Spectators Who Spectate and Voters Who Vote

Tao Wang and Ruqu Wang

Economics Department, Queen’s University
Kingston, Ontario K7L 3N6, Canada

October 2011

Abstract

In this paper, we examine the paradox of voting, which assesses that, no one should vote in large elections, as the probability of one vote making a difference is too small compared to the cost of voting. We incorporate modern psychological and sociological findings in the classical rational choice model and model elections as similar to spectator sport events. Suspense and group effects are two important determining factors in attracting spectators in sports and voters in elections. We find that in a large election, the voting rate is significantly different from zero. Moreover, the voting rate is higher for the candidate who has more supporters. Using the polls before the election as a proxy for the number of supporters a candidate has, we analyze the 2004 US Presidential Election data, and the results support our theory.

Keywords: voting, turnout rate, rational choice, suspense, surprise effect, group effect

JEL Classification #: D03, D71, D72, D74

*We would like to thank Dan Usher and the participants at the 2008 International Workshop on Economic Theory in Shanghai for helpful comments. Ruqu Wang’s research is supported by the Social Sciences and Humanities Research Council of Canada.
1 Introduction

The “Paradox of Voting” has interested many economists and political scientists for decades. A bountiful literature has spawned on this area. The basic assumption in this literature is the rationality of voters. The initial formulation of the “rational choice model” by Downs (1957) and Tullock (1967) uses an expected utility framework; taking the cost of voting into account, voters vote if the expected net benefit from voting is positive, and abstain if it is negative. However, their model has led to the pathological conclusion that it is irrational to vote in large elections. A voter’s vote matters only if it is decisive, i.e., if it breaks a tie or it brings a tie. In large elections, the probability of a vote being decisive is extremely small. Therefore, if voters vote for the purpose of casting a decisive vote, then any reasonably small cost should prevent the voters from voting. This is the “Paradox of Voting.”

Trying to resolve this Paradox of Voting, Riker and Ordeshook (1968) extended the initial rational choice model by adding a term $D$ to a voter’s utility to capture the vote’s satisfaction from voting. Later, Barry (1976) and Blais (2000), among others, argue that the $D$ term can also be the sense of duty for a voter. Either way, this (positive) $D$ term seems to be very helpful in explaining the high voter turnout rate. However, it has received many critics, such as those in Mueller (2003), who criticizes that these $D$ terms save the rational choice model but destroy its predictive power.

In this paper, we continue to build on the principles of the rational choice model but assign a new meaning to the $D$ term by making use of modern psychological and sociological findings. By doing so, we predict a high voter turnout as well as restore the predictive power of the rational choice model. The findings that we will make use of is the literature studying spectatorship in sporting events. Sporting events provide spectators with suspense and excitement, the exact reasons why people pay to watch them. We consider an election as nothing too different. First, most of the sport competitions consist of two teams, and sometimes more teams. Elections are similar. Most of the time, there are also two major candidates, and sometimes more. Second, suspense is high when the
two competing teams are equally strong, and it is low when one team dominates. It is exactly the same in elections with two candidates. Last, spectators enjoy the excitement significantly more from watching the contests in stadiums rather than from learning the results from TVs or newspapers, especially when the team they support wins. Even if their team loses, they still benefit from the comforting from surrounding team supporters. Elections are much alike. Voters also obtain enjoyment from going to the polls and later being surrounded by other voters voting for the same candidate.

According to this psychology and sociology literature, people derive utility from participating in or watching contests (O’Keeffe, Viscusi and Zeckhauser, 1984). A spectator’s enjoyment is proved to be a monotonic function of the suspense about the outcome of the game (Berlyne, 1960; Zillmann, 1991). It is also proved that people should experience intensified enjoyment with great suspense in the result (Zillmann, 1983, 1996), while such enjoyment-intensifying excitation is, of course, nonexistent in lopsided (i.e., one side dominating the other side) games (Gan, et al, 1997). Lopsided games are viewed as providing resolution too early, and thus rendering unnecessary the wait for resolution at game’s end. Close games, on the other hand, have to be attended to in full because only their actual end provides resolution. The above findings have obvious implications in voting. Voter turnouts are much higher in close elections than in lopsided elections. We make use of this theory in the modeling.

We also make use of the disposition theory of sports spectatorship (Zillmann, Bryant, & Sapolsky, 1989; Zillmann & Paulus, 1993), which leaves no doubt about the fact that enjoyment of an athletic contest, especially of its outcome, hinges on favorable dispositions toward the winning team. Meanwhile, Laverie and Arnett (2000) demonstrate that attachment and involvement influence fan attendance in identity related activities, such as a sporting event. Research suggests that fans may become loyal to a particular team because of the enjoyment associated with spectating (Zillman, Bryant, & Sapolsky, 1989) or because of the sense of identity felt by associating themselves with a particular group (Wann & Branscombe, 1990). The latter authors demonstrate that higher fan identification can lead to an increase in the likelihood of basking-in reflected-glory (BIRGing) and a decrease in the likelihood to cutting-off-reflected failure (CORF). They conclude
that die-hard fans believe that being a "fan" is an important part of who they are, and, as a result, they persevere, even when their team is unsuccessful.\footnote{The identity theory has been subjected to empirical examination in leisure, sociology, and consumer research (cf., Stryker and Serpe, 1982; Serpe, 1987; Hoelter, 1983; Kleine et al., 1993; Laverie, 1998; Shamir, 1992). Those people with high identity salience are more likely to participate in identity related activities (Kleine et al., 1993; Laverie, 1998). Therefore, it is posited that those who have high identity salience for being a fan of a sports team will attend games more often than those low in identity salience.}

In this paper, we make use of the above findings and add suspense and excitement to the rational choice model to replace the $D$ term. We first add suspense. Suspense is modeled as the amount of surprise in the election. Let $p$ be the probability of one’s preferred candidate winning. Then if that candidate does win, $(1 - p)$ is defined as the amount of surprise in terms of probability. In this case, a voter receives a positive surprise of $S^+$ in terms of utility, and thus the amount of surprise is $(1 - p)S^+$. Note that this positive surprise happens with probability $p$, and thus the expected positive surprise is $p(1 - p)S^+$. Similarly, if one’s preferred candidate loses, there is a negative surprise $pS^-$, and it happens with probability $(1 - p)$. So the expected negative surprise is $(1 - p)pS^-$. So the total expected surprise for the voter is $p(1 - p)(S^+ + S^-)$, where $S^+ > 0$, $S^- < 0$ and $S^+ + S^- > 0$.\footnote{In this paper, we assume that every voter has the same $S^+$ and $S^-$. In reality, each voter could have different $S^+$ and $S^-$, some of whom with $S^+ + S^- > 0$ and some of whom with $S^+ + S^- < 0$. In this case, those with $S^+ + S^- < 0$ would be less likely to vote. The analysis in this paper applies to those with $S^+ + S^- > 0$ only.}

We now add excitement to the model. Excitement is modeled as a group effect. With probability $p$, one’s preferred candidate wins. In this case, the voter obtains a utility of $W$. Similarly, with probability $(1 - p)$, the preferred candidate loses. In this case, the voter obtains a utility of $L$, where $L < W$.

