Do Polls Affect Turnout?  
An Uncertainty Principle for Election Surveys

Marc Davignon
mdavign1@binghamton.edu
William B. Heller
wheller@binghamton.edu
Ekrem Karakoc
ekarakoc@binghamton.edu
Department of Political Science, Binghamton University

Abstract

Election campaigns often look more like horse races than the precursor to basic decisions about who sets the policy agenda for society. Pollsters survey public opinion in an attempt to predict what the voting public will do when it comes to filling out actual ballots. Political partisans jockey for position, sometimes complaining that polls are skewed in their opponents’ favor. Only after the polls close and votes are tallied is it possible to tell how accurate pollsters’ predictions were. But what if, as complaints about inaccurate polling implicitly suggest, voter decisions to vote at all are affected not only by the usual suspects—information, education, preferences, political awareness, etc.—but also by how poll results affect their beliefs about outcomes? In this paper we examine how election-as-horse-race information from opinion surveys should affect who chooses to turn out and who chooses to stay home rather than vote. We argue that common understandings in the literature of the turnout decision imply a number of distinct hypotheses about the relationship between poll results and outcomes, which we test using polling and election data from US presidential elections.
Political polling during campaigns is designed to take the pulse of the electorate. Ideally, a good poll should reveal what the outcome of an election would be if it were held at the moment the poll was taken. At the same time, it is well understood that expectations about election outcomes are dynamic, changing over time in response to campaigns, information related to campaigns, and the unavoidable exogeneities of the real world. Polls—at least early polls—consequently often are skewed one way or another, as Nate Silver documented in a 5 November fivethirtyeight blog article. Why they are skewed is less clear.

Silver suggests in the aforementioned blog post that at least part of the explanation for skewed polls might be pollster “‘herding’—the tendency of polls to mirror one another’s results rather than being independent.” In itself, however, such herding should show up as skewing only to the extent that it is biased, which calls into question the utility of the polling enterprise. We suggest an alternative explanation for skewed polls that is consistent with the herding explanation but goes beyond it to generalize to all polls. We argue, in a nutshell, that polls are skewed precisely because they provide information to campaigns and to voters, who then can modify their behavior on the basis of what the polls tell them.

Using all available polls for the 2000 to 2012 presidential elections, we take a first-cut run at a set of hypotheses derived from distinct assumptions about what motivates voters both to bother to go to the polls and to vote for one candidate over another. We lay out these hypotheses

---

1 “The Polls Were Skewed Toward the Democrats” [http://fivethirtyeight.com/features/the-polls-were-skewed-toward-democrats], last accessed 20 April 2015)
in the next section and test them in the following section. In all, our results suggest at the very least that there is a systematic relationship between election surveys and actual election results; the Heisenberg Principle applies in the social as well as the natural sciences.

ARGUMENT AND HYPOTHESES

Surveys are informative. First and foremost, they tell voters and candidates alike who is ahead and who is behind—or, perhaps more accurately, how the election in question would turn out if the poll itself were the election. Second, depending on the questions asked, they also might provide insight about how and why voters are thinking about voting. And third, as part and parcel of these points, surveys predict who will turn out to vote and who will not.

Information, of course, is meaningless unless people pay attention to it. And it is irrelevant unless it has at least the potential to affect behavior. Campaign polls are no exception. It follows that if candidates and potential voters respond to the information in campaign polls in systematic ways, most polls should be skewed. In essence, the measurement and dissemination of survey results should in hindsight render false the information they provide. How that happens depends on what people do with that information.

Candidates (and their parties) and voters alike should use the information they get from surveys to adjust their behavior. How that plays into subsequent surveys or actual election results depends on what they are trying to maximize. As far as candidates are concerned, we expect them either to adjust their campaign strategies (if polls show them to be behind where they expect) or carry on as they had been (if polls show them to be doing at least as well as they expect) as long as other candidates do the same. If they do so optimally—that is, if when candidates adjust their campaigns they succeed in making themselves more attractive to voters—
then the next time voters are sampled those who did poorly should do better, all else equal, thus narrowing the gap between front runners and the rest of the pack.

