

Note

**A positive correlation between turnout and plurality does not refute the rational voter model**

AMIHAI GLAZER & BERNARD GROFMAN

School of Social Sciences, University of California, Irvine, CA 92717, U.S.A.

**Abstract.** Many papers have tested the prediction of the rational voter model that, *ceteris paribus*, turnout will be low when potential voters expect the winner's plurality to be large. The appropriate null hypothesis, however, is unclear. We show that statistical models of voting in which each voter's decision of whether to vote does *not* vary with the expected plurality can nonetheless generate data which lead to both positive and negative correlations between turnout and plurality.

The extensive empirical literature on the relation between turnout and political competition assumes that under the rational actor model of voting turnout is higher the smaller the expected plurality of the winning candidate (Downs, 1957; Ferejohn and Fiorina, 1975; Foster, 1984; Gray, 1976; Grofman, 1983; Patterson and Caldeira, 1983; Settle and Abrams, 1976; Tollison *et al.*, 1975). This conclusion appears to follow from the assumption that a person's expected benefit from voting is

$$PU + (B - C), \tag{1}$$

where  $P$  is the probability that a shift of a single vote will change the election outcome,  $U$  is a voter's benefit from having his preferred candidate win,  $C$  is the non-instrumental cost of voting (e.g., time spent at the polls), and  $B$  is the non-instrumental benefit of voting (e.g., "psychic" satisfaction for expressing solidarity with a candidate or position). A rational person would vote if this expected net benefit is positive. A person who expects the election to be close will assign  $P$  a high value, and thus, *ceteris paribus*, is more likely to vote. Thus, for the electorate as a whole, an election expected to be close should induce a high level of turnout. The additional assumption that the actual results of an election correlate with peoples' prior expectations about the results yields the prediction that turnout is higher the closer the election results.

A closer examination of the theory shows, however, that rational behavior

need not imply high turnout in close elections. Consider an election in which most constituents see little difference between the candidates. The value of  $U$  will be small; the probability that any one person prefers one candidate over the other is about  $1/2$ . Statistical arguments show that for any given level of total turnout the probability,  $P$ , that any one vote is decisive is larger the more evenly divided is the electorate in its preferences. Thus, in terms of expression (1) a race between similar candidates makes  $P$  uncommonly large and  $U$  uncommonly small. In contrast, suppose most voters agree that one of the candidates is better than the other; the better one may be viewed as more effective or as more in tune with the voters' views. The election is therefore likely to give the winner (almost certainly the better candidate) a large plurality. For a given level of turnout this means that the probability,  $P$ , that any one vote is decisive is low. To say that voters see one candidate as better than the other is to say that they think the value of  $U$  is large. Combining these two effects implies that the value of  $PU$  used in expression (1) can be either large or small when voters see much difference between the candidates. Thus, close elections (where most constituents are indifferent between the candidates) may have either higher or lower turnout than landslide elections (where most constituents agree that one candidate is far better than the other).

We further study the appropriate null hypothesis by examining the correlations between turnout and plurality generated by statistical models in which no voter considers the probability that his vote will be decisive. These models can be viewed as generating the null hypothesis which the researcher can compare to his empirical results. If the null hypothesis is clear – if, say, turnout and plurality are positively correlated only under rational voting – then empirical data can shed light on the validity of that model. If instead models without rational voting can generate either positive or negative correlation between turnout and plurality, then tests of the rational voting model are problematic.

Related criticisms of statistical studies are made by Cox (1988). He shows that if the level of turnout is used to define both the dependent variable (the percentage margin of the eligible persons in a district who vote), then a spurious correlation arises between these two variables. This statistical problem can be avoided by using the raw (rather than the percentage) margin of victory. The criticisms we make differ from those made by Cox; most also apply if raw instead of percentage values are used.

We shall consider in turn three statistical models which generate correlations between turnout and plurality. Model 1 supposes that a random number of potential voters decide to vote, and that once at the voting booth a random process determines which candidate each person supports. This

model, which is in the spirit of models often used to calculate the probability of a decisive vote, generates a negative correlation between turnout and plurality. The next two models assume that potential voters know their preferences before they go to the polling booth. Model 2 supposes that Democrats and Republicans have the same probability of voting. Since, however, the event of a particular person voting is a random variable, the levels of turnout among these two types of persons are random variables as well. This model can generate a negative correlation between turnout and plurality. The third model resembles the previous one in considering two types of voters. But instead of viewing any person as voting with some specified probability, Model 3 supposes that the distribution of voting costs differs across elections or districts. Since some people are certain to vote while others are not, the fraction of Democrats and Republicans who vote can vary across districts or elections. This model can generate both positive and negative correlations between turnout and plurality.