One major conclusion of a model with the above features is that the voting rate in a large election will be significantly different from zero, regardless of the probabilities of winning for one’s preferred candidate. Another major conclusion is that the voting rate is higher for the candidate with more supporters, as he is more likely to win and provide
higher utility gains to his supporters. This is in contrast to most of the papers in the literature which predict the opposite. Here, the voting rate for a candidate is defined as the percentage of his supporters incurring the cost to go to the poll stations.

We use the 2004 US Presidential Election data to test the theory empirically. The first result can be tested easily. Voter’s voting rate is obviously much higher than zero. For the second result, the data shows that in over 72.5% of the cases, the candidate with "more supporters" in that state has a higher voting rate. We approximate the number of a particular candidate’s supporters by using the poll numbers right before the election. This approximation is quite preliminary, but it should provide a statistics for the amount of potential voters who support a particular candidate. (These voters may or may not go to vote depending on their voting costs.)

We run a regression to test the relationship between the voting rate difference and the indicator function of the supporters difference. The result indicates that these two variables are positively related and the coefficient is statistically significant at the 90% confidence interval. This provides another evidence to the conclusion that the voting rate is higher for the candidate with more supporters. (We are currently using the 2008 US Presidential Election data to re-do the tests.)

The rest of this paper is organized as follows. In Section 2, we review the literature on the classical rational choice model and some of its extensions. In Section 3, we analyze a model with two candidates. In Section 4, we extend the model to $k$ candidates. In Section 5, we use the 2004 US Presidential Election data to test our theory. Section 6 contains the conclusion.

2 Literature Review

This literature review is divided into five subsections. The first subsection is on the classical rational choice model and the rest are on the extensions to the classical model.
2.1 The Classical Rational Choice Model

The earliest rational choice theory is believed to be proposed by Downs (1957) and Tullock (1967). They analyze a model of an election with two candidates in which a voter decides to vote or abstain. The voter votes if the expected net benefit $R$ from voting is positive, where

$$R = PB - C.$$ 

The parameter $P$ is the probability that a voter’s vote is decisive, that is, his vote is going to either make a tie or break a tie. The parameter $B$ is the benefit difference in utilities that he receives from the success of his preferred candidate over his less preferred one. The parameter $C$ is the cost of voting.

The problem is that this parameter $P$ is very small, that is, close to zero, in large elections. To see this, let the total number of voters $N$ be an odd number and $p$ be the probability that a voter votes for candidate 1. Then, $P$ is given by

$$P = \left( \frac{N-1}{N-2} \right) p^{\frac{N-1}{2}} (1 - p)^{\frac{N-1}{2}}. \quad (1)$$

This binomial expression has been approximated in various ways (see Beck, 1975; Margolis, 1977; Linehan and Schrodt, 1979; Owen and Grofman, 1984). For instance, the approximation by Owen and Grofman (1984) is given by

$$P = \frac{2e^{-2(N-1)(p-\frac{1}{2})^2}}{\sqrt{2\pi (N-1)}}. \quad (2)$$

It demonstrates that $P$ is inversely related to $N$ (the total number of electorates) and $p - \frac{1}{2}$ (candidate competitiveness). To be more precise, $P$ is relatively flat in the neighborhood of $p = \frac{1}{2}$, but it slopes away quite sharply thereafter. For example, if $N = 100,000,000$ and $p = \frac{1}{2}$, we have $P \approx 0.00008$. With the same $p$, the larger the $N$ is, the smaller the $P$ is. On the other hand, if $p = 0.6$, then even if $N$ is as low as 1000, $P$ is in the order of $10^{-11}$. In this case, even if $C = 1$, $B$ has to be in the order of $10^{11}$ for the voter to vote.

Essentially, $P$ is infinitesimal if $N$ is large, whereas the probability of being hit by a car on the way to the polls could be higher than that. This illustrates the paradox of
voting – if people vote with the purpose of influencing the outcome of an election, then even a small cost of voting should dissuade them from voting. However, the empirical results show that the turnout rate is usually as high as 70% in large elections such as the presidential elections (Blais, 2000). In view of this, the earliest rational choice model is not satisfactory in explaining the high turnout. What this model can explain is only the marginal increase of the turnout rate rather than the total.

2.2 The Cost Argument

Many theorists have suggested that the cost of voting is nil (Olson, 1971, p. 164; Niemi, 1976; Smith, 1975; Hinich, 1981; Palfrey and Rosenthal 1985; Aldrich, 1993). The cost of voting is often considered to include the opportunity cost, namely, the time spent on collecting the information about the election, going to the polls, etc., and other negative externalities, such as bad weather or a schedule conflict. If the cost is indeed negligible, then although the term $P$ is minuscule, $PB$ is still greater than $C$. In this sense, the rational choice theory is useful in explaining the high turnout phenomenon.

The question is: Is $C$ really negligible? We have seen that when voting cost goes up, turnout goes down. Therefore voting cost does matter. Downs (1957) has also pointed out that the cost of acquiring the information about candidates is costly. Some other scholars also support the argument that voting is costly (Borgers, 2004). One of Downs’ solutions to the information problem is ideology (see Hinich and Munger 1994), and using British Election Study data, Larcinese (2000) has demonstrated that ideological motivations affect information acquisition and both influence turnout (see also Lassen 2005). But again these results show that costs have marginal effects (the less ideologically motivated the higher the cost of information acquisition) rather than show it is rational to vote.
2.3 The Benefit Argument

As is shown in Subsection 2.1, if parameter $B$ is massive, the rational choice model can be justified. It is often observed that the more important the election is, the more voters the election attracts. This is consistent with the rational choice model, since the importance of the election is positively correlated to the benefits at stake. The only concern is that whether $B$ is really this large in actual elections.

The ethical voter hypothesis might also give us a solution. Hudson and Jones (1993) examine the concept of ethical voter and present an estimate of altruism in their model. Altruism has been modelled as an externality (e.g. see Becker (1974)). An altruist is assumed to derive utility not only from his personal consumption of goods and services, but also from the welfare of other individuals. In this way, it can make $B$ higher. Whether we can get altruism to make $B$ high enough to cover even low costs of voting still partially depends upon the value of $P$.

2.4 The Probability of Winning Argument

One approach to fix the classical rational choice model is to change the calculation of $P$, or to get rid of it completely. Kahneman (et al.1974) suggests that people do not understand the true distribution of $P$ and so overestimate their decisiveness. Blais (2000) provides evidence that individuals massively overestimate the probability of their decisiveness. But even with these overestimations, the probability of a voter being decisive is still small.

Ferejohn and Fiorina (1974) attempt to explain the high turnout rate by proposing the minimax-regret hypothesis, which does not require people to supply any (objective or subjective) estimates of $P$ and voters are risk-aversers who try to minimize the maximum of regret they can suffer. Under this hypothesis, people do not calculate the actual expected payoff for each strategy, but the regret instead. There are two relevant states of the world to be considered: $S_I$, a voter’s vote is not decisive (the outcome of the
election is independent of whether one votes); $S_D$, a voter’s vote is decisive (by voting the individual brings one’s preferred candidate to win either by breaking a tie and making a runoff). In each state, one can choose to vote or to abstain. If the election outcome is independent of one’s vote yet one vote, one has a regret of $C$ which is the cost that one incurred by voting. If in the same state $S_I$, one abstains, then one has no regret. Whereas, in the state one’s vote is decisive, $S_D$, if one votes, one has no regret. However, if one abstains, one’s regret is given by $B - C$, where $B$ is the net benefit one can gain by producing the victory of one’s preferred candidate. Assume $B$ is at least double the costs of voting, $C$. The payoff is shown in the table below (Mueller, 2003).

