Institute for Policy Research
Northwestern University
Working Paper Series

WP-09-09

Probabilistic Polling and Voting in the 2008 Presidential Election: Evidence from the American Life Panel

Adeline Delavande
Department of Economics, Universidade Nova de Lisboa
RAND Corporation

Charles F. Manski
Faculty Fellow, Institute for Policy Research
Professor of Economics
Northwestern University

Version: September, 2009

DRAFT

Abstract

This paper reports new empirical evidence on probabilistic polling, which asks persons to state in percent-chance terms the likelihood that they will vote and for whom. Before the 2008 presidential election, seven waves of probabilistic questions were administered bi-weekly to participants in the American Life Panel. Actual voting behavior was reported after the election. We find that responses to the verbal and probabilistic questions are well-aligned ordinally. Moreover, the probabilistic responses predict actual voting behavior, beyond what is possible using verbal responses alone. The probabilistic responses have more predictive power in early August and the verbal responses have more power in late October. However, throughout the sample period, one can predict voting behavior better using both types of responses than either one alone. Studying the longitudinal pattern of responses, we segment respondents into those who are *consistently pro- Obama, consistently anti-Obama, and undecided/vacillators*. Membership in the consistently pro or anti Obama group is an almost perfect predictor of actual voting behavior, while the undecided/vacillators group has more nuanced voting behavior. We find that treating the ALP as a panel improves predictive power: current and previous polling responses together provide more predictive power than current responses alone.

1. Introduction

Pollsters have long asked persons to verbally express uncertainty about their voting intentions. Consider, for example, the New York Times/CBS News (NYT/CBS) presidential poll. In fall 2008, respondents were asked[1]

V1. How likely is it that you will vote in the 2008 election for President this November -- would you say you will definitely vote, probably vote, probably not vote, or definitely not vote in the election for President?

V2. If the 2008 presidential election were being held today and the candidates were Barack Obama, the Democrat, and John McCain, the Republican, would you vote for Barack Obama or John McCain?

Although question V2 does not explicitly permit a person to express uncertainty, some respondents volunteered that they were undecided. They were then asked this follow up question:

V3. Well as of today, do you lean more toward Barack Obama or more toward John McCain ?

When persons did respond toV2, they were asked this follow up to gauge the certitude of their preference:

V4. Is your mind made up or is it still too early to say for sure?

Probabilistic polling (Manski, 2002) is an alternative to verbal questioning that asks persons to state, in percent-chance terms, the likelihood that they will vote and for whom. The objective is to provide readily interpretable, interpersonally comparable, quantitative measures of the uncertainty that persons perceive about their future voting behavior. Consider, for example, a person who responds to question V1that she will

[1] http://graphics8.nytimes.com/packages/pdf/politics/20080918_POLL.pdf

"probably vote," states "Obama" in response to V2, and declares that "it is too early to say for sure" when asked V4. This person clearly expresses some uncertainty about her future voting behavior, but the verbal questions posed in the NYT/CBS poll permit her to give only a vague sense of her perceptions. In response to probabilistic polling questions, this person might state that she perceives a 75 percent chance of voting and that, conditional on voting, she sees a 60 percent chance that she will vote for Obama. These responses provide a precise report of her voting intentions. They imply that, at the date of the interview, the respondent perceives an unconditional probability of 0.45 that she will vote for Obama.

To see the potential advantages of probabilistic polling, consider the efforts that pollsters now make to classify respondents as *likely/unlikely* voters and as *decided/undecided* in their candidate preference. The specific method varies, but most pollsters divide the electorate into two groups, those who are likely to vote and those who are unlikely. They similarly segment respondents into groups who have decided to support particular candidates and a residual group who remain undecided.

These efforts to classify potential voters are problematic, because there is no way to predict with certainty who will vote and for whom. Considering attempts to define likely voters, the pollster Mark Mellman put it this way in the September 8, 2004 issue of *The Hill*, writing [2]

> "Likely" and "unlikely" are probability statements. A likely voter has, say, an 80 percent chance of voting. An unlikely voter has a 20 percent chance of showing up to the polls. Thus, out of every 100 likely voters, 20 will not show up, while 20 of every 100 unlikely voters will. Polling only likely voters skews the sample, systematically excluding a group that will show up in some meaningful numbers on Election Day.

When Mellman writes that polling only likely voters "skews the sample," he recognizes that persons deemed likely and unlikely to vote may differ systematically in the votes that they will actually cast. This possibility

[2] www.hillnews.com/mellman/090804.aspx.

inevitably makes surveys of likely voters controversial.[3]

It seems evident that pollsters should assign voting probabilities to members of the electorate, rather than classify them as likely/unlikely and decided/undecided. A central objective of election polls is to predict election outcomes. Probabilistic polling provides self-reported voting probabilities. Pollsters may use the responses directly or may combine them with other information to develop their own probabilistic predictions of voting. This paper shows how.

Although the potential advantages of probabilistic polling are transparent, practical experience has been scant. There has been a conventional wisdom among pollsters that respondents to election polls would be unable or unwilling to respond informatively to questions asking for probabilistic predictions of their voting behavior. This conventional wisdom has inhibited conduct of empirical research that might shed light on the matter.

This paper reports new empirical evidence. During the three months before the 2008 presidential election, we administered probabilistic polling questions bi-weekly to participants in the American Life Panel (ALP), the RAND Corporation's ongoing longitudinal internet survey. To familiarize respondents with probabilistic polling, we began with this introduction:[4]

> In this interview, we will ask you questions about the upcoming general election for President of the United States. The presidential election is scheduled for Tuesday, November 4, 2008. Many of the questions ask you to think about the percent chance that something will happen in the future. The percent chance can be thought of as the number of chances out of 100. You can use any number between 0 and 100. For example, numbers like: 2 and 5 percent may be "almost no chance", 20

---

[3] In 2004, the likely-voter model used by the Gallup organization drew particular attention for its possible skew towards Republican voters. See, for example, the October 6, 2004 article of Farhad Manjoo in *Salon* (www.salon.com/tech/feature/2004/10/06/polling/index_np.html) and the retrospective assessment of Traugott (2005). The Gallup model was also criticized for exaggerating the volatility of voter preferences during the 2000 presidential campaign (Erikson, Panagopoulos, and Wlezein, 2004).

[4] This introduction paraphrases one previously used in the Survey of Economic Expectations (SEE) to elicit probabilistic expectations of various future events from respondents. See Section 2 for discussion of SEE.

percent or so may mean "not much chance", a 45 or 55 percent chance may be a "pretty even chance", 80 percent or so may mean a "very good chance", and a 95 or 98 percent chance may be "almost certain."

We then asked

P1. What is the percent chance that you will vote in this year's presidential election?

P2.[5] Barack Obama is the Democratic candidate and John McCain is the Republican candidate. If you do vote in the presidential election , what do you think is the percent chance that you will vote for

Barack Obama (Democrat)___ % John McCain (Republican)___ % Someone else___ %

We also administered verbal questions with wording similar to V1 through V3; our versions of V2 and V3 permitted respondents to express a preference for some candidate other than Obama and McCain. In mid-November, following the election, we returned to respondents and asked whether they had voted and, if so, for whom.

Section 2 places our work in context, referencing the large recent literature on measurement of probabilistic expectations in surveys and the few previous applications of probabilistic polling. Section 3 describes the data we collected from the ALP respondents, whom we interviewed bi-weekly in seven waves prior to the election and then again after the election.

Our analysis begins in Section 4, where we study the data wave by wave, treating the ALP as a traditional poll composed of repeated cross-sectional samples. We find that responses to the verbal and probabilistic questions are well-aligned ordinally. Responses to the verbal questions ordinally predict actual voting behavior. Responses to the probabilistic questions predict voting behavior not just ordinally but

[5] Prior to the Republican convention, question P2 began with the words "Suppose that." As in conventional polls, we randomized the order of the names of the two major candidates when administering the verbal and probabilistic candidate preference questions.

quantitatively as well. These findings establish the face validity of the ALP data and show how probabilistic polling refines our understanding of voting intentions.

Section 5 studies the data as a panel, examining the longitudinal pattern of responses. An important feature of the ALP data is that we can examine how the voting intentions of individual respondents evolve over the three months before the election. With this in mind, we focus on the panel members who responded to question P2 in all seven pre-election waves and who reported their actual voting behavior after the election. We segment respondents into three broad response groups, whom we call *consistently pro-Obama*, *consistently anti-Obama*, and *undecided/vacillators*. We find that membership in the consistently pro or anti Obama group is an almost perfect predictor of actual voting behavior. The undecided/vacillators group has more nuanced voting behavior.

Section 6 addresses what for some readers will be a bottom-line question about probabilistic polling: Do probabilistic responses predict actual voting behavior, beyond what is possible using verbal responses alone. In the ALP context, the answer is clearly positive. We also find that verbal responses have their own predictive power. Thus, one can predict voting behavior better using both types of response than either one alone. If one were required to choose between the two types of question, the probabilistic responses have more predictive power in early August and the verbal responses have more power in late October. Finally, we find that treating the ALP as a panel improves predictive power. Using current and previous responses provides more predictive power than using current responses only.

2. Related Literature

2.1. Survey Research Eliciting Probabilistic Expectations

There long was a conventional wisdom among survey researchers that typical respondents will not or cannot respond informatively to percent-chance questions about future events. Hence, the standard practice was to measure uncertainty verbally. For example, the General Social Survey (GSS) has used this question to elicit respondent perceptions about future job loss (Davis and Smith, 1994):

Thinking about the next twelve months, how likely do you think it is that you will lose your job or be laid off – very likely, fairly likely, not too likely, or not at all likely?

