DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

THE MATHEMATICS AND STATISTICS OF VOTING POWER

Andrew Gelman
Columbia Univerisity

Jonathan N. Katz
California Institute of Technology

Francis Tuerlinckx
California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 1141

October, 2002

The Mathematics and Statistics of Voting Power*

Andrew Gelman† Jonathan N. Katz‡
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 have usually 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. Formation of coalitions has features of the prisoner's dilemma game, in that a coalition increases voting power for its members but decreases the average voting power of the electorate. We then 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.

Keywords: Banzhaf index, coalitions, cooperation, decisive vote, elections, electoral college, political science, prisoner's dilemma, trees.

*We thank Peter Dodds, Amit Gandhi, Yuval Peres, Hal Stern, Jan Vecer, and Tian Zheng for helpful discussions. This work was supported in part by grants SES-9987748 and SES-0084368 of the U.S. National Science Foundation.

†Department of Statistics, Columbia University, New York, NY, 10027; gelman@stat.columbia.edu, http://www.stat.columbia.edu/~gelman/.

‡Division of the Humanities and Social Sciences, D.H.S.S. 228-77, Pasadena, CA 91125; jkatz@caltech.edu, http://jkatz.caltech.edu.

§Department of Statistics, Columbia University, New York, NY, 10027; tuerlinckx@stat.columbia.edu

1 Introduction

Consider an election with four voters representing constituencies of unequal size, who are given weights of 12, 9, 6, and 2, respectively. Each voter must choose between two parties, and the party wins that receives a higher total weighted vote. This simple example illustrates two potential problems with weighted voting. First, 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. Second, 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 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. As the above example illustrates, these tasks are not trivial even in small elections.

This paper reviews some of the mathematical findings on voting power and places 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. In Section 4, 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 5 concludes with an assessment of the connections between voting power and political representation.

The main results presented in this paper are:

1. For the random voting model, we derive results for optimal coalition sizes and the effects on voting power of various coalition structures.

2. We show a connection between voting power and the formation of coalitions, which leads to a prisoner's dilemma in which coalitions are locally beneficial but have negative results for society as a whole.

3. We discuss the general relevance of voting power in politics and the mistaken recommendations that have arisen from simplistic voting power calculations.

4. We 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. This paper focuses on weighted majority voting (as in the example in the first paragraph of Section 1) 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).

We shall also consider bloc voting, in which subsets of the electorate can voluntarily form binding coalitions, so that all the votes for a bloc are assigned to the party that wins a majority of the vote within the bloc. (Recall that we assume two parties and no abstention throughout.) The blocs can themselves be nested. We assume that tied elections at all levels are decided by coin flips.

Voting power measures have been developed in the 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) give a recent review. This paper 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, for example, 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, filibusters, and so forth (Shapley and Shubik, 1954). Luce and Raiffa (1957) discuss power indexes as an application of game theory, following Shubik (1953).

2.2 Definitions of voting power

Voting power can and has been defined in a variety of ways (see Saari and Sieberg, 1999, and Heard and Swartz, 1999); in this paper 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, $v = (v_1, \ldots, v_n)$ 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.

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

$$\text{power}_i = \text{Power of voter } i = \Pr(R = +1|v_i = +1) - \Pr(R = +1|v_i = -1). \tag{1}$$

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

A related measure that is sometimes considered is the probability of being satisfied with the election outcome (Straffin, 1978, Heard and Swartz, 1999):

$$\begin{aligned}\Pr(\text{satisfied}) &= \text{Probability that voter } i \text{ is satisfied} \\ &= \Pr(R = v_i) \\ &= \Pr(R = +1|v_i = +1)\Pr(v_i = +1) + \Pr(R = -1|v_i = -1)\Pr(v_i = -1). \end{aligned} \tag{2}$$

If $\Pr(v_i = +1) = \frac{1}{2}$, then the measures (1) and (2) are essentially equivalent, because

$$\begin{aligned}\text{power}_i &= \Pr(R = +1|v_i = +1) + \Pr(R = -1|v_i = -1) - 1 \\ &= \frac{1}{2}\Pr(\text{satisfied}) - 1 \qquad (\text{if } \Pr(v_i = +1) = \tfrac{1}{2}). \end{aligned} \tag{3}$$

In general, however, the power and satisfaction measures differ and address slightly different concerns. For example, consider an electoral system in which 90% of the voters pick one candidate and 10% pick the other. Then 90% of the voters are satisfied with the outcome, but the voters are essentially powerless at the individual level. Politically, this means that the winning candidate does not need to woo the voters. Thus, getting the desired outcome is not the same as having power or influence. In general there are many ways of measuring the fairness of an electoral system (see, for example, Gelman, 2002); we recognize that the results on voting power will not directly apply to other measures. We are working in the tradition of game theory in which voting power summarizes one aspect of the multi-player election game.

We shall work with the following situations:

1. For simple *weighted voting*, we assume n voters with weights $w_i, i = 1, \ldots, n$, and an aggregation rule $R(v) = \text{sign}(\sum_{i=1}^n w_i v_i)$; that is, the winner is determined by weighted majority. We assume throughout that ties are decided by coin flips.

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.

3. For *coalition-formation*, we assume n voters with equal weights that are allowed to form arbitrary coalitions. The coalitions can be nested so that the entire electorate has a tree structure. At the bottom of the tree, elections are decided by majority rule, with a winner-take-all rule for each coalition.

2.3 Claims in the literature and connections to empirical data

Social scientists have studied both the theoretical and 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, and 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 4.1).

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 have 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 1000 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

$$\Pr(\text{tie election}) \approx \frac{f_{\overline{V}}(0)}{n}, \tag{4}$$

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 4.1). 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 4.1, 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. In addition, as we discuss in Section 3.3, even this simplified model has some interesting features.

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 semi-decisive 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

$$\text{for weighted voting:} \quad \text{power}_i = \Pr(|V_{-i}| < w_i) + \frac{1}{2}\Pr(|V_{-i}| = w_i). \tag{5}$$

From the random voting model, we can immediately derive that $\mathrm{E}(V_{-i}) = 0$ and $\mathrm{sd}(V_{-i}) = \sqrt{\sum_{k \neq i} w_k^2} = \sqrt{W^2 - w_i^2}$. If certain regularity conditions hold—if the number of voters is large enough, and 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 (5) by $\Phi(\frac{w_i}{\sqrt{W^2 - w_i^2}}) - \Phi(\frac{-w_i}{\sqrt{W^2 - w_i^2}})$, where Φ is the cumulative normal distribution function.

If $w_i^2 << W_i^2$, 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, power$_i \approx \sqrt{\frac{2}{\pi}}\frac{w_i}{W}$. 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 paper 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_j'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

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

under the random voting model. Unless n is very small, this can be approximated as $\sqrt{\frac{2}{\pi n_j}}$, which is a special case of (4).