<table>
<thead>
<tr>
<th>States</th>
<th>$S_I$</th>
<th>$S_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vote</td>
<td>$C$</td>
<td>0</td>
</tr>
<tr>
<td>Abstain</td>
<td>0</td>
<td>$B - C$</td>
</tr>
</tbody>
</table>

It is clearly shown in the table that if one votes, the maximum regret is $C$, and if one abstains, the maximum regret is $B - C$. Given $B \geq 2C$, we must have

$$\min\{C, B - C\} = C,$$

which means that one’s strategy should be to vote.

However, this approach is also problematic. The minimax regret strategy leads to bizarre behavior. Suppose there are multiple candidates running in an election. According to the minimax regret criterion, one should envisage the worst scenario, that the candidate one likes the least wins, without taking into account her probability of winning and vote if that regret outweighs the cost of voting. Thus the presence of an extremist candidate should substantially increase the turnout, whatever the degree of support she enjoys (Mueller, 2003).
2.5 The Satisfaction from Voting Argument

Riker and Ordeshook (1968) extend the basic model by introducing the $D$ term, the benefit of satisfaction from voting, which alone could outweigh the cost of voting. And the calculation of expected benefit from voting becomes

$$R = PB + D - C.$$  

They claim that the elements in $D$ are the following satisfactions: the satisfaction from compliance with the ethic of voting, which is positive if the citizen votes, negative if he does not; the satisfaction from affirming allegiance to the political system; the satisfaction from affirming a partisan preference (voting gives the citizen the chance to stand up and be counted for the candidate he supports); the satisfaction of deciding, going to the polls; the satisfaction of affirming one’s efficacy in the political system (the theory of democracy asserts that individuals and voting are meaningful and for most people the only chance to fulfill this role is the voting booth).

There have been many arguments over the meanings of $D$. Barry (1976) postulates that it is duty that gets people to the polls. Blais (2000) also supports that it is duty to vote because there have been evidence showing that most voters are regular voters who vote nearly every year. Obviously, the $D$ term is most helpful to explain the high turnout, yet there have been many critics, such as Mueller’s (2003) criticism that without a theory explaining the origin, strength, and extent of an individual’s sense of civic duty, merely postulating a sense of civic duty ‘saves’ rational egoism by destroying its predictive content.

Our paper intends to give a new understanding of $D$ and restore the predictive power of the rational choice model.
3 A Model of Two Candidates

Suppose that there are 2 candidates, candidate 1 and candidate 2, in an election. There are two groups of voters, group 1 and group 2. Voters in group 1 all prefer candidate 1, and voters in group 2 all prefer candidate 2. Other than the cost of voting, which will be explained in more detail later, all voters in group 1 are the same. Likewise, all voters in group 2 are the same except their cost of voting.

Consider a representative voter in group 1. If he does not vote, the probability of candidate 1 winning is \( p \), and of candidate 2 winning is \( 1 - p \). His utility is

\[
U_1 = p\pi_1 + (1 - p)\pi_2,
\]

where \( \pi_1 \) is the utility he obtains if candidate 1 wins, and \( \pi_2 \) is his utility if candidate 2 wins. Since group 1 voters prefer candidate 1, \( \pi_1 > \pi_2 \). On the other hand, if he votes, the probability of candidate 1 winning increases to \( \hat{p} \).

Furthermore, we assume there are two kinds of effects taking place. One effect is referred to as the surprise effect, which the voter derives from participating in the election with suspense in the result. The suspense in essence produces the surprise effect, which has a positive component \( S^+ > 0 \) if his preferred candidate (candidate 1) wins, and a negative component \( S^- < 0 \) if candidate 1 loses. We assume that \( |S^+| > |S^-| \), as is evidenced in the psychology and sociology literature. The other effect is the group effect, which captures the utility he gets from being part of the group who support the same candidate as he does. By voting rather than silent support, he can enjoy higher utility cheering together with the group if candidate 1 wins, since he associates himself with and contributes to the triumph of candidate 1. This utility is denoted by \( W \). On the other hand, he can claim comfort from the group if candidate 1 loses. This utility is denoted by \( L \). We assume that \( W > L \), because people derive intensified enjoyment from winning.

As in the classical rational choice model, we assume that a voter incurs the cost of voting \( C_1 \). This \( C_1 \) is assumed to be uniformly distributed on \([0, 1]\). Assume that each voter's cost is distributed independently. The utility from voting for a representative
A voter in group 1 is then given by:

\[ \hat{U}_1 = \hat{p} \left( \pi_1 + (1 - \hat{p}) S^+ + W \right) + (1 - \hat{p}) \left( \pi_2 + \hat{p} S^- + L \right) - C_1. \]

Rearranging it, we have

\[ \hat{U}_1 = \hat{p} \pi_1 + (1 - \hat{p}) \pi_2 + \hat{p} (1 - \hat{p}) \left( S^+ + S^- \right) + \hat{p} W + (1 - \hat{p}) L - C_1. \]

To simplify this formula, let \( S = S^+ + S^- > 0 \). Then

\[ \hat{U}_1 = \hat{p} \pi_1 + (1 - \hat{p}) \pi_2 + \hat{p} (1 - \hat{p}) S + \hat{p} W + (1 - \hat{p}) L - C_1. \]

Note that a voter votes if \( \hat{U}_1 \geq U_1 \), and does not vote if \( \hat{U}_1 < U_1 \). Since

\[ \hat{U}_1 - U_1 = \left[ \hat{p} \pi_1 + (1 - \hat{p}) \pi_2 + \hat{p} (1 - \hat{p}) S + \hat{p} W + (1 - \hat{p}) L - C_1 \right] - \left[ p \pi_1 + (1 - p) \pi_2 \right] \]
\[ = (\hat{p} - p) (\pi_1 - \pi_2) + \hat{p} (1 - \hat{p}) S + \hat{p} W + (1 - \hat{p}) L - C_1. \]

(4)

The first term on the right hand side of (4) is identical to the term \( PB \) in the literature of rational voter theory, i.e \( P = \hat{p} - p \) is the probability of his vote being decisive, and \( B = \pi_1 - \pi_2 \) is the difference in utilities when different candidates win. This term is negligible if the number of voters is large, as we discussed in the literature review section. In the rest of the analysis, we will omit it to simply the calculation. Therefore, the critical value of \( C \) for the indifferent voter is given by

\[ \hat{C}_1 = \hat{p} (1 - \hat{p}) S + \hat{p} W + (1 - \hat{p}) L. \]

Similarly, we can consider a representative voter in group 2. Let \( S^- \), \( S^+ \), \( S \), \( W \), and \( L \) represent the corresponding values for a voter in group 2. Let \( \tilde{p} \) denote the probability of candidate 1 winning if he does not vote. If he votes (for candidate 2), the probability of candidate 1 winning decreases from \( \hat{p} \) to \( \tilde{p} \). Let \( \pi'_1 \) and \( \pi'_2 \) be his benefit from candidate 1 winning and candidate 2 winning respectively, where, of course, \( \pi'_1 < \pi'_2 \) for group 2 voters. Denote \( C_2 \) as the cost of voting for a voter in group 2. This \( C_2 \) is assumed
to be uniformly distributed on $[0, 1]$ and each voter’s cost is distributed independently. Therefore,

$$
\hat{U}_2 = \hat{p} \left( \pi'_1 + (1 - \hat{p})S^- + L \right) + (1 - \hat{p}) \left( \pi'_2 + \hat{p}S^+ + W \right) - C_2
$$

$$
= \hat{p}\pi'_1 + (1 - \hat{p})\pi'_2 + \hat{p}(1 - \hat{p})S + (1 - \hat{p})W + \hat{p}L - C_2.
$$

