Political Science* **38** 514–554.

GELMAN, A., KING, G. and BOSCARDIN, W. J. (1998). Estimating the probability of events that have never occurred: When is your vote decisive? *J. Amer. Statist. Assoc.* **93** 1–9.

GOOD, I. J. and MAYER, L. S. (1975). Estimating the efficacy of a vote. *Behavioral Science* **20** 25–33.

HEARD, A. D. and SWARTZ, T. B. (1999). Extended voting measures. *Canad. J. Statist.* **27** 173–182.

KATZ, J. N., GELMAN, A. and KING, G. (2002). Empirically evaluating the electoral college. In *Rethinking the Vote: The Politics and Prospects of American Election Reform* (A. N. Crigler, M. R. Just and E. J. McCaffery, eds.). Oxford Univ. Press.

LEECH, D. (2002). An empirical comparison of the performance of classical (voting) power indices. *Political Studies* **50** 1–22.

LUCE, R. D. and RAIFFA, H. (1957). *Games and Decisions: Introduction and Critical Survey*, Chap. 12. Wiley, New York.

MANN, I. and SHAPLEY, L. S. (1960). The a priori voting strength of the electoral college. Memo, RAND. Reprinted (1964) in *Game Theory and Related Approaches to Social Behavior: Selections* (M. Shubik, ed.). Wiley, New York.

MARGOLIS, H. (1977). Probability of a tie election. *Public Choice* **31** 135–138.

MARGOLIS, H. (1983). The Banzhaf fallacy. *American Journal of Political Science* **27** 321–326.

MERRILL, S., III (1978). Citizen voting power under the Electoral College: A stochastic model based on state voting patterns. *SIAM J. Appl. Math.* **34** 376–390.

MULLIGAN, C. B. and HUNTER, C. G. (2001). The empirical frequency of a pivotal vote. Working Paper 8590, National Bureau of Economic Research.

NATAPOFF, A. (1996). A mathematical one-man one-vote rationale for Madisonian Presidential voting based on maximum individual voting power. *Public Choice* **88** 259–273.

NOAH, T. (2000). Faithless elector watch: Gimme "equal protection." *Slate*, Dec. 13 (slate.msn.com/?id=1006680).

OWEN, G. (1975). Evaluation of a presidential election game. *American Political Science Review* **69** 947–953.

PAGE, B. I. and SHAPIRO, R. Y. (1992). *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences*. Univ. Chicago Press.

PENROSE, L. S. (1946). The elementary statistics of majority voting. *J. Roy. Statist. Soc.* **109** 53–57.

POOLE, K. T. and ROSENTHAL, H. (1997). *Congress: A Political-Economic History of Roll Call Voting*. Oxford Univ. Press.

RABINOWITZ, G. and MACDONALD, S. E. (1986). The power of the states in U.S. Presidential elections. *American Political Science Review* **80** 65–87.

RIKER, W. H. and ORDESHOOK, P. C. (1968). A theory of the calculus of voting. *American Political Science Review* **62** 25–42.

RIPLEY, B. D. (1981). *Spatial Statistics*. Wiley, New York.

SAARI, D. G. and SIEBERG, K. K. (1999). Some surprising properties of power indices. Discussion Paper 1271, Center of Mathematical Economics, Northwestern Univ.

SHAPLEY, L. S. (1953). A value for n-person games. In *Contributions to the Theory of Games* **2** (H. W. Kuhn and A. W. Tucker, eds.) 307–317. Princeton Univ. Press.

SHAPLEY, L. S. and SHUBIK, M. (1954). A method for evaluating the distribution of power in a committee system. *American Political Science Review* **48** 787–792.

SNYDER, J. M., TING, M. M. and ANSOLABEHERE, S. (2001). Legislative bargaining under weighted voting. Technical Report, Dept. Economics, MIT.

STRAFFIN, P. D. (1978). Probability models for power indices. In *Game Theory and Political Science* (P. Ordeshook, ed.) 477–510. New York Univ. Press.

STROMBERG, D. (2002). Optimal campaigning in Presidential elections: The probability of being Florida. Technical report, IIES, Stockholm Univ.

UHLANER, C. J. (1989). Rational turnout: The neglected role of groups. *American Journal of Political Science* **33** 390–422.

WATTS, D. J., DODDS, P. S. and NEWMAN, M. E. J. (2002). Identity and search in social networks. *Science* **296** 1302–1305.

WHITTLE, P. (1956). On the variation of yield variance with plot size. *Biometrika* **43** 337–343.