Forecasting Elections: Voter Intentions versus Expectations*

David Rothschild
Yahoo! Research

David@ReseachDMR.com
www.ResearchDMR.com

Justin Wolfers
The Wharton School, University of Pennsylvania
Brookings, CEPR, CESifo, IZA and NBER
jwolfers@wharton.upenn.edu
www.nber.org/~jwolfers

Abstract

In this paper, we explore the value of an underutilized political polling question: who do you think will win the upcoming election? We demonstrate that this expectation question points to the winning candidate more often than the standard political polling question of voter intention: if the election were held today, who would you vote for? Further, the results of the expectation question translate into more accurate forecasts of the vote share and probability of victory than the ubiquitous intent question. This result holds, even if we generate forecasts with the expectations of only Democratic voters or only Republican voters and compare those forecasts to forecasts created with the full sample of intentions. Our structural interpretation of the expectation question shows that every response is equivalent to a multi-person poll of intention; the power of the response is that it provides information about the respondent's intent, as well as the intent of her friends and family. This paper has far reaching implications for all disciplines that use polling.

This draft: June 23, 2011

Latest draft: www.ResearchDMR.com/RothschildExpectations

Keywords: Polling, information aggregation, belief heterogeneity

JEL codes: C53, D03, D8

* The authors would like to thank Alex Gelber, Andrew Gelman, Sunshine Hillygus, and Marc Meredith, for useful discussions, and seminar audiences at AAPOR, the CBO, University of Pennsylvania's political science department, and Wharton for comments.

I. Introduction

Since the advent of scientific polling in the 1930s, political pollsters have asked people whom they intend to vote for; occasionally, they have also asked who they think will win. Our task in this paper is long overdue: we ask which of these questions yields more accurate forecasts. That is, we contrast the predictive power of the questions probing voter *intention* with questions probing *expectation*. Judging by the attention paid by pollsters, the press, and campaigns, the conventional wisdom appears to be that polls of voter intention are more accurate than polls of voter expectation.

Yet there are good reasons to believe that the expectation question is more informative. Survey respondents may possess much more information about the upcoming political race than the voting intention question allows them to answer. At a minimum, they know their own current voting intention, so the information set feeding into their expectation will be at least as rich as that captured by the voting intention question. But beyond this, they may also have information about the current voting intentions, both preference and probability of voting, of their friends and family. So too, they have some sense of the likelihood that today's expressed intention will be changed before it ultimately becomes an Election Day vote. Our research is motivated by idea that the richer information embedded in this expectation data may yield more accurate forecasts.

We find robust evidence that expectation-based forecasts yield more accurate predictions of election outcomes. By comparing the performance of these two questions only when they are asked in exactly the same survey, we effectively difference out the influence of other factors. Our primary dataset consists of all of the Presidential Electoral College races from 1952 to 2008, where both the intention and expectation question are asked. In 268 of the 345 polls, either both the intention and expectation question point to the winner or neither does. But in the 77 cases in which one points to the winner and the other does not, the expectation question points to the winner 60 times, while the intention question points to the winner only 17 times. That is, 78% of the time that these two approaches disagree, the expectation data is correct. We can also assess the relative accuracy of the two methods by assessing the extent to which each can be informative in forecasting the vote share and the probability of victory; we find that relying on voter expectation rather than intention data yield substantial and statistically significant increases in forecasting accuracy. Our findings remain robust to correcting for an array of known biases in voter intention data.

The better performance of forecasts created with expectation question, versus intention question, data varies somewhat, depending on the specific context. The expectation-based question is particularly

valuable when small samples are involved. The intuition for this result comes from a simple thought experiment. In our primary dataset, we have 13,208 individual respondents providing their intention and expectation in 345 different races; 58.0% of respondents intend to vote for the winning candidate, while 68.5% expect that candidate to win. Thus, if we survey only one voter, expectation will outperform intention 10.5% of the time. It is unclear ex-ante which type of question relatively benefits from increases in the days before the election, although both are less accurate as less information is available.

One strand of literature this paper addresses is the emerging documentation that prediction markets tend to yield more accurate forecasts than polls (Wolfers and Zitzewitz, 2004; Berg, Nelson and Rietz, 2008). More recently, Rothschild (2009) has updated these findings in light of the 2008 Presidential and Senate races, showing that forecasts based on prediction markets yielded systematically more accurate forecasts of the likelihood of Obama winning each state than did the forecasts based on aggregated intention polls compiled by the website FiveThirtyEight.com and another more transparent intention poll-based forecast created by the author. One hypothesis for this superior performance is that prediction markets—by asking traders to bet on outcomes—effectively ask a different question, eliciting the expectations rather than intentions of participants. If correct, this suggests that much of the accuracy of prediction markets could be obtained simply by polling voters on their expectations, rather than intentions.

These results also speak to yet another strand of research, the historical question about the value of scientific polling and representative samples (Robinson, 1937). Begun prior to the advent of scientific polling and renewed most recently with the rise of cellphones as well as use of online survey panels, this debate is again of contemporary significance. Surveys of voting intentions rely heavily on polling representative cross-sections of the electorate. By contrast, as we demonstrate in this paper, surveys of voter expectation can still be quite accurate, even when drawn from non-representative samples. Again, the logic of this claim comes from the difference between asking about expectations, which should not systematically differ across demographic groups, and asking about intentions, which should. Again, the connection to prediction markets is useful, as Berg and Rietz (2006) shows that these have yielded accurate forecasts, despite drawing from an unrepresentative pool of overwhelmingly white, male, highly educated, high income, self-selected traders.

While direct voter expectation questions about electoral outcomes have been virtually ignored by political forecasters, they have received some interest from psychologists. In particular, Granberg and Brent (1983) document wishful thinking, in which people's expectation about what will occur is positively correlated with what they want to happen. Thus, people who intend to vote Republican are also

more likely to predict a Republican victory. This same correlation is also consistent with voters preferring the candidate they think will win, as in bandwagon effects, or gaining utility from being optimistic. We re-interpret this correlation through a rational lens, in which the respondents know their own voting intention with certainty and have knowledge about the voting intentions of their friends and family. Insights from this structural interpretation of the data both explain the power of the expectation data and, by revealing the relationship between intention and expectation, may help us identify even more efficient translations of these two sets of raw data into the underlying values of the election.

More accurate forecasts will provide researchers a tool for capturing the impact of campaigns on elections; this is currently a difficult question to address, as there is great variation in both the slope and values of currently utilized forecasts of elections. Our method will also allow for forecasts of campaigns that are currently too difficult or costly to poll. Beyond understanding the effect of campaigns on elections, the findings in this paper are important as forecasts affect races, as resources are allocated by the progress of the campaign (Mutz, 1995), and voters themselves act strategically in certain contexts (Irwin and Holsteyn, 2002). Political forecasts also have consequences beyond politics, as individual companies and markets react to the probability of different outcomes (Imai and Shelton, 2010).

We believe that our findings have substantial applicability in other forecasting contexts. Market researchers ask variants of the voter intention question in an array of contexts; as they read this paper, they can seamlessly substitute a product launch in place of an election, the preference for one product over another in place of voter intention, and the consumer expectation of sales for one product over another in place of voter expectation. Likewise, indices of consumer confidence are partly based on the stated purchasing intentions of consumers, rather than their expectations about the purchase conditions for their community. The same insight that motivated our study—that people also have information on the plans of others—is also likely relevant in these other contexts. Thus, it seems plausible that other types of research may also benefit from paying greater attention to people's expectations than to their intentions.

In Section II, we describe our first cut of the data, illustrating the relative success of the two approaches to predicting the winner of elections. In Section III, we create a naïve translation of the raw data into forecasts of vote share. In Section IV, we generate a more efficient translation of the raw data into forecasts of vote share. In Section V, we determine the accuracy of the polls in creating probabilities of victory. In Section VI, we test our forecasts with 2008 data. In Section VII, we examine the accuracy of expectation-based forecasts produced with non-random samples of respondents. In Section VIII, we assess the methods derived in this paper from the primary data source, on a secondary data source. In Section IX, we provide a structural interpretation of the response to the expectation question.