So

$$
\hat{U}_2 - U_2 = (\hat{p} - \bar{p}) \left( \pi'_2 - \pi'_1 \right) + \hat{p}(1 - \hat{p})S + (1 - \hat{p})W + \hat{p}L - C_2.
$$

Again, omit the first term $(\hat{p} - \bar{p}) (\pi'_2 - \pi'_1)$, as it is small relative to the rest of the terms. The critical value of $\hat{C}_2$ for the indifference voter is given by

$$
\hat{C}_2 = \hat{p}(1 - \hat{p})S + (1 - \hat{p})W + \hat{p}L.
$$

Denote the size of group $i$ by $N_i$ and the voting rate of the voters in group $i$ by $r_i$, $i = 1, 2$. This $r_i$ is the probability of a voter in group $i$ going to vote. Note that $C_i$ is distributed according to uniform distribution on $[0, 1]$. We have $r_1 = \hat{C}_1$, i.e., $r_1 = \hat{p}(1 - \hat{p})S + \hat{p}W + (1 - \hat{p})L$, and $r_2 = \hat{C}_2$, i.e., $r_2 = \hat{p}(1 - \hat{p})S + (1 - \hat{p})W + \hat{p}L$.

To simplify notation, let $q = \hat{p}$. Therefore,

$$
r_1 = q (1 - q) S + qW + (1 - q) L,
$$

and

$$
r_2 = q (1 - q) S + (1 - q)W + qL.
$$

**Proposition 1** Given $S \geq 0$, $W > 0$ and $L > 0$, the voting rate will be significantly different from zero regardless of the value of $q$.

**Proof.** Since $r_1 \geq \min \{L, W\}$, and $r_2 \geq \min \{L, W\}$, it is obvious that both $r_1$ and $r_2$ will be greater than zero. $\blacksquare$

The intuition for this proposition is as follows. Even if $q = 0$ or $q = 1$, that is, when the surprise effect is zero, the voting rate is still strictly greater than zero. If $q \neq 0$ and
$q \neq 1$, i.e., the surprise effect is not equal to zero, the voting rate could be higher (it depends on the value of $W, L$ and $S$). And the surprise effect is the greatest if $q = \frac{1}{2}$, when the election is the closest.

Denote the actual number of voters voting for candidate 1 and candidate 2 by random variables $n_1$ and $n_2$ respectively, which can be approximated by normal distributions. Let random variable $X_i$ be the indicator function for voter $i$ in group 1 voting for candidate 1; if he votes, the value is 1; if not, the value is 0. Therefore,

$$X_i = \begin{cases} 0, & \text{with probability } 1 - r_1, \\ 1, & \text{with probability } r_1. \end{cases}$$

Then

$$E(X_i) = 0 \cdot (1 - r_1) + 1 \cdot r_1 = r_1$$

and

$$Var(X_i) = E(X_i^2) - [E(X_i)]^2 = 1^2 \cdot r_1 - r_1^2 = r_1 - r_1^2$$

Similarly, define random variable $Y_i$ as the indicator function for voter $i$ in group 2 voting for candidate 2. If he votes, the value is 1; if not, the value is 0. Therefore,

$$Y_i = \begin{cases} 0, & \text{with probability } 1 - r_2, \\ 1, & \text{with probability } r_2. \end{cases}$$

Similarly, $E(Y_i) = r_2$ and $Var(Y_i) = r_2 - r_2^2$.

Note that

$$n_1 = \sum_{i=1}^{N_1} X_i,$$

and

$$n_2 = \sum_{i=1}^{N_2} Y_i.$$
Therefore,
\[ E(n_1) = E\left( \sum_{i=1}^{N_1} X_i \right) = N_1 E(X_i) = N_1 r_1, \]
and
\[ \text{Var}(n_1) = \text{Var}\left( \sum_{i=1}^{N_1} X_i \right) = N_1 \text{Var}(X_i) = N_1 \left( r_1 - r_1^2 \right). \]

Similarly,
\[ E(n_2) = N_2 r_2, \]
and
\[ \text{Var}(n_2) = N_2 \left( r_2 - r_2^2 \right). \]

Define random variable \( D \equiv n_2 - n_1 \). We have
\[ E(D) = E(n_2) - E(n_1) = N_2 r_2 - N_1 r_1, \]
and
\[ \text{Var}(D) = \text{Var}(n_2) + \text{Var}(n_1) = N_1 \left( r_1 - r_1^2 \right) + N_2 \left( r_2 - r_2^2 \right). \]

Let \( \mu = N_1 r_1 - N_2 r_2 \) and \( \sigma^2 = N_1 \left( r_1 - r_1^2 \right) + N_2 \left( r_2 - r_2^2 \right) \). Then when \( N_1 \) and \( N_2 \) are large, the random variable \( Z = \frac{D - \mu}{\sigma} \) approximately follows standard normal distribution, i.e. \( Z \sim N(0, 1) \).

Note that \( q \) is the probability of candidate 1 winning. So,
\[ q = \Pr\{n_1 > n_2\} = \Pr\{D < 0\} = \Pr\{Z < -\frac{\mu}{\sigma}\} = \Phi\left(-\frac{\mu}{\sigma}\right), \]
i.e.,
\[ q = \Phi\left(\frac{N_1 r_1 - N_2 r_2}{\sqrt{N_1 \left( r_1 - r_1^2 \right) + N_2 \left( r_2 - r_2^2 \right)}}\right). \tag{5} \]

In the following proposition, we show that the larger group has a higher probability of voting (i.e., higher turnout rate).
Proposition 2 If $N_1 > N_2$, then $r_1 > r_2$. Therefore, $E(n_1) > E(n_2)$.

Proof. Suppose first that $L = W$, then $r_1 = r_2$. In this case, because $N_1 > N_2$, we have
\[ r_1 N_1 - r_2 N_2 > 0. \] (6)
From (5), we know that $q > \frac{1}{2}$.

Now consider the general case of $L \leq W$. Since
\[ r_1 = q (1 - q) S + qW + (1 - q) L \] (7)
and
\[ r_2 = q (1 - q) S + (1 - q) W + qL, \]
we have
\[ r_1 - r_2 = (2q - 1) (W - L). \] (8)
Noting $q > \frac{1}{2}$ when $L = W$, we have
\[ \left. \frac{\partial (r_1 - r_2)}{\partial L} \right|_{L=W} = \left. \left[ 2 \frac{\partial q}{\partial L} (W - L) - (2q - 1) \right] \right|_{L=W} = 1 - 2q < 0. \]

Suppose at any point $L^* < W$ the curve $r_1 - r_2$ intersects the $L$ axis again, i.e. $r_1 - r_2 = 0$. Thus, from (8), we must have
\[ q = \frac{1}{2}. \]
However, if $r_1 - r_2 = 0$, inequality (6) holds, and from (5), it implies $q > \frac{1}{2}$. This is a contradiction. Hence at any point $L^* < W$, the curve cannot intersect the $L$ axis again (See Figure 1). Therefore, given $N_1 > N_2$, $r_1 - r_2 > 0$, for all $L < W$.