Of course, all else rarely is equal. Candidate strategies alone do not determine trends in polls and, ultimately, elections. Most importantly, voters also should react to poll results. How they do so, we think, depends on why they would consider voting in the first place. If polls set the conditions that make them skewed ex-post, therefore, their expected effects necessarily depend on voter motivations. In what follows we examine the different implications of various fundamental assumptions about what makes voters tick. We consider first the implications of different voting motivations in isolation and is if all potential voters shared the same basic motivation (but not preferences), though voters in the real world undoubtedly make more complex (if less conscious) calculations. In all cases, it is worth noting, we assume that voters see voting as a costly activity and roughly understand the relationship between the competitiveness of a race and the probability that their individual votes will affect outcomes. We assume as well, perhaps heroically, that campaigns work not by changing voters’ minds, but rather by affecting their propensity to show up at the polls.\footnote{Modern polls take into account the question of whether a given respondent is likely to vote in reported results.} We assume finally—and our results suggest—that polls are accurate inasmuch they accurately sample the population and measure respondents’ vote preferences to indicate how the election would turn out if it were held at that moment.

Whatever motivates voters to turn out or stay home on election day should affect both aggregate turnout and election results. We explore both these aspects of elections, with hypotheses that address each.
**Citizen Duty** Riker and Ordeshook (1968, 1973) asked why anyone would bother to vote in the first place. The explanation, the famous “D term” fits nicely with the normative view that democratic government is a privilege that requires all citizens to contribute to the common good by voting. In a nutshell, voters go to the polls because they see voting as part of the cost of living in a free society (Aldrich 1993).

If voters go to the polls because they feel obliged to, the implications are straightforward. First, of course, turnout should be fairly uniform and in line with what campaign surveys predict. And second, given our assumption that voters are less fickle than at least “independents” in the middle are portrayed to be, predictions from surveys should accurately reflect vote outcomes as well as turnout. This observation leads directly to our baseline hypothesis:

**Hypothesis 1 (Citizen Duty):** The results of surveys taken in the leadup to elections will accurately reflect both actual vote outcomes and turnout, which will be relatively invariant across elections.

The expectation with respect to turnout for Hypothesis 1 is illustrated by the horizontal line at the top of Figure 1.

**Bandwagoning** Some scholars have suggested that voters like the feeling of being on the winning side and so will pile on to whatever candidate looks likely to do best (McAllister and Studlar 1991; Grofman, Owen and Collet 1999). Voters who subscribe to this sort of motivation thus should be more likely to vote the clearer it is what the election outcome will be. In terms of turnout, the prediction for bandwagoning behavior is the opposite of the prediction for efficacy (see below).
Hypothesis 2 (Bandwagoning—Turnout): *Voters should be more likely to go to the polls the clearer it is which candidate is favored to win.*

The turnout prediction for bandwagoning is illustrated by the dashed-line U-shaped curve in Figure 1. And because any bandwagoning-induced increase in turnout would be due to voters flocking to support the expected winner, higher turnout should go hand in hand with more lopsided election results. Predicted lopsided results from electoral surveys are the catalyst for a cascade of support for the predicted winner. This is our next hypothesis.

Hypothesis 3 (Bandwagoning—Results): *The more lopsided are survey predictions with respect to election outcomes, the more bandwagoning voters will flock to the polls and to the expected winner, making polls ex-post look skewed toward losers.*

**Efficacy** One reasonable alternative to the Citizen Duty hypothesis is the notion that voters are more likely to go to the polls, all else equal, when they think their votes are likely to matter. When polls indicate that election results are likely to be highly skewed, for instance, any given voter can calculate that her vote will not affect the outcome of the election and thus should be less likely to vote. Turnout should be lowest at the extremes of predicted vote share, where the outcome appears to be a foregone conclusion, and highest where the election is most competitive. Our second hypothesis follows directly:

Hypothesis 4 (Efficacy—Turnout): *Turnout should be lowest when elections are uncompetitive and highest when they are competitive.*

The upshot of this second hypothesis is that where election polls identify uncompetitive races turnout should be lower, as illustrated by the upside-down U shaped (solid-line) curve in Figure 1. But while polls might be skewed in terms of turnout, voters who vote purely
on the basis of efficacy should stay away from the polls equally whether their preferred party is predicted to win or to lose, so electoral surveys should accurately predict election outcomes. This is our next hypothesis.