It is instructive to compare the wording of this question with polling question V1. Whereas the GSS uses the four phrases (very likely, fairly likely, not too likely, not at all likely) to express various degrees of certitude, question V1 uses another four phrases (definitely, probably, probably not, definitely not). Responses to the GSS and the NYT/CBS poll do not reveal how respondents interpret these phrases. When different respondents to the GSS state "fairly likely," they may or may not mean the same thing. When a person states "probably" in response to question V1, it is not clear how she would have responded if the GSS phrases had been used instead.

The conventional wisdom began to break down in the early 1990s, particularly among economists who perform survey research. One concern was that verbal questions yield at most ordinal measures of respondent beliefs. Another was that the responses may not be interpersonally comparable. These concerns led to empirical research aiming to assess the viability of probabilistic questioning, using a percent-chance format.

Since the early 1990s, survey researchers have accumulated substantial constructive experience with

probabilistic questions, using them to learn how persons perceive various aspects of their future. Manski (2004) reviews the history of thought in several social and behavioral science disciplines, describes the emergence of the modern empirical literature, summarizes a spectrum of applications, and calls attention to open issues. Delavande and Rohwedder (2008), Hurd (2009), and Delavande, Giné, and McKenzie (2009) review specific aspects of the recent research.

We give here a brief description of the major American platforms for methodological exploration and substantive research, with representative citations to completed empirical studies. Beginning in 1992 and continuing through the present, the longitudinal Health and Retirement Study (HRS) has regularly elicited probabilistic expectations of retirement, bequests, and mortality from multiple cohorts of older Americans (Hurd and McGarry, 1995, 2002; Hurd, Smith, and Zissimopoulos, 2004). From 1994 through 2002, the nationwide Survey of Economic Expectations (SEE) asked repeated cross sections of persons to state the percent chance that they will lose their jobs, have health insurance coverage, or be victims of crime in the year ahead, and also to give their income expectations (Dominitz and Manski, 1997a, 1997b; Manski and Straub, 2000).[6]

From 1997 on, the National Longitudinal Survey of Youth 1997 has periodically queried youth about the chance that they will become a parent, be arrested, or complete schooling in the future (Fischhoff *et al.*, 2000; Dominitz, Manski, and Fischhoff, 2001; Lochner, 2007). Probabilistic expectations of stock market returns have been elicited in several surveys, including SEE, HRS, and the Michigan Monthly Survey (Dominitz and Manski, 2004, 2007; Hurd, 2009). We have learned from these and other surveys that most

---

[6] It is instructive to compare the SEE question on job loss with the one in the GSS and with question P1. The SEE question was

I would like you to think about your employment prospects over the next 12 months. What do you think is the percent chance that you will lose your job during the next 12 months?

This question replaces the four verbal phrases in the GSS with a request for the percent chance of job loss. The format of the question is similar to probabilistic polling question P1.

people have little difficulty, once the concept is introduced to them, using subjective probabilities to express the likelihood they place on future events relevant to their lives.

2.2. Previous Probabilistic Polling Studies

Over the years, occasional researchers have independently suggested versions of probabilistic polling and have conducted exploratory studies. The earliest related work that we are aware of was performed by Meier (1980) and Meier and Campbell (1979), who used a seven-point scale to elicit voting expectations. Maas, Steenbergen, and Saris (1990) analyzed probabilities of voting for particular parties reported by Dutch voters in 1986. Burden (1997) analyzed data collected in Ohio in 1986 and 1988 eliciting probabilities that persons would vote for particular candidates in upcoming state and federal elections. Hoek and Gendall (1993, 1997) elicited voting probabilities in elections in New Zealand. For many reasons, these isolated studies have not sufficed to evaluate the merits of probabilistic polling. Nevertheless, they are instructive in some respects. An Appendix in Manski (2002) discusses in detail the work of Burden (1997) and Hoek and Gendall (1997).

Our work on probabilistic polling began with brief discussions of the idea in Manski (1990, 2000), followed by a more lengthy appraisal in Manski (2002). The last article reported a pilot study of 50 respondents in a Chicago suburb performed in the weeks preceding the year 2000 American presidential election. The questions posed were analogous to P1 and P2. The findings were encouraging but the sample was much too small and idiosyncratic to permit firm conclusions about the relative merits of probabilistic polling and verbal questioning in practice.

3. The ALP Data

The American Life Panel is a national longitudinal survey of persons of age 18 and older, begun by RAND in 2006.[7] Most participants were previously interviewed by the Michigan Monthly Survey, a major American survey of consumer attitudes. The Monthly Survey, which interviews its respondents once or twice at six month intervals, asks outgoing sample members if they would be willing to participate in internet surveys. The ALP recruits participants from those who give any response except "no, certainly not."[8] About ninety percent of the respondents have their own internet access. RAND provides a Web TV to the remaining ten percent.

The ALP sampling process yields a wide spectrum of participants. However, as is typical with internet surveys, respondents over-represent some demographic groups relative to others. Table 1 describes the composition of the 1,814 participants who responded to at least one of our pre-election surveys and to the post-election survey. Relative to the population of the United States, the respondents whose data we analyze are more often female (57 percent), non-Hispanic white (89 percent), middle-aged (41 percent with age 50-64), and college educated (45 percent with 16 or more years of schooling).

Whereas the demographic composition of the ALP clearly differs from that of the population of eligible voters, the panel may more closely approximate the composition of the sub-population who actually vote. Fully 90 percent of the respondents reported in our post-election survey that they had voted for president. This compares with a national turnout estimated to be about 62 percent.[9]

We think that the high turnout reported by APL respondents stems mainly from the fact that the panel

[7] The ALP is documented at https://mmicdata.rand.org/alp/index.php/Main_Page

[8] About 80 percent of the ALP participants have been recruited in this manner. The remaining 20 percent are a "snowball sample," their names having been suggested by current participants.

[9] See http://elections.gmu.edu/Turnout_2008G.html for this estimate of the actual turnout.

over-represents racial/ethnic, age, and schooling groups who vote at a higher rate than the population at large. However, a contributing reason may be overreporting of voting. Belli, Traugott, and Beckmann (2001) compared self-reports of voting on the National Election Survey (NES) with administrative data from voting records. They found that the self-reports of the NES respondents exceeded actual voting rates by 7.9 to 14.2 percent in the period 1964–1990. We conjecture that ALP self-reports are more accurate than NES ones, in part because ALP participants have a long-term attachment to the survey and in part because we queried them about their voting immediately after the election rather than several months later. However, we cannot exclude the possibility that some overreporting may have occurred.

Of the ALP respondents who reported that they voted, 50 percent stated that they voted for Obama, 48 percent for McCain, and 2 percent for another candidate. This makes the voting composition of our sample a bit more Republican than the actual presidential vote, which was 53 percent for Obama, 46 percent for McCain, and 1 percent for others.

Non-representativeness of the population of eligible voters may, for some types of analysis, be a shortcoming of the ALP. However, the ability to interview panel members repeatedly is a strong advantage. Traditional polls are repeated cross-sections, drawing new samples each time they go into the field. Hence, one cannot study the evolution of respondent voting intentions over time. Nor can one compare the voting intentions that persons state before an election with their actual voting behavior. The ALP makes it possible to do all of this.

We administered questions P1 and P2 on seven pre-election waves of the ALP. Each wave began when panel members received an email message asking them to access a web page to respond to a new survey. Participants could respond anytime until the next wave was fielded, about two weeks later. Our pre-election questions appeared on these waves, with the opening date in parentheses: Wave 38 (August 4), Wave 40 (August 18) , Wave 42 (September 2), Wave 44 (September 15), Wave 47 (September 29), Wave 49 (October 13), Wave 51 (October 27). Respondents reported their actual voting behavior soon after the

election, in Wave 52 (November 7).[10]

4. Verbal Responses, Probabilistic Responses, and Voting Behavior

We begin our analysis by comparing verbal and probabilistic voting intentions with each another, and with subsequent voting behavior. Section 4.1 examines the likelihood of voting. Section 4.2 considers candidate preference. Throughout this section we treat the ALP as a traditional poll that surveys repeated cross-sections of the population. Sections 5 and 6 exploit the longitudinal data.

4.1. Voting Likelihood

Table 2 cross-tabulates the responses to questions V1 and P1. Considering each wave of the ALP separately, the table presents data for respondents who answered V1 and P1 and who also reported their actual voting behavior after the election. The response patterns are very similar across waves. For concreteness we focus on Wave 44, which was fielded in the middle of the sampling period.

There are 1,474 respondents who reported their likelihood of voting verbally and probabilistically in wave 44, and whether they actually voted in the post-election wave. Most respondents (81 percent) state that they will definitely vote. Of the others, (10, 3, 5) percent state that they will (probably, probably not, definitely not) vote. Persons who state that they will (definitely, probably, probably not, definitely not) vote respectively report a (99, 73, 23, 3) mean percent chance of voting. The corresponding medians are (100, 80, 20, 0) percent chance of voting. Thus, the table shows a strong ordinal correspondence in the verbal and

[10] Wave 51 ended after the November 4 date of the election. We only use responses submitted through November 3.

probabilistic likelihoods of voting.