Because $E(n_1) = N_1 r_1$ and $E(n_2) = N_2 r_2$, it also implies that $E(n_1) > E(n_2)$. ■

The intuition for this proposition is as follows. A candidate with a larger group of supporters has a higher probability of winning. In addition, because $W > L$, winning provides more incentive for supporters to vote. As the supporters from the winning side derive more utility from voting, more of them (in % term) will vote. Of course, this implies that they are even more likely to win.
Suppose now that there are $k$ candidates in an election. There are $k$ groups of voters. Voters in group $i$ all prefer candidate $i$, $i = 1, 2, \ldots, k$. Other than the cost of voting, all voters in group $i$ are the same.

Consider a voter in group $i$. If he does not vote, the probability of candidate $j$ winning is $p_j$, and $\sum_{j=1}^{k} p_j = 1$. This representative voter’s utility is

$$U_i = \sum_{j=1}^{k} p_j \pi_j,$$

where $\pi_j$ is the benefit he obtains if $j$ wins, $j = 1, 2, \ldots, k$. Assume $\pi_i > \max_{j \neq i} \{\pi_j\}$ since this voter prefers candidate $i$. Again, assume that $C_i$, the cost of voting for a voter in group $i$, is distributed independently and uniformly on $U[0, 1]$. 

4 A Model of $k$ Candidates

Figure 1: Proof of Proposition 2
If this voter votes, the probability of $j$ winning changes to $\hat{p}_j$, and $\sum_{j=1}^{k} \hat{p}_j = 1$. Other assumptions are the same as in the two candidate model. Then, the utility if he votes is given by

$$\hat{U}_i = \sum_{j=1}^{k} \hat{p}_j \pi_j + \hat{p}_i \left[ (1 - \hat{p}_i) S^+ + W \right] + (1 - \hat{p}_i) \left( \hat{p}_i S^- + L \right) - C_i.$$ 

Rearranging, we have

$$\hat{U}_i = \sum_{j=1}^{k} \hat{p}_j \pi_j + \hat{p}_i (1 - \hat{p}_i) S + \hat{p}_i W + (1 - \hat{p}_i) L - C_i,$$

where $S = S^+ + S^-$. Therefore,

$$\hat{U}_i - U_i = \sum_{j=1}^{k} \left( \hat{p}_j - p_j \right) \pi_j + \hat{p}_i (1 - \hat{p}_i) S + \hat{p}_i W + (1 - \hat{p}_i) L - C_i.$$ 

The first term on the RHS will be omitted since it is usually extremely small in large elections. Again, let $q_i = \hat{p}_i$. The critical value of $C_i$ (and thus also the voting rate for group $i$) is given by

$$\hat{C}_i = r_i = q_i (1 - q_i) S + q_i W + (1 - q_i) L.$$ 

According to the analysis in the previous section, we can approximate the distribution of $n_i$, the actual number of electorates voting for $i$ by a normal distribution

$$n_i \sim N \left( N_i r_i, N_i \left( r_i - r_i^2 \right) \right).$$

We obtain the following proposition, which is similar to Proposition 2.

**Proposition 3** If $N_1 > N_2 > \ldots > N_k$, then $r_1 > r_2 > \ldots > r_k$. Therefore, $E(n_1) > E(n_2) > \ldots > E(n_k)$. 

17
**Proof.** Only need to show that if \( N_i > N_j \), then \( r_i > r_j, \forall i > j \).

Since

\[
r_i - r_j = [q_i (1 - q_i) S + q_i W + (1 - q_i) L] - [q_j (1 - q_j) S + q_j W + (1 - q_j) L]
\]

\[
= (q_i - q_j) [(1 - q_i - q_j) S + (W - L)].
\]

(9)

Define \( P_{st} \equiv \Pr(n_s > n_t) \). Then

\[
q_i = \Pr(n_i > n_m, \forall m \neq i)
\]

\[
= \prod_{m \neq i} \Pr(n_i > n_m)
\]

\[
= P_{ij} \cdot \prod_{m \neq i,j} P_{im}.
\]

Similarly,

\[
q_j = \Pr(n_j > n_m, \forall m \neq j)
\]

\[
= \prod_{m \neq j} \Pr(n_j > n_m)
\]

\[
= P_{ji} \cdot \prod_{m \neq j,i} P_{jm}
\]

\[
= (1 - P_{ij}) \cdot \prod_{m \neq j,i} P_{jm}.
\]

Suppose first that \( S = 0 \) and \( W = L \). In this case, from (9) we must have

\[
r_i - r_j = 0
\]

(10)

Because \( N_i > N_j \), we have

\[
r_i N_i - r_j N_j > 0.
\]

(11)

and

\[
N_i (r_i - r_i^2) > N_j (r_j - r_j^2)
\]

(12)
Therefore, from (11), we have
\[ P_{ij} = \Phi \left( \frac{N_i r_i - N_j r_j}{\sqrt{N_i (r_i - r_i^2) + N_j (r_j - r_j^2)}} \right) > \Phi(0) = \frac{1}{2}. \] (13)

and from (12), we have
\[ \frac{N_m r_m}{\sqrt{N_i (r_i - r_i^2) + N_m (r_m - r_m^2)}} < \frac{N_m r_m}{\sqrt{N_j (r_j - r_j^2) + N_m (r_m - r_m^2)}} \] (14)

We can show that
\[ \frac{N_i r_i}{\sqrt{N_i (r_i - r_i^2) + N_m (r_m - r_m^2)}} > \frac{N_j r_j}{\sqrt{N_j (r_j - r_j^2) + N_m (r_m - r_m^2)}} \] (15)

This is because
\[ (15) \iff \frac{N_i}{\sqrt{N_i (r_i - r_i^2) + N_m (r_m - r_m^2)}} > \frac{N_j}{\sqrt{N_j (r_j - r_j^2) + N_m (r_m - r_m^2)}} \]
\[ \iff \frac{N_i^2}{N_j^2} \frac{N_i}{\sqrt{N_i (r_i - r_i^2) + N_m (r_m - r_m^2)}} > \frac{N_j}{\sqrt{N_j (r_j - r_j^2) + N_m (r_m - r_m^2)}} \]
\[ \iff N_i [N_j (r_j - r_j^2) + N_m (r_m - r_m^2)] - N_j [N_i (r_i - r_i^2) + N_m (r_m - r_m^2)] > 0 \]
\[ \iff N_i N_j (r_i - r_i^2) (N_i - N_j) + N_m (r_m - r_m^2) (N_i^2 - N_j^2) > 0. \]

Then, from (14) and (15), we have
\[ \frac{N_i r_i - N_m r_m}{\sqrt{N_i (r_i - r_i^2) + N_m (r_m - r_m^2)}} > \frac{N_j r_j - N_m r_m}{\sqrt{N_j (r_j - r_j^2) + N_m (r_m - r_m^2)}} \]

Therefore, \( \forall m \neq i, j, \)
\[ P_{im} = \Phi \left( \frac{N_i r_i - N_m r_m}{\sqrt{N_i (r_i - r_i^2) + N_m (r_m - r_m^2)}} \right) > \Phi \left( \frac{N_j r_j - N_m r_m}{\sqrt{N_j (r_j - r_j^2) + N_m (r_m - r_m^2)}} \right) = P_{jm}. \] (16)

Hence, from (13) and (16), we have
\[ q_i = P_{ij} \cdot \prod_{m \neq i,j} P_{im} > (1 - P_{ij}) \cdot \prod_{m \neq j,i} P_{jm} = q_j. \] (17)
Now allow $S$ to increase from 0, and $W$ to increase from $L$, continuously. We shall show that $q_i - q_j = 0$ will never happen.

