*Public Choice* **98:** 431–446, 1999. 


# Voter turnout: How much can we explain?*

JOHN G. MATSUSAKA[1] & FILIP PALDA[2]
[1]*Marshall School of Business, University of Southern California, Los Angeles, CA 90089-1421, U.S.A.;* [2]*E.N.A.P., University of Quebec, Montreal, Quebec, Canada H2L 4Z1*

Accepted 11 March 1997

**Abstract.** This paper evaluates the ability of common explanatory variables to predict who votes. Logit voting regressions are estimated with more than three dozen explanatory variables using survey and aggregate data for the 1979, 1980, 1984, and 1988 Canadian national elections. We find that the usual demographic variables such as age and education, and contextual variables such as campaign spending have significant effects on the probability of voting, but the models have low $R^2$'s and cannot predict who votes more accurately than random guessing. We also estimate regressions using past voting behavior as a predictor of current behavior, and find that although the explanatory power rises it remains low. This suggests that the difficulty in explaining turnout arises primarily from omitted time-varying variables. In some sense, then, it appears that whether or not a person votes is to a large degree random. The evidence provides support for the rational voter theory, and is problematic for psycho/sociological approaches.

## 1. Introduction

Why do some people vote and others abstain? This question is central to the study of public choice, and attracts an impressive amount of research attention. A healthy empirical literature has discovered a long list of variables that impact voter turnout at the margin, such as age, education, and income. However, the overall explanatory power of these variables (as opposed to their marginal effect) is a neglected research issue. For example, although it is well-documented that each year of schooling increases the probability that a citizen votes, little attention is paid to measuring how much of the observed variation in turnout can be explained by education differences. Our goal in this paper is to make just such a quantitative assessment. We attempt to document to what extent the variables identified in the voting literature can explain the variation in turnout. In doing so, we are implicitly trying to put a

* An earlier version of this paper circulated under the title, "Is Voting a Random Behavior?" We thank Nolan McCarty, Jeffrey Smith, an anonymous referee, and workshop participants at the University of Chicago for helpful comments. We gratefully acknowledge the financial support of the Bradley Foundation (through a grant to the Stigler Center at the University of Chicago) and the University of Chicago.

number on how much is really known about why some people vote and others abstain.

The empirical focus is on four consecutive Canadian national elections that were held in 1979, 1980, 1984, and 1988. Separate turnout regressions are estimated for each year using survey data merged with aggregate data. The regressions include over three dozen explanatory variables, many of which are assigned prominent roles in the voting literature. The explanatory power of the variables is assessed in several ways, with $R^2$-type goodness-of-fit measures, by comparing actual participation with model forecasted participation, and by comparing the average estimated turnout probabilities of voters and abstainers. We find that existing variables can explain no more than 15% of turnout variation. Although individuals with certain demographic characteristics have higher propensities to vote (more educated and older people, for example), and contextual factors such as campaign expenditures have significant positive effects on an individual's likelihood of voting, the overall ability of these variables to organize the data is weak.

One reason the regressions might fail to explain turnout is because they omit some difficult-to-measure variables, such as an individual's sense of citizen duty (as suggested, for example, by Riker and Ordeshook, 1968). To assess the seriousness of unobserved person-specific factors, we also estimate regressions that include past voting behavior as an explanatory variable. The idea is that if there are time-stationary factors leading a person to vote, whether or not he voted in the previous elections should be a good predictor of his turnout decision in the present election. The regressions that include the past turnout variable can account for more of the turnout variation than those without it, approaching 30% in some cases. This suggests that the weak explanatory power of the basic regressions is due in part to omitted time-stationary variables. However, it also implies that the remaining unexplained portion of the variance in turnout – more than 70% – is due to unobserved non-stationary variables. Such variables might be so numerous and idiosyncratic as to be unmeasurable for practical purposes. If so, the individual decision whether or not to vote is observationally equivalent to random behavior.

The fact that the individual voting decision appears to be more-or-less random can be taken as evidence in support of the rational voter theory. As Aldrich (1993) argues, the costs and benefits of voting are likely to be small for the typical person. The turnout decision for the rational voter then hinges on relatively minor factors, such as traffic and the weather perhaps. Because these factors are too numerous and minor to appear in voting databases, turnout at the individual level should be unpredictable and variable. In this sense, the results provide a partial rebuttal to recent criticisms of the rational voter theory, in particular, to the claim that it fails empirically.[1] On the other hand, the relatively minor role of time-stationary variables seems problematic for one of the more prominent alternatives to the rational voter theory, the so-called psycho/sociological approach, which implies a high degree of consistency in behavior across time.