### **Model 1: Turnout and preferences are random**

A plausible statistical model of voting behavior let any one person vote with some given exogenous probability, so that in any particular election the level of turnout is a random variable. A given person who votes is supposed to cast his vote for the Democratic rather than the Republican candidate with probability  $1/2$ . The expected share of the vote won by the Democratic candidate is therefore  $1/2$ , but because of random effects the fraction can be higher or lower than that. Statistical theory proves that the deviation between the expected and the actual proportions of people who vote for the Democratic candidate is greater the smaller the size of the sample analyzed, that is, the smaller the number of persons who vote. Because high turnout causes a small sampling variance, turnout and the proportionate plurality of the winner will be negatively correlated.

More precisely, let  $N$  be the number of people who decide to vote. Let  $p$  be the probability that a person who does vote votes for the Democratic candidate. Then the probability that the Democratic candidate receives  $D$  votes is

$$\binom{N}{D} p^D (1-p)^{N-D}. \quad (2)$$

The expected proportionate plurality in an election is

$$\sum_{D=0}^N |D/N - 1/2| \binom{N}{D} p^D (1-p)^{N-D}. \quad (3)$$

Consider values of  $N$  equal to 50,000, 60,000 . . . 190,000, 200,000. For each value of  $N$  we use the normal approximation to expression (3) to find the expected plurality. The model states that plurality is a function of turnout. But the data generated can be used to estimate a regression equation with turnout as the dependent variable and plurality as a proportion of total turnout as the explanatory one. We obtain

$$N = 329,385 - 170 \times 10^6 \text{ Plurality},$$

with a  $t$ -statistic on the *Plurality* variable of  $-12.9$ . This is a strong effect indeed: an increase in the share of the vote from 0.501 to 0.5011 is associated with a decrease of 17,000 in the expected number of voters. Empirical works find much smaller effects. For example, in their study of congressional elections Silberman and Durden (1975) estimate a decline of turnout of only 5 voters for the change in plurality described above. A model which treats the relation between turnout and plurality as a mere statistical artifact predicts a far greater effect than do the empirical data often presented in support of the rational voter model!

The essence of this statistical effect is that the larger the sample size the smaller the variance in the proportion of people who vote for a particular candidate. But for reasonably large sample sizes an increase in the sample size causes only a small reduction in the variance. Precisely because this effect is small, a regression equation estimates large effects. Small pluralities are associated with small variances which in turn must be associated with very large sample sizes.

Most empirical studies neglect this effect in asking whether the coefficient on the plurality variable is statistically different from zero. Researchers should instead ask whether the coefficient is statistically different from that predicted by a model which assumes only random effects. If the model we used is accepted, then the question is whether the coefficient on the plurality variable differs from  $-170$  million.

The model just described generates implausibly small pluralities (for example, when  $N$  is 100,000 the expected value of the plurality is 130 votes). This means that the correlation between turnout and plurality will be close to zero if one candidate is, on average, far more popular than the other. Models 2 and 3 described below do not suffer from this limitation. In addition, the fully specified rational voter models of Ledyard (1984), and of Palfrey and Rosenthal (1985) also predict small pluralities. And slight variations in the

assumptions of Model 1 can generate larger expected values for the plurality. Thus, the population of 150,000 potential voters can be viewed as divided into 150 groups of 1,000 each, where members of each group behave identically (that is, either all members turn out or none do, and all members of each group support the same candidate). Causes of such common behavior among members of groups include pressures from neighbors, local weather conditions, and the distance of residents from the polling booth. The statistical model would have  $N$  equal 150 rather than 150,000, so the winner receives a large fraction of the vote with non-negligible probability.

## Model 2: Random samples drawn from populations of supporters

Recall that we wish to determine the relationship between turnout and plurality in a model where no individual considers the probability that his vote is decisive. Model 1 assumed that each person first decides whether to vote, and then decides which party to support. Here we assume that the population of potential voters consists of Democrats and Republicans. The probability that a person votes is exogenously fixed. The number of Democrats who actually vote will thus be a random variable, with a distribution described by the binomial distribution. Similar statements apply to the number of Republican voters.