Suppose not, i.e., $q_i - q_j = 0$ at some point. Then, from (9), we have (10). However, if (10) holds, as shown above, $q_i > q_j$. This is a contradiction. Therefore, $q_i - q_j$ stays positive all the time. Equation (9) implies $r_i - r_j$ is also positive all the time, i.e., $r_i > r_j$.

From this, we conclude that $N_1 > N_2 > ... > N_k$ implies $r_1 > r_2 > ... > r_k$. Because $E(n_i) = N_i r_i$, it also implies $E(n_1) > E(n_2) > ... > E(n_k)$. $\blacksquare$

The intuition behind this proposition is similar to the intuition for Proposition 2.

5 Empirical Study

Using the 2004 United States Presidential Election data (see Table 1), we want to test our prediction on the relationship between the total number of eligible voters and the voting (turnout) rate of different groups.
<table>
<thead>
<tr>
<th>State</th>
<th>Total Eligible (in 000s)</th>
<th>Bush Poll %</th>
<th>Kerry Poll %</th>
<th>Bush Actual Votes</th>
<th>Kerry Actual Votes</th>
<th>Bush Vote %</th>
<th>Kerry Vote %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>3257</td>
<td>57%</td>
<td>39%</td>
<td>1,176,394</td>
<td>693,933</td>
<td>62.50%</td>
<td>36.80%</td>
</tr>
<tr>
<td>Alaska</td>
<td>434</td>
<td>57%</td>
<td>30%</td>
<td>190,889</td>
<td>110,025</td>
<td>61.10%</td>
<td>35.50%</td>
</tr>
<tr>
<td>Arizona</td>
<td>3508</td>
<td>56%</td>
<td>41%</td>
<td>1,104,294</td>
<td>893,524</td>
<td>54.90%</td>
<td>44.40%</td>
</tr>
<tr>
<td>Arkansas</td>
<td>1942</td>
<td>51%</td>
<td>46%</td>
<td>572,898</td>
<td>469,953</td>
<td>54.30%</td>
<td>44.50%</td>
</tr>
<tr>
<td>California</td>
<td>20693</td>
<td>43%</td>
<td>54%</td>
<td>5,509,826</td>
<td>6,745,485</td>
<td>44.40%</td>
<td>54.30%</td>
</tr>
<tr>
<td>Colorado</td>
<td>3109</td>
<td>50%</td>
<td>47%</td>
<td>1,101,255</td>
<td>1,001,732</td>
<td>51.70%</td>
<td>47.00%</td>
</tr>
<tr>
<td>Connecticut</td>
<td>2409</td>
<td>42%</td>
<td>52%</td>
<td>693,826</td>
<td>857,488</td>
<td>43.90%</td>
<td>54.30%</td>
</tr>
<tr>
<td>Delaware</td>
<td>579</td>
<td>41%</td>
<td>50%</td>
<td>171,660</td>
<td>200,152</td>
<td>45.80%</td>
<td>53.30%</td>
</tr>
<tr>
<td>DC</td>
<td>390</td>
<td>11%</td>
<td>78%</td>
<td>21,256</td>
<td>202,970</td>
<td>9.30%</td>
<td>89.20%</td>
</tr>
<tr>
<td>Florida</td>
<td>11469</td>
<td>48%</td>
<td>50%</td>
<td>3,964,522</td>
<td>3,583,544</td>
<td>52.10%</td>
<td>47.10%</td>
</tr>
<tr>
<td>Georgia</td>
<td>5866</td>
<td>55%</td>
<td>43%</td>
<td>1,914,254</td>
<td>1,366,149</td>
<td>58.00%</td>
<td>41.40%</td>
</tr>
<tr>
<td>Hawaii</td>
<td>852</td>
<td>46%</td>
<td>45%</td>
<td>194,191</td>
<td>231,708</td>
<td>45.30%</td>
<td>54.00%</td>
</tr>
<tr>
<td>Idaho</td>
<td>948</td>
<td>59%</td>
<td>30%</td>
<td>409,235</td>
<td>181,098</td>
<td>68.40%</td>
<td>30.30%</td>
</tr>
<tr>
<td>Illinois</td>
<td>8640</td>
<td>42%</td>
<td>54%</td>
<td>2,346,608</td>
<td>2,891,989</td>
<td>44.50%</td>
<td>54.80%</td>
</tr>
<tr>
<td>Indiana</td>
<td>4435</td>
<td>58%</td>
<td>39%</td>
<td>1,479,438</td>
<td>969,011</td>
<td>59.90%</td>
<td>39.30%</td>
</tr>
<tr>
<td>Iowa</td>
<td>2136</td>
<td>47%</td>
<td>50%</td>
<td>751,957</td>
<td>741,898</td>
<td>49.90%</td>
<td>49.20%</td>
</tr>
<tr>
<td>Kansas</td>
<td>1850</td>
<td>60%</td>
<td>37%</td>
<td>736,456</td>
<td>434,993</td>
<td>62.00%</td>
<td>36.60%</td>
</tr>
<tr>
<td>Kentucky</td>
<td>2969</td>
<td>59%</td>
<td>38%</td>
<td>1,069,439</td>
<td>712,733</td>
<td>59.60%</td>
<td>39.70%</td>
</tr>
<tr>
<td>Louisiana</td>
<td>3218</td>
<td>52%</td>
<td>44%</td>
<td>1,102,169</td>
<td>820,299</td>
<td>56.70%</td>
<td>42.20%</td>
</tr>
<tr>
<td>Maine</td>
<td>1007</td>
<td>44%</td>
<td>52%</td>
<td>330,201</td>
<td>396,842</td>
<td>44.60%</td>
<td>53.60%</td>
</tr>
<tr>
<td>Maryland</td>
<td>3678</td>
<td>43%</td>
<td>54%</td>
<td>1,024,703</td>
<td>1,334,493</td>
<td>42.90%</td>
<td>55.90%</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>4497</td>
<td>36%</td>
<td>50%</td>
<td>1,071,109</td>
<td>1,803,800</td>
<td>36.80%</td>
<td>61.90%</td>
</tr>
<tr>
<td>Michigan</td>
<td>7177</td>
<td>47%</td>
<td>50%</td>
<td>2,313,746</td>
<td>2,479,183</td>
<td>47.80%</td>
<td>51.20%</td>
</tr>
<tr>
<td>Minnesota</td>
<td>3645</td>
<td>45%</td>
<td>51%</td>
<td>1,346,695</td>
<td>1,445,014</td>
<td>47.60%</td>
<td>51.10%</td>
</tr>
<tr>
<td>Mississippi</td>
<td>2049</td>
<td>51%</td>
<td>42%</td>
<td>684,981</td>
<td>457,766</td>
<td>59.40%</td>
<td>39.70%</td>
</tr>
<tr>
<td>Missouri</td>
<td>4106</td>
<td>50%</td>
<td>45%</td>
<td>1,455,713</td>
<td>1,259,171</td>
<td>53.30%</td>
<td>46.10%</td>
</tr>
<tr>
<td>Montana</td>
<td>687</td>
<td>57%</td>
<td>36%</td>
<td>266,063</td>
<td>173,710</td>
<td>59.10%</td>
<td>38.60%</td>
</tr>
<tr>
<td>Nebraska</td>
<td>1215</td>
<td>61%</td>
<td>32%</td>
<td>512,814</td>
<td>254,328</td>
<td>65.90%</td>
<td>32.70%</td>
</tr>
<tr>
<td>Nevada</td>
<td>1477</td>
<td>50%</td>
<td>45%</td>
<td>418,690</td>
<td>397,190</td>
<td>50.50%</td>
<td>47.00%</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>948</td>
<td>49%</td>
<td>48%</td>
<td>331,237</td>
<td>340,511</td>
<td>48.90%</td>
<td>50.20%</td>
</tr>
<tr>
<td>State</td>
<td>% for Bush</td>
<td>% for Kerry</td>
<td>Bush Votes (in thousands)</td>
<td>Kerry Votes (in thousands)</td>
<td>Bush %</td>
<td>Kerry %</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
<td>------------</td>
<td>---------------------------</td>
<td>---------------------------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>New Jersey</td>
<td>42%</td>
<td>54%</td>
<td>1,670,003</td>
<td>1,911,430</td>
<td>46.20%</td>
<td>52.90%</td>
<td></td>
</tr>
<tr>
<td>New Mexico</td>
<td>48%</td>
<td>51%</td>
<td>376,930</td>
<td>370,942</td>
<td>49.80%</td>
<td>49.00%</td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>37%</td>
<td>54%</td>
<td>2,962,567</td>
<td>4,314,280</td>
<td>40.10%</td>
<td>43.60%</td>
<td></td>
</tr>
<tr>
<td>North Carolina</td>
<td>53%</td>
<td>45%</td>
<td>1,961,166</td>
<td>1,525,849</td>
<td>56.00%</td>
<td>43.60%</td>
<td></td>
</tr>
<tr>
<td>North Dakota</td>
<td>55%</td>
<td>35%</td>
<td>196,651</td>
<td>111,052</td>
<td>62.90%</td>
<td>35.50%</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>49%</td>
<td>47%</td>
<td>2,859,764</td>
<td>2,741,165</td>
<td>50.80%</td>
<td>48.70%</td>
<td></td>
</tr>
<tr>
<td>Oklahoma</td>
<td>64%</td>
<td>34%</td>
<td>959,792</td>
<td>503,966</td>
<td>56.00%</td>
<td>34.40%</td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td>47%</td>
<td>50%</td>
<td>866,831</td>
<td>943,163</td>
<td>47.20%</td>
<td>51.30%</td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>48%</td>
<td>49%</td>
<td>2,793,847</td>
<td>2,938,095</td>
<td>48.40%</td>
<td>50.90%</td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>41%</td>
<td>54%</td>
<td>169,046</td>
<td>259,760</td>
<td>38.70%</td>
<td>59.40%</td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td>54%</td>
<td>43%</td>
<td>937,974</td>
<td>661,699</td>
<td>58.00%</td>
<td>40.90%</td>
<td></td>
</tr>
<tr>
<td>South Dakota</td>
<td>52%</td>
<td>42%</td>
<td>232,584</td>
<td>149,244</td>
<td>58.00%</td>
<td>38.40%</td>
<td></td>
</tr>
<tr>
<td>Tennessee</td>
<td>58%</td>
<td>40%</td>
<td>1,384,375</td>
<td>1,036,477</td>
<td>56.80%</td>
<td>43.20%</td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>59%</td>
<td>37%</td>
<td>4,526,917</td>
<td>2,832,704</td>
<td>61.10%</td>
<td>38.20%</td>
<td></td>
</tr>
<tr>
<td>Utah</td>
<td>69%</td>
<td>24%</td>
<td>663,742</td>
<td>241,199</td>
<td>71.50%</td>
<td>26.00%</td>
<td></td>
</tr>
<tr>
<td>Vermont</td>
<td>40%</td>
<td>53%</td>
<td>121,180</td>
<td>184,067</td>
<td>38.80%</td>
<td>58.90%</td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td>51%</td>
<td>47%</td>
<td>1,716,959</td>
<td>1,454,742</td>
<td>53.70%</td>
<td>45.50%</td>
<td></td>
</tr>
<tr>
<td>Washington</td>
<td>47%</td>
<td>51%</td>
<td>1,304,894</td>
<td>1,510,201</td>
<td>45.60%</td>
<td>52.80%</td>
<td></td>
</tr>
<tr>
<td>West Virginia</td>
<td>51%</td>
<td>43%</td>
<td>423,778</td>
<td>326,541</td>
<td>56.10%</td>
<td>43.20%</td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td>44%</td>
<td>51%</td>
<td>1,478,120</td>
<td>1,489,504</td>
<td>49.30%</td>
<td>49.70%</td>
<td></td>
</tr>
<tr>
<td>Wyoming</td>
<td>65%</td>
<td>29%</td>
<td>167,629</td>
<td>70,776</td>
<td>68.90%</td>
<td>29.10%</td>
<td></td>
</tr>
</tbody>
</table>