The plan of the paper is the following. Section 2 discusses the data and the logit regressions. Section 3 reports a number of measures of their overall explanatory power. Section 4 considers the effect of weather. Section 5 tries to assess why prediction is difficult by including past participation in the regressions. Section 6 contains concluding comments.

## 2. Data

We constructed four data sets, one for each election, 1979, 1980, 1984, and 1988. The precise data sources are given in the Appendix. In brief, we began with individual-level survey data from the Canadian National Election Studies. We then matched information to each observation about (1) electoral conditions, drawn from the *Report of the Chief Electoral Officer Respecting Election Returns*, (2) campaign spending, drawn from the *Report of the Chief Electoral Officer Respecting Election Expenses*, and (3) district characteristics, drawn from the Census of Canada.

Canadian data have two desirable properties for the purposes of this study. First, voter registration is automatic at the age of 18. We do not have to worry that variation in turnout is being driven by differences in registration requirements, as would be the case in the United States. Second, a ballot cast in a Canadian national election pertains only to that national election. As opposed to U.S. elections, where a ballot contains the names of tens of individuals running for numerous races, we can be sure a Canadian voter is at the polling place specifically to vote on the national election. Thus, when we include campaign spending and election closeness as contextual variables, we do not need to worry if we are considering the right race.

One limitation of the survey data is that self-reported turnout rates exceed actual turnout rates. The actual turnout rate for the 1979 election was 76% while the sample rate was 91%. The corresponding numbers were 69% and 90% for 1980, 75% and 87% for 1984, and 75% and 90% for 1988. While this is troubling, the severity of the problem is comparable to other studies; for example, Ashenfelter and Kelley (1975) report a 74% turnout in their sample from the 1972 presidential election compared to the official estimate of 56%. Vote validation studies for the United States indicate that false voters differ from the population at large; in particular, they tend to be more educated and older (Silver, Anderson, and Abramson, 1986). The main danger associated with non-representative sample respondents and false responses is the possibility of biased regression coefficients. This is a serious concern, but there is some evidence from vote validation studies that it is unlikely to have a material affect on most research that makes use of survey data (Sigelman, 1982; Anderson and Silver, 1986).

A separate regression is estimated for each year. We attempted to incorporate all of the explanatory variables that are important according to the voting literature and then some. The variables are defined in the Appendix. There are slight variations in the variables across years due to idiosyncracies in the data sources. Most of them are self-explanatory, such as a person's age, education, income, occupation, sex, and marital status. The "margin" variable proxies for election competitiveness (or the probability that one vote is decisive). The measure we use was introduced into the literature by Barzel and Silberberg (1973): it is the difference between the votes of the top two finishers as a percentage of their combined votes.[2] Canada has three viable political parties so we also include margin measures between the winning and third place parties. The regressions also contain variables representing campaign activity – a dummy for whether or not a person was contacted by a campaign worker and per capita campaign spending in the person's district. Several variables describe the district's demographics, such as its mean income; these allow for the possibility that an individual's turnout decision is influenced by the characteristics of his neighbors. We include the mean turnout rate in the person's district; this is intended to capture unobserved district-specific determinants of turnout. Finally, we include region dummies for Quebec, Ontario, and the western provinces.

The regressions do not include disposition variables, such as a person's interest in politics, sense of citizen duty, and sense of political efficacy.[3] These variables purport to measure an individual's subjective attitudes, but there are reasons to be skeptical whether they actually do. First, there is the possibility that respondents rationalize their voting decisions with their answers to disposition questions. For example, Bishop, Oldendick, and Tuchfarber (1984) asked respondents whether or not they voted and how interested they were in politics; responses to the political interest question varied considerably with the order in which the questions were asked. Second, the responses to disposition questions vary in ways that suggest they are not recovering what they claim to be measuring. For example, Abramson, Silver, and Anderson (1987) shows that responses to citizen duty questions vary dramatically with question order. If there is such a thing as a person's sense of citizen duty, and it is being accurately measured, the value should not depend on something as irrelevant as question order.

The regressions also ignore issues and the spatial positions of the parties, but this is without cost. Because we focus on national elections, the issues

*Table 1.* Goodness-of-fit measures for logit voting regressions

| Measure | 1979 | 1980 | 1984 | 1988 |
|---|---|---|---|---|
| McFadden $R^2$ | 0.136 | 0.137 | 0.145 | 0.114 |
| Pseudo-$R^2$ | 0.103 | 0.101 | 0.101 | 0.084 |
| OLS $R^2$ | 0.079 | 0.082 | 0.107 | 0.073 |
| OLS $\overline{R}^2$ | 0.048 | 0.031 | 0.095 | 0.056 |
| Observations | 1,090 | 684 | 2,594 | 1,698 |

*Note*. Each column reports four goodness-of-fit measures for a logit regression of voting. The Appendix defines the independent variables and reports the estimated coefficients. The goodness-of-fit measures are defined in the text. The election year of the model is indicated at the head of the column. The last row reports the number of observations.

and parties facing each voter are the same. Thus, they cannot explain cross-sectional differences in turnout.