Hypothesis 5 (Efficacy—Outcomes): *Election surveys should be unbiased.*

**Proximity (Partisanship)** Many scholars have suggested that citizens are more likely to make an effort to get to the polls if they feel strongly about one party or another (feelings need not be positive to goad voters to action; Abramowitz and Webster 2015). This is not something that our abstract conception of electoral surveys should affect. Voters feel the way they feel about the parties on the ballot and vote accordingly. We might as a general rule expect higher turnout where there are more parties, since the existence of more parties increases the chances that any given voter will be fairly near to some party (and, on the negative side) fairly distant from others. We would thus expect turnout to increase with the number of parties but, as in the case of voting as citizen duty, neither outcomes nor turnout for any given election should be affected by opinion surveys.³

**Partisan Efficacy (Efficacy + Proximity)** Our promise to consider distinct motivations for voting in isolation notwithstanding, we think a combination of efficacy and proximity considerations is both plausible and useful to consider as a stand-alone category. Basically, the idea is that rational voters should be more likely to vote the more reasonable it is to believe that their votes might matter—that is, when the election looks like it might be competitive. Adding partisan considerations adds useful nuance, however, in the following senses. First, disillusionment: when a voter’s preferred party clearly is well-behind in the polls, the incentive

³ A reasonable test of the pure proximity motivation would require examining multiparty systems (cf. Kedar 2009). We leave this for future work.
to vote should be very low. The only way that going to the polls would be useful would be if many of the expected winner’s supporters were to decide to stay home (about which see below), and even then, if the predicted margin of victory is large enough, no amount of turnout from supporters of the expected loser is likely to affect the outcome. Second, caution: when a voter’s preferred party clearly is well-ahead in the polls, the temptation to stay home rather than pay the cost of going out to vote might loom large. However, where the expected losers in this circumstance are unlikely to be able to turn the election around the expected winners run the risk of tossing the election if they simply rest on their laurels. Here there is a strategic consideration, which we think should be more compelling for potential winners than for potential losers—the more likely is turnout to be low, the more likely any given vote will matter. Thus, voters whose candidates are predicted to do well should lean more on their partisan preferences (as well as the realization that the election is theirs to lose) than on the pure cost-benefit analysis of efficacious voting in deciding whether to go the polls. This yields two further hypotheses:

Hypothesis 6 (Partisan Efficacy—Turnout): The more lopsided are predicted election results, the lower turnout should be.

Hypothesis 7 (Partisan Efficacy—Outcomes): Lopsided polls should in hindsight be skewed against winning candidates.

Hypotheses 6 and 7 together imply one more hypothesis. Poll predictions affect turnout and, through turnout, elections. When polls predict a lopsided elections, fewer people bother to vote. But those who expect to be on the losing side, by the assumptions underpinning this formulation of partisan efficacy, should stay home more than those who expect their party to win. Our final hypothesis thus is interactive.
Hypothesis 8 (Partisan Efficacy—Outcomes and Turnout): *The lower is turnout, the more polls should appear ex-post to have been skewed against winning candidates.*

This last hypothesis, as noted, builds on the reason that voters bother turnout out. The motivation is about competitiveness, but there also is an element of something else—it could be risk aversion (not voting for the expected winner slightly reduces the probability of victory, all else equal), for instance, or something akin to bandwagoning (voting for a sure-thing winner is more attractive than voting for a certain loser)—that skews the dropoff in turnout due to lack of competitiveness to the detriment of the expected loser (see the dashed “Partisan Efficacy” curve in Figure 1).