II. Simple Forecasting of the Winner

Our primary dataset consists of the American National Election Studies (ANES) cumulative data file. In particular, we are interested in responses to two questions:

Voter Intention: *Who do you think you will vote for in the election for President?*

Voter Expectation: *Who do you think will be elected President in November?*

These questions are typically asked one month prior to the election. Throughout this paper, we treat elections as two-party races, and so discard responses involving professed intention to vote for or expectation of victory for third-party candidates. In order to keep the sample sizes comparable, we only keep respondents with valid responses to both the intention and expectation questions and adjust the individual response with the ANES provided weights. When we describe the "winner" of an election, we are thinking about the outcome that most interests forecasters, which is who takes office (and so we describe George W. Bush as the winner of the 2000 election, despite his losing the popular vote).

At the national level, both questions have been asked since 1952, and to give a sense of the basic patterns, we summarize these data in Table 1.

Table 1: Forecasting the Winner of the Presidential Races

Year	Race	%Expect the winner	%Intended to vote for winner	%Reported voting for winner	Actual result: % voting for winner	N
1952	Eisenhower beat Stevenson	56.0%	56.0%	58.6%	55.4%	1,135
1956	Eisenhower beat Stevenson	76.4%	59.2%	60.6%	57.8%	1,161
1960	Kennedy beat Nixon	45.0%	45.0%	48.4%	50.1%	716
1964	Johnson beat Goldwater	91.0%	74.1%	71.3%	61.3%	1,087
1968	Nixon beat Humphrey	71.2%	56.0%	55.5%	50.4%	844
1972	Nixon beat McGovern	92.5%	69.7%	68.7%	61.8%	1,800
1976	Carter beat Ford	52.6%	51.4%	50.3%	51.1%	1,320
1980	Reagan beat Carter	46.3%	49.5%	56.5%	55.3%	870
1984	Reagan beat Mondale	87.9%	59.8%	59.9%	59.2%	1,582
1988	GHW Bush beat Dukakis	72.3%	53.1%	55.3%	53.9%	1,343
1992	Clinton beat GHW Bush	65.2%	60.8%	61.5%	53.5%	1,541
1996	Clinton beat Dole	89.6%	63.8%	60.1%	54.7%	1,274
2000	GW Bush beat Gore	47.4%	45.7%	47.0%	49.7%	1,245

2004	GW Bush beat Kerry	67.9%	49.2%	51.6%	51.2%	921
2008	Obama beat McCain	65.7%	56.6%	56.5%	53.7%	1,632
	Simple Average:	**68.5%**	**56.7%**	**54.6%**	**57.5%**	**18,471**

Notes: Table summarizes authors' calculations, based on data from the American National Election Studies, 1952-2008. Sample restricted to respondents whose responses to both the expectation and intention questions listed the two major candidates.

Each method can be used to generate a forecast of the most likely winner, and so we begin by assessing how often the majority response to each question correctly picked the winner. The first column with data on Table 1 shows that the winning Presidential candidate was expected to win by a majority of respondents in 12 of the 15 elections, missing Kennedy's narrow victory in 1960, Reagan's election in 1980, and G.W. Bush's controversial win in 2000. The more standard voter intention question performed similarly, correctly picking the winning candidate in one fewer election. The only election in which the two approaches pointed to different candidates was 2004, in which a majority of respondents correctly expected that Bush would win, while a majority intended to vote for Kerry. So far we have been analyzing data from the pre-election interviews. In the third column we summarize data from post-election interviews which also ask which candidate each respondent ultimately voted for. The data in this column reveal the influence of sampling error, as a majority of the people sampled in 1960 and 2000 ultimately did vote for the losing candidate.

The last line of this table summarizes, and on average, 68.5% of all voters correctly expected the winner of the Presidential election, while 56.7% intended to vote for the winner. These averages give a hint as to the better performance of expectation-based forecasts. Taken literally, they say that if one forecasted election outcomes based on a random sample of one person in each election, asking about voter expectations would predict the winner 68.5% of the time, compared to 56.7%, when asking about voter intentions. More generally, in small polls, sampling error likely plays a larger role in determining whether a majority of respondents intend to vote for the election winner, than in whether they correctly forecast the winner. We will develop this insight at much greater length, in section IV.

The analysis in Table 1 does not permit strong conclusions, and indeed, it highlights two important analytic difficulties. First, we have very few national Presidential elections, and so the data will permit only noisy inferences. Second, our outcome measure—asking whether a method correctly forecasted the winner—is a very coarse measure of the forecasting ability of either approach to polling. Thus, we will proceed in two directions. First, we will exploit a much larger number of elections by analyzing data from the same surveys on who respondents expect to win the Electoral College votes of their state. And second, we will proceed to analyzing the forecasting performance of each approach

against other measures: their ability to match the two-party preferred vote share and their forecast of the probability of victory.

We begin with the state-by-state analysis, analyzing responses to the state-specific voter expectation question:

Voter Expectation (state level): *How about here in [state]. Which candidate for President do you think will carry this state?*

We compare responses to this question with the voter intention question described above. Before presenting the data, there are four limitations of these data worth noting. First, the ANES does not survey people in every state, and so in each wave, around 35 states are represented. Second, this question was not asked in the 1956-68 and 2000 election waves. We do not expect either of these issues to bias our results toward favoring either intention- or expectation-based forecasts. Third, the sample sizes in each state can be small. Across each of these state elections, the average sample size is only 38 respondents, and the sample size in a state ranges from 1 to 246. In section IV we will see that this is an important issue, as the expectation-based forecasts are relatively stronger in small samples. Fourth, while the ANES employs an appropriate sampling frame for collecting nationally representative data, it is not the frame that one would design were one interested in estimating state-specific aggregates, as these samples typically involve no more than a few Primary Sampling Units. Despite these limitations, this data still presents an interesting laboratory for testing the relative efficacy of intention versus expectation-based polling. All told we have valid ANES data from 10 election cycles (1952, 1972-1996, and 2004-2008), and in each cycle, we have data from between 28 and 40 states, for a final sample size of 345 races.[2]

Table 2 summarizes the performance of our two questions at forecasting the winning Presidential candidate in each state. Again, we use a very coarse performance metric, simply scoring the proportion of races in which the candidate who won a majority in the relevant poll ultimately won in that state; the voter expectation question is the first column with data and the usual voter intention question is the second column with data. (When a poll yields a fifty-fifty split, we score it as half of a correct call.) All told, the voter expectation question predicted the winner in 279 of these 345 races, compared with 239 correct calls for the voter intention question. A simple difference in proportions test reveals that these differences are clearly statistically significant (z=3.63). Of course the forecast errors of each approach may be correlated across states within an election cycle, and so a more conservative approach would note that the voter expectation question outperformed the voter intent question in 8 of the 10 election cycles

[2] Only the 311 pre-2008 elections are used in Sections III, IV, and V to test and calibrate our models. Section VI then runs the model, with coefficients created with pre-2008 data, on 2008 data.

and tied in 2008. The difference in that tenth cycle (1972) was that Nixon won a tight race in Minnesota. More to the point, this difference in forecasting performance is large.

Table 2: Forecasting the Presidential Election, by State

Year	Proportion of states where the winning candidate was correctly predicted by a majority of respondents to:		Number of states surveyed
	Expectation question	Intention question	
1952	74.3%	58.6%	35
1972	97.4%	100%	38
1976	80.3%	77.6%	38
1980	57.7%	41.0%	39
1984	86.7%	68.3%	30
1988	88.3%	56.7%	30
1992	89.4%	77.3%	33
1996	75.0%	67.5%	40
2004	89.3%	67.9%	28
2008	76.5%	76.5%	34
Totals:	**279 of 345 correct**	**239 of 345 correct**	**Difference:**
Average:	**80.9%**	**69.3%**	**11.6%*** **
(Standard error)	**(3.8)**	**(5.4)**	**(3.2)**

Notes: ***, **, and * denote statistically significant coefficients at the 1%, 5%, and 10%, respectively. (Standard errors in parentheses are clustered by year).

At this point, our analysis has been quite crude—only analyzing whether the favored candidate won. This approach has the virtue of transparency, but it leaves much of the variation in the data—such as variation in the winning margin—unexamined. Thus we now turn to analyzing the accuracy of the forecasted vote shares derived from both intention and expectation data. We will also add some structure to how we are thinking about these data. At this point we drop 2008 from the dataset, in order to have some "out of-sample" data to review in Section VI.