Examination of the (0.25, 0.50, 0.75) quantiles of the distribution of response to the percent-chance question shows that the verbal phrases "probably" and "probably not" encompass much more varied quantitative perceptions of voting likelihood than do the phrases "definitely" and "definitely not." The three quantiles of the response distribution shown in Table 1 are (100, 100, 100) for persons who state they will definitely vote and (0, 0, 0) for those who state they will definitely not vote. The three quantiles are (50, 80, 90) for persons who state they will probably vote and (10, 20, 40) for those who state they will probably not vote. Thus, the phrases "definitely" and "definitely not" map well into the extreme percentages 100 and 0. However, the phrases "probably" and "probably not" summarize wide ranges of probabilistic beliefs.

The final column of Table 2 gives the fraction of respondents who report after the election that they actually voted. Of those stating in Wave 44 that they will (definitely, probably, probably not, definitely not) vote, the fractions who later report that they did vote are (0.99, 0.75, 0.23, 0.05). These voting rates are remarkably close to the average and median percent-chances of voting that respondents stated a month and a half before the election. As we explore the data further, we will repeatedly find that the probabilistic responses predict well the actual voting behavior of the ALP respondents.

4.2. Candidate Preference

Table 3 cross-tabulates the responses to questions V2 and P2. Considering each wave of the ALP separately, the table presents data for respondents who answered V2 and P2 and who also reported their actual voting behavior after the election. We again focus on Wave 44 for specificity. The main findings that we discuss here are common to all waves.

We have usable data for 1,461 respondents to Wave 44.[11] Of these, 44 percent state that they would vote for Obama, 47 percent for McCain, and 3 percent for another candidate if the election were held today. Of the 6 percent who choose not to answer V2, (3, 2, 1) percent state that they lean towards (Obama, McCain, someone else).

The table shows a strong ordinal correspondence in the verbal and probabilistic candidate preferences. Persons who state that they would vote for (Obama, McCain, someone else) if the election were held today respectively report a (92, 6, 7) mean and (100, 0, 0) median percent chance of voting for Obama, conditional on voting. Persons who state that they lean towards (Obama, McCain, someone else) respectively report a (57, 40, 16) mean and (50, 50, 0) percent chance of voting for Obama. The last set of results is intriguing, showing that the median respondent leaning to Obama and the median respondent leaning to McCain both place a 50 percent chance on voting for Obama.

The penultimate column of Table 3 gives the fraction of respondents who report after the election that they actually voted. Two features are noteworthy. First, persons who state a definite candidate preference vote more frequently than those who only state that they lean to a candidate. Of the persons who state in Wave 44 that they would vote for (Obama, McCain, someone else) if the election were held today, the fractions who later report voting are (0.94, 0.93, 0.52) percent. The corresponding fractions for those who only lean to a candidate are considerably lower, being (0.86, 0.73, 0.24). The second noteworthy feature, evident in the voting fractions just cited, is that persons who prefer or lean to one of the two major-party candidates vote much more frequently than do those who prefer or lean to another candidate.

The final column gives the fraction of voting respondents who report after the election that they voted for Obama. Almost all those who state a preference for a major-party candidate in Wave 44 later report voting for the preferred candidate. Those who state only that they lean to a candidate tend to vote for

[11] Observations are usable when respondents report their verbal and probabilistic candidate preference in a given wave, and report whether they voted (and if so, for whom) in the post-election wave.

this candidate, but significant minorities vote for another candidate.

Table 4 views candidate preferences from the perspective of probabilistic polling. For each wave of the ALP, the left panel of the table segments respondents by their response to question P2, focusing on the percent chance of voting for Obama. For example, the row marked [70, 80) considers persons who state a 70 to 79 percent chance of voting for Obama, if they do vote. The right panel segments respondents by their unconditional stated likelihood of voting for Obama, given by the product of their responses to P1 and P2. Here, the row marked [70, 80) considers persons for whom the product of the responses to P1 and P2 lies in the interval [70, 80).

Comparing the first and last column of each panel, we find a close correspondence between the subjective voting probabilities that persons express in mid-September and their subsequent voting behavior in early November. The correspondence at the extremes was already apparent in Table 3—the persons who state a 90 to 100 (0 to 10) percent chance of voting for Obama overlap strongly with those who state they would vote for Obama (McCain) if the election were held today. Observe that a larger fraction of respondents report extreme probabilities in wave 51, just prior to the election, than in earlier waves. This indicates that candidate preferences tended to sharpen as the election neared.

The value of probabilistic polling relative to verbal questioning manifests itself in the intermediate rows of Table 4. Whereas verbal questioning only coarsely partitions "undecided" voters into those who lean to one candidate or another, probabilistic polling quantifies the degree of certitude of the candidate preference. A person who states that she leans to Obama might state a 50, 60, or 75 percent chance of voting for Obama. One who states that she leans to McCain might state a 25, 40, or 50 percent chance of voting for Obama. It is important to distinguish persons by the magnitude of their subjective probabilities. Table 4 shows that actual voting behavior has a strong positive association with the stated percent chance of voting for Obama. The table shows occasional exceptions, but this should be expected given the small to moderate sizes of the group with intermediate values of the subjective probabilities.

Consider Wave 44. The right panel shows that, among those who place a 50 percent unconditional chance on voting for Obama, the fraction who later report voting for Obama is 0.56. Of those who place a 60 to 69 (31 to 40) percent chance on voting for Obama, the fraction who do vote for him is 0.77 (0.34).

To provide further perspective, Figure 1 plots a kernel nonparametric regression estimate of the relationship between responses to question P2 and subsequent candidate choice. The figure graphically displays the data summarized in the left panel of Table 4D. The x-axis, which ranges over the interval 0–1, gives a person's response to P2 divided by 100. The y-axis gives the predicted probability of voting for Obama, conditional on voting. Most obviously, the estimated relationship is strongly monotone increasing. A closer look shows that the predicted probability increases slowly for P2-responses below about 0.3, then rises roughly linearly for responses in the range 0.3 – 0.7, after which the predicted probability stays close to one.

Results such as these, which are typical of all waves of the survey, show that the elicited subjective probabilities of voting are not only strong ordinal but quantitative predictors of actual voting behavior.

5. Longitudinal Response Patterns

The ALP data enable us to examine how the voting intentions of individual respondents evolve over the three months before the election. With this in mind, we now focus on the 867 panel members who responded to question P2 in all seven pre-election waves and who reported their actual voting behavior after the election.

Table 5 segments respondents into three broad response groups, with further division into sub-groups. The top group of 362 *consistently pro-Obama* persons state at least a 50 percent chance of voting for Obama in all seven pre-election waves and sometimes state a higher likelihood. The bottom group of 419

consistently anti-Obama persons never report more than a 50 percent chance of voting for Obama and sometimes report less than 50 percent. The middle group of 86 *undecided/vacillators* are neither consistently pro nor anti Obama before the election. We use the hybrid term undecided/vacillators because this group contains at least two types of persons. The word "undecided" describes persons who consistently state close to an even chance of voting for Obama, while "vacillators" describes ones who waver over time between high and low subjective probabilities of voting for Obama.

We find that membership in the consistently pro or anti Obama group is an almost perfect predictor of actual voting behavior. Of the consistently pro-Obama persons who actually vote in the election, voting for Obama is unanimous among those who always state at least a 70 percent chance of doing so and is nearly unanimous (all but one respondent) among those who always state at least a 60 percent chance. Of the 41 respondents who sometimes state a 50 to 59 percent chance of voting for Obama, 38 persons vote and 34 of these vote for Obama.

Symmetrically, persons who are consistently anti-Obama before the election essentially never vote for him. Of those who never state more than a 40 percent chance of voting for Obama, no one votes for him. Of the 56 respondents who sometimes state a 41 to 50 percent chance of voting for Obama, 39 actually vote and just one of them votes for Obama.

This leave the undecided/vacillators. This group is relatively small, comprising about ten percent of the respondents studied in Table 5. However it is potentially highly important because the pro and anti-Obama groups are close in size. If the 2008 election had been held among the ALP respondents, the voting behavior of the undecided/vacillator group would have determined its outcome (59 percent of the voters in the group voted for Obama). It is reasonable to conjecture that the party conventions, presidential debates and other events of the election campaign mainly affected the voting intentions and behavior of this group.

To scrutinize the undecided/vacillator respondents more closely, Table 6 presents the full time-series of responses for all group members other than the one person who reported a 50 percent chance of voting

for Obama on every wave. Inspection of the table makes clear the distinction between undecided persons and vacillators. The term "undecided" clearly fits respondent #5 whose seven responses to question P2 are (50, 60, 40, 20, 70, 60, 60). The term "vacillator" fits respondent #25, whose response sequence is (0, 25, 50, 0, 100, 100, 100).

The considerable heterogeneity of response sequences in Table 6 is intriguing. One would like to understand the thought processes that yield sequences as different as those of respondents #5 and #25. However, the ALP data only measure expectations, not explain them. Explanation would require more intensive data collection, probing respondents about how they form the expectations they report.

6. Using Polling Responses and Respondent Attributes to Predict Voting Behavior

Sections 4 and 5 showed that responses to probabilistic polling questions have considerable power to predict the overall voting behavior of the ALP panel. Responses to verbal questions also have considerable predictive power. Verbal questioning of repeated cross sections has been the norm in polling for many years. One might argue that traditional polling practices should continue unless the responses to probabilistic questions and the longitudinal data in the ALP significantly enhance our ability to predict voting behavior, beyond what is possible with traditional polls.