![Figure 1: Patterns of Voter Turnout by Voter Motivation (Supporters of Party P)](image)

**DATA AND VARIABLES**

In order to test these hypotheses, we used all available presidential general election public opinion polls conducted in all states from 2000 to 2012. While looking at the Roper Center for Public Opinion Research and other sources for the polls essential for our study, we found the
FiveThirtyEightPolitics website as the one with the most comprehensive polls dataset. We wished to gather all polls for all states and all years, however these polls were not available to us in this dataset. While gathering the polls, our focus has been on the US presidential elections in all states. In some states, there were multiple polls from the same company or others while in others one or no poll was conducted. If there were multiple polls available, we took the mean score for a particular state. All polls gathered were conducted the final three weeks of the campaign. For a particular state, the polls on the same date or close-by date were collected and then their mean score was calculated. The names of the polling agencies and their dates can be accessed at the FiveThirtyEightPolitics website.4

Our first dependent variable is turnout. To reiterate, we are not interested in explaining turnout per se; rather, we are looking for systematic patterns in turnout (and results) that can be attributed to an information effect of surveys. To wit, hypotheses 2, 4, and 6 provide competing predictions—based on different presumptions about what motivates voters to get to the polls—about variations in turnout as a consequence of electoral surveys. For these hypotheses, the independent variable of interest is the predicted advantage to one candidate or the other in terms of margins, i.e., the absolute difference between the predicted vote percentages for major-party presidential candidates. To reiterate: if voters are interested simply in piling onto the winner’s bandwagon, we should expect turnout to increase with predicted margins of victory as it gets easier to identify who will win. If voters are interested in affecting the outcome of the election, by contrast (the efficacy hypothesis), turnout should be highest when predicted margins are lowest. Hypotheses 4 and 6 differ only in that hypothesis 6 predicts that the decline in turnout should level off as supporters of the predicted winner heed the call to vote to protect their

victory, while supporters of the predicted loser stay home. Hypothesis 4 predicts (implausibly) that at the limit where one candidate is predicted to win 100 percent of the vote no one should show up at the polls at all.

Our second dependent variable is the percentage of the votes for a major-party Presidential candidate. Note again that we are not interested here in explaining one or the other party’s vote share, but rather whether and how vote share is affected by pre-election surveys. To this end, the independent variable of interest is expected vote share for the Republican or Democratic candidate. In our estimations we use results for Democrats, but the analogous results for Republicans are identical. The partisan efficacy hypothesis suggests that polls should be biased against winners, so that whatever party’s candidate is predicted to win should do better in the actual election than in the polls. This effect should be greater the larger the predicted margin of victory, with the relevant observable implication being that election polls should in hindsight look skewed in favor of the eventual losers—just as, at least in fivethirtyeight’s estimation, happened in the 2014 US Congressional elections and the May parliamentary elections in the UK. The independent variable of interest here is the party in question’s predicted vote percentage in the polls.

RESULTS

Our argument boils down to a simple statement built on xx assumptions. The assumptions are three: first, that polls are accurate at the moment they are taken; second, that individual voters do not switch their support from one candidate in polls to another at election time; and third, that voters use information from polls in deciding whether to bother voting at all. It follows that predictions from election surveys affect election outcomes by influencing voters’ decisions whether to turn out and vote. The argument thus suggests a two-stage model, which has the
added advantage of addressing both bias in the turnout variable and general endogeneity concerns.

For the first stage, predicting turnout, we instrument using a number of usual-suspect variables identified in Geys’ (2006) meta-analysis of the determinants of turnout (Table 1). Our first stage thus uses population, home ownership, population density, and lagged turnout, all variables found in the surveyed literature to be important predictors of turnout. We also add, in line with the argument spelled out above, the absolute value of the survey-predicted winning candidate’s margin of victory. For the second stage of our panel time-series, two-stage least-squares model we interact our predicted turnout with the two-party vote share (as predicted in pre-election surveys) of the Democratic candidate.5

Table 1: Instrumenting Turnout

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poll Predicted Margin of Victory (Absolute Value)</td>
<td>-0.0729**</td>
<td>0.0355</td>
</tr>
<tr>
<td>Population</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Home ownership</td>
<td>0.3611***</td>
<td>0.3611</td>
</tr>
<tr>
<td>Population Density</td>
<td>0.0017</td>
<td>0.0017</td>
</tr>
<tr>
<td>Turnout (lagged)</td>
<td>0.6287***</td>
<td>0.1275</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.2373</td>
<td>4.185</td>
</tr>
<tr>
<td>R²</td>
<td>0.666</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>129</td>
<td></td>
</tr>
</tbody>
</table>