The parameter estimates are not our concern, so we merely report them in the Appendix. Suffice it to say that there are no surprises, except perhaps for the margin variables which are never significant.[4] Otherwise, the estimates are broadly consistent with any number of similar studies.

## 3. Explaining the variation of participation

This section assesses how much of the overall (within-sample) variability in voter turnout can be explained by our set of independent variables. Table 1 reports four goodness-of-fit measures for each of the models. The models differ by year, as indicated at the heading of each column, and by explanatory variables, as indicated in the Appendix. The first two measures are $R^2$ analogs for maximum likelihood estimates, the McFadden $R^2$ and the pseudo-$R^2$, as defined in Maddal (1983):

$$\text{McFadden } R^2 = 1 - \frac{\log L_1}{\log L_0};$$
$$\text{Pseudo} - R^2 = \frac{L_1^{2/n} - L_0^{2/n}}{1 - L_0^{2/n}}.$$

Here $L_1$ is the value of the likelihood function when maximized with respect to all 36+ parameters and a constant, $L_0$ is the value of the likelihood

*Table 2*. Mean estimated turnout probability for voters and abstainers

| | 1979 | 1980 | 1984 | 1988 |
|---|---|---|---|---|
| Mean estimated turnout probability, voters | 0.92 | 0.91 | 0.89 | 0.91 |
| Mean estimated turnout probability, abstainers | 0.81 | 0.80 | 0.75 | 0.82 |
| Difference in means | 0.11 | 0.11 | 0.14 | 0.09 |

*Note*. An estimated turnout probability was computed for each person in the sample using the coefficients from the logit regressions reported in the Appendix. The election year of the model is indicated at the head of the column. "Voters" are respondents who voted in the indicated election; "abstainers" are respondents who did not vote.

function when maximized only with respect to a constant, and n is the sample size. Both measures compare the maximum value of the likelihood function attained by the full model with that of an intercept-only model. The third and fourth measures are the $R^2$ and $\overline{R}^2$ from a linear probability model using OLS with the same explanatory variables as the logit regressions. All four measures have a theoretical range of 0 to 1, where 0 indicates no explanatory power. None of these measures actually represent the fraction of explained variance. Their primary purpose is to compare model specifications, but we think their absolute magnitudes are of interest when presented in conjunction with the other measures in this section.

The models share a poor fit. The McFadden $R^2$ ranges from 0.114 to 0.145, the Pseudo-$R^2$ ranges from 0.083 to 0.103, and the OLS $R^2$ ranges from 0.073 to 0.107. Substantially more explanatory power is expected from such a large list of theoretically relevant variables. Notwithstanding that cross-sectional survey regressions are often plagued with low $R^2$'s, Table 1 suggests that the available variables do not explain why people vote.

The accuracy with which the models approximate the observed data is assessed in a different way in Table 2. Each person's estimated probability of voting is computed using the estimated coefficients. The mean of these probabilities is then calculated separately for voters and abstainers. If the models have a good explanatory power, the two probabilities should be far apart. The row labeled "Difference in means" is the average estimated probability for voters less that for abstainers. Perfect prediction would give a difference of 1 while a complete absence of predictive power would give a difference of 0.

Table 2 tells the same story as Table 1. The explanatory power of the estimated models is weak. The difference between the mean estimated probability for voters and abstainers hovers around 0.10 for both years. The models have the toughest time picking out non-voters; the average estimated probability

*Table 3*. Percentage of people whose predicted behavior is consistent with actual behavior

| | 1979 | 1980 | 1984 | 1988 |
|---|---|---|---|---|
| % consistent, overall | 90.9 | 89.5 | 87.2 | 89.9 |
| % consistent, voters | 99.9 | 99.8 | 99.3 | 99.7 |
| % consistent, abstainers | 2.0 | 0.0 | 7.9 | 2.4 |
| % actual turnout in sample | 90.8 | 89.6 | 86.8 | 90.0 |
| Improvement over naive classification | 0.01 | –0.01 | 0.03 | –0.01 |

*Note*. An estimated turnout probability was computed for each person using the logit coefficients in the Appendix. A person was a "predicted voter" if his estimated probability was greater than 0.5, and a "predicted abstainer" otherwise. Behavior was classified as "consistent" if a predicted voter actually voted or a predicted abstainer actually abstained. The column headings identify the election year of the model. "Voters" and "abstainers" are respondents who actually voted in the indicated election. "Improvement over naive classification" is defined in the text.

for abstainers is never lower than 0.75; recall that perfect prediction for this subsample would yield a value of 0.