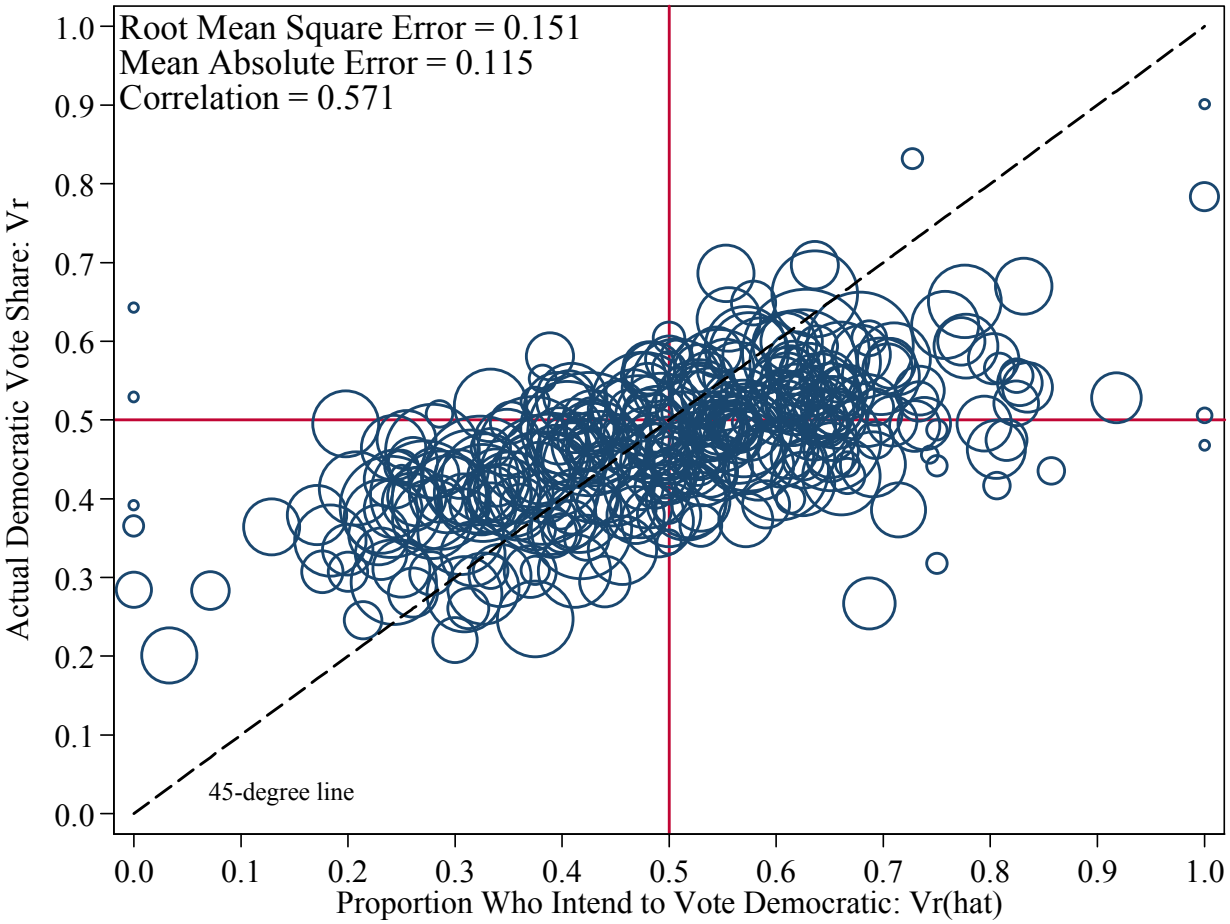
III. Simple (or Naïve) Forecasting of Vote Share

Our goal is to use the state-by-state ANES data to come up with forecasts of the two-party vote share in each of the R state×year races in our dataset. In this section, we analyze the data the way they are typically used—interpreting a poll that says that $\widehat{v}_r$ percent of *sample* respondents will vote for a candidate in state-year race r as a forecast that this candidate will win v_r percent of votes among the entire *population*. That is, we follow the norm among pollsters and make our projections as if the sample

moments represent population proportions. Likewise, we interpret a poll that says that $\widehat{x_r}$ percent of sample respondents expect a candidate to win as a forecast that x_r percent of the population expect that candidate to win. While this may sound obvious, in fact raw polling data rarely represent optimal forecasts. Thus, we refer to the projections in this section as "naïve forecasts", and in section IV we will describe how our raw polling data can be adjusted to create efficient forecasts.

Our focus is on predicting each candidate's share of the two-party vote, and we begin by analyzing data on voter intention. Figure 1 plots the relationship between the actual Democratic vote share in each state-year race, and the proportion of poll respondents who plan to vote for the Democratic candidate. There are two features of these data to notice. First, election outcomes and voter expectations are clearly positively correlated—that is, these polls are informative. But second, the relationship is by no means one-for-one, and these relatively small polls of voter intention are only a noisy measure of the true vote share on Election Day.

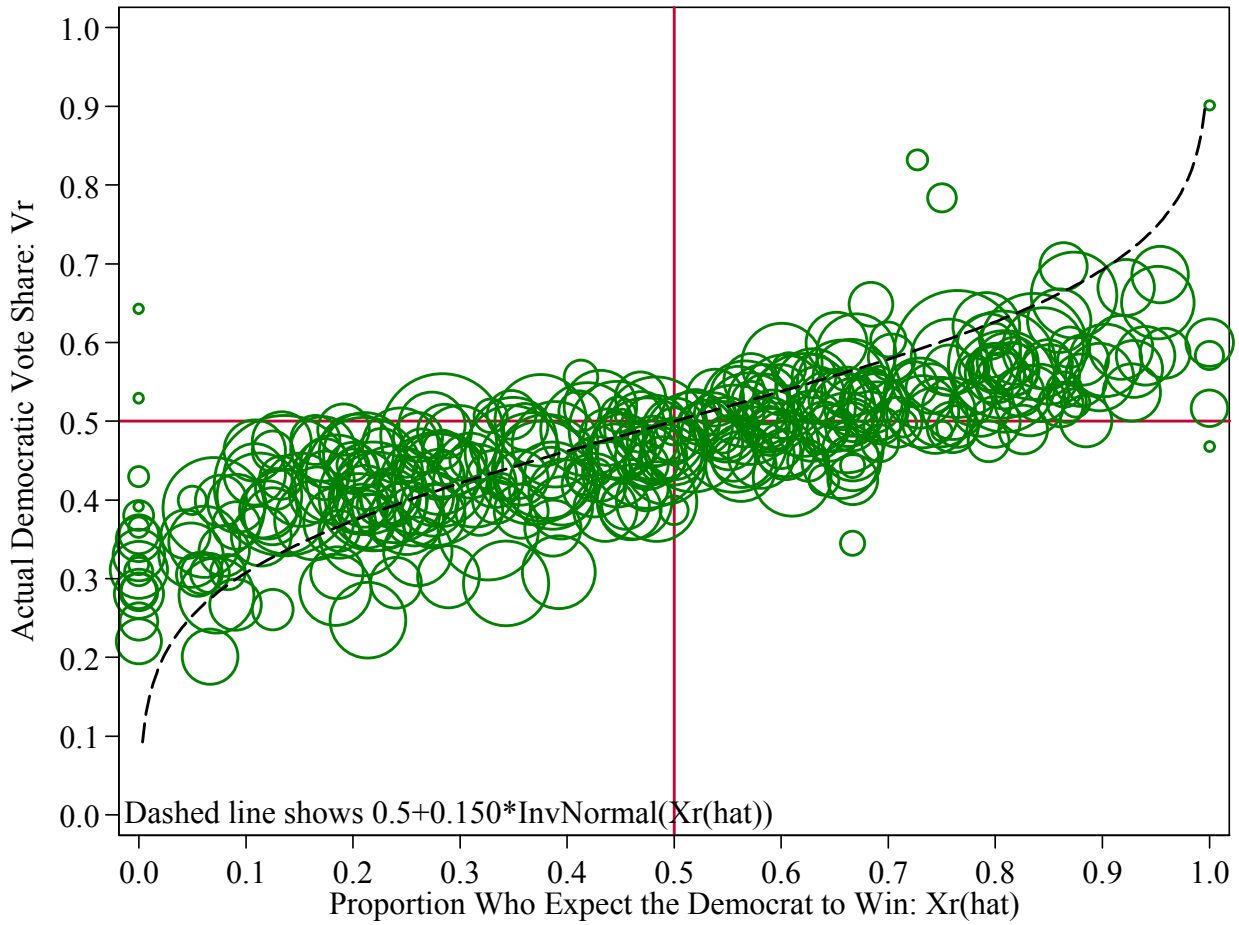
Figure 1: Naïve Voter Intention Forecast and Actual Vote Share



Notes: Each point shows a separate state-year cell in a Presidential Electoral College election; the size of each point is proportional to the number of survey respondents. Both voter intention and election outcomes refer to shares of the total votes cast for the two major parties. There are a total of n=311 elections, as the 2008 data is not included.