To address the question of predictive power, Table 7 reports a variety of best linear predictors (BLPs) of voting for Obama among the ALP respondents who actually vote. Each BLP is estimated by least squares and robust standard errors are presented beneath the parameter estimates. The bottom row of the table gives the root mean square error (RMSE) of the predictions made with the estimated predictor

function.[12] We also estimated binary logit models. The pattern of findings is similar, so we do not present them here.

In principle, we would like to perform a fully nonparametric analysis, allowing the verbal and probabilistic responses to interact flexibly with one another and with respondent attributes. However, such an analysis would be very complex to interpret. We will, however, briefly discuss kernel nonparametric regression estimates of the type shown in Figure 1.

Sub-tables A, B, and C respectively use as predictors the polling data obtained in Wave 38 alone, in Waves 38 and 44, and in Waves 38, 44, and 51. Thus, Table 7A predicts voting behavior using polling data available in early August. Table 7B uses data available in early August and mid-September. Table 7C uses data available in early August, mid-September, and late October.[13]

The left panel of each sub-table treats the ALP as a repeated cross-sectional survey, as we did in Section 4. This panel uses as predictors only the verbal and probabilistic responses obtained in the current wave. The middle and right panels of Tables 7B and 7C treat the ALP as a panel, uses the polling responses obtained in the current and earlier waves. The right panel of each sub-table adds respondent attributes as predictors of voting behavior.

---

[12] The RMSE is defined as follows. Suppose that an estimation sample contains J respondents. Let $y_j = 1$ if person j votes for Obama and $y_j = 0$ otherwise. Let $p(x_j)$ be a real-valued function predicting the person's vote, using polling-response and attribute data contained in the covariates x_j. The mean square error of the predictor is the group average of the squared prediction errors $[y_j - p(x_j)]^2$. The RMSE is the square root of the mean square error.

[13] There are many variants on Table 7 that we could produce, but we choose not to for reasons of space and ease of exposition. We could produce tables that use all seven waves of ALP data, but they would be more cumbersome to present and would only add marginal information. Instead of taking the outcome to be voting for Obama, we could take it to be a trinomial variable indicating whether a person votes for Obama, McCain, or Someone Else. We see no need to use this more complex outcome because a negligible fraction of the respondents vote for Someone Else. We could produce tables that predict voting for Obama among all ALP respondents, including those who do not actually vote. Predicting voting among all respondents would be more cumbersome, because we would need to include as predictor variables the verbal and probabilistic responses on likelihood of voting. We choose not to do this because the great majority of respondents vote, making turnout a minor issue in the ALP data. Thus, we think that Table 7 suffices to show the main results of interest.

Examination of the table reveals multiple striking findings, which we discuss below.

Predictive Power of Probabilistic Polling

The responses to the probabilistic questions clearly enhance one's ability to predict voting behavior, beyond what is possible using verbal responses alone. In every prediction scenario, the predicted probability of voting for Obama increases substantially with the respondent's probabilistic preference for Obama given in response to question P2. Moreover, the parameter estimates are statistically precise. The probabilistic responses have predictive power even though the responses to the verbal questions V2 and V3 are also used as predictors throughout. Addition of respondent attributes as predictors does not diminish their predictive power at all.

The RMSE measures the average accuracy of the predictions. For concreteness, consider the predictions made using the ALP as a repeated cross-section. The RMSEs in Waves 38, 44, and 51are 0.220, 0.209, and 0.136 respectively. Suppose that one uses only the verbal responses as predictors, rather than the verbal and probabilistic responses together. Then the corresponding RMSEs for the three waves are 0.238, 0.217, and 0.141.[14] Thus, using both responses as predictors significantly reduces prediction errors.

Observe that, whether all polling responses or only the probabilistic ones are used as predictors, predictive power increased moderately from early August to mid-September and then dramatically by late October. The qualitative finding of increased predictive power as the election nears is not surprising, but we think it useful to quantify the magnitude of the change through the RMSE.

Predictive Power of Verbal Polling

The responses to the verbal questions have their own predictive power, beyond what one achieves using the probabilistic responses alone. This finding holds across scenarios and the estimates are usually

[14] The predictor estimates are not shown in Table 7, but are available from the authors.

statistically precise. Continuing the comparison of RMSEs when the ALP is used as a repeated cross-section, if one uses only the probabilistic responses as linear predictors, the RMSEs for the three waves are 0.234, 0.227, and 0.164.[15] Thus, neither the verbal nor the probabilistic responses are sufficient statistics for one another. Both contribute when performing linear prediction of voting behavior.

It is a necessary caveat to state that the verbal responses contribute to *linear* prediction. Figure 1 showed that the association between responses to question P2 and voting choices is nonlinear, having a small positive slope for low response values (0–0.3), a large positive slope in the mid-range (0.3– 0.7), and essentially zero slope for responses values above 0.7. Using the probabilistic responses as linear predictors does not permit this nonlinearity to express itself and, hence, undervalues the predictive power of the probabilistic responses.

Suppose that one uses only the probabilistic responses as predictors and computes kernel nonparametric regression estimates. The RMSEs for the three waves turn out to be 0.225, 0.219, and 0.155. These results substantially improve those obtained using the probabilistic responses as linear predictors.

Choosing Between Probabilistic and Verbal Polling

It is revealing to compare the RMSEs obtained using the probabilistic responses alone with those obtained using the verbal responses alone. Suppose that one were required to choose between the two types of question. We find that the probabilistic responses yield more accurate predictions in Wave 38 (RMSE 0.225 vs. 0.238) and the verbal responses in Wave 51 (RMSE 0.141 vs. 0.155). The two predictors are essentially equally accurate in the intermediate Wave 44 (RMSEs 0.219 and 0.217).

These results indicate that in 2008 probabilistic polling was more informative prior to mid-September and verbal questioning thereafter. We did not anticipate observing this pattern but it seems reasonable ex post. The strength of probabilistic polling is that it enables respondents to fully express uncertainty about

[15] The predictor estimates are available from the authors.

their voting intentions. We have earlier found that respondent uncertainty was greatest early in the 2008 campaign and lessened as the election neared.[16]

Predictive Power of Panel Data

We find that treating the ALP as a panel rather than as a repeated cross section has predictive power. Tables 7B and 7C show that using a person's current and previous polling responses to predict voting behavior enables more accurate predictions than using the current responses alone. Whereas the RMSEs in Waves 44 and 51 using only current-wave polling responses are 0.209 and 0.136, they decrease to 0.195 and 0.131 when current and previous-wave responses are used.

Predictive Power of Respondent Attributes

The respondent attributes add essentially no predictive power beyond that in the polling responses. Recall that Table 1 showed considerable variation in voting behavior by gender, age, and schooling when these attributes were used as predictors without regard to polling responses. Table 7 shows that a person's polling responses are a sufficient statistic for his or her attributes when computing best linear predictors. The parameter estimates for the respondent attributes are close to zero. Using the attributes as predictors leaves the RMSEs of prediction essentially unchanged, reducing them by only 0.001 each wave.

Again, it is a necessary caveat to confine this conclusion to linear prediction. It may be that personal attributes have some predictive power when interacted with the polling responses rather than used additively as in the specifications of Table 7.

---

[16] Further evidence on the temporal pattern of predictive power appears in the parameter estimates of Table 7. Focus on the left panel, which is most straightforward to interpret. The coefficient on "Probability of Voting for Obama" weakens over time, from 0.647 in Wave 38 to 0.454 in Wave 44 to 0.351 in Wave 51. The coefficients on a verbal preference for McCain correspondingly strengthen over time.

7. Conclusion

We see this paper as making important contributions to the design and analysis of election polls. We have performed the first large scale nationwide application of probabilistic polling to a presidential election. The ALP enables longitudinal analysis of voting intentions and realizations, which has no precedent.

Our comparison of the verbal and probabilistic responses should remove any skepticism about the viability of probabilistic polling. We have found that probabilistic and verbal response both contribute significantly to prediction of voting behavior. These results are highly encouraging for future applications.

The main open issue concerns implementation of probabilistic polling in national surveys that are more representative of the entire electorate. One may ask whether the broad diversity of Americans would be as comfortable with probabilistic polling as we have found the ALP respondents to be. We cannot answer this question with certainty, but we can point to the HRS and SEE as relevant evidence. These surveys have successfully administered many probabilistic expectations questions to broad populations in telephone and face-to-face surveys. Probabilistic polling questions are simple to administer and understand, indeed more so than some of the questions asked in the HRS and SEE. Hence, we think it reasonable to move forward.

References

Belli, R., M. Traugott, and M. Beckmann (2001), "What Leads to Voting Overreports? Contrasts of Overreports to Validated Voters and Admitted Nonvoters in the American National Election Studies," *Journal of Official Statistics*, 17, 479-498.

Burden, B. (1997), "Deterministic and Probabilistic Voting Models," *American Journal of Political Science*, 41, 1150-1169.

Delavande, A. and S. Rohwedder (2008), "Eliciting Subjective Probabilities in Internet Surveys," *Public Opinion Quarterly*, 72, 866-891.

Delavande, A., X. Giné, and D. McKenzie (2009), "Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence," World Bank Research Working Paper No. 4824.

Dominitz, J. and C. Manski (1997a), "Perceptions of Economic Insecurity: Evidence from the Survey of Economic Expectations," *Public Opinion Quarterly*, 61, 261-287.