Yet another way to assess the explanatory power of the logit regressions is to use the estimated turnout probabilities to predict who votes and who abstains. The procedure is the following: if an individual's estimated turnout probability is greater than 0.5 he is classified as a *predicted voter*, otherwise he is a *predicted abstainer* (other cutoff points gave similar results). Each individual's predicted classification is then compared with his actual behavior.

Table 3 reports the percentage of people in the full sample whose predicted and actual behavior are the same. We also report the percentage of observations with consistent predicted and actual behavior for the voters-only and abstainers-only subsamples. The actual within-sample turnout percentage for each year is also reported. This is the lower bound on the prediction accuracy; naively classifying each person as a voter results in a consistency percentage equal to the actual turnout percentage. The relatively high within-sample actual turnout percentage leaves little for the logit regressions to improve on. If a completely naive prediction can achieve 90% accuracy then in a sense only 10% remains to be explained. The last row of the table, "Improvement over naive classification," measures the fraction of this unexplained turnout that is explained by the regression. For example, if 90% of the sample voted and the regression classifies 91% correctly, then improvement is (91–90)/(100–90) = 0.1. Perfect prediction gives an improvement of 1.

The models correctly classify the voting behavior of about 90% of the people in the sample. This is a high level of accuracy, but no better than what is achieved by simply predicting that everyone is a voter. The improvement compared to the naive classification ranges from –0.01 to 0.03. These results are humbling; the naive classification essentially does as good a job as the 36+ variable regressions. The model does well on overall prediction; it thus leans heavily on sensitivity, correctly predicting the positives. The success rate in predicting voters is always over 99%; the cutoff of 0.5 is almost the same as simply classifying everyone as a voter. The specificity, the success rate in predicting abstainers, suffers, as one would expect; in the best model, 1984, only 7.9% of abstainers are classified correctly.

The results of Tables 1–3 can be summarized in a simple way. Despite inclusion of a large number of theoretically relevant variables, the logits have almost no predictive power. Evidently the estimated models leave a large part of the voting story untold.

These findings are not out of line with estimates in other studies. A notable early econometric study by Silver (1973) investigated participation in the 1960 presidential election. In Silver's full equation, estimated with ordinary least squares, he reported an $R^2$ of 0.34. Given that his equation included a number of psychological variables that have big effects on $R^2$'s, this is a comparable number. In their study of the 1972 presidential election, Wolfinger and Rosenstone (1980) estimated a probit that was able to predict correctly 72.9% of their sample. The overall turnout rate in their sample was 66.7%. In terms of Table 3, their model had an improvement compared to a naive classification of 0.186, which is better than we find but still rather low.

## 4. Weather

According to conventional wisdom, weather conditions have an effect on voter turnout. None of the preceding models incorporate weather variables. In order to assess the importance of this omission, we collected a variety of weather-related variables for the 1980 election and re-estimated the model. We chose 1980 because weather effects are most likely to show up under extreme circumstances, and that was the only winter election in our sample. The variables are the mean, minimum, and maximum temperature on election day, the mean, minimum, and maximum deviation of the election day temperature from the monthly mean, and the amount of precipitation.[5] The point estimates (not reported) indicate quantitatively significant marginal effects in some cases, but none of the coefficients are different from zero at conventional levels of statistical significance. This is not out of line with the findings of Knack (1994) that weather has a small effect on turnout in U.S. national elections.

Table 4 reports for this model the various goodness-of-fit measures introduced in Tables 1–3. Consideration of the weather appears to improve slightly

*Table 4.* Goodness-of-fit measures, mean estimated turnout probabilities, and consistency of predicted with actual behavior when weather variables are included in the 1980 model

| | |
|---|---|
| McFadden $R^2$ | 0.158 |
| Pseudo-$R^2$ | 0.117 |
| OLS $R^2$ | 0.102 |
| OLS $\overline{R}^2$ | 0.041 |
| Mean estimated turnout probability, voters | 0.91 |
| Mean estimated turnout probability, abstainers | 0.78 |
| Difference in means | 0.13 |
| % consistent, overall | 89.6 |
| % consistent, voters | 99.5 |
| % consistent, abstainers | 4.3 |
| % actual turnout in sample | 89.6 |
| Improvement over naive classification | 0.00 |
| Observations | 670 |

*Note*. The numbers are based on a 1980 model with seven weather variables and the variables listed in the Appendix. The top panel reports four goodness-of-fit measures as in Table 1. In the second panel, each person was assigned an estimated turnout probability based on the estimated logit coefficients, as in Table 2. In the third panel, each person was predicted to be a voter if his estimated probability was 0.5 or greater, and was predicted to be an abstainer otherwise; the numbers indicate the percentage of people whose predicted behavior matched their actual behavior, as in Table 3. The last row reports the number of observations.

the ability of the model to explain the variation in participation. But the basic conclusion stands: the model has little predictive power.