Note: * p < .1, ** p< .05, *** p < .01

The results of the first stage of the model provide support for the theory that the greater the margin of victory, the lower turnout will be, as the coefficient for poll predicted margin of victory is both negative and significant. This result clearly mitigates against hypotheses 1 (Citizen Duty) and 2 (Bandwagoning), most strongly with respect to the latter given its

5 As noted above, the results are essentially identical for each party.
prediction of a positive relationship between turnout and expected margin of victory. Home ownership and lagged turnout appear not to have a statistically significant effect. The evidence thus far suggests instead that turnout is driven by a desire to affect outcomes—that is, efficacy concerns. On the turnout side, then, hypotheses 4 and 6 therefore are still in the running.

For the second stage of the regression, we use Predicted Democratic Vote Share from electoral surveys and the predicted values from our turnout model to predict the actual vote share for the Democratic candidate at the election (Table 2). The results here are interesting and provide telling support for all three Partisan Efficacy hypotheses (hypotheses 6, 7, and 8). First, the coefficient for Predicted Democratic Vote Share is fairly large, positive, and statistically significant. Of course, by itself this tells us little. To make sense of the results, we turn to the interaction of instrumented turnout and Predicted Democratic Vote Share. Figure 2 below shows a marginal-effects plot of Predicted Democratic Vote Share given turnout (for output, see Appendix). As can be seen in the figure, when turnout is very low the effect of predicted vote share is significantly greater than unity (albeit with a large standard error). As turnout increases, the coefficient for predicted vote share becomes indistinguishable from 1 (at turnout somewhere above 60 percent).

The results from Table 2 and Figure 2 suggest several interesting observations. First, the effect of poll predictions is consequential: at turnout of 55 percent—not far from actual turnout in the post-Nixon era—the candidate predicted to win in the polls winning candidate can expect to see a vote premium of about 14 percent above the prediction. Second, as turnout increases—which, to reiterate, it does when elections are predicted to be more competitive—survey predictions perfectly predict outcomes. This latter point supports the assumption that polls accurately measure outcomes, all else equal. And third and most important, the combination of
low turnout and skewed (toward losers) polls strongly supports the argument that voters use the information from pre-election surveys not so much to determine how they will vote but rather whether to do so.

Table 2: Turnout as a Determinant of Vote Share

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turnout (Instrumented)</td>
<td>0.00231</td>
<td>0.00281</td>
</tr>
<tr>
<td>Predicted Democratic Vote Share</td>
<td>1.4711***</td>
<td>0.32864</td>
</tr>
<tr>
<td>Turnout * Predicted Democratic Vote Share</td>
<td>-0.00599</td>
<td>0.00559</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.19014</td>
<td>0.16489</td>
</tr>
<tr>
<td>R²</td>
<td>0.955</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>128</td>
<td></td>
</tr>
</tbody>
</table>

Note: * p < .1, ** p < .05, *** p < .01

Figure 2: Marginal effects graph of poll predictions on actual results, given turnout

Our results for turnout align nicely with hypotheses for both Efficacy and Partisan Efficacy. The results for the effect of poll predictions on vote outcomes strongly indicate that Efficacy alone is not what motivates voters. Hypothesis 5, which indicates a constant coefficient
of 1 for Predicted Democratic Vote Share, is rejected. Rather, in line with hypotheses 6, 7, and 8, Efficacy plus a desire to ensure the “right” outcome, i.e., partisanship, is what gets voters off the couch and out to the ballot box. Erstwhile voters who expect their preferred candidate to lose stay at home in droves; those who expect to win also might be less inclined to go out and vote, but they do so at a greater rate than their counterparts who expect to lose.