In Figure 2 we show the relationship between voter expectations—the share of voters who expect the Democrat to win that state's presidential ballot—and the vote share he actually garnered. This plot reveals that there is a close relationship between election outcomes and voter expectations, and typically the candidate who most respondents expect to win, does in fact win. That is, most of the data lie in either the Northeast or Southwest quadrants, a fact also evident in Table 2. Equally, the relationship between voter expectations and vote shares does not appear to be linear. Indeed, it should seem obvious that a statement that two-thirds of voters expect Obama to win does not—without adding further structure—immediately correspond to any particular forecast about his likely vote share. Thus we now turn to assessing how to tease out the forecast of vote shares implicit in these data.

Figure 2: Voter Expectation and Actual Vote Share



We begin by characterizing how people respond to the question probing their expectations about the likely winner. If each poll respondent views an unbiased noisy signal, x^{*i}_{r} of a candidate's final vote

share—where the superscript i serves as a reminder that we are analyzing individual responses, and the asterisk reminds us that this is an unobserved latent variable—then:

$$x^{*i}_r = v_r + \epsilon^i_r \text{ and } \epsilon^i_r \sim N(0, \sigma^2_\epsilon) \quad [1]$$

where ϵ^i_r is an idiosyncratic error reflecting the respondent's imperfect observation.[3] We assume that this noise term is drawn from a normal distribution, and its variance is constant across both poll respondents, and across elections. In turn, if poll respondents describe themselves as expecting a specific candidate to win if this noisy signal suggests that this candidate will win at least half the vote, then we will observe voter expectations as follows:

$$x^i_r = \begin{cases} 1 & if\ x^{*i}_r = v_r + \epsilon^i_r > 0.5 \\ 0 & if\ x^{*i}_r = v_r + \epsilon^i_r < 0.5 \end{cases} \quad [2]$$

Consequently the probability that an individual respondent says that they expect a candidate to win is $\Phi\left(\frac{v_r - 0.5}{\sigma_\epsilon}\right)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. That is, equations [1] and [2] together imply that we can estimate σ_ϵ from a simple probit regression explaining whether the respondent expected a candidate to win, by a variable describing the extent by which the vote share garnered by that candidate exceeds the 50% required to win:

$$x^i_r = \frac{1}{\sigma_\epsilon}(v_r - 0.5 + \epsilon^i_r) \quad [3]$$

This regression yields an estimate of $1/\widehat{\sigma_\epsilon} = 6.661$ with a standard error allowing for within-state-year correlated errors of 0.385 (n=11,548), which implies that $\widehat{\sigma_\epsilon} = 0.150$, with a standard error of 0.0089 (estimated using the delta method). For now, we simply note that this estimate provides a link between election results, and the proportion of the population who expect the Democrat to win, x_r:

$$E[x_r] = Prob\left(v_r + \epsilon^i_r > 0.5\right) = \Phi\left(\frac{v_r - 0.5}{\sigma_\epsilon}\right) \quad [4]$$

When the voting population is large then the noise induced by ϵ^i in the mapping between the population parameters v_r and x_r is negligible,[4] and so it follows that:

[3] Part of this error may be due to the fact that the election is still a month away; thus ϵ^i includes variation due to voters who may later change their minds.

[4] As we will see in the next section, the noise term ϵ^i_r is relevant to the mapping between the population parameter x_r and the sample estimate $\widehat{x_r}$.

$$x_r \approx \Phi\left(\frac{v_r - 0.5}{\sigma_\epsilon}\right) \quad [5]$$

From this, we can back out the implied expected vote share by inverting this function:

$$E[v_r|x_r] \approx 0.5 + \sigma_\epsilon \Phi^{-1}(x_r) \quad [6]$$

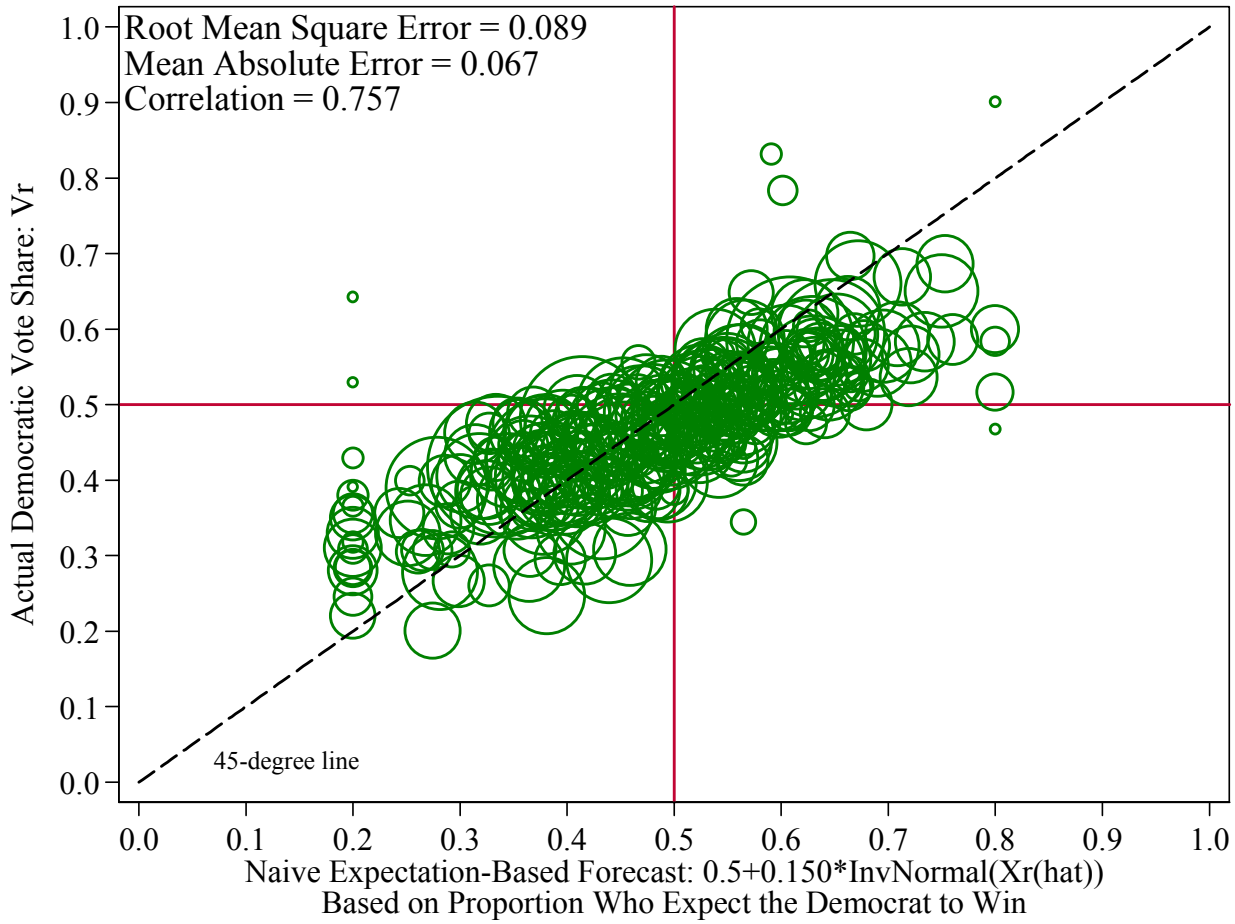
This function is also shown as a dashed line in Figure 2, based on our estimated value of $\widehat{\sigma_\epsilon} = 0.150$. To be clear, this is the appropriate mapping only if we know the true population proportion who expect a particular candidate to win, x_r. While this assumption is clearly false, our goal here is to provide a forecast comparable to the naïve forecast of voter intentions, and so in both cases we use the mapping between survey proportions and forecasts that would be appropriate in the absence of sampling variation. Indeed, in Figure 2 many of the extreme values of the proportion of voters expecting a candidate to win likely reflect sampling variation. That is, our transformation of voter expectations data into vote shares is clearly not estimated as a line of best fit, as elections are rarely as lopsided as the dashed line suggests. This feature roughly parallels the observation that in Figure 1; elections are rarely as lopsided as suggested by small samples of voter intentions. We will explore this feature of the data in greater detail in the next section when we evaluate efficient shrinkage-based estimators. But for now we will call this simple transformation of voter expectation our "naïve" expectation-based forecast.