## 5. Why explanation is difficult

The inability to explain within-sample variation in turnout in a model that includes almost all of the key variables that have been identified in the literature is somewhat surprising. It suggests one of two things. Either one or more important explanatory variables are missing, or voting is an intrinsically random behavior and hence unpredictable (these may be saying the same thing).

There is a simple way to test if the regressions are missing important variables that are constant over time, such as a person's sense of citizen duty. If missing variables are a problem and they are more-or-less constant across time, then they will be incorporated in a person's past voting behavior. A person's past voting behavior should then allow good predictions of future behavior. Intuitively, if an unobserved variable like citizen duty is in fact a primary determinant of participation then if a person's sense of duty drives him to vote in one election it is likely to drive him to vote in the next.

To test this, the models were re-estimated with a dummy variable added indicating whether or not an individual voted in the previous national election (the election preceding 1979 was in 1974). Past voting behavior is important; the coefficient on the dummy variable (not reported) is statistically significant at better than the 1% level in all regressions. And the marginal effects are large; a person who voted in the previous election is 12% to 18% more likely to vote in the current election.[6]

Table 5 reports the corresponding goodness-of-fit measures, average estimated probabilities for voters and non-voters, and correct prediction percentages, as in Tables 1–4. The explanatory power of the model substantially increases with inclusion of the past vote dummy variable. The McFadden $R^2$ rises from 0.136 to 0.230 in the 1979 model, 0.158 to 0.207 in the 1980 weather model, 0.145 to 0.192 in the 1984 model, and 0.114 to 0.163 in the 1988 model – an average improvement of about 0.05. The pseudo-$R^2$ also increases, but the magnitudes are smaller. The OLS $R^2$ jumps 4% in the 1979 and 1988 models, and 5% in the 1980 and 1984 models. The difference between the mean estimated turnout probability of voters and abstainers increases by 16%, 5%, 4%, and 4%, respectively, in the 1979, 1980, 1984, and 1988 models. The consistency, on the other hand, does not improve much. However, the consistency of abstainers rises by 9.1% in 1979, 10.9% in 1980, and 13% in 1984.

Two results stand out. First, it seems clear that missing time-stationary variables are an important component of the inability to explain why people vote. But what is more important are missing non-stationary variables, or perhaps equivalently, randomness. Most Canadians are inclined to vote, but whether or not a person makes the trip to the polls might depend on factors as trivial and unsystematic as his mood on election day, how busy he is and hence how forgetful, the traffic, the amount of paperwork at the office, and so on.

## 6. Conclusion

This study evaluates the ability of logit voting regressions to explain the cross-sectional variation in turnout. The explanatory power is weaker than might be expected, especially because the study is designed to maximize the predictive ability in two ways. First, we study Canadian elections where

*Table 5.* Goodness-of-fit measures, mean estimated turnout probabilities, and consistency of predicted with actual behavior when a dummy variable for past turnout is included in all models

| | 1979 | 1980 | 1980w | 1984 | 1988 |
|---|---|---|---|---|---|
| McFadden $R^2$ | 0.230 | 0.185 | 0.207 | 0.192 | 0.164 |
| Pseudo-$R^2$ | 0.185 | 0.140 | 0.158 | 0.137 | 0.123 |
| OLS $R^2$ | 0.120 | 0.130 | 0.152 | 0.157 | 0.117 |
| OLS $\overline{R}^2$ | 0.048 | 0.080 | 0.091 | 0.145 | 0.100 |
| Mean estimated turnout probability, voters | 0.94 | 0.92 | 0.92 | 0.89 | 0.91 |
| Mean estimated turnout probability, abstainers | 0.77 | 0.77 | 0.74 | 0.71 | 0.78 |
| Difference in means | 0.27 | 0.15 | 0.18 | 0.18 | 0.13 |
| % consistent, overall | 92.8 | 90.3 | 90.1 | 88.5 | 90.2 |
| % consistent, voters | 99.3 | 98.5 | 98.3 | 98.5 | 99.3 |
| % consistent, abstainers | 11.1 | 14.9 | 15.2 | 20.9 | 7.2 |
| % actual turnout | 92.6 | 90.1 | 90.1 | 87.1 | 90.1 |
| Improvement over naive classification | 0.03 | 0.02 | 0.00 | 0.11 | 0.01 |
| Observations | 484 | 678 | 664 | 2,530 | 1,685 |