Total: 197005

62,041,268 59,028,548 50.73% 48.27%

These data are categorized by states in an alphabetical order. And the state names are indicated in the first column. The second column is the total number of eligible voters and the number is in thousands. (See http://www.census.gov/population/socdemo/voting/cps2004/tab04a.xls)

The third and fourth columns are poll results. These poll results are based on the polls done by major pollster companies in the US from May through November of 2004. Considering there are many poll results regarding one state, we choose the one that is done closest to the election date by major pollster companies. (See http://www.electoralvote.com/2004/pastpolls.html)

The fifth and sixth columns are the actual number of voters voting for the two major candidates, Bush and Kerry, respectively. Following the notations in the previous sections, the actual number of voters voting for Bush is denoted by \( n_1 \), and Kerry by \( n_2 \). The last two columns are the percentage of voters voting for each candidate. (See
The numbers of total eligible voters supporting Bush ($N_1$) and Kerry ($N_2$) are not directly obtainable. We use the number of total eligible voters times the percentage as in the poll results in each state to approximate $N_1$ and $N_2$. This makes sense because the poll results indicate the supporting rate for each candidate in each state before the election. These supporters may or may not go to the poll station and vote. The turnout (voting) rate of a candidate’s supporters is calculated as the actual votes the candidate receives ($n_1$) divided by the estimated total number of supporters ($N_1$). Therefore, in Table 2, we use $r_1 = n_1/N_1$ to denote Bush’s voting rate and $r_2 = n_2/N_2$ to denote Kerry’s voting rate.
Table 2: Voting Rates for Different Groups of Voters