Let the probability that  $x$  Democrats vote be given by the function  $D(x)$ . Let  $v$  be the probability that a person votes, and let  $ND$  potential voters be Democrats. The probability that exactly  $x$  persons turn out to vote for the Democratic candidate is

$$D(x) = \binom{ND}{x} v^x (1-v)^{ND-x}. \quad (4)$$

A similar expression applies to the probability distribution,  $R(x)$ , of Republican turnout.

If Democrats outnumber Republicans, an election with a large turnout is most likely associated with many Democratic votes. But note also that a change in the number of Republican voters has a larger effect on the proportion of the vote received by the winner than does a change in the number of Democratic voters. (The fraction of the vote received by the Democratic candidate is  $D/(D+R)$ , where  $D$  is the number of Democratic votes and  $R$  the number of Republican votes. The value of  $d[D/(D+R)]/dR$  is  $-1/(D+R)^2$ , while the value of  $d[D/(D+R)]/dD$  is  $1/(D+R) - 1/(D+R)^2$ ; the absolute value of the second expression is less than that

*Table 1.* Relation between turnout and plurality: Random samples drawn from populations of supporters

| Mean turnout | Fraction Democrats | Coefficient on plurality | <i>t</i> -statistic |
|--------------|--------------------|--------------------------|---------------------|
| 0.8          | 0.5                | 4,200                    | 0.5                 |
| 0.8          | 0.7                | -128,700                 | -38                 |
| 0.8          | 0.9                | -136,500                 | -66                 |
| 0.9          | 0.5                | 6,000                    | 0.6                 |
| 0.9          | 0.7                | -137,730                 | -38                 |
| 0.9          | 0.8                | -154,000                 | -67                 |

of the first for values of  $D$  greater than  $R$ .) An increase in turnout therefore has an ambiguous effect on expected plurality, and numerical calculations are necessary to determine the sign and magnitude of the effect.

For simulation purposes let there be 2,000 jurisdictions, each with 150,000 eligible voters. For each jurisdiction we draw a random normal variable for the number of Democrats who vote and another random variable for the number of Republicans who vote. We thereby obtain for each jurisdiction a value for aggregate turnout and a value for the difference between the fractions of the voters who supported the winner and the loser. We use these values to estimate a regression equation in which the dependent variable is turnout and the explanatory variable is the plurality of the winner as a fraction of total turnout. Table 1 reports some of the results: Mean Turnout is the value of  $v$  used in expression (4); Fraction Democrats is  $ND/150,000$ . The two columns on the right give the regression coefficient and the  $t$ -statistic obtained from a linear regression estimated from the simulated data. For example, suppose a person votes with probability 0.8, and that out of the 150,000 potential voters in the population 70% are Democrats and 30% are Republicans (these are the numbers reported in the second line of Table 1). Actual turnout and votes for each candidate are random variables. The regression results show that a one point increase in the fraction of the voters who vote for the winner is associated with a turnout which is smaller by 128,700. As in Model 1, turnout and plurality are negatively related even though no person considers the probability that his vote will be decisive.

### **Model 3: Correlated turnout among supporters**

The previous model assumed that the probability that a particular person votes is fixed and identical for supporters of both parties. An alternative assumption is that each individual votes if and only if his cost of voting is below some critical value. Different persons may have different costs of

voting, so that only some persons choose to vote. The model is interesting if, in addition, the distribution of voting costs differs for Democrats and Republicans, and across jurisdictions or elections. This change in the assumptions generates different correlations between turnout and plurality.

To motivate the assumptions, suppose as before that potential voters are either Democrats or Republicans. A person votes if the cost of voting is less than some critical value. This assumption differs from the rational voter assumption because here the voter does not compare the costs of voting to the benefits. Let the cost of voting to individual  $i$ , a Democrat, be  $c_i + d$ ; let the cost of voting to individual  $j$ , a Republican, be  $c_j + r$ . Both  $c_i$  and  $c_j$  are random variables that can take different values for different persons. In each election the values of  $r$  and  $d$  are fixed, though they can vary across elections. That is, the costs of voting for Republicans are positively correlated, and similarly for Democrats. For example, most Democrats may live in the northern part of a state while most Republicans live in the southern part. A rain storm is unlikely to hit both parts of the state simultaneously, so that the costs of voting on election day will differ systematically between Democrats and Republicans.