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

$$\text{for two-stage random voting:} \quad \text{power}_i \approx \frac{w_j}{\sqrt{n_j}}, \tag{6}$$

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 4.1.

3.3 Coalitions

Voting power is strongly linked to the theory of coalitions in cooperative games (see Luce and Raiffa, 1957). We present some results here for the random voting model and then return in Section 4.3 to discuss how coalitions could be studied under more general models.

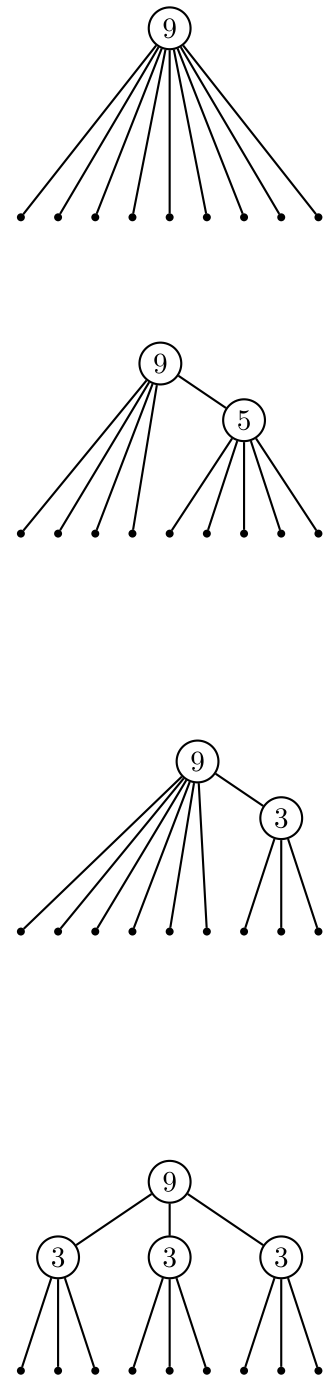
3.3.1 Effects of coalitions on individual and average voting power

The starting point and fundamental result is that a group of voters can increase their individual voting powers by forming a *coalition*, which we define here as a binding agreement to pool their votes so that they all go to the winner in the coalition. Coalitions can be nested, so that at each level, the already-computed votes of subcoalitions are counted as weighted votes to determine the coalition winner.

Individual voting power. Figure 1 illustrates an example with various systems of coalitions for an election of 9 voters. The top level of the tree is itself a sort of coalition, in that the total vote is reduced to a simple $+1$ or -1 to determine a final winner. These trees (and the accompanying calculations) illustrate the benefits of joining a coalition—and they also illustrate the negative side: voters who are left out of a coalition tend to do worse than if no coalitions had been formed at all.

For example, consider scenario C, in which 3 voters are in a coalition and the other 6 vote independently. Then how likely is your vote to be decisive? If you are in the coalition, it is first necessary that the other 2 voters in the coalition be split; this happens with probability 1/2. Next, your coalition's 3 votes are decisive in the entire election, which occurs if the remaining 6 voters are divided 3-3 or 4-2; this has probability $\frac{50}{64}$. The voting power any of the 3 voters in the coalition is then $\frac{1}{2} \cdot \frac{50}{64} = 0.391$. What if you are not in the coalition? Then your vote will be decisive if the remaining votes are split 4-4, which occurs if the 5 unaffiliated voters (other than you) are split 4-1 in the direction opposite to the 3 voters in the coalition. The probability of this happening is $\binom{5}{1}2^{-5} = 0.156$. Compared to the simple majority system (scenario A in the figure), you have more voting power if you are in the coalition and less if you are outside. The average voting power is $\frac{3}{9} \cdot 0.391 + \frac{6}{9} \cdot 0.156 = 0.234$, which is lower than under majority voting.

Average voting power. One way to study the total effect of coalitions 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 coalition system. Figure 1 illustrates this point: the coalitions benefit their members but lower the average probability of decisiveness.



A. No Coalitions

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

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

Average Pr(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:

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

A voter not in the coalition can never be decisive:

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

Average Pr(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:

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

A voter not in the coalition is decisive with probability:

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

Average Pr(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 is decisive if others in the coalition are split 1-1 and the other two coalitions are split 1-1:

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

Average Pr(Voter is decisive) = 0.250

Figure 1: *An example of four different systems of coalitions with 9 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).*

Payoffs (in terms of your own voting power) from joining a coalition

	Have other voters formed coalitions?	
Your option	No	Yes
Stay alone	Moderate	Very low
Join a coalition	High	Low

Figure 2: *Schematic of the prisoner's dilemma facing an individual voters when deciding whether to form coalitions. Joining a coalition increases your voting power, whatever the behavior of the other voters. However, if all the voters join coalitions, all of them have low voting power. Thus, a suboptimal outcome is assured if all players act rationally.*

To prove the general result, we start with identity (3), which shows that, 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 in coalitions magnifies small differences (for example, 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 21 wins) vs. 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.

3.3.2 Forming coalitions as a prisoner's dilemma

Forming coalitions is beneficial to those who do it, but is negative to "society" as a whole, at least in terms of average voting power. From a political perspective, this is reminiscent of the prisoner's dilemma (see, for example, Luce and Raiffa, 1957), a game in which the behavior that benefits each player in the game has negative consequences for all the players. The situation for voters is illustrated in Figure 2. To follow the usual terminology from game theory, to refuse to join a coalition is "cooperating," in the sense that this refusal is cooperative behavior with respect to the general population of voters. Conversely, joining a coalition is "uncooperative" behavior relative to the general electorate, who are being excluded from the coalition.

As Figure 1 illustrates, it can be beneficial to join a coalition—especially if other voters have already done so. This raises the question, what sorts of coalition structures can arise spontaneously in a voting system? More formally, consider a set of n voters with no coalition structure (as in scenario A of Figure 2). Now allow groups of voters to join coalitions, with the rule that a set of voters will join a coalition only if the voting power increases for *each of the voters in the coalition*. Thus, we could move from scenario A to scenario C in Figure 2. From scenario C, we could then move to scenario D, since this transition benefits the six voters who would be joining the new coalitions.

We can think of this process as a walk in the space of trees. We shall refer to agreements as "locally beneficial" if the probability of decisiveness (voting power) increases for all the voters making the agreement. Possible agreements—that is, moves in tree space—include:

1. A set of separate voters forming a coalition
2. A coalition disbanding or dividing into sub-coalitions
3. A set of coalitions forming a super-coalition (without destroying their internal structure)
4. A set of coalitions merging into a single larger coalition.

In general, there can be more than one possible locally beneficial move from any given tree. For example, starting from scenario A in Figure 2, a move to either B or C is locally beneficial. And the move to B, for example, is not unique either, since any subset of 5 voters could form the coalition.