Dominitz, J. and C. Manski (1997b), "Using Expectations Data to Study Subjective Income Expectations," *Journal of the American Statistical Association*, 92, 855-867.

Dominitz, J. and C. Manski (2004), "How Should We Measure Consumer Confidence?" *Journal of Economic Perspectives*, 18, 51-66.

Dominitz, J. and C. Manski (2007), "Expected Equity Returns and Portfolio Choice: Evidence from the Health and Retirement Study," *Journal of the European Economic Association*, 5, 369-379.

Dominitz, J, C. Manski, and B. Fischhoff (2001), "Who are Youth *At-Risk*?: Expectations Evidence in the NLSY-97," in R. Michael (editor), *Social Awakenings: Adolescents' Behavior as Adulthood Approaches*, New York: Russell Sage Foundation, 230-257.

Erikson, R., C. Panagopoulos, and C. Wlezien (2004), "Likely (and Unlikely) Voters and the Assessment of Campaign Dynamics," *Public Opinion Quarterly*, 68, 588-601.

Fischhoff, B., A. Parker, W. de Bruin, J. Downs, C. Palmgren, R. Dawes, and C. Manski (2000), "Teen Expectations for Significant Life Events," *Public Opinion Quarterly*, 64, 189-205.

Hoek, J. and P. Gendall (1993), "A New Method of Predicting Voting Behaviour," *Journal of the Market Research Society*, 35, 361-371.

Hoek, J. and P. Gendall (1997), "Factors Affecting Poll Accuracy: An Analysis of Undecided Respondents," *Marketing Bulletin*, 8, 1-14.

Hurd, M. and K. McGarry (1995), "Evaluation of the Subjective Probabilities of Survival in the Health and Retirement Study," *Journal of Human Resources*, 30, S268-S292.

Hurd, M. and K. McGarry (2002), "The Predictive Validity of Subjective Probabilities of Survival," *The*

Economic Journal, 112, 966-985.

Hurd, M., J. Smith, and J. Zissimopoulos (2004), "The Effects of Subjective Survival on Retirement and Social Security Claiming." *Journal of Applied Econometrics,* 19, 761-775.

Hurd, M. (2009), "Subjective Probabilities in Household Surveys," *Annual Review of Economics*, 1, 543-564.

Lochner, L. (2007), "Individual Perceptions of the Criminal Justice System," *American Economic Review*, 97, 440-460.

Maas, K., M. Steenbergen, and W. Saris (1990), "Vote Probabilities," *Electoral Studies* 9, 91-107.

Manski, C. (1990), "The Use of Intentions Data to Predict Behavior: A Best Case Analysis," *Journal of the American Statistical Association* 85, 934-940.

Manski, C. (2000), "Why Polls Are Fickle," Op-Ed article, *The New York Times*, October 16, A27.

Manski, C. (2002), "Probabilistic Polling," in J. Manza, F. Cook, and B. Page, editors, *Navigating Public Opinion: Polls, Policy, and the Future of American Democracy*, Oxford: Oxford University Press, 251-271.

Manski, C. (2004) , "Measuring Expectations," *Econometrica*, 72, 1329-1376.

Manski, C. and J. Straub (2000), "Worker Perceptions of Job Insecurity in the Mid-1990s: Evidence from the Survey of Economic Expectations," *Journal of Human Resources* 35, 447-479.

Meier, K. (1980), "Rationality and Voting: A Downsian Analysis of the 1972 Election," *Western Political Quarterly*, 33, 38-49.

Meier, K. and J. Campbell (1979), "Issue Voting: An Empirical Examination of Individually Necessary and Jointly Sufficient Conditions," *American Political Quarterly*, 7, 21-50.

Traugott, M. (2005), "The Accuracy of the National Preelection Polls in the 2004 Presidential Election," *Public Opinion Quarterly*, 69, 642-654.

TABLE 1: RESPONDENT ATTRIBUTES AND VOTING

	Sample Size	Percent of Sample	Fraction who vote and report candidate	Fraction of voters who vote for Obama	Fraction of voters who vote for McCain	Fraction of voters who vote for other candidate
all persons	1,814	100	0.90	0.50	0.48	0.02
male	788	43	0.90	0.45	0.53	0.03
female	1,026	57	0.89	0.54	0.44	0.02
non-Hisp. white	1,611	89	0.90	0.46	0.51	0.02
non-Hisp. black	110	6	0.90	0.98	0.02	0.00
Hispanic	60	3	0.85	0.53	0.45	0.02
other	33	2	0.88	0.52	0.41	0.07
age 18-34	300	17	0.75	0.51	0.46	0.03
age 35-49	495	27	0.90	0.46	0.52	0.02
age 50-64	735	41	0.92	0.52	0.45	0.03
age 65+	284	16	0.97	0.48	0.50	0.01
schooling 0-12	337	19	0.71	0.46	0.52	0.02
schooling 13-15	653	36	0.90	0.43	0.55	0.02
schooling 16+	824	45	0.97	0.56	0.42	0.02

TABLE 2: VERBAL AND PERCENT-CHANCE LIKELIHOOD OF VOTING

Verbal Response	Sample Size	Percent of Sample	Percent-Chance Response				Fraction who vote
			Mean	0.25 quantile	median	0.75 quantile	
WAVE 38							
definitely	1,064	82	99	100	100	100	0.99
probably	125	10	73	50	80	90	0.81
probably not	52	4	25	5	20	50	0.31
definitely not	64	5	2	0	0	0	0.05
Total sample	1,305	100	89	98	100	100	0.90
WAVE 40							
definitely	1,060	81	99	100	100	100	0.99
probably	136	10	76	60	80	90	0.81
probably not	50	4	23	5	20	30	0.30
definitely not	58	4	0	0	0	0	0.02
Total sample	1,304	100	89	98	100	100	0.90
WAVE 42							
definitely	1,134	83	99	100	100	100	0.99
probably	127	9	73	50	80	90	0.76
probably not	36	3	28	10	25	50	0.25
definitely not	73	5	5	0	0	0	0.07
Total sample	1,370	100	90	99	100	100	0.90
WAVE 44							
definitely	1,211	82	99	100	100	100	0.99
probably	151	10	73	50	80	90	0.75
probably not	39	3	23	10	20	40	0.23
definitely not	73	5	3	0	0	0	0.05
Total sample	1,474	100	90	99	100	100	0.90
WAVE 47							
definitely	1,245	82	99	100	100	100	0.99
probably	142	9	73	50	80	90	0.73
probably not	50	3	22	5	15	40	0.28
definitely not	75	5	4	0	0	0	0.07
Total sample	1,512	100	90	99	100	100	0.90
WAVE 49							
definitely	1,225	84	99	100	100	100	1.00
probably	118	8	74	50	80	95	0.70
probably not	38	3	19	0	20	30	0.21
definitely not	78	5	4	0	0	0	0.05
Total sample	1,459	100	90	99	100	100	0.90
WAVE 51							
definitely	1,308	85	100	100	100	100	0.99
probably	108	7	75	55	80	90	0.72
probably not	35	2	25	10	20	50	0.23
definitely not	92	6	2	0	0	0	0.05
Total sample	1,543	100	90	99	100	100	0.90

TABLE 3: VERBAL AND PROBABILISTIC CANDIDATE PREFERENCE

Verbal Response	Sample Size	Percent of Sample	Percent-Chance of Voting for Obama Conditional on voting				Voting Behavior	
			Mean	0.25 quantile	median	0.75 quantile	Fraction voting	Fraction voting for Obama, conditional on voting
WAVE 38								
Obama	561	43	91	90	100	100	0.95	0.96
Lean to Obama	33	3	53	50	50	60	0.85	0.79
Lean to McCain	47	4	35	20	50	50	0.81	0.21
McCain	559	43	6	0	0	0	0.93	0.05
Someone else	56	4	9	0	0	18	0.68	0.18
Lean to someone else	34	3	21	0	12.5	45	0.32	0.27
Total sample	1,290	100	46	0	40	100	0.91	0.50
WAVE 40								
Obama	565	44	92	90	100	100	0.95	0.96
Lean to Obama	33	3	56	50	50	60	0.91	0.73
Lean to McCain	37	3	36	25	42	50	0.78	0.14
McCain	576	44	6	0	0	1	0.93	0.06
Someone else	54	4	8	0	0	4	0.56	0.27
Lean to someone else	30	2	16	0	7.5	25	0.37	0.18
Total sample	1,295	100	46	0	40	100	0.91	0.50
WAVE 42								
Obama	584	43	92	90	100	100	0.94	0.96
Lean to Obama	40	3	56	50	50	60	0.82	0.70
Lean to McCain	34	3	31	15	34	50	0.71	0.17
McCain	626	46	6	0	0	0	0.93	0.06
Someone else	47	3	11	0	0	5	0.57	0.30
Lean to someone else	23	2	12	0	0	25	0.26	0.33
Total sample	1,354	100	45	0	40	100	0.90	0.49
WAVE 44								
Obama	639	44	92	95	100	100	0.94	0.97
Lean to Obama	37	3	57	50	50	60	0.86	0.72
Lean to McCain	33	2	40	35	50	50	0.73	0.21
McCain	681	47	6	0	0	0	0.93	0.05
Someone else	50	3	7	0	0	2	0.52	0.19
Lean to someone else	21	1	16	0	0	40	0.24	0.40
Total sample	1,461	100	46	0	40	100	0.90	0.49