*Note*. The numbers were computed using parameter estimates generated from voting regressions that include a dummy variable equal to 1 if a person voted in the previous election, and the variables listed in Appendix. The top panel reports four goodness-of-fit measures as in Table 1. In the second panel, each person was assigned an estimated turnout probability based on the estimated logit coefficients, as in Table 2. In the third panel, each person was predicted to be a voter if his estimated probability was 0.5 or greater, and was predicted to be an abstainer otherwise; the numbers indicate the percentage of people whose predicted behavior matched their actual behavior, as in Table 3. The column headings identify the election year of the model; 1980w is the model that also includes weather variables.

registration is automatic so that it is not necessary to account for differences in the cost of registration (unlike in the United States). Second, we include over three dozen explanatory variables in the regressions, among them election closeness, an individual's age, education, income, marital status, religion, native language, employment status, occupation, and union membership; his community's size, education, age, religion, and language; campaign spending in his district, whether he was contacted in the course of the campaign by a party official, phone, or mail; regional dummies; and local weather conditions – in short, a preponderance of the explanatory variables identified in the voting literature. This study is not the first to note that our understanding of why people vote is somewhat less than we might hope (see Aldrich and Simon, 1986), but we believe it is the first to quantify systematically the severity of the problem.

Better turnout predictions apparently require identification of additional cost and benefit variables. To investigate whether the missing variables are constant or changing over time, a dummy variable for past turnout is included as an explanatory variable in the regressions. We find that past voting behavior is a significant predictor of present voting behavior, indicating the presence of unobserved time-stationary variables, such as the sense of citizen duty. This highlights the importance of understanding why some people have a sense of duty (or whatever it is) and not others, and suggests the need for richer measures of these dispositions.

However, most of the inability to predict who votes appears to come from non-stationary factors. This is good news for the rational voter theory. As Aldrich (1993) argues, both the benefits and costs of voting are small. Consequently, we expect the turnout decision to be sensitive to small variations in benefits and costs. If these small effects are difficult to measure, turnout should appear to be for the most part random. It should also be random if voters are following mixed strategies, as suggested by game-theoretic rational voter models such as Palfrey and Rosenthal (1985). Randomness appears to be bad news for psycho/sociological explanations of voter turnout, which link participation to individual attitudes, social norms, culture, and the like, that are expected to be persistent over time (for example, see Campbell, Converse, Miller, and Stokes, 1960; Verba and Nie, 1972; and more recently Knack, 1992). In order to fit the data, a psycho/sociological explanation would have to maintain that an individual's sense of citizen duty, attachment to the community, and so on, are highly variable over time.

More generally, our results open the door to the possibility that turnout is driven by idiosyncratic costs like the weather, the traffic, personal health, and so on. The diversity of these costs and the difficulty in measuring them may mean that predicting who votes is ultimately infeasible. From a research point of view, perhaps individual voting must be approached as a fundamentally random behavior.

Less nihilistically, the results suggest to us two profitable paths for future research. One is the search for new explanatory variables and away from reliance on traditional demographics like education and income. The other path is toward the study of aggregated voting behavior (in which, for example, the unit of observation is a district rather than an individual), where individual idiosyncracies will cancel each other and allow the estimation of models with greater explanatory power.

## Notes

1. The results are also consistent with game-theoretic turnout models where voters play mixed strategies, for example, Palfrey and Rosenthal (1985).
2. We also estimated (but do not report) regressions using other competitiveness measures such as the "closeness" variable of Chapman and Palda (1983), Cox (1988), and Cox and Munger (1989), and ex ante measures that use only information available prior to the election. None of the alternatives altered the estimates in an important way.
3. However, to the extent that these variables are constant across time, they will be picked up by our past voting dummy; see Section 5.
4. As we show in our literature survey in Matsusaka and Palda (1993), the absence of a relation between margin and propensity to vote is not unusual in the voting literature. In our view, this is entirely consistent with the rational voter theory; we should not expect measurable responses to changes in nearly infinitesimal probabilities. See the discussions in Matsusaka (1993) and Matsusaka and Palda (1993).
5. The data were taken from *Monthly Record: Meteorological Observations*, February 1980, published by Environment Canada.
6. The derivative of the probability with respect to the dummy evaluated at the mean is 0.123 in the 1979 model, 0.178 in the 1980 weather model, 0.180 in the 1984 model, and 0.137 in the 1988 model.

## References


Abramson, P.R., Silver, B.D. and Anderson, B.A. (1987). The effects of question order in attitude surveys: The case of the SRC/CPS citizen duty items. *American Journal of Political Science* 31: 900–908.

Aldrich, J.H. and Simon, D.M. (1986). Turnout in American national elections. In S. Long (Ed.), *Research in micropolitics*, Vol. 1. Greenwich, Connecticut: JAI Press.