We can thus imagine rational voters moving through the space of trees, making locally beneficial agreements that lead to complex coalition structures that, as in scenario D of Figure 2, leave everyone worse off. These structures would not themselves be stable, however; for example, if the three coalitions in scenario D merge, they will return to scenario A. The actual behavior of the process depends on the moves that are allowed in tree space and the rules determining which locally beneficial agreements are made.

This nontransitivity of allowable moves between trees implies that there is no "objective function" that is increased by locally beneficial rules. This is related to the principle in economics that a group of agents can form a cartel that is Pareto-optimal within the group but has a negative utility for the larger economy of which it is a part.

3.3.3 Voting power with simple coalitions

To understand the potential benefits of coalitions for voting power, we begin with some relatively simple cases.

Forming a single coalition of size m. Consider a system of n separate voters, and suppose m of these form a coalition. Under random voting, each of the m individuals in

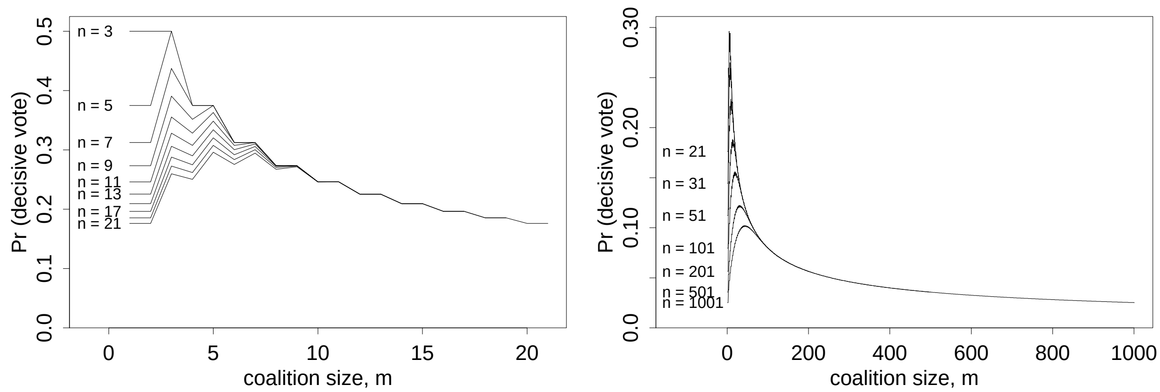


Figure 3: *Probability of decisiveness (voting power) for a voter in a coalition of m, within a general population of n voters that are otherwise not divided into coalitions. Each line on the graphs shows the voting power as a function of m for a given value of n. The two plots are on different scales. The random voting model is assumed.*

the coalition then has voting power,

$$\text{power}_i = \binom{m-1}{[(m-1)/2]} 2^{-(m-1)} \left(\Pr\left(\frac{n}{2} - m < x < \frac{n}{2}\right) + \frac{1}{2}\Pr\left(x = \frac{n}{2} - m\right) + \frac{1}{2}\Pr\left(x = \frac{n}{2}\right) \right), \quad (7)$$

where x has a binomial $(n-m, 1/2)$ distribution and represents the number of votes of "+1"among the $n-m$ voters not in the coalition. We in fact only need to evaluate (7) for odd values of n. If n is even, then voting power is unchanged if we evaluate at $n+1$, since the extra vote can only break a tie, which we are assuming would be done randomly anyway.

For any n, we can then evaluate (7) to see the potential gain in voting power from joining a coalition in an otherwise atomized electorate. Figure 3 shows voting power as a function of m, for each of several population sizes n. For clarity, low values of n are shown on the left plot and high values on the right plot. For any n, the minimum voting power is for $m = 1$ (no coalition), and, equivalently, $m = n$ (one large coalition). We can also see that it is never a good idea to have a coalition with an even number of members: if m is even, it is always as good or better to be in a coalition of size $m - 1$.

Figure 3 also shows that the optimal coalition size increases slowly with n. To explore this more fully, we display in Figure 4 a plot of optimal coalition size vs. population size. The dotted line on the graph shows an asymptotic form for large n that we derive here. We begin by approximating the two factors in (7) using the normal distribution:

$$\text{for large } n: \quad \text{power}_i \approx \sqrt{\frac{2}{\pi m}} \left(2\Phi\left(\frac{m}{\sqrt{n-m}}\right) - 1 \right). \quad (8)$$

Let $m_{\text{opt}}(n)$ be the value of m that maximizes (8) given n. To determine the behavior of m_{opt} for large n, we first note that $m_{\text{opt}}/n \to 0$ as $n \to \infty$ (because once m gets larger than $O(\sqrt{n})$, the cumulative normal density saturates and the $\frac{1}{\sqrt{m}}$ factor causes power$_i$

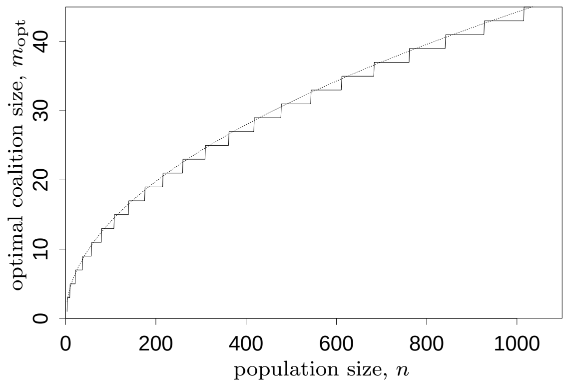


Figure 4: *Optimal coalition size m_{opt}, for maximizing voting power under random voting, in a population of n voters that are otherwise not divided into coalitions. The dotted line shows the approximation $m_{\text{opt}} = 1.4\sqrt{n}$. The solid line has jumps because the actual m_{opt} must be an integer and is always odd.*

to decline as m increases further). We can then approximate (8) by,

$$\text{for large } n: \quad \text{power}_i \approx \sqrt{\frac{2}{\pi m}}\left(2\Phi\left(\frac{m}{\sqrt{n}}\right) - 1\right) = n^{-1/4}\sqrt{\frac{2}{\pi R}}\,(2\Phi(R) - 1)\,, \qquad (9)$$

where $R = m/\sqrt{n}$. To maximize (9) given n, we need only optimize the factor $\sqrt{\frac{2}{\pi R}}\,(2\Phi(R) - 1)$, which we can do numerically: the maximum is 0.565 and is achieved at $R = 1.40$.

Thus, for large n, the optimal m is approximately $1.4\sqrt{n}$, and the voting power from being in such a coalition is approximately $0.57\,n^{-1/4}$. This approximation is in fact good for small n also, as can be seen in Figure 4. The voting power for the coalition member can be compared to the approximate voting power of $\sqrt{2\pi n} = 0.80\,n^{-1/2}$ if there is no coalition. Optimal coalition-formation allows power to decline as $O(n^{-1/4})$ rather than $O(n^{-1/2})$.