The one remaining difficulty is that in 22 races (7 percent of races), either 0% or 100% of survey respondents expect the Democrat to win, and so equation [6] does not yield a specific forecast. In these cases we (somewhat arbitrarily) infer that the candidate is expected to win 20% or 80% of the vote, respectively.[5] Given the extreme nature of these inferences, we regard these assumptions as unfavorable to the expectations-based forecast. Even so, we obtain qualitatively similar results when imputing expected vote shares of 0% and 100% instead. (Section IV provides a more satisfactory treatment of this issue.)

Figure 3 plots our naïve expectations-based forecasts of vote shares against the actual election results. These forecasts are clustered along the 45-degree line, suggesting that they are quite accurate.

[5] Our rationale was simply that these are the nearest round numbers that ensure that the implied forecast is monotonic in the proportion of respondents expecting a particular candidate to win. (Across the 289 other elections, the minimum was 24% and the maximum was 76%.)

Figure 3: Naïve Expectation-Based Forecast and Actual Vote Share



In Table 3 we provide several simple comparisons of forecast accuracy. The first two rows show that the expectation-based forecast yields both a root mean squared error and mean absolute error that is significantly less than the intention-based forecast. The third row shows that the expectation-based forecast is also the more accurate forecast in 65% of these elections. The significance in these advantages for expectation-based forecasts is shown in the corresponding final column. In the fourth row, we examine the correlation coefficient. One might be concerned that the better performance of the expectation-based forecasts reflects the fact that they rely on an estimated parameter, σ_ϵ, and thus they use up one more degree of freedom. That is, our estimated value of σ_ϵ "tilts" the expectations data so that the implied forecasts lie along the 45-degree line. The correlation coefficient effectively both tilts the data and shifts it up and down, so as to maximize fit. Thus, it arguably puts each forecast on something closer to an equal footing. Even so, the expectation-based forecasts are also more highly correlated with actual vote shares than are the intention-based forecasts.

Table 3: Comparing the Accuracy of Naïve Forecasts of Vote Shares

	Raw Voter Intention: $\widehat{v_r}$	Transformed Voter Expectation: $0.5 + 0.150\Phi^{-1}(\widehat{x_r})$	Test of Equality
Root Mean Squared Error	0.151 (0.008)	0.089 (0.005)	t_{344}=7.05 (p<0.0001)
Mean Absolute Error	0.115 (0.006)	0.067 (0.003)	t_{344}=8.81 (p<0.0001)
How often is forecast closer?	35.0% (2.7)	65.0% (2.7)	t_{344}=5.37 (p<0.0001)
Correlation	0.571	0.757	
Encompassing regression: $v_r = \alpha + \beta_v Intention_r + \beta_x Expectation_r$	0.058** (0.026)	0.480*** (0.035)	
Optimal weights: $v_r = \beta Intention_r + (1-\beta) Expectation_r$	8.5%** (3.7)	91.5%*** (3.7)	

Notes: ***, **, and * denote statistically significant coefficients at the 1%, 5%, and 10%, respectively. (Standard errors in parentheses). These are assessments of forecasts of the Democrat's share of the two-party vote in n=311 elections. Comparisons in the third column test the equality of the measures in the first two columns. In the encompassing regression, the constant $\hat{\alpha} = 0.207^{***}$ (se=0.013).

We also show a Fair-Shiller (1989 and 1990) regression, attempting to predict election outcomes on the basis of a constant, and our two alternative forecasts. The expectation-based forecast has a large and extremely statistically significant weight, as does the constant. But it does not fully encompass the information in the intention-based forecast, which still receives a statistically significant, albeit small weight. Finally, we estimate the optimal weighted average of these two forecasts, which puts greater than 90% of the weight on the expectation-based forecast.

These tests, of course, are all based on the common but problematic assumption that our sample data can be interpreted as representing population moments. We now turn to generating statistically efficient forecasts instead, and will repeat our forecast evaluation exercise based on these adjusted forecasts.

IV. Efficient Poll-Based Forecasts of Vote Share

An important insight common to both Figure 1 and Figure 3 is that in those elections where the Democrats are favored, the final outcome typically does favor Democrats, but by less than suggested by the naïve forecasts based on either voter intentions or voter expectations. Likewise, when the poll favors Republicans, the Democrats do tend to lose, but again, by less than suggested by our naïve forecasts. In

fact, this is a natural implication of sampling error, because our extreme polls likely reflect sampling variation, and so it is unsurprising that they are not matched by extreme outcomes. It is this observation that motivates our use of shrinkage estimators in this section—"shrinking" the raw estimates of the proportion intending to vote for one candidate or another toward a closer race. This idea is widely understood by political scientists (Campbell, 2000), but is typically ignored in media commentary. We will also adjust for any biases (or "house effects") in these data.

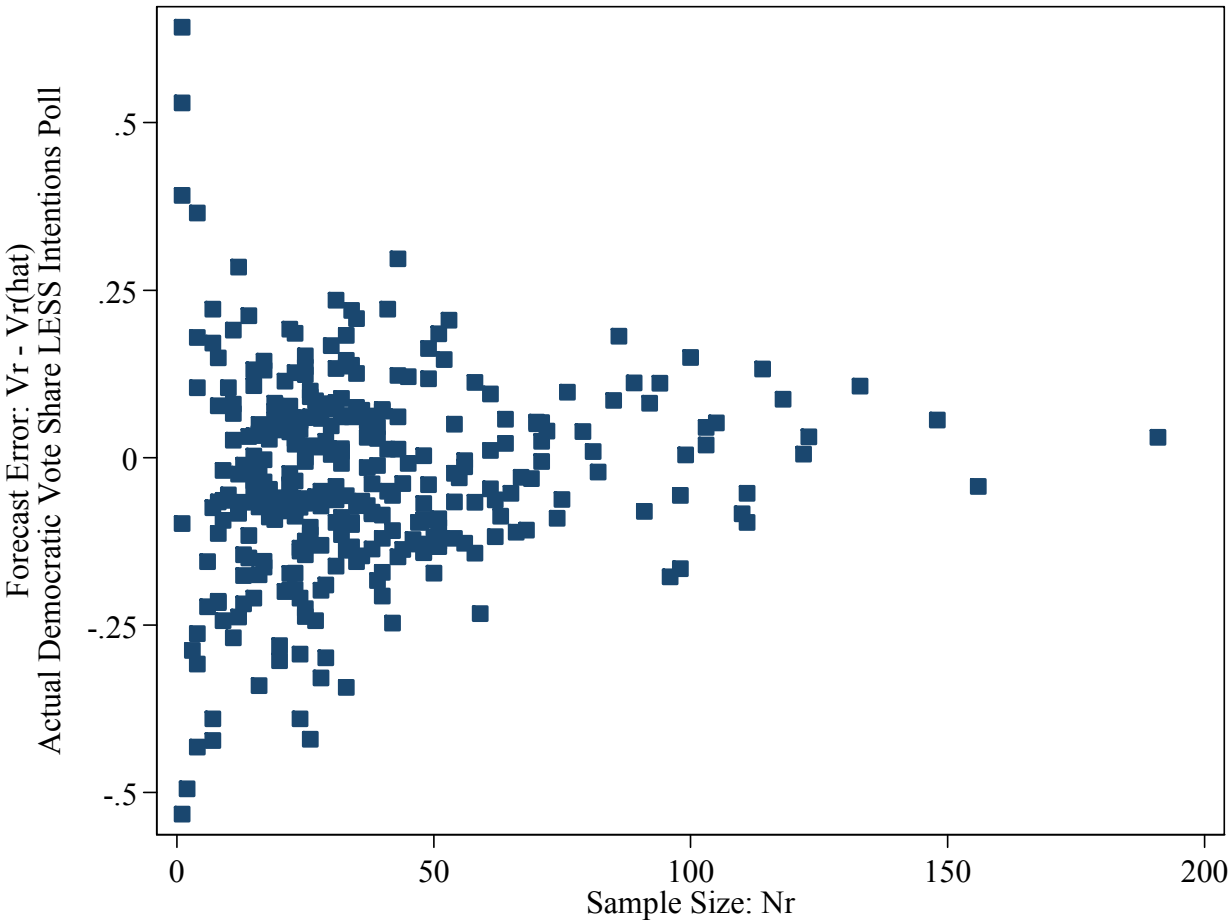
In the following discussion it is important to distinguish between the actual vote share won by the Democrat, v_r, from the sample proportion who intend to vote for the candidate, $\widehat{v_r}$, and the optimal intention-based forecast, $E[v_r|\widehat{v_r}]$. Likewise, we distinguish the sample proportion who expect a candidate to win, $\widehat{x_r}$, from the population proportion, x_r, and our optimal expectation-based forecast of the vote share, $E[v_r|\widehat{x_r}]$. Because we are only analyzing respondents with valid expectation and intent data, each forecast will be based on the same sample size, n_r.