Since the costs of voting are random variables correlated among Democrats and Republicans respectively, the fraction of Democrats who vote will differ from the fraction of Republicans who vote. Let  $D(x)$  be the probability that a fraction  $x$  of the Democrats vote. This probability is assumed to be normally distributed with mean 0.5 and standard deviation 0.1. Let the probability that a fraction,  $x$ , of the Republicans vote also be normally distributed with mean 0.5 and standard deviation 0.1; the population of eligible voters consists of 90,000 Democrats and 60,000 Republicans. In any particular election turnout among Democrats can differ from turnout among Republicans.

We use the procedure described in the previous section to estimate a regression equation in which turnout is the dependent variable and plurality is the explanatory variable. The coefficient for plurality is 28,800, with a  $t$ -statistic of 3.6. That is, an increase in the winner's proportionate vote from, say, 0.5 to 0.55 is associated with an increase in turnout of 1,440, or an increase in turnout of about 2 percentage points. Recall that this effect appears not because any voter is more likely to vote the closer he expects the election to be, but simply as a statistical artifact from the assumed random behavior of voters.

The opposite effect occurs if, in contrast to the earlier assumptions, the constituency is equally divided between Democrats and Republicans. Repeating the procedure described above with the new values of the parameters shows that the value of the coefficient for plurality is  $-47,500$  with a  $t$ -statistic of  $-4.6$ . High turnout will be associated with small pluralities.

An additional effect should be considered. Suppose  $D(x)$  has a very high

variance while  $R(x)$  has a small variance. Then high aggregate turnout will be associated with high Democratic turnout; low turnout will be associated with low Democratic turnout. If Republicans are a minority in the population, then plurality and turnout will be negatively correlated. If Republicans are a majority in the population, turnout and plurality will be positively correlated (cf. De Nardo, 1980; Tucker *et al.*, 1986).

These examples do not purport to explain fully the relationship between turnout and plurality. Individuals may be more likely to vote the closer they expect the election to be. If, however, the effects discussed above have some bite, then interpreting the empirical evidence for the rational voter model is difficult: even if voters do not weigh the costs and benefits of voting, turnout and plurality may be negatively correlated, so that finding a statistically significant coefficient does not support the rational voter model. Or it can be that in the absence of rational voting plurality and turnout are positively correlated. Finding a zero coefficient on plurality then provides strong evidence for the rational voter model.

## References

- Cox, Gary W. (1988) Closeness and turnout: a methodological note. *Journal of Politics* 50: 768–775.
- De Nardo, J. (1980) Turnout and the vote – The joke’s on the Democrats. *American Political Science Review*, 74(2): 406–420.
- Downs, Anthony (1957) *An Economic Theory of Democracy*. New York: Harper and Row.
- Ferejohn, John A. and Fiorina, Morris P. (1975) Closeness counts only in horseshoes and dancing. *American Political Science Review* 69(3): 920–925.
- Foster, Carroll B. (1984) The performance of rational voter models in recent presidential elections. *American Political Science Review* 78(3): 678–690.
- Gray, Virginia (1976) A note on competition and turnout in the American states. *Journal of Politics* 38(1): 153–158.
- Grofman, Bernard (1983) Models of voter turnout: An idiosyncratic review. *Public Choice* 41: 55–61.
- Ledyard, John O. (1984) The pure theory of large two-candidate elections. *Public Choice* 44: 7–41.
- Palfrey, Thomas R. and Rosenthal, Howard (1984) A strategic calculus of voting. *Public Choice*, 41: 7–53.
- Palfrey, Thomas R. and Rosenthal, Howard (1985) Voter participation and strategic uncertainty. *American Political Science Review* 79(1): 62–78.
- Patterson, Samuel C. and Caldeira, Gregory A. (1983) Getting out the vote: Participation in gubernatorial elections. *American Political Science Review* 77(3): 675–689.
- Settle, Russell F. and Abrams, Burton A. (1976) The determinants of voter participation: A more general model. *Public Choice* 27: 81–89.
- Silberman, J. and Durden, G. (1975) The rational behavior theory of voter participation: The evidence from congressional elections. *Public Choice* 23: 101–108.



Tollison, Robert, Crain, Mark, and Paulter, Paul (1975) Information and voting: An empirical note. *Public Choice* 24: 43-49.

Tucker, H.J. Vedlitz, A., and De Nardo, J. (1986) Does heavy turnout help Democrats in presidential elections? *American Political Science Review* 80(4): 1291-1304.