All voters in coalitions. Now suppose that all n voters, knowing the above result, arrange themselves into coalitions of size m. Then there will be n/m such coalitions (assuming n is large enough that we can ignore that the numbers will not divide evenly). If we assume that both m and n/m increase with n, then

$$\text{with two levels of coalitions:} \quad \text{power}_i \approx \sqrt{\frac{2}{\pi m}}\sqrt{\frac{2}{\pi(n/m)}} = \frac{2}{\pi\sqrt{n}} = 0.65\,n^{-1/2}.$$

Thus, if all the voters form coalitions, they all become worse off than if they had stayed apart (in which case they would each have power$_i = 0.80\,n^{-1/2}$).

Fully nested coalitions of size 3. In the most extreme case of coalition formation, consider an electorate of $n = 3^d$ voters that are arranged in coalitions of 3, that are themselves arranged in coalitions of 3, and so forth, to d levels. Then all the n voters are symmetrically-situated, and a given voter is decisive if the other 2 voters in his or her local coalition are split—this happens with probability $\frac{1}{2}$—and then the next two local coalitions must have opposite preferences—again, with a probability of $\frac{1}{2}$—and so on up to the top level. The probability that all these splits happen, and thus the individual voter is decisive, is $\left(\frac{1}{2}\right)^d = n^{-\log_3 2} = n^{-0.63}$, which is worse than the $O(n^{-1/2})$ result from simple majority voting.

3.3.4 More complex coalitions and weighted voting

In general, to evaluate the potential benefits of joining a coalition, then you must know the configuration of the other voters within the electorate. More specifically, if you are in a subtree that has n_{subtree} voters, and you are considering joining or leaving a coalition within that subtree, you must know the coalition structures of the others in the subtree. This will be necessary in order to determine the probability that your coalition is decisive within that subtree. At that point, the total vote of $\pm n_{\text{subtree}}$ propagates up the tree, and the structure of the remaining $n - n_{\text{subtree}}$ voters are irrelevant for the purpose of determining the proportional change in voting power from joining a coalition within the subtree.

Local calculations of changes in voting power. We assume that m voters will join a coalition only if it is locally beneficial, that is, if it increases the voting power for each of them (see Section 3.3.2). As noted just above, determining this increment of voting power in general requires calculation for all the other voters in the subtree. For a more tractable approximation, we consider the following local calculation of relative voting power.

Suppose that you are one of a small number m of voters in a subtree who are considering forming a coalition. Let V_{others} be the total vote in the rest of the subtree (applying winner-take-all rules within coalitions). The local calculation assumes that there are enough other distinct voters and voting blocs in the subtree that the distribution of V_{others} in the range $[-m, m]$ is approximately uniform. Then we can approximate the change in voting power by counting your expected influence: the expected number of votes you will swing alone or in the coalition.

Benefits from joining a coalition of 2 or 3. For example, suppose you are considering forming a coalition with one other voter, so that $m = 2$. If you stay apart, your influence is 2 votes (the effect of changing from -1 to $+1$). If you join a coalition, then your combined vote will be ± 2, but there is only a $\frac{1}{2}$ chance that your vote will swing this (because your vote will either make or break a tie). So your expected influence is 2 votes, and joining the coalition gives no benefit.

Now consider your influence if you join with two other voters, so that $m = 3$. Your vote is decisive if the other two are split, which under random voting has a probability of $\frac{1}{2}$, and if this happens there will be a vote swing of 6. So your vote has an expected influence of 3. Thus, joining the coalition increases your voting power by an estimated factor of 3/2.

Similar calculations show the estimated gains from joining larger coalitions. These calculations are valid as long as m is small compared to the standard deviation of V_{others}.

Weighted voting and coalitions of coalitions. Local calculations of approximate voting power can also be done for weighted voting. For example, consider again potential coalitions of size 2 or 3, but this time of weighted voters.

It is clear that it never makes sense for a coalition of 2: if the voters have equal weights, the earlier calculation applies, and if the weights are unequal, then the voter with lower weight will always be outvoted and will have no power in the coalition.

For the same reason, three voters with weights w_1, w_2, w_3 should consider joining a coalition only if their weights satisfy the triangle inequality: that is, $w_1 < w_2 + w_3$, and so forth. In this case, we can locally approximate the potential benefit of joining. Suppose you are voter 1, so that staying apart gives you an influence of $2w_1$ on the total vote in your subtree. If you join the coalition, there is a $\frac{1}{2}$ chance your vote will be swing the entire group; your expected influence is thus $\frac{1}{2} \cdot 2(w_1 + w_2 + w_3)$. The gain in expected influence from joining is then $w_2 + w_3 - w_1$, which we already know is positive from the triangle inequality condition. The stability of coalitions of 3 is thus robust and holds under weighted voting (as long as neither of the three voters dominates the other two).

The same calculations apply when considering whether it is beneficial for a set of coalitions to form a super-coalition. In this case, each of coalition acts as a weighted voter in determining voting power. If the coalitions each gain voting power, then so do the individual voters within.

4 Models with dependence and unequal probabilities

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. Progress has been made with all these approaches, but much work still remains.

In Section 4.1 we discuss the empirical failings of the random voting model and why this has major implications for voting power. Section 4.2 discusses more realistic probability models for votes, and in Section 4.3 we consider the possibilities for coalition formation when vote probabilities are unequal.

4.1 Empirical results on vote distributions and voting power

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 more than much more one-tenth of a percentage point.

Before discussing potential model improvements, we consider here the voting power implications of empirical problems with the random voting 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 (4) to approximate the probability that a single voter is decisive within district j as,

$$\Pr(\text{a vote is decisive within district } j) \approx \frac{f_j(0)}{n_j}, \qquad (10)$$

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

$$\Pr(\text{a vote is decisive within district } j) \approx \frac{1}{n_j \text{sd}(\overline{V}_j)}, \qquad (11)$$

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

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.