We will begin by analyzing forecasts based on standard polls of voter intentions, and will then turn to analyzing how voter expectations might improve these forecasts.

Interpreting Voter Intentions

Our goal is to find the mapping between our raw voter intentions data, and the forecast which minimizes the mean squared forecast error. The usual approach—of fitting an OLS regression line—involves shrinking each estimate back toward the grand mean, using the signal-to-noise ratio. (This is why the least-squares estimator of an errors-in-variables model yields a regression coefficient that is shrunk by a factor related to the signal-to-noise ratio in the explanatory variable.) The difficulty is that in our setting, the sample size varies widely across each race, and as Figure 4 illustrates, so too does the noise underpinning each observation.

Figure 4: Sample Size and Forecast Errors in the Intent Poll



Our point of departure in this section is to note that our sample estimate of the proportion of voters intending to vote for a candidate is a noisy estimate—and possibly also biased—estimate of the election outcome:

$$\widehat{v_r} = v_r + b + e_r, \text{ where } e_r \sim N(0, \sigma_{e_r}^2) \qquad [7]$$

Where and b is a bias term which picks up the pro-Democrat house effect in ANES polls, and e_r is the noise term.[6] In particular, notice the r subscript on the variance of this noise term, which reflects the fact that sampling variability will vary with the characteristics (particularly, sample size) of a specific poll. Assuming that $E[v_r e_r] = 0$ we get the following familiar result:

$$E[v_r | \widehat{v_r}] = \mu_v + \frac{\sigma_v^2}{\sigma_{\hat{v}}^2} (\widehat{v_r} - b - \mu_v) \qquad [8]$$

[6] We also tested for an anti-incumbent party bias, but found it to be small and statistically insignificant.

where μ_v and σ_v^2 are the mean and variance of the Democratic vote share, across all the races in our dataset. Forming an optimal intentions-based forecast requires estimating each of these parameters. We estimate the average vote share of Democrats directly from our sample: $\widehat{\mu_v} = \frac{1}{R}\sum v_r = 0.468$ $(se = 0.005)$, and the variance of the Democrat vote share is: $\widehat{\sigma_v^2} = \frac{1}{R-1}\sum(v_r - \widehat{\mu_v})^2 = 0.0089$. Likewise, it is easy to estimate the bias term, $\hat{b} = \sum(\widehat{v_r} - v_r)/R = 0.031$ $(se = 0.008)$. (This bias in the ANES is reasonably large, statistically significant, and to our reading, has not previously been documented; even when we cluster results by year, the bias remains statistically significant.) All that remains is to sort out the variance of the polls, which can be broken into: $\sigma_{\hat{v}}^2 = \sigma_v^2 + \sigma_e^2$, where $\sigma_e^2 = E[e_r^2] = E[(v_r - \widehat{v_r} - b)^2]$.[7]

There are two sources of error to consider. First, these polls are typically taken one month prior to the election,[8] and many voters may change their minds in the final weeks of the campaign. Hence while we are sampling from a population where v_r percent of respondents will ultimately vote for the Democrat, but when they are polled one month prior, an extra τ_r percent intend to vote Democratic. Second, sampling error plays an important role, particularly in small samples. Assuming that these two sources of error are orthogonal so that $E[\tau_r(v_r - \widehat{v_r} - b - \tau_r)] = 0$, the variance of the polling error can be decomposed as:

$$\sigma_e^2 = E[(v_r - \widehat{v_r} - b)^2] = E[(v_r - (v_r + \tau_r))^2] + E\left[\left((v_r + \tau_r) - (\widehat{v_r} - b)\right)^2\right] \qquad [9]$$

where the first term in the above expression reflects the variability in "true" public opinion between polling day and Election Day (and hence is unrelated to the size of our samples). Our data does not have any useful time variation, and so we simply use the estimate of $\widehat{\sigma_\tau^2} = 0.001$, from Lock and Gelman (2010) (who estimate this as a function of time before the election; we use their fitted value for one-month before Election Day). The second term reflects sampling variability, and because the poll result $\widehat{v_r}$ is the mean of a binomial variable with mean $v_r + \tau_r$, this second term can be expressed as:

$$E[(v_r - \tau_r - b - \widehat{v_r})^2] = \frac{(v_r + \tau_r)(1 - v_r - \tau_r)}{n_r^{eff}} \approx \frac{0.25}{n_r/\gamma_r^v} \qquad [10]$$

The numerator of this expression is the product of the vote shares of the two parties, if the election were held on polling day. For most elections (and particularly competitive contests), the term $(v_r + \tau_r)(1 - v_r - \tau_r) \approx 0.25$, and so we use this approximation. (We obtain similar results when we

[7] Any variance in the bias term from cycle to cycle would be included in the variance of the polling error.
[8] In fact, the polls are taken fairly uniformly over the two months prior to the election.

plug in actual vote shares or vote shares from the previous election instead; our approximation has the virtue of being usable in a real-time forecasting context.) The denominator, n_r^{eff} denotes the effective sample size of the specific poll in race r. If the sample were a simple random sample, the effective sample size would be exactly equal to the actual sample size. But the American National Election Studies uses a complex sample design, polling only in a limited number of primary sampling units. The "design effect" γ_r^v corrects for the effects of the intra-cluster correlation within these sampling units. (The subscript r serves as a reminder that it varies with the sample size of specific poll—for instance, it is one when $n_r = 1$—and the superscript v serves as a reminder that the design effect varies, and this is the design effect for voter intentions). Unfortunately published estimates of γ_r^v are based on design effects in national samples, and so they can't be applied to our analysis of state samples. Moreover, the public release files in the ANES files do not contain sufficient detail about the sampling scheme to allow us to estimate these design effects directly. A standard approach to estimating γ_r^v is by the so-called "Moulton factor", $\gamma_r^v = 1 + (n_r - 1)\rho$, where ρ is the intra-cluster correlation coefficient (Moulton, 1990).[9] In what follows, we assume that ρ is constant across states and time. While we lack the details on sampling clusters to estimate ρ directly, we can estimate it indirectly. Figure 4 highlights the underlying variation, showing a consistent pattern of errors varying with the sample size of a poll. Our identification comes from the fact that this pattern is shaped by ρ—the higher is ρ, the less quickly the variance of $(v_r - \widehat{v_r})$ declines with sample size. Thus, we return to equation [8], plug in the values for $\widehat{\mu_v}$, $\widehat{\sigma_v^2}$, $\widehat{\sigma_t^2}$ and $\hat{b}$ and estimate ρ directly by running non-linear least-squares on the following regression:

$$v_r = \widehat{\mu_v} + \frac{\widehat{\sigma_v^2}}{\widehat{\sigma_v^2} + \widehat{\sigma_t^2} + \frac{1 + (n_r - 1)\rho}{4n_r}} (\widehat{v_r} - \hat{b} - \widehat{\mu_v}) \qquad [11]$$

which yielded an estimate of $\hat{\rho} = 0.030$ (with a standard error of 0.0080), which implies an average design effect of $\widehat{\gamma^v} = 2.09$. Thus, returning to equation [8], our MSE-minimizing forecast based on the voter intentions data, is:

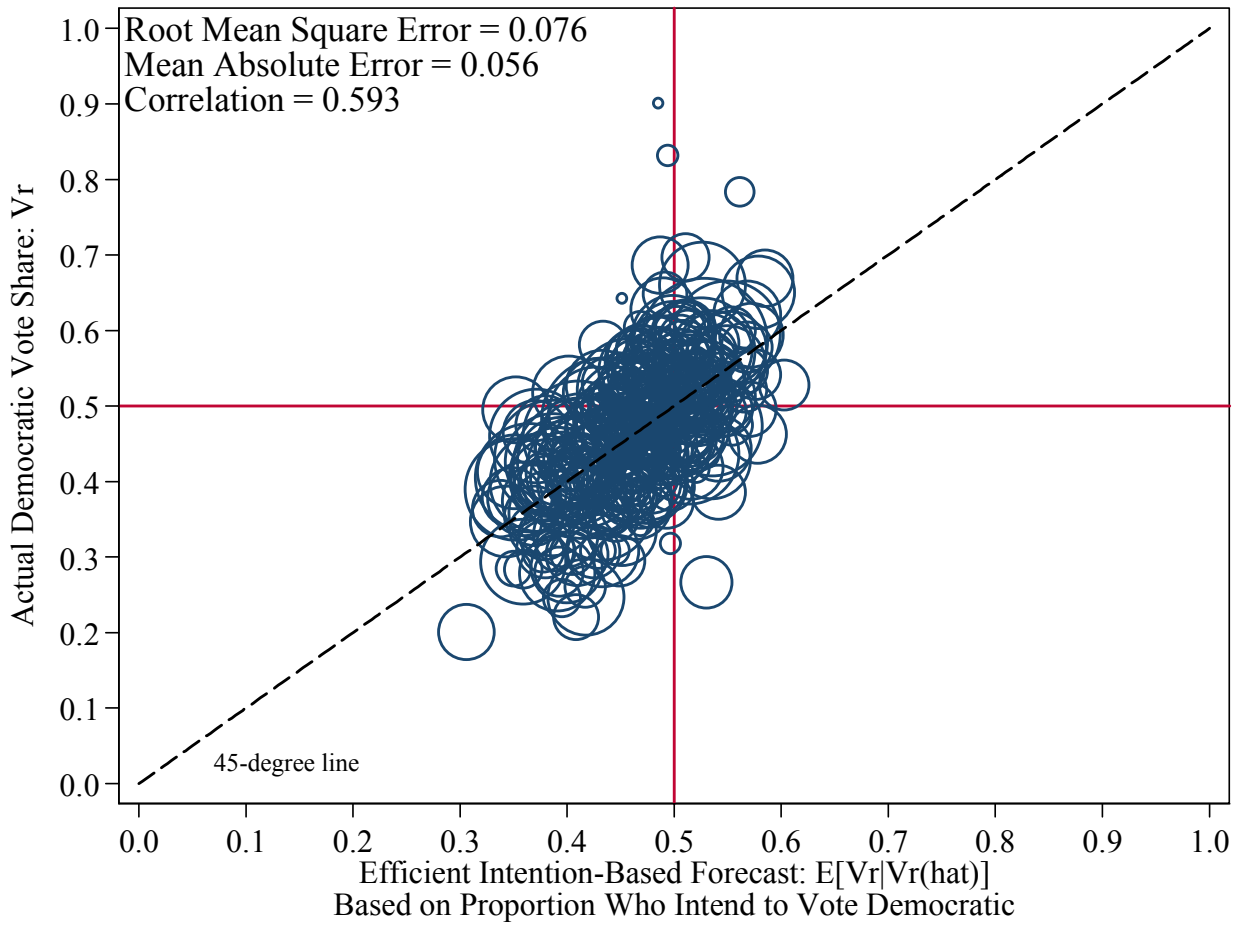
$$E[v_r|\widehat{v_r}] = \mu_v + \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2} (\widehat{v_r} - b - \mu_v)$$

$$= 0.468 + \frac{0.0089}{0.0089 + 0.0001 + \frac{1 + (n_r - 1)0.030}{4n_r}} (\widehat{v_r} - 0.031 - 0.468)$$

[9] A remaining difficulty is that while we assume that the total observations for each state come from a single cluster, in fact some of the larger states include multiple primary sampling units. Thus, our approach is appropriate if we think of ρ as the intra-cluster correlation, where a "cluster" is the set of sampled addresses within a state.

Given the data in our sample, the shrinkage estimator (the coefficient on the de-meaned and de-biased intent poll) averages 0.33 (which corresponds closely with the average slope seen in Figure1, but it ranges from 0.03 (in a race with only one survey respondent) to 0.47.

Figure 5: Efficient Intention-Based Forecasts and Actual Vote Share



In Figure 5, we show the relationship between our optimal intention-based forecast and actual vote share. These adjusted intention-based forecasts are clearly more accurate than the naïve forecasts numbers: the forecasts lie along the 45-degree line, and both the mean absolute error and the root mean squared error are about half that found Figure 1. We now turn to finding the most efficient transformation of our voter expectation data.

Interpreting Voter Expectations

In our previous analysis in Section III, we transformed data on voter expectations into vote share forecasts based on $E[v_r|x_r]$. But taking the sample variability seriously means that we are trying to figure out $E[v_r|\widehat{x_r}]$. As an intermediate step, we begin by estimating the $E[x_r|\widehat{x_r}]$. Once again, we turn

to a shrinkage estimator in order to generate efficient estimates of x_r, given our small sample estimates, $\widehat{x_r}$. As in equation [8],

$$E[x_r|\widehat{x_r}] = \mu_x + \frac{\sigma_x^2}{\sigma_{\hat{x}}^2}\,(\widehat{x_r} - b - \mu_x) \qquad [12]$$

where μ_x is the mean across all elections of the proportion of the population who expect the Democrat to win; σ_x^2 of the corresponding variance, while $\sigma_{\hat{x}}^2$ is the variance of the corresponding sample estimator; and as before, we have a bias parameter, b which allows for the possibility that these data oversample people who expect Democrats to win.

The key difficulty in working with data on voter expectations rather than voter intentions is that while we do ultimately observe how the entire population votes, we never observe what the whole population expects. Thus in estimating population parameters, we will rely heavily on the mapping in equation [5] between population vote shares and population expectations.[10] Using this insight, we estimate $\widehat{\mu_x} = \sum \Phi\left(\frac{v_r - 0.5}{\widehat{\sigma_\epsilon}}\right)/R = 0.427\ (se = 0.005)$. Likewise we estimate the bias term $\hat{b} = \sum(\widehat{x_r} - \Phi\left(\frac{v_r - 0.5}{\widehat{\sigma_\epsilon}}\right))/R = 0.043\ (se = 0.009)$. This bias term represents the increased probability of an individual expects a Democrat to win, its importance cannot be directly compared to the intention data, where the bias has a meaningful impact in the naïve intention-based forecast of vote share.