Verbal Response	Sample Size	Percent of Sample	Percent-Chance of Voting for Obama Conditional on voting				Voting Behavior	
			Mean	0.25 quantile	median	0.75 quantile	Fraction voting	Fraction voting for Obama, conditional on voting
WAVE 47								
Obama	679	45	93	95	100	100	0.94	0.98
Lean to Obama	36	2	51	50	50	55	0.78	0.64
Lean to McCain	34	2	37	30	49	50	0.85	0.34
McCain	664	44	6	0	0	0	0.93	0.03
Someone else	58	4	8	0	0	5	0.52	0.30
Lean to someone else	29	2	22	0	2	50	0.38	0.00
Total sample	1,500	100	47	0	49	100	0.90	0.50
WAVE 49								
Obama	686	47	93	95	100	100	0.94	0.96
Lean to Obama	25	2	51	50	50	55	0.72	0.50
Lean to McCain	30	2	37	25	47	50	0.70	0.10
McCain	616	43	4	0	0	0	0.94	0.02
Someone else	57	4	8	0	0	1	0.56	0.09
Lean to someone else	31	2	18	0	0	50	0.26	0.25
Total sample	1,445	100	48	0	50	100	0.91	0.50
Wave 51								
Obama	718	47	95	100	100	100	0.94	0.98
Lean to Obama	16	1	56	50	50	50	0.62	0.70
Lean to McCain	22	1	40	20	50	50	0.82	0.06
McCain	687	45	4	0	0	0	0.93	0.01
Someone else	71	5	4	0	0	0	0.54	0.13
Lean to someone else	24	2	19	0	10	40	0.29	0.00
Total sample	1,538	100	48	0	43	100	0.90	0.50

TABLE 4A: PROBABILISTIC RESPONSES IN WAVE 38

Actual Voting by Conditional Likelihood of voting for Obama

Percent chance of voting for Obama, if you do vote	Sample Size	Percent of Sample	Average Stated Likelihood of Voting	Fraction who actually vote and Report Candidate	Fraction who vote for Obama, conditional on voting
[90, 100]	440	34	95	0.97	0.99
[80, 90)	29	2	86	0.86	0.92
[70, 80)	37	3	91	0.92	0.94
[60, 70)	33	3	92	0.94	0.84
[51, 60)	11	1	83	0.82	0.67
50	81	6	79	0.81	0.56
(40, 50)	13	1	84	0.85	0.55
(30, 40]	33	3	94	0.85	0.25
(20, 30]	31	2	85	0.81	0.12
(10, 20]	33	3	78	0.85	0.00
[0, 10]	549	43	87	0.88	0.04
total sample	1,290	100	89	0.91	0.50

Actual Voting by Unconditional Likelihood of Voting for Obama

Percent chance of voting for Obama	Sample Size	Percent of Sample	Fraction who vote for Obama
[90, 100]	398	31	0.98
[80, 90)	33	3	0.97
[70, 80)	29	2	0.90
[60, 70)	37	3	0.86
[51, 60)	17	1	0.59
50	50	4	0.58
(40, 50)	23	2	0.48
(30, 40]	40	3	0.38
(20, 30]	43	3	0.26
(10, 20]	33	3	0.03
[0, 10]	586	45	0.03
total sample	1,289	100	0.45

TABLE 4B: PROBABILISTIC RESPONSES IN WAVE 40

Actual Voting by Conditional Likelihood of voting for Obama

Percent chance of voting for Obama, if you do vote	Sample Size	Percent of Sample	Average Stated Likelihood of Voting	Fraction who vote	Fraction who vote for Obama, conditional on voting
[90, 100]	458	34	96	0.97	0.98
[80, 90)	39	3	85	0.85	0.94
[70, 80)	30	2	92	0.93	0.89
[60, 70)	40	3	88	0.90	0.89
[51, 60)	7	1	98	1.00	0.43
50	67	5	79	0.79	0.55
(40, 50)	16	1	91	0.88	0.50
(30, 40]	44	3	86	0.86	0.34
(20, 30]	46	3	79	0.80	0.22
(10, 20]	24	2	76	0.79	0.00
[0, 10]	566	42	88	0.89	0.03
total sample	1,337	100	90	0.91	0.49

Actual Voting by Unconditional Likelihood of Voting for Obama

Percent chance of voting for Obama	Sample Size	Percent of Sample	Fraction who vote for Obama
[90, 100]	417	31	0.98
[80, 90)	38	3	0.89
[70, 80)	35	3	0.91
[60, 70)	33	2	0.88
[51, 60)	16	1	0.56
50	40	3	0.52
(40, 50)	28	2	0.46
(30, 40]	42	3	0.38
(20, 30]	51	4	0.22
(10, 20]	32	2	0.19
[0, 10]	604	45	0.03
total sample	1,336	100	0.45

TABLE 4C: PROBABILISTIC RESPONSES IN WAVE 42

Actual Voting by Conditional Likelihood of voting for Obama

Percent chance of voting for Obama, if you do vote	Sample Size	Percent of Sample	Average Stated Likelihood of Voting	Fraction who vote	Fraction who vote for Obama, conditional on voting
[90, 100]	465	34	96	0.96	0.99
[80, 90)	40	3	84	0.85	0.97
[70, 80)	30	2	90	0.93	0.93
[60, 70)	28	2	80	0.89	0.72
[51, 60)	7	1	99	1.00	0.57
50	71	5	80	0.79	0.62
(40, 50)	16	1	88	0.94	0.60
(30, 40]	34	3	82	0.85	0.38
(20, 30]	40	3	87	0.82	0.12
(10, 20]	36	3	81	0.81	0.24
[0, 10]	587	43	89	0.89	0.02
total sample	1,354	100	90	0.90	0.49

Actual Voting by Unconditional Likelihood of Voting for Obama

Percent chance of voting for Obama	Sample Size	Sample Perncent	Fraction who vote for Obama
[90, 100]	426	31	0.99
[80, 90)	39	3	0.95
[70, 80)	35	3	0.89
[60, 70)	21	2	0.76
[51, 60)	17	1	0.71
50	46	3	0.57
(40, 50)	22	2	0.59
(30, 40]	30	2	0.40
(20, 30]	52	4	0.17
(10, 20]	43	3	0.23
[0, 10]	623	46	0.02
total sample	1,354	100	0.44

TABLE 4D: PROBABILISTIC RESPONSES IN WAVE 44

Actual Voting by Conditional Likelihood of voting for Obama

Percent chance of voting for Obama, if you do vote	Sample Size	Percent of Sample	Average Stated Likelihood of Voting	Fraction who vote	Fraction who vote for Obama, conditional on voting
[90, 100]	512	35	96	0.96	0.98
[80, 90)	37	3	86	0.92	0.91
[70, 80)	31	2	90	0.90	1.00
[60, 70)	34	2	86	0.94	0.78
[51, 60)	13	1	98	1.00	0.92
50	74	5	78	0.77	0.53
(40, 50)	16	1	89	0.88	0.71
(30, 40]	42	3	83	0.86	0.31
(20, 30]	41	3	84	0.85	0.23
(10, 20]	31	2	76	0.74	0.09
[0, 10]	631	43	89	0.88	0.02
total sample	1,462	100	90	0.90	0.49

Actual Voting by Unconditional Likelihood of Voting for Obama

Percent chance of voting for Obama	Sample Size	Percent of Sample	Fraction who vote for Obama
[90, 100]	470	32	0.97
[80, 90)	37	3	0.97
[70, 80)	34	2	0.91
[60, 70)	30	2	0.77
[51, 60)	20	1	0.85
50	43	3	0.56
(40, 50)	29	2	0.66
(30, 40]	35	2	0.34
(20, 30]	57	4	0.23
(10, 20]	37	3	0.14
[0, 10]	670	46	0.02
total sample	1,462	100	0.45

TABLE 4E: PROBABILISTIC RESPONSES IN WAVE 47

Actual Voting by Conditional Likelihood of voting for Obama

Percent chance of voting for Obama, if you do vote	Sample Size	Percent of Sample	Average Stated Likelihood of Voting	Fraction who vote	Fraction who vote for Obama, conditional on voting
[90, 100]	553	37	96	0.96	0.99
[80, 90)	35	2	85	0.86	0.97
[70, 80)	32	2	94	0.94	0.87
[60, 70)	33	2	81	0.88	0.90
[51, 60)	16	1	93	0.94	0.73
50	79	5	76	0.77	0.46
(40, 50)	21	1	90	0.90	0.47
(30, 40]	40	3	88	0.92	0.19
(20, 30]	35	2	82	0.86	0.17
(10, 20]	28	2	81	0.75	0.10
[0, 10]	630	42	88	0.88	0.02
total sample	1,502	100	90	0.90	0.50

Actual Voting by Unconditional Likelihood of Voting for Obama

Percent chance of voting for Obama	Sample Size	Percent of Sample	Fraction who vote for Obama
[90, 100]	507	34	0.99
[80, 90)	32	2	0.97
[70, 80)	41	3	0.80
[60, 70)	28	2	0.89
[51, 60)	22	1	0.77
50	42	3	0.52
(40, 50)	36	2	0.47
(30, 40]	41	3	0.17
(20, 30]	40	3	0.20
(10, 20]	40	3	0.12
[0, 10]	672	45	0.02
total sample	1,501	100	0.45