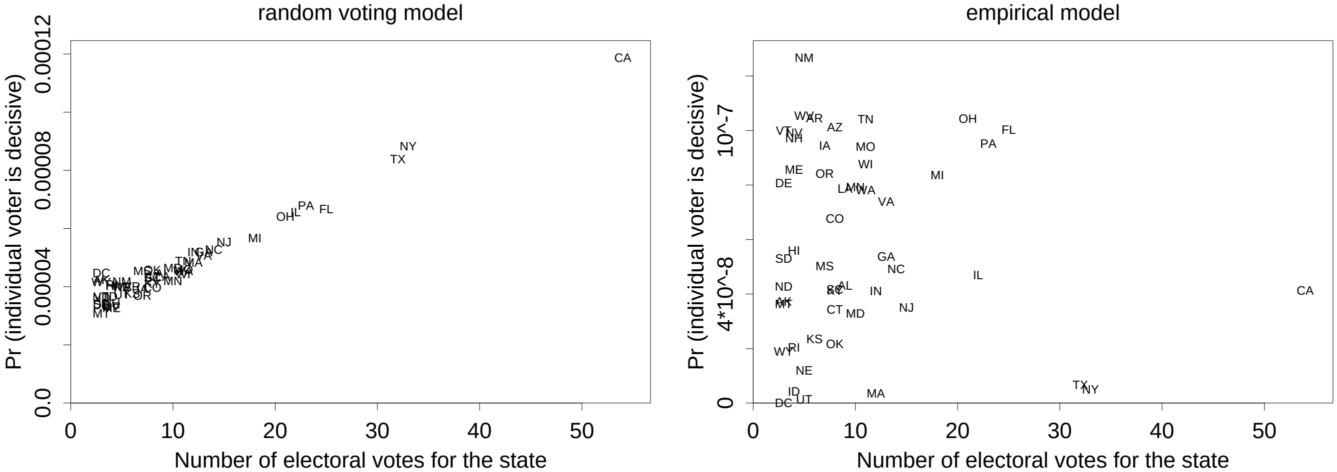


Figure 5: *Voting power for individuals in the 2000 U.S. Presidential election, by state: (a) under the random voting model, (b) under an empirical model based on randomly perturbing the actual election outcomes with variation at the national, regional, state, and local levels. Voters in large states have disproportionate power under the random voting model but not under the empirical model.*

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 (6) 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 5.

Thus, the general conclusion in the voting power literature is that the Electoral College benefits voters in large states. For example, Banzhaf (1968) claims 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) claim 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) come to similar conclusions. This impression has also made its way into the popular press; for example, Noah (2000) states, "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).

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 are much more likely to be close. This assumption can be tested with data. For example, Figure 6 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)

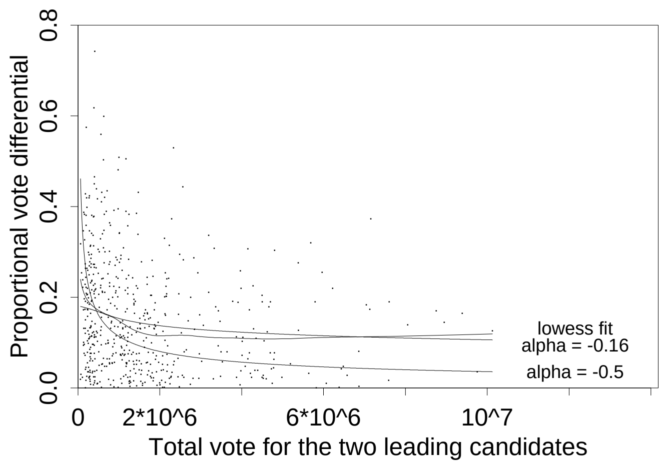


Figure 6: *The margin in state votes for President as a function of the number of voters n_j in the state: each dot represents a different state and election year from 1960–2000. The margins are proportional; for example, a state vote of 400,000 for the Democratic candidate and 600,000 for the Republican 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.*

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 relation much weaker than $1/\sqrt{n_j}$. As the scale of the graph makes clear, it is in practice 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.

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 5 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, $\overline{V}_i^{\text{obs}}$,

$$\overline{V}_i^{\text{sim}} = \overline{V}_i^{\text{obs}} + \epsilon^{\text{nation}} + \epsilon_k^{\text{region}} + \epsilon_j^{\text{state}} + \epsilon_i^{\text{district}}. \quad (12)$$

The $\overline{V}_i^{\text{sim}}$ values are then summed up within each state j to get a simulation of the state-level vote differentials, $\overline{V}_j$. The error terms in the simulation (12) 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 5b, 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 (10), 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 ϵ^{nation} and $\epsilon_k^{\text{region}}$ 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.

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 7 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.

The Electoral College and average voting power. As discussed at the end of Section 3.3.1, 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.

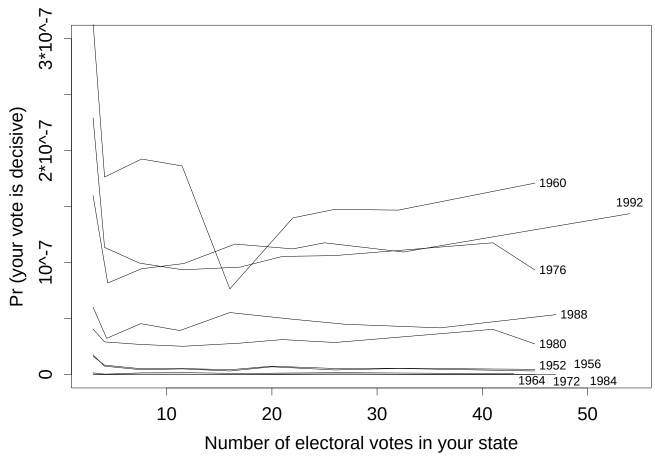


Figure 7: *The average probability of a decisive vote as a function of the number of electoral votes in the voter's state, for each U.S. Presidential election from 1952–1992 (excluding 1968, when a third party won in some states). The probabilities are calculated based on a forecasting model that uses 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.*

Using the model (12) based on perturbing district-by-district outcomes (as was used to calculate the values displayed in Figure 5b), 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.

Empirical evidence from other elections. Analyses of electoral data from U.S. Congress, U.S. state legislatures, and European national elections have also 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 5b)—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, for example, Felsenthal, Maoz, and Rapoport, 1993, Heard and Swartz, 1999, and Leech, 2000).

4.2 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 (10), 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, and Chamberlain and Rothchild, 1981.)

The next step is give a dependence structure to the voters' probabilities p_i. It makes sense to build this dependence upon existing relations 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 (for example, states, legislative districts, and precincts). It might also be appropriate to include structures based on non-nested predictors. For example, voters have similarities based on demographics as well as geography; non-nested models have also been used to capture "small-world" phenomena in social networks (Watts, Dodds, and Newman, 2002).

Sections 4.2.1 discusses an approach based on directly modeling dependence of individual votes, and Section 4.2.2 presents a model using correlated latent variables.