**Conclusion**

The question underlying this paper—do campaign polls matter?—is simple, yet of a kind too often unasked. On its face, polls *should* matter. Campaigns and news organizations pour money into them, and political junkies and observers with more casual interest in politics alike pay attention to what they report. At the same time, however, they seem divorced from the phenomenon they report: like horse-race calling or play-by-play announcing in sport, pre-election polls are treated as outside the action. But are they really removed from the action, no more than a source of entertainment and thrills for the politically aware, or are they something more?

We address this question by looking at polls as sources of information for voters. Rather than focusing on whether surveys (and other information available during campaign season) might affect how people vote or how they view candidates (see, e.g., Dobrzynska et al. 2003; McAllister and Studlar 1991; Navazio 1977; Weaver 1996), we instead examine whether polls affect people’s incentive to vote in the first place. To this end we formulate 8 hypotheses built from various assumptions about why people vote. We test these hypotheses using survey data and state-level election results from US presidential elections from 2000 to 2012 and find strong evidence that voters are influenced by polls—not in whom they vote for, but rather in whether they vote at all.
These findings jibe with suggestions that voters make use of survey information in other contexts. The question of whether early returns from eastern states in the US might affect turnout in the west is a perennial concern, for instance. And in three-candidate races strategic voters might use predicted election results to vote strategically (cf., e.g., Cox 1997; McAllister and Studlar 1991). Where there are only two candidates, strategic voting is impossible; the decision to vote, however, still remains; and where voting is at all costly, making an effort when the outcome is a foregone conclusion makes little sense. The possibility of strategy kicks in with the question of who bothers to vote at all, and the data support our presumption that voters who expect their candidates to win are more likely to turn out to vote than voters whose candidates are predicted to lose.

In all, our evidence suggests that polls are on average accurate with respect to how people would vote (i.e., if the poll were the election it would be spot on), but the existence of polls makes their accuracy for predicting election results suspect. Simply put, the information they provide affects who votes, hence outcomes. Were polls wildly wrong—if they predicted a lopsided victory for a candidate whose “true” support were less than the opposing candidate—our data indicate that they could conceivably manufacture a victory for the “wrong” candidate, or at least a more competitive election than either that predicted or what would have occurred in the absence of polling. More likely, however, polls get the expected competitiveness and probable winner right, and voter response to information from those same polls cements and (in the case of noncompetitive elections) magnifies their predictions.
Table 3. Marginal Effects of Poll-Predicted Two Party Vote Share on Actual Two Party Vote Share, Given Levels of Turnout

| Turnout | dy/dx  | Std. Err. | z    | P>|z| | [95% Conf.Interval] |
|---------|--------|-----------|------|-----|-------------------|
| 40      | 1.23144| .1088458  | 11.31| 0.000| 1.018106   1.444774 |
| 45      | 1.201486| .0828401  | 14.50| 0.000| 1.039122   1.36385 |
| 50      | 1.171532| .0586582  | 19.97| 0.000| 1.056564   1.2865  |
| 55      | 1.141578| .0397782  | 28.70| 0.000| 1.063614   1.219542 |
| 60      | 1.111624| .0358753  | 30.99| 0.000| 1.04131    1.181939 |
| 65      | 1.08167 | .050547   | 21.40| 0.000| .9826      1.180741 |
| 70      | 1.051716| .0733906  | 14.33| 0.000| .9078735   1.195559 |
| 75      | 1.021762| .0988968  | 10.33| 0.000| .8279283   1.215597 |
| 80      | .9918086| .125452   | 7.91 | 0.000| .7459272   1.23769 |

Table 4  Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Two Party Vote Share</td>
<td>179</td>
<td>.4918</td>
<td>.0914</td>
<td>.2537</td>
<td>.7299</td>
</tr>
<tr>
<td>Poll Predicted Two Party Vote Share</td>
<td>178</td>
<td>.4885</td>
<td>.0810</td>
<td>.2554</td>
<td>.6467</td>
</tr>
<tr>
<td>Turnout</td>
<td>179</td>
<td>60.294</td>
<td>7.150</td>
<td>44.2</td>
<td>78.4</td>
</tr>
</tbody>
</table>
References