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

Anderson, B.A. and Silver, B.D. (1986). Measurement and mismeasurement of the validity of the self-reported vote. *American Journal of Political Science* 30: 771–785.

Ashenfelter, O. and Kelley, S., Jr. (1975). Determinants of participation in presidential elections. *Journal of Law and Economics* 18: 695–733.

Barzel, Y. and Silberberg, E. (1973). Is the act of voting rational? *Public Choice* 16: 51–58.

Bishop, G.F., Oldendick, R.W. and Tuchfarber, A.J. (1984). Interest in political campaigns: The influence of question order and electoral context. *Political Behavior* 6: 159–169.

Campbell, A., Converse, P.E., Miller, W.E. and Stokes, D.E. (1960). *The American voter*. New York: Wiley.

Chapman, R.G. and Palda, K.S. (1983). Electoral turnout in rational voting and consumption perspectives. *Journal of Consumer Research* 9: 337–346.

Cox, G.W. (1988). Closeness and turnout: A methodological note. *The Journal of Politics* 50: 768–775.

Cox, G.W. and Munger, M.C. (1989). Closeness, expenditures, and turnout in the 1982 U.S. House elections. *American Political Science Review* 83: 217–231.

Knack, S. (1992). Civic norms, social sanctions, and voter turnout. *Rationality and Society* 4: 133–156.

Knack, S. (1994). Does rain help the Republicans?: Theory and evidence on turnout and the vote. *Public Choice* 79: 187–209.

Maddala, G.S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge, UK: Cambridge University Press.

Matsusaka, J.G. (1993). Election closeness and voter turnout: Evidence from California ballot propositions. *Public Choice* 76: 313–334.
Matsusaka, J.G. and Palda, F. (1993). The Downsian voter meets the ecological fallacy. *Public Choice* 77: 855–878.
Palfrey, T.R. and Rosenthal, H. (1985). Voter participation and strategic uncertainty. *American Political Science Review* 79: 62–78.
Riker, W.H. and Ordeshook, P.C. (1968). A theory of the calculus of voting. *American Political Science Review* 62: 25–42.
Sigelman, L. (1982). The nonvoting voter in voting research. *American Journal of Political Science* 26: 47–56.
Silver, B.D., Anderson, B.A. and Abramson, P.R. (1986). Who overreports voting? *American Political Science Review* 80: 613–624.
Silver, M. (1973). A demand analysis of voting costs and voting participation. *Social Science Research* 2: 111–124.
Verba, S. and Nie, N.H. (1972). *Participation in America: Political democracy and social inequality*. New York: Harper & Row.
Wolfinger, R. and Rosenstone, S.J. (1980). *Who votes?* New Haven: Yale University Press.

**Appendix: Variable definitions, data sources, and estimated logit coefficients**

| Individual variables | Source | 1979 | 1980 | 1984 | 1988 |
|---|---|---|---|---|---|
| Age in years | 1,2,3 | 0.09[b] | 0.07 | 0.60[b] | 0.06 |
| Age-squared $\times 10^{-2}$ | 1,2,3 | –0.08[b] | –0.06 | –0.03 | –0.04 |
| Education – years of education for 1979 and 1980; categorical variable with range 0–6 for 1984, and 1–9 for 1988 | 1,2,3 | 0.16[c] | 0.10[a] | 0.34[c] | 0.02 |
| Family income – categorical variable with range 1–8 for 1979 and 1980, 1–11 for 1984, and 1–9 for 1988 | 1,2,3 | –0.01 | 0.03 | 0.13[c] | 0.11[b] |
| Dummy = 1 if married | 1,2,3 | 0.29 | 0.23 | 0.48[c] | 0.37[a] |
| Dummy = 1 if male | 1,2,3 | 0.39 | 0.68[b] | 0.28[a] | –0.12 |
| Dummy = 1 if native Canadian | 1,2,3 | –0.13 | 0.28 | 0.01 | 0.59[b] |
| Dummy = 1 if Catholic | 1,2,3 | 0.07 | 0.12 | 0.16 | –0.28 |
| Frequency of church attendance – categorical variable with range 0–4 for 1979 and 1980, 1–5 for 1984, and 0–8 for 1988 | 1,2,3 | 0.18[b] | –0.02 | 0.16[c] | 0.17[c] |
| Dummy = 1 if person or member of family belonged to a union | 1,2,3 | 0.48[a] | 0.02 | –0.30[a] | 0.27 |
| Dummy = 1 if person spoke French at home | 1,2,3 | –0.84 | –0.14 | –0.37 | 0.05 |
| Length of residence in province – categorical variable with range 1–4 for 1979 and 1980 (1 is longer than 4), and 1–4 for 1984 (1 is shorter than 4) | 1,2 | –0.18 | –0.10 | 0.19[b] | – |
| Community size – categorical variable with range 1–9 for 1979 and 1980 (1 is large, 9 is small), and 1–6 for 1984 | 1,2 | –0.03 | 0.00 | –0.04 | – |
| Dummy = 1 if unemployed | 1,2,3 | 0.80 | 0.12 | –0.16 | 0.39 |
| Dummy = 1 if retired | 1,2,3 | –0.43 | –0.94[a] | –0.04 | 0.46 |
| Dummy = 1 if student | 1,3 | –0.76 | –0.68 | – | 4.61 |
| Dummy = 1 if farmer | 1,2,3 | –0.84 | –0.65 | –0.11 | –0.50 |
| Dummy = 1 if person was a professional, owner, manager, or executive | 1,2,3 | –0.53 | –0.11 | 0.37[a] | 0.28 |
| Dummy = 1 if person worked as a skilled or unskilled laborer | 1,2,3 | –0.64[b] | –0.36 | 0.08 | –0.28 |
| Dummy = 1 if person was contacted in person by a campaign worker before election | 1,2 | 0.11 | 1.00[b] | –0.51[c] | – |
| Dummy = 1 if person was contacted by mail by a campaign before election | 1,3 | 0.39 | 0.54[a] | – | 0.24 |
| Dummy = 1 if person was contacted by phone by a campaign worker before election | 1 | 0.61 | 0.59 | – | – |
| Dummy = 1 if person was contacted by mail or phone by a campaign before election | 2 | – | – | 0.80[c] | – |
| Dummy = 1 if person was contacted in person or by phone by a campaign worker before election | 3 | – | – | – | 0.54[c] |