4.2.1 Discrete modeling of dependence of votes

One idea for modeling votes, 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 $e^{-\sum_{ij} c_{ij} v_i v_j}$, 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

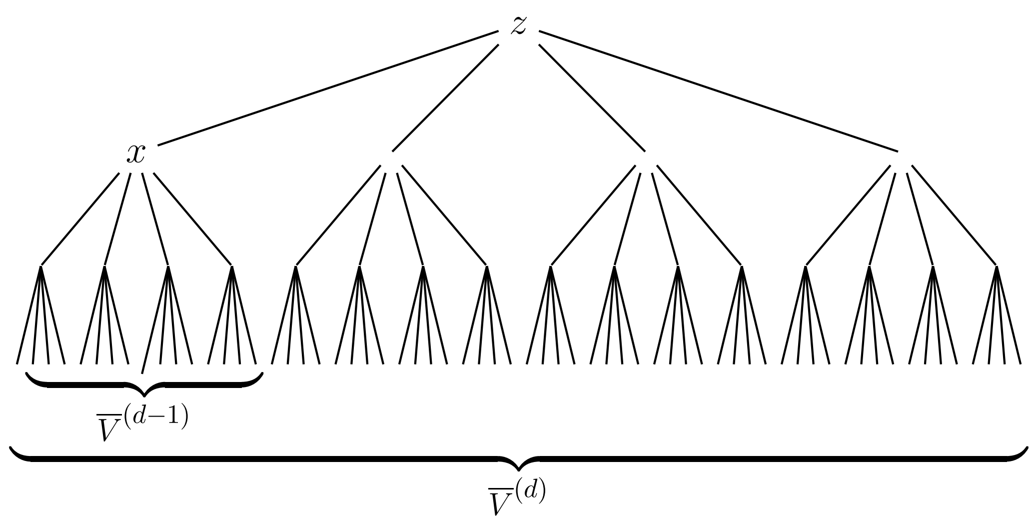


Figure 8: *Notation for the derivation of the standard deviation of the average vote differential, $\overline{V}^{(d)}$, for the Ising model on k^d voters in a tree structure. This tree, unlike that in Figure 1, represents dependence between voters,* not *a coalition structure. Here, z represents the 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.*

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 et al., 2000).

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 8); for the leaf nodes this variable represents a vote, and 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.

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 4.1, 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 $\text{sd}(\overline{V}_j) \propto n_j^{-0.5}$, whereas empirical electoral data show much weaker declines of the order of $\text{sd}(\overline{V}_j) \propto n_j^{-\alpha}$, with powers estimated at lower values such as $\alpha = 0.15$ (see Figure 6).

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 8. For an Ising-model tree of depth d, let $z = \pm 1$ be the equally-likely values at the root and $\overline{V}_d$ 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

$$\begin{aligned}
\mathrm{var}(\overline{V}^{(d)}) &= \mathrm{E}(\mathrm{var}(\overline{V}^{(d)}|z)) + \mathrm{var}(\mathrm{E}(\overline{V}^{(d)}|z)) \\
&= \frac{1}{2}\left(\mathrm{var}(\overline{V}^{(d)}|z=+1) + \mathrm{var}(\overline{V}^{(d)}|z=-1)\right) + \\
&\quad \frac{1}{2}\left((\mathrm{E}(\overline{V}^{(d)}|z=+1) - 0)^2 + \mathrm{E}((\overline{V}^{(d)}|z=-1) - 0)^2\right) \\
&= \mu_d^2 + \sigma_d^2,
\end{aligned} \tag{13}$$

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

$$\mu_d = \mathrm{E}(\overline{V}^{(d)}|z), \qquad \sigma_d^2 = \mathrm{var}(\overline{V}^{(d)}|z).$$

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,

$$\mu_0 = \mathrm{E}(\overline{V}^{(0)}|z=+1) = 1, \qquad \sigma_0^2 = \mathrm{var}(\overline{V}^{(0)}|z=+1) = 0.$$

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 8). Then we can use the symmetry of the underlying model to obtain,

$$\begin{aligned}
\mu_d &= \mathrm{E}(\mathrm{E}(\overline{V}^d|x)|z=+1) \\
&= (1-\pi)\cdot\mu_{d-1} + \pi\cdot(-\mu_d) \\
&= (1-2\pi)\mu_{d-1} \\
&= (1-2\pi)^d,
\end{aligned} \tag{14}$$

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:

$$\begin{aligned}
\sigma_d^2 &= \frac{1}{k}\left(\mathrm{E}(\mathrm{var}(\overline{V}^d|x)|z=+1) + \mathrm{var}(E(\overline{V}^d|x)|z=+1)\right) \\
&= \frac{1}{k}\left(\sigma_{d-1}^2 + 4\pi(1-\pi)\mu_{d-1}^2\right),
\end{aligned} \tag{15}$$

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 (14) and insert (13) to obtain,

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

and combining with (13) into (**??**) yields,

$$\mathrm{var}(\overline{V}^{(d)}) = \xi^d + (1-\xi)k^{-d}\sum_{i=0}^{d}(\xi k)^i, \tag{16}$$

where, for convenience, we have defined

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

We evaluate (15) 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 (15) as,

$$\begin{aligned}\sigma_d^2 &= \xi^d + (1-\xi)k^{-d}\frac{1-(\xi k)^d}{1-\xi k} \\ &\approx \left(\frac{1-\xi}{1-\xi k}\right)\frac{1}{n} \quad \text{for large } n.\end{aligned} \tag{17}$$

  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 6.

- If $\xi k = 1$, then (15) becomes,

$$\begin{aligned}\sigma_d^2 &= \xi^d + (1-\xi)k^{-d}\cdot d \\ &= \frac{1}{n}\left(1+\left(1-\frac{1}{k}\right)\log_k n\right) \\ &\approx \left(1-\frac{1}{k}\right)\frac{1}{n}\log_k n \quad \text{for large } n.\end{aligned} \tag{18}$$

  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 (15) becomes,

$$\begin{aligned}\sigma_d^2 &= \xi^d + (1-\xi)k^{-d}\frac{(\xi k)^d - 1}{\xi k - 1} \\ &\approx \frac{k-1}{k-1/\xi}\xi^d \quad \text{for large } n. \qquad (19)\end{aligned}$$

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 (18) as,

$$\sigma_d^2 \approx \left(\frac{k-1}{k-k^{2\alpha}}\right) n^{-2\alpha} \quad \text{for large } n. \qquad (20)$$

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

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 4.1). 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 6 can be fit by $\alpha = 0.16$. For $k = 2, 3, 10, 100$, the best-fitting π's are $0.05, 0.08, 0.15, 0.26, 0.33$.

Next, one can imagine setting the parameter k so that the coefficient $\sqrt{\frac{k-1}{k-k^{2\alpha}}}$ from (19) matches the coefficient c in the fitted curve, $\text{sd}(\overline{V}_j) \approx c n_j^{-\alpha}$. However, this second fitting step is not so effective, because the coefficient in (19) has a narrow range of possibilities. For example, for $\alpha = 0.16$, $\sqrt{\frac{k-1}{k-k^{2\alpha}}}$ 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 6 is 1.74.