<table>
<thead>
<tr>
<th>State</th>
<th>Bush ($N_1$)</th>
<th>Kerry ($N_2$)</th>
<th>$r_1$</th>
<th>$r_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>1,856,490</td>
<td>1,270,230</td>
<td>0.633666</td>
<td>0.546305</td>
</tr>
<tr>
<td>Alaska</td>
<td>247,380</td>
<td>130,200</td>
<td>0.771643</td>
<td>0.852727</td>
</tr>
<tr>
<td>Arizona</td>
<td>1,964,480</td>
<td>1,438,280</td>
<td>0.56213</td>
<td>0.621245</td>
</tr>
<tr>
<td>Arkansas</td>
<td>990,420</td>
<td>893,320</td>
<td>0.578439</td>
<td>0.56075</td>
</tr>
<tr>
<td>California</td>
<td>8,897,990</td>
<td>11,174,220</td>
<td>0.619221</td>
<td>0.603665</td>
</tr>
<tr>
<td>Colorado</td>
<td>1,554,500</td>
<td>1,461,230</td>
<td>0.70843</td>
<td>0.68554</td>
</tr>
<tr>
<td>Connecticut</td>
<td>1,011,780</td>
<td>1,252,680</td>
<td>0.685748</td>
<td>0.684523</td>
</tr>
<tr>
<td>Delaware</td>
<td>237,390</td>
<td>289,500</td>
<td>0.725114</td>
<td>0.691371</td>
</tr>
<tr>
<td>DC</td>
<td>42,900</td>
<td>304,200</td>
<td>0.495478</td>
<td>0.667226</td>
</tr>
<tr>
<td>Florida</td>
<td>5,505,120</td>
<td>5,734,500</td>
<td>0.720152</td>
<td>0.62491</td>
</tr>
<tr>
<td>Georgia</td>
<td>3,226,300</td>
<td>2,522,380</td>
<td>0.593328</td>
<td>0.541611</td>
</tr>
<tr>
<td>Hawaii</td>
<td>391,920</td>
<td>383,400</td>
<td>0.495486</td>
<td>0.604351</td>
</tr>
<tr>
<td>Idaho</td>
<td>3,628,800</td>
<td>4,665,600</td>
<td>0.646662</td>
<td>0.636772</td>
</tr>
<tr>
<td>Illinois</td>
<td>2,572,300</td>
<td>1,729,650</td>
<td>0.575142</td>
<td>0.560235</td>
</tr>
<tr>
<td>Indiana</td>
<td>1,003,920</td>
<td>1,068,800</td>
<td>0.743921</td>
<td>0.694661</td>
</tr>
<tr>
<td>Iowa</td>
<td>1,110,000</td>
<td>684,500</td>
<td>0.663474</td>
<td>0.635649</td>
</tr>
<tr>
<td>Kansas</td>
<td>1,751,710</td>
<td>1,128,220</td>
<td>0.610511</td>
<td>0.631732</td>
</tr>
<tr>
<td>Kentucky</td>
<td>1,673,360</td>
<td>1,415,920</td>
<td>0.658656</td>
<td>0.57934</td>
</tr>
<tr>
<td>Louisiana</td>
<td>443,080</td>
<td>523,640</td>
<td>0.74524</td>
<td>0.75853</td>
</tr>
<tr>
<td>Maine</td>
<td>1,581,540</td>
<td>1,986,120</td>
<td>0.647915</td>
<td>0.67191</td>
</tr>
<tr>
<td>Maryland</td>
<td>1,618,920</td>
<td>2,248,500</td>
<td>0.601619</td>
<td>0.802224</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>3,373,190</td>
<td>3,588,500</td>
<td>0.685922</td>
<td>0.690869</td>
</tr>
<tr>
<td>Michigan</td>
<td>1,640,250</td>
<td>1,858,950</td>
<td>0.82103</td>
<td>0.777328</td>
</tr>
<tr>
<td>Minnesota</td>
<td>1,044,990</td>
<td>860,580</td>
<td>0.65549</td>
<td>0.531927</td>
</tr>
<tr>
<td>Mississippi</td>
<td>2,053,000</td>
<td>1,847,700</td>
<td>0.709066</td>
<td>0.68148</td>
</tr>
<tr>
<td>Missouri</td>
<td>391,590</td>
<td>247,320</td>
<td>0.679443</td>
<td>0.702369</td>
</tr>
<tr>
<td>Montana</td>
<td>741,150</td>
<td>388,800</td>
<td>0.691917</td>
<td>0.654136</td>
</tr>
<tr>
<td>Nebraska</td>
<td>738,500</td>
<td>664,650</td>
<td>0.566947</td>
<td>0.597593</td>
</tr>
<tr>
<td>Nevada</td>
<td>464,520</td>
<td>455,040</td>
<td>0.713074</td>
<td>0.74831</td>
</tr>
</tbody>
</table>
It is shown in the above table that 37 out of 51 states are consistent with our prediction that the candidate with more supporting voters has higher turnout rate in that state. That is, over 72.5% of the states are consistent with the above prediction.

Furthermore, we run a simple regression of

\[ R_i = \alpha I_i + \varepsilon_i \]

for the 51 states, where the dependent variable \( R_i \) is equal to \( r_1 - r_2 \), \( \alpha \) is the parameter to be estimated, and \( \varepsilon_i \) is the error term. Considering there may be errors in the approximation of the variables \( N_1 \) and \( N_2 \), we use an indicator function \( I \) in the regression, where \( I = 1 \) if \( N_1 > N_2 \), and \( I = -1 \) if \( N_1 < N_2 \).

The results of the regression are given by Table 3.

The coefficient \( \alpha \) is positive and statistically significant at the 90% confidence level. This demonstrates that there is a positive relationship between \( R \) and \( I \). This is another evidence supporting our theory.
Table 3: Regression Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.016184</td>
<td>0.009252</td>
<td>1.749355</td>
<td>0.0864</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.028999</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.028999</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>0.066069</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>0.218259</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>66.70858</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6 Conclusion

In this paper, we reexamine the rational choice hypothesis. The classical rational choice model by Downs (1957) and Tullock (1967) propose a "calculus of voting", which in fact, can only explain the marginal increase in the turnout rate rather than the high turnout rate as a whole. Riker and Ordeshook (1968) extend the model by introducing a $D$ term, which is described as satisfaction from voting, or sense of duty. In this way, the model can explain well the high turnout rate. However, it also destroys the predictive contents of the model.

In this paper, we build a two-candidate model. The framework of this model is similar to the rational choice model. It not only restores the predictive power of the rational choice model, but also explains the high turnout rates. Based on relevant psychological and sociological findings, two factors are added to our model: a surprise effect and a group effect. With these two features, the voting rate is significantly different from zero. In addition, we find that for the candidate who has more supporting voters, the turnout rate for him is higher and the expected number of voters is larger as well. This model is then extended to a $k$-candidate case. For this model, we have similar results as in the 2-candidate case.

The empirical study is based on the 2004 US Presidential Election data. The results of the tests support our prediction that voting rate is not zero. And in 37 out of 51
states, the voting rate is higher for the candidate who has more supporters and also he wins that state.

In the modeling, we concentrate on the psychological and sociological features and do not consider the ‘duty’ or other voting satisfaction factor. However, we still think they may play some role in determining the turnout rate. We also omit the $PB$ term, which still contributes to some marginal increase in the turnout rate.

In the empirical study section, we use the data for the 2004 US Presidential Election only. (We are currently updating the analysis using the 2008 US Presidential Election data.) This is a very preliminary empirical investigation. We plan to use election data for more countries and from more years to support our theory.

References