The numerator of the shrinkage estimator in equation [12] is the variance of the population expectation across all elections, and it is also quite straightforward to estimate: $\widehat{\sigma_x^2} = \frac{1}{R-1}\sum(x_r - \widehat{\mu_x})^2 = 0.0450$. All that remains is to sort out the denominator of the shrinkage estimator, $\sigma_{\hat{x}}^2$, which is equal to the underlying variation in population expectations across elections, plus sampling variability. Because different elections have different sample sizes, this denominator varies across elections. Again, because we are asking about binary outcome—whether or not you expect the Democrat to win in your state—we know the functional form of the relevant sampling error. Thus:

$$\sigma_{\hat{x}}^2 = \sigma_x^2 + \frac{x_r(1 - x_r)}{n_r^{eff}} \approx \frac{0.25}{n_r/\gamma_r^x} \qquad [13]$$

where, the approximation follows because the product of the population proportions expecting each candidate, $x_r(1 - x_r) \approx \Phi\left(\frac{v_r - 0.5}{\widehat{\sigma_\epsilon}}\right)\Phi\left(\frac{0.5 - v_r}{\widehat{\sigma_\epsilon}}\right)$, and in turn, $\Phi\left(\frac{v_r - 0.5}{\widehat{\sigma_\epsilon}}\right)\Phi\left(\frac{0.5 - v_r}{\widehat{\sigma_\epsilon}}\right) \approx 0.25$ because most

[10] We are yet to adjust our standard errors for the extra uncertainty generated because we are using an estimate of $\widehat{\sigma_\epsilon}$.

elections are competitive (and $\widehat{\sigma_\epsilon}$ is not too small).[11] As before, n_r^{eff} is the effective sample size, and γ_r^x is the relevant design effect, and we apply the "Moulton factor" to estimate $\gamma_r^x = 1 + (n_r - 1)\rho_x$. The x superscript on the intra-cluster correlation reminds us that there is no reason to expect the intra-cluster correlation in voter expectations to be similar to that for voter intentions. Thus, our remaining challenge is to estimate the intra-class correlation in voter expectations. As we did with the voter intentions data, we will use an indirect approach to estimating ρ_x. That is, we plug the results in equations [12] and [13] back in to equation [6] to get a simple vote share forecasting equation, and find the value that yields the best overall fit:

$$E[v_r|\widehat{x_r}] = 0.5 + \widehat{\sigma_\epsilon}\Phi^{-1}\left(\widehat{\mu_x} + \frac{\widehat{\sigma_x^2}}{\widehat{\sigma_x^2} + \frac{1 + (n_r - 1)\rho_x}{4n_r}}\left(\widehat{x_r} - \hat{b} - \widehat{\mu_x}\right)\right) \qquad [14]$$

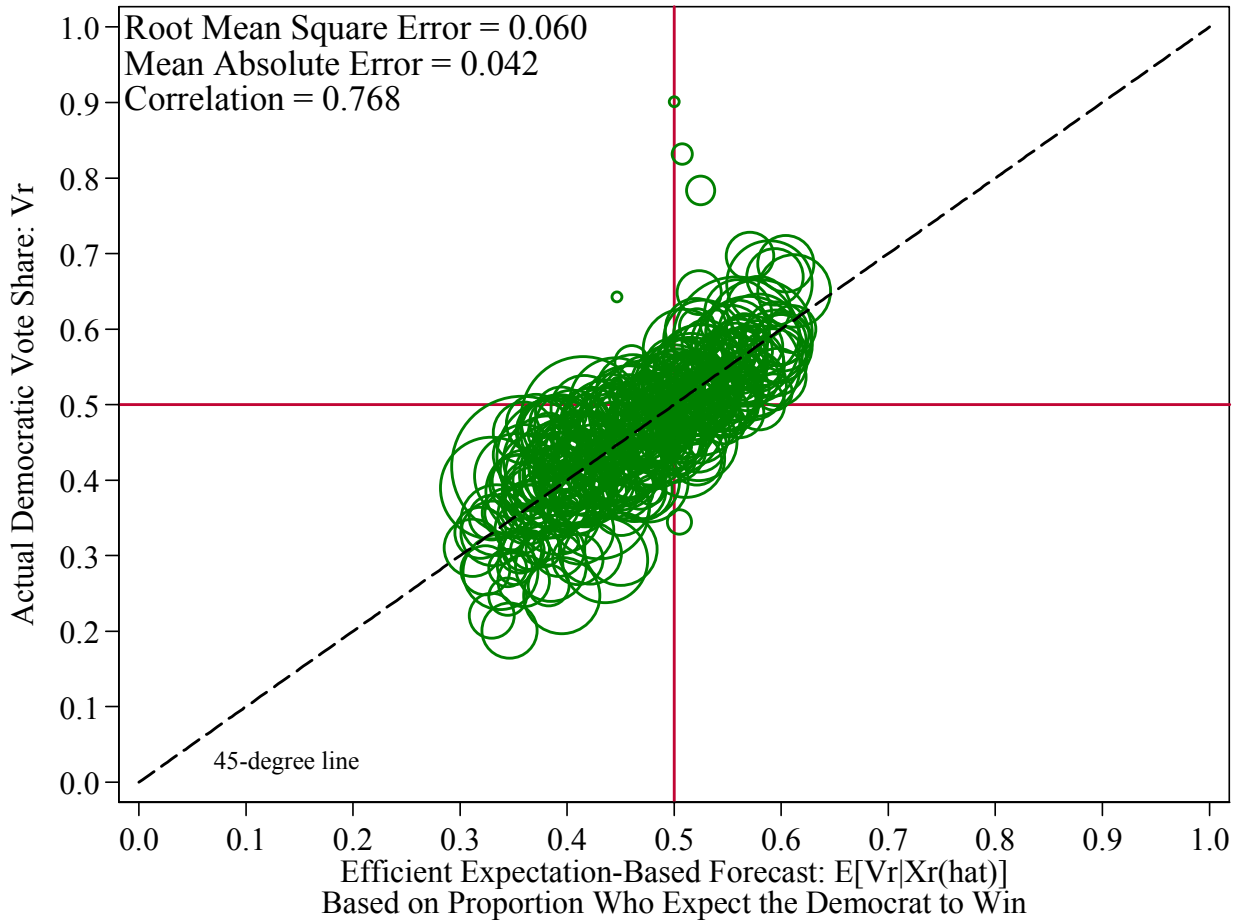
We fit this equation using non-linear least squares, and it yields an estimate of $\widehat{\rho_x} = 0.045$ (se $=$ 0.010, which in turn implies an average design effect $\widehat{\gamma^x} = 2.62$, and that the shrinkage estimator which has an average value of 0.65, ranges from 0.14 to 0.76. Thus, our MSE-minimizing forecast of the Democrats vote share, based only the voter expectations data is:

$$E[v_r|\widehat{x_r}] = 0.5 + 0.150\Phi^{-1}\left(0.440 + \frac{0.040}{0.040 + \frac{1 + (n_r - 1)0.045}{4n_r}}(\widehat{x_r} - 0.043 - 0.427)\right)$$

In Figure 6, we show the relationship between our efficient expectation-based forecast and actual election outcomes. These adjusted expectation-based forecasts are clearly very accurate, and appropriately scaled: the forecasts lie along the 45-degree line.

[11] Notice that equation [13] involves the product of the *population* proportions who expect each candidate to win, and it is this product that ≈0.25. As previously noted, we don't actually observe these population proportions, but we do observe the population vote shares, and using the transformation in equation [5], we confirm that $\Phi\left(\frac{v_r - 0.5}{\widehat{\sigma_\epsilon}}\right)\left(1 - \Phi\left(\frac{0.5 - v_r}{\widehat{\sigma_\epsilon}}\right)\right) \approx 0.25$. In our small samples, $\widehat{x_r}(1 - \widehat{x_r})$ can diverge quite significantly from 0.25.

Figure 6: Efficient Expectation-Based Forecast and Election Outcomes



Forecast evaluation

Turning to Table 4, we see that these efficient forecasts based on voter expectations are more accurate than efficient intention-based forecasts. Again, the first two rows show that the expectation-based forecasts yield both a root mean squared error and mean absolute error that is less than the intention-based forecasts by a statically significant amount. The third row shows that the expectation-based forecasts are also the more accurate forecast in 63% of these elections, very similar to the naïve approach. The expectation-based forecasts are still more highly correlated with actual vote shares than are the intention-based forecasts by a sizable margin. In the Fair-Shiller regression, the expectation-based forecasts have a much larger weight than the intention-based forecast; both are statistically significant, so the intention-based data is providing some unique information. Finally, an optimally weighted