Even if the estimated c from this dataset 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.

4.2.2 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 non-nested model have the form, $p_i = \Phi((X\beta)_i)$, with geographic and demographic predictors X. This sort of model is consistent with understanding of 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, 1987), with the additional analytical difficulties that arise from the nonlinear logistic transformation.

A stochastic model on a tree of voters. A starting point for theoretical exploration, by analogy to the Ising models discussed in Section 4.2.1, 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 $\epsilon \sim \mathrm{N}(0, \tau^2)$ 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 (that is, 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.

The distribution of the average of n votes. As in Section 4.2.1, 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:

$$\mathrm{var}(\overline{V}^{(d)}) = \frac{1}{k^d}\left(1 + \sum_{\delta=1}^{d}(k-1)k^{\delta-1}\rho_\delta\right), \tag{21}$$

where ρ_δ is the correlation between the votes v_i at two leaves a distance δ apart in the tree. In deriving (20), 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

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

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 $\mathrm{N}(0,1)$ 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,

$$\mathrm{var}(\overline{V}^{(d)}) = \frac{1}{k^d}\left(1 + \frac{2}{\pi}(k-1)\sum_{\delta=1}^{d} k^{\delta-1} \arcsin\left(\frac{(D+1-\delta)\tau^2}{(D+1)\tau^2+1}\right)\right). \tag{22}$$

Because of the power of k, the last terms (with higher values of δ) will dominate in the summation in (21). For these higher terms, the expression $\frac{(D+1-\delta)\tau^2}{(D+1)\tau^2+1}$ will be close to zero, and so the arcsine can be approximated by the identity function. We use this approximation in order to gain understanding of the behavior of $\mathrm{var}(\overline{V}^{(d)})$, knowing that we can compare to the exact formula (21) at any point. Approximating $\arcsin(x)$ by x yields,

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

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

$$\mathrm{var}(\overline{V}^{(d)}) \approx \frac{2}{\pi}\frac{\tau^2}{(D+1)\tau^2+1}\left(D + \frac{k}{k-1} - d\right). \tag{23}$$

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

$$\mathrm{var}(\overline{V}^{(d)}) \approx a - b \log n, \tag{24}$$

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 (23) to (21) and found the approximation to be essentially exact.

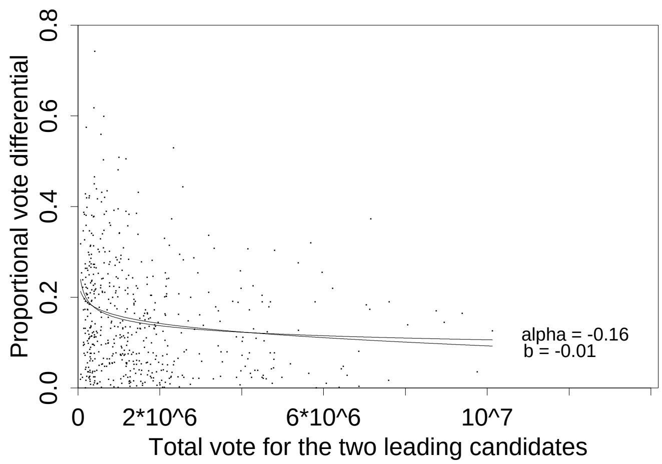


Figure 9: *The margin in state votes for President as a function of the number of voters n_j in the state, repeated from Figure 6. The best-fit lines of the form cn_j^α and $\sqrt{a - b \log n_j}$ are displayed. The power law is consistent with the Ising model described in Section 4.2.1, and the logarithmic form is consistent with the random walk model described in Section 4.2.2. Both fit the data much better than the $1/\sqrt{n_j}$ curve predicted by the random voter model (see Figure 6)*

Fitting the model to electoral data. The model (23) 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 6 have n ranging from 60,000 to 10 million. The best-fit line of the form (23) to these data is $\text{var}(\overline{V}^{(d)}) = 0.20 - 0.0113 \log n$. Assuming a normal distribution, this implies $\text{E}(|\overline{V}^{(d)}|) = 0.8\sqrt{0.20 - 0.0113 \log n}$, which we display in Figure 9 along with the previously-fit 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, $a = 0.020$, $b = 0.0113$, to D, k, and τ in (22). 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.

4.3 Formation and dissolution of coalitions

We have just discussed models of correlated votes given an existing structure on the voters. We now wish to consider the opposite problem: studying how voters can spontaneously form structure through mutually-beneficial coalitions. One way to understand the behavior of coalitions would be through some sort of simulation, thinking of voters as cellular automata that form structures of local agreements.

In going beyond random voting, we need a model in which voters have unequal p_i's and some sort of spatial structure (so that voters have neighbors with whom they can confer and consider joining coalitions). The model can be set up on a purely theoretical basis; for example, by using a Poisson process to place n voters on a two-dimensional space, assigning values of z_i from a Gaussian process, and then determining votes v_i with the rule, $\Pr(v_i = 1) = \Phi(z_i)$. Another approach, more appropriate to models of voting in a legislature, is to set up probabilities and a correlation structure based on data from roll-call votes. In this case, the voting options of -1 and -1 should be assigned consistently across the issues being voted on, as in Poole and Rosenthal (1997).

No benefit from joining a coalition of size 2. The next step is to generalize the results of Section 4.3 to go beyond the random voting model. This research can go in many directions; we illustrate here for the simple problem of evaluating the feasibility of coalitions of size 2 and 3. We shall evaluate the decisions based on the local calculations of expected influence, as described in Section 3.3.4.

Suppose you are a voter with probability p_1 deciding whether to join a coalition with another voter with probability p_2. If you stay apart, your influence is 2 votes (a change from -1 to $+1$). If you join the coalition, the expected total vote from the coalition is $2p_2$ if you vote $+1$ and $-2(1-p_2)$ if you vote -1, and so the difference—the expected influence—is still 2. There is, once again, no benefit to forming a coalition of size 2.

Potential benefit from joining a coalition of size 3. Now suppose you are a voter with weight w_1 and probability p_1 deciding whether to join a coalition with two other voters with weights w_2, w_3 and probabilities p_2, p_3. To determine the benefit from joining a coalition, note that if the other two voters agree, then your vote will have no influence. Your expected influence through the coalition is thus,

$$\begin{aligned}\text{expected influence in coalition} &= (w_1+w_2+w_3)\Pr(\text{voters 2 and 3 disagree}) \\ &= 2(w_1+w_2+w_3)(p_2(1-p_3)+p_3(1-p_2)). \quad (25)\end{aligned}$$

(Here we are assuming the voters are independent given the probabilities p_i. If not, the probability of disagreement can be calculated from the joint distribution.)