TABLE 4F: PROBABILISTIC RESPONSES IN WAVE 49

Actual Voting by Conditional Likelihood of voting for Obama

Percent chance of voting for Obama, if you do vote	Sample Size	Percent of Sample	Average Stated Likelihood of Voting	Fraction who vote	Fraction who vote for Obama, conditional on voting
[90, 100]	559	39	97	0.97	0.99
[80, 90)	27	2	82	0.78	0.90
[70, 80)	39	3	94	0.95	0.89
[60, 70)	28	2	80	0.82	0.83
[51, 60)	17	1	94	0.94	0.69
50	66	5	65	0.65	0.40
(40, 50)	11	1	100	1.00	0.36
(30, 40]	24	2	74	0.79	0.05
(20, 30]	34	2	77	0.74	0.08
(10, 20]	21	1	90	0.90	0.11
[0, 10]	619	43	90	0.89	0.02
total sample	1,445	100	91	0.91	0.50

Actual Voting by Unconditional Likelihood of Voting for Obama

Percent chance of voting for Obama	Sample Size	Percent of Sample	Fraction who vote for Obama
[90, 100]	527	36	0.98
[80, 90)	23	2	0.87
[70, 80)	38	3	0.89
[60, 70)	25	2	0.80
[51, 60)	23	2	0.70
50	29	2	0.38
(40, 50)	20	1	0.40
(30, 40]	26	2	0.23
(20, 30]	41	3	0.15
(10, 20]	28	2	0.11
[0, 10]	665	46	0.02
total sample	1,445	100	0.45

TABLE 4G: PROBABILISTIC RESPONSES IN WAVE 51

Actual Voting by Conditional Likelihood of voting for Obama

Percent chance of voting for Obama, if you do vote	Sample Size	Percent of Sample	Average Stated Likelihood of Voting	Fraction who vote	Fraction who vote for Obama, conditional on voting
[90, 100]	632	41	97	0.97	0.99
[80, 90)	25	2	81	0.80	1.00
[70, 80)	22	1	86	0.86	0.95
[60, 70)	21	1	77	0.76	0.69
[51, 60)	10	1	90	0.80	0.88
50	55	4	61	0.62	0.26
(40, 50)	4	0	100	1.00	0.50
(30, 40]	27	2	79	0.78	0.24
(20, 30]	18	1	84	0.78	0.07
(10, 20]	26	2	83	0.85	0.05
[0, 10]	698	45	88	0.89	0.01
total sample	1,538	100	91	0.90	0.50

Actual Voting by Unconditional Likelihood of Voting for Obama

Percent chance of voting for Obama	Sample Size	Percent of Sample	Fraction who vote for Obama
[90, 100]	597	39	0.99
[80, 90)	27	2	1.00
[70, 80)	23	1	0.91
[60, 70)	14	1	0.64
[51, 60)	17	1	0.71
50	27	2	0.33
(40, 50)	9	1	0.33
(30, 40]	27	2	0.15
(20, 30]	25	2	0.08
(10, 20]	30	2	0.13
[0, 10]	742	48	0.01
total sample	1,538	100	0.45

TABLE 5: PROBABILISTIC RESPONSE PATTERNS ACROSS SURVEYS, RESPONDENTS WITH COMPLETE DATA

percent chance of voting for Obama, if you do vote	Number of Respondents	Fraction who vote	Fraction voting for Obama, conditional on voting
Consistently Pro-Obama			
all [90, 100]	248	0.98	1.00
all [80, 100] and some [80, 89]	23	0.87	1.00
all [70, 100] and some [70, 79]	24	1.00	1.00
all [60, 100] and some [60, 69]	26	0.96	0.96
all [50, 100], some (50, 59]	41	0.93	0.89
Undecided/Vacillators			
all 50	1	0.00	NA
some [51, 100] and some [0, 49]	85	0.87	0.59
Consistently Anti-Obama			
all [0, 50], some [41, 50)	56	0.70	0.03
all [0, 40] and some [31, 40]	22	0.82	0.00
all [0, 30] and some [21, 30]	18	0.89	0.00
all [0, 20] and some [11, 20]	20	0.85	0.00
all [0, 10]	303	0.93	0.00
Total Sample	867	0.92	0.49

TABLE 6: TIMES SERIES OF PROBABILISTIC RESPONSES FOR UNDECIDED/VACILLATORS

Respondent Number	Percent Chance of voting for Obama, if you do vote							Voting Behavior	
	Wave 38	Wave 40	Wave 42	Wave 44	Wave 47	Wave 49	Wave 51	vote	vote for Obama, conditional on voting
1	51	49	49	10	45	49	49	Yes	Yes
2	100	100	100	100	0	0	100	No	.
3	0	0	100	100	100	100	100	Yes	Yes
4	50	45	50	50	50	55	15	Yes	No
5	50	60	40	20	70	60	60	Yes	Yes
6	100	100	20	100	100	85	100	Yes	Yes
7	80	60	50	40	40	90	100	Yes	Yes
8	52	47	55	70	60	65	80	Yes	Yes
9	50	50	90	100	100	50	0	Yes	No
10	100	100	40	55	100	100	100	Yes	Yes
11	100	100	40	100	99	100	100	Yes	Yes
12	1	4	2	100	90	96	100	Yes	No
13	99	100	50	1	25	50	0	Yes	No
14	100	100	40	0	25	50	0	Yes	Yes
15	98	95	95	99	99	0	100	Yes	Yes
16	75	40	100	100	100	100	95	Yes	Yes
17	49	50	75	90	95	99	100	Yes	Yes
18	51	55	45	45	45	20	10	Yes	No
19	50	0	75	50	95	85	85	No	.
20	100	40	100	100	50	100	100	Yes	Yes
21	0	0	0	0	80	100	100	No	.
22	0	100	90	50	90	90	50	No	.
23	40	50	20	60	40	60	40	Yes	Yes
24	50	45	50	55	70	95	100	Yes	Yes
25	0	25	50	0	100	100	100	Yes	Yes
26	60	55	40	40	40	50	80	Yes	Yes
27	100	100	50	25	50	50	75	Yes	Yes
28	40	45	50	50	50	55	60	Yes	No
29	0	0	100	0	0	0	0	Yes	No
30	60	60	40	45	60	100	100	Yes	Yes
31	60	40	90	90	100	100	100	Yes	Yes
32	50	50	50	50	60	55	30	Yes	No
33	50	45	75	70	80	90	90	Yes	Yes
34	20	30	25	50	75	25	25	Yes	No
35	40	60	70	70	75	80	90	Yes	Yes
36	0	0	100	100	100	0	0	Yes	No
37	10	50	90	95	98	100	100	Yes	Yes
38	25	20	25	25	40	40	60	Yes	No
39	30	0	0	0	30	75	50	Yes	No
40	10	0	5	0	0	60	0	Yes	No

41	0	0	0	0	10	80	100	Yes	Yes
42	0	50	50	75	90	50	50	No	.
43	40	40	45	45	50	55	55	Yes	Yes
44	60	60	50	40	45	60	52	Yes	No
45	50	51	48	49	51	50	40	Yes	No
46	75	25	33	33	60	60	92	Yes	Yes
47	75	60	50	40	50	65	75	Yes	Yes
48	100	100	95	97	48	0	50	Yes	Yes
49	100	25	50	50	70	90	90	Yes	Yes
50	50	50	50	50	50	60	25	Yes	No
51	60	50	50	0	0	30	0	Yes	No
52	50	90	25	50	25	25	25	Yes	No
53	50	25	25	25	45	70	70	Yes	Yes
54	0	0	0	0	95	20	2	Yes	No
55	100	100	60	40	40	30	0	Yes	No
56	100	80	0	95	30	1	1	Yes	No
57	100	100	100	80	55	20	20	No	.
58	10	60	50	50	50	40	60	No	.
59	40	80	60	70	50	60	50	No	.
60	50	10	20	30	20	100	100	Yes	Yes
61	60	25	30	25	50	30	25	Yes	Yes
62	35	50	45	51	51	60	100	Yes	Yes
63	45	45	20	49	49	51	51	Yes	Yes
64	50	65	45	20	50	65	50	Yes	Yes
65	60	80	100	50	50	100	0	Yes	No
66	75	80	50	50	40	80	40	Yes	Yes
67	50	50	60	50	45	50	50	Yes	No
68	40	93	88	88	80	66	77	Yes	Yes
69	0	0	80	100	80	0	100	Yes	Yes
70	80	4	10	9	0	0	0	Yes	Yes
71	0	100	0	50	0	0	0	Yes	No
72	50	40	0	50	50	100	100	Yes	No
73	0	0	0	95	0	0	0	Yes	No
74	80	20	80	90	80	80	100	No	.
75	0	0	0	0	0	0	100	Yes	Yes
76	45	35	45	45	55	55	55	Yes	Yes
77	20	10	100	100	25	75	75	No	.
78	0	0	0	0	0	70	50	Yes	Yes
79	2	10	10	60	80	80	90	Yes	Yes
80	30	25	20	15	40	55	45	Yes	No
81	60	60	60	40	50	60	60	No	.
82	70	30	20	0	7	0	0	Yes	No
83	50	50	75	50	50	0	0	Yes	No
84	50	75	60	60	45	50	50	Yes	Yes
85	100	0	0	0	0	0	0	Yes	No

TABLE 7A: PREDICTING THE OBAMA VOTE WITH WAVE 38 DATA
(best linear prediction under square loss, robust standard errors in brackets)