| District variables | Source | 1979 | 1980 | 1984 | 1988 |
|---|---|---|---|---|---|
| Margin – difference between votes for winner and runner-up, divided by their combined votes | 4 | 0.05 | 0.70 | 0.47 | 0.27 |
| Margin – difference between votes for winner and third-place finisher, divided by their combined votes | 4 | 0.44 | 1.14 | –0.71 | –0.05 |
| Proportion educated – number of people older than 15 with some high school education divided by number of registered voters | 4,6,7 | –1.18 | 3.81 | –1.78 | –1.56 |
| Proportion in labor force – number of people older than 15 in the labor force divided by number of registered voters | 4,6,7 | –1.64 | 3.29 | –0.25 | 0.13 |
| Proportion Catholic – number of Catholics divided by population | 6 | 1.13 | –2.53 | –0.14 | – |
| Proportion native – number of native Canadians divided by population | 6,7 | 0.69 | –3.09 | –1.17 | 0.61 |
| Proportion French-speaking – number of French-speakers divided by population | 6,7 | –0.68 | 1.68 | 0.44 | –0.71 |
| Population growth – population in 1981 minus population in 1976, divided by population in 1976 | 6 | 0.57 | –2.38 | –0.01 | – |
| Average male income plus average female income (in $1,000s), divided by two | 6,7 | 0.10 | –0.10 | –0.02 | 0.02 |
| Per capita campaign spending – combined spending of Liberal, PC, and NDP parties for 1979 and 1980, combined spending of all parties for 1984 and 1988 | 5,6 | 1.16[b] | 1.03 | –0.04 | 0.03 |
| Turnout rate – votes cast divided by number of registered voters | 4 | 2.16 | 1.61 | 6.03[c] | 3.25 |
| Dummy = 1 if person lived in Quebec | 1,2,3 | 0.21 | 0.19 | –0.21 | 1.30[c] |
| Dummy = 1 if person lived in Ontario | 1,2,3 | 0.47 | –0.51 | –0.55[b] | 0.30 |
| Dummy = 1 if person lived in Alberta, British Columbia, Manitoba, or Saskatchewan | 1,2,3 | 0.24 | –1.31[b] | –0.25 | 0.53 |
| Constant | – | –5.40[c] | –5.10 | –5.41[c] | –4.39[a] |

*Note*. The numbers were computed using parameter estimates generated from voting regressions that include a dummy variable equal to 1 if a person voted in the previous election, and the variables listed in Appendix. The top panel reports four goodness-of-fit measures as in Table 1. In the second panel, each person was assigned an estimated turnout probability based on the estimated logit coefficients, as in Table 2. In the third panel, each person was predicted to be a voter if his estimated probability was 0.5 or greater, and was predicted to be an abstainer otherwise; the numbers indicate the percentage of people whose predicted behavior matched their actual behavior, as in Table 3. The column headings identify the election year of the model; 1980w is the model that also includes weather variables.