In any case, the local calculation states that you should join the coalition if this expected influence (24) is greater than the expected gain of $2w_1$ if you were to vote alone; that is,

$$\text{join if} \quad (w_1+w_2+w_3)(p_2(1-p_3)+p_3(1-p_2)) - w_1 > 0.$$

After some algebra, this can be rewritten as,

$$\text{join if} \quad (1-2p_2)(1-2p_3) < 1 - 2\frac{w_1}{w_1+w_2+w_3}. \quad (26)$$

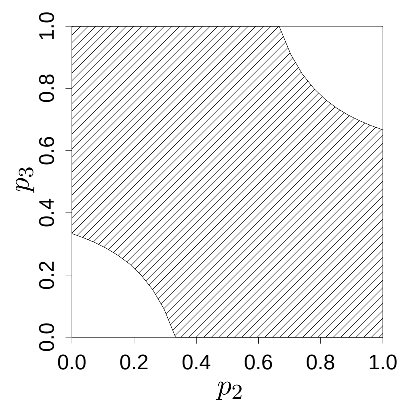


Figure 10: *Suppose three voters, with equal weights and probabilities p_1, p_2, p_3 of voting $+1$, are considering a forming a coalition. The shaded area on the graph shows the region of values of (p_2, p_3) for which it is beneficial for voter 1 to join the coalition. The coalition will form only if (p_1, p_2) and (p_2, p_3) are also in the shaded area.*

(The right side must be positive or else voter 1 is dominant and voters 2 and 3 would have no motivation to join the coalition.) For all three voters to be willing to join the coalition, condition (25) must also hold with the other two permutations of the indexes $\{1, 2, 3\}$.

For example, if $w_1 = w_2 = w_3$, then (25) becomes $(1 - 2p_2)(1 - 2p_3) < \frac{1}{3}$. The shaded region of Figure 10 shows the conditions on (p_2, p_3) where this condition holds. The points in the shaded region correspond to high probabilities that voters 2 and 3 will disagree, which is when it is beneficial for voter 1 to join the coalition. A sufficient (but not necessary) condition for this condition to hold for all three voters is that p_1, p_2, and p_3 all be in the range $(0.5 \pm 1/\sqrt{12}) = (0.21, 0.79)$. This is a fairly broad range of probabilities, indicating the robust strength of coalitions of size 3 when weights are equal.

Larger coalitions. For larger coalitions, we can continue to use expected influence to approximately determine whether a potential coalition is locally beneficial. For example, consider m voters with equal weights, potentially unequal probabilities p_i, and independent votes given the p_i's. (As discussed earlier, it is reasonable to assume independence if enough structure is built into the probabilities p_i.). Under the local calculation, the coalition is possible if the expected influence of any vote within the coalition is larger than what could be gained by voting alone. This depends on the probability that the total vote is tied, which in turn is strongly dependent on the expected vote differential, $\mathrm{E}(\overline{V}) = \frac{1}{m}\sum_{i=1}^{m}(1 - 2p_i)$. If $\mathrm{E}(\overline{V})$ is close to 0, then it can be very beneficial to join (see Section 3.3.3). However, if $|\mathrm{E}(\overline{V})| >> \mathrm{sd}(\overline{V})$, then the probability of a tie will be too low for the coalition to be viable.

Thus, a large group of voters who agree will *not* increase their individual voting power by joining a coalition, but individuals in a more heterogeneous group may benefit (in terms of voting power) from joining together. These results can change if we study the probability of satisfaction rather than voting power, since the identity (3) holds in general only if $p_i = \frac{1}{2}$.

5 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 Section 4, 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.

5.1 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 paper'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.

5.2 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.

References

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

Banzhaf, J. R. (1965). Weighted voting doesn't work: a mathematical analysis. *Rutgers Law Review* **19**, 317–343.

Banzhaf, J. R. (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.

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. *Journal of Statistical Physics* **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 allocation 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 Rothchild, M. (1981). A note on the probability of casting a decisive vote. *Journal of Economic Theory* **25**, 152–162.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74**, 829–836.

Colantoni, C. S., Levesque, T. J., and Ordeshook, P. C. (1975). Campaign resource allocation under the electoral college (with discussion). *American Political Science Review* **69**, 141–161.

Edlin, A., Gelman, A., and Kaplan, N. (2002). Rational voting. Technical report, Department of Statistics, Columbia University.

Evans, W., Kenyon, C., Peres, Y., and Schulman, L. J. (2000). Broadcasting on trees and the Ising model. *Annals of Applied Probability* **10**, 410–433.

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**, 1–28.

Felsenthal, D. S., and Machover, M. (1998). *The Measurement of Voting Power*. Edward Elgar Publishing.

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.

Ferejohn, J., and Fiorina, M. (1974). The paradox of not voting: a decision theoretic analysis. *American Political Science Review* **68**, 525.

Friedgut, E., and Kalai, G. (1996). Every monotone graph property has a sharp threshold. *Proceedings of the American Mathematical Society* **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*, to appear.

Gelman, A., Katz, J. N., and Bafumi, J. (2002). Why standard voting power indexes don't work. Technical report, Department of Statistics, Columbia University.

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 model for 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? *Journal of the American Statistical Association* **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. *Canadian Journal of Statistics*, **27**, 177–186.

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*, ed. A. N. Crigler, M. R. Just, and E. J. McCaffery. Oxford University Press.

Leech, D. (2002). An empirical comparison of the performance of classical power indices. *Political Studies*, to appear.

Luce, R. D., and Raiffa, H. (1957). *Games and Decisions: Introduction and Critical Survey*, chapter 12. Harvard University Press.

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

Margolis, H. (1977). Probability of a tie election. *Public Choice* **31**, 134–137.

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

Merrill, S. (1978). Citizen voting power under the Electoral College: a stochastic model based on state voting patterns. *SIAM Journal of Applied Mathematics* **34**, 376–390.

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

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*. University of Chicago Press.

Penrose, L. S. (1946). The elementary statistics of majority voting. *Journal of the Royal Statistical Society* **109**, 53–57.

Poole, K. T., and Rosenthal, H. (1997). *Congress: A Political-Economic History of Roll Call Voting*. Oxford University 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. (1987). *Spatial Statistics*. New York: Wiley.

Saari, D. G., and Sieberg, K. K. (1999). Some surprising properties of power indices. Northwestern University Center of Math Econ Discussion Paper #1271.

Shapley, L. S. (1953). A value for n-person games. In *Contributions to the Theory of Games II*, ed. H. W. Kuhn and A. W. Tucker, 307–317. Annals of Mathematics Studies **28**. Princeton University 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, Department of Economics, Massachusetts Institute of Technology.

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

Stromberg, D. (2002). Optimal campaigning in Presidential elections: the probability of being Florida. Technical report, IIES, Stockholm University, Sweden.

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). Search and identity in small-world networks. *Science*, to appear.

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