	Predictors	
	Current Candidate Preferences (N = 1169)	Current Candidate Preferences, Respondent Attributes (N = 1167)
Wave 38 verbal preference (response to V2/V3, Obama as default)		
Lean to Obama	0.059	0.061
	[0.080]	[0.081]
Lean to McCain	-0.414	-0.414
	[0.083]	[0.082]
McCain	-0.348	-0.348
	[0.082]	[0.082]
Someone else	-0.479	-0.485
	[0.111]	[0.110]
Lean to someone else	-0.383	-0.364
	[0.131]	[0.130]
Wave 38 probabilistic preference (response to P2 divided by 100)		
Probability of voting for Obama	0.647	0.640
	[0.085]	[0.085]
Probability of voting for Someone Else	0.313	0.329
	[0.138]	[0.138]
Respondent Attributes		
female		-0.004
		[0.013]
non-Hispanic black		0.047
		[0.019]
Hispanic		-0.064
		[0.072]
other race		-0.074
		[0.050]
age 35-49		0.006
		[0.028]
age 50-64		0.008
		[0.027]
age 65+		0.022
		[0.029]
13-15 years of schooling		-0.019
		[0.022]
16+ years of schooling		0.005
		[0.022]
Constant	0.358	0.359
	[0.083]	[0.088]
Root mean square error	0.220	0.219

TABLE 7 B: PREDICTING THE OBAMA VOTE WITH WAVE 38 and 44 DATA
(best linear prediction under square loss, robust standard errors in brackets)

	Predictors		
	Current Candidate Preferences (N = 1321)	Current and Past Candidate Preferences (N = 1092)	Current and Past Candidate Preferences, Respondent Attributes (N = 1090)
Wave 44 verbal preference (response to V2/V3, Obama as default)			
Lean to Obama	-0.090	-0.088	-0.077
	[0.089]	[0.089]	[0.086]
Lean to McCain	-0.538	-0.424	-0.424
	[0.094]	[0.108]	[0.108]
McCain	-0.525	-0.416	-0.411
	[0.077]	[0.088]	[0.088]
Someone else	-0.526	-0.245	-0.230
	[0.129]	[0.164]	[0.164]
Lean to someone else	-0.349	-0.300	-0.302
	[0.223]	[0.189]	[0.184]
Wave 44 probabilistic preference (response to P2 divided by 100)			
Probability of voting for Obama	0.454	0.179	0.185
	[0.080]	[0.087]	[0.088]
Probability of voting for Someone Else	0.162	-0.054	-0.064
	[0.144]	[0.192]	[0.192]
Wave 38 verbal preference (response to V2/V3, Obama as default			
Lean to Obama		0.103	0.101
		[0.073]	[0.073]
Lean to McCain		-0.152	-0.155
		[0.087]	[0.087]
McCain		-0.123	-0.128
		[0.080]	[0.080]
Someone else		-0.248	-0.263
		[0.107]	[0.108]
Lean to someone else		-0.177	-0.172
		[0.130]	[0.128]
Wave 38 probabilistic preference (response to P2 divided by 100)			
Probability of voting for Obama		0.290	0.287
		[0.086]	[0.087]
Probability of voting for Someone Else		0.170	0.188
		[0.127]	[0.129]
Respondent Attributes			
female			-0.010
			[0.012]
non-Hispanic black			0.018
			[0.014]

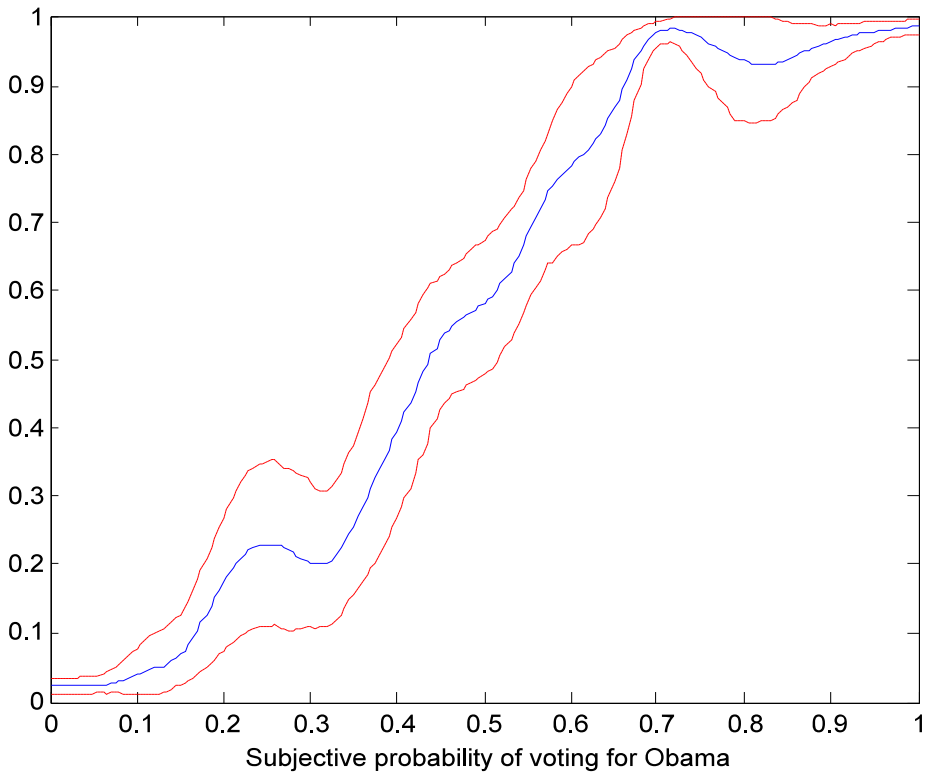
Hispanic			-0.089
			[0.057]
other race			-0.050
			[0.043]
age 35-49			0.001
			[0.028]
age 50-64			0.006
			[0.026]
age 65+			0.010
			[0.028]
13-15 years of schooling			-0.013
			[0.022]
16+ years of schooling			-0.016
			[0.022]
Constant	0.549	0.548	0.562
	[0.077]	[0.096]	[0.099]
Root mean square error	0.209	0.195	0.194

TABLE 7C: PREDICTING THE OBAMA VOTE WITH WAVE 38, 44, AND 51 DATA
(best linear prediction under square loss, robust standard errors in brackets)

	Predictors		
	Current Candidate Preferences, Respondent Attributes (N = 1391)	Current and Past Candidate Preferences, Respondent Attributes (N = 1022)	Current and Past Candidate Preferences, Respondent Attributes (N = 1020)
Wave 51 verbal preference (response to V2/V3, Obama as default)			
Lean to Obama	-0.143	0.026	0.029
	[0.150]	[0.158]	[0.161]
Lean to McCain	-0.730	-0.735	-0.732
	[0.077]	[0.080]	[0.081]
McCain	-0.644	-0.615	-0.613
	[0.085]	[0.112]	[0.112]
Someone else	-0.614	-0.658	-0.649
	[0.122]	[0.108]	[0.107]
Lean to someone else	-0.782	-0.772	-0.763
	[0.063]	[0.093]	[0.092]
Wave 51 probabilistic preference (response to P2 divided by 100)			
Probability of voting for Obama	0.351	0.235	0.238
	[0.087]	[0.112]	[0.112]
Probability of voting for Someone Else	0.104	0.110	0.101
	[0.104]	[0.103]	[0.102]
Wave 44 verbal preference (response to V2/V3, Obama as default)			
Lean to Obama		-0.017	-0.013
		[0.040]	[0.041]
Lean to McCain		-0.126	-0.125
		[0.065]	[0.065]
McCain		-0.064	-0.060
		[0.069]	[0.070]
Someone else		0.108	0.112
		[0.105]	[0.105]
Lean to someone else		-0.090	-0.100
		[0.074]	[0.071]
Wave 44 probabilistic preference (response to P2 divided by 100)			
Probability of voting for Obama		-0.069	-0.069
		[0.056]	[0.057]
Probability of voting for Someone Else		-0.069	-0.068
		[0.140]	[0.140]
Wave 38 verbal preference (response to V2/V3, Obama as default)			
Lean to Obama		0.035	0.037
		[0.030]	[0.031]
Lean to McCain		-0.061	-0.061
		[0.064]	[0.064]

McCain		-0.047	-0.047
		[0.056]	[0.057]
Someone else		-0.140	-0.145
		[0.068]	[0.069]
Lean to someone else		0.065	0.070
		[0.101]	[0.102]
Wave 38 probabilistic preference (response to P2 divided by 100)			
Probability of voting for Obama		0.114	0.115
		[0.065]	[0.066]
Probability of voting for Someone Else		0.110	0.122
		[0.081]	[0.085]
Respondent Attributes			
female			0.002
			[0.009]
non-Hispanic black			0.031
			[0.016]
Hispanic			-0.008
			[0.019]
other race			-0.030
			[0.022]
age 35-49			0.003
			[0.023]
age 50-64			0.003
			[0.020]
age 65+			0.011
			[0.020]
13-15 years of schooling			-0.002
			[0.013]
16+ years of schooling			0.001
			[0.014]
Constant	0.645	0.720	0.710
	[0.085]	[0.094]	[0.096]
Root mean square error	0.136	0.131	0.130

FIGURE 1: Probability of voting for Obama conditional
on probabilistic preference for Obama in Wave 44



Note: The x-axis gives the response to question P2, divided by 100. The conditional probabilities are estimated by kernel regression using the Gaussian kernel with Silverman's rule-of-thumb bandwidth (1,170 observations). The solid curve gives the regression estimate. The dashed curves give bootstrapped 95 per cent confidence intervals (500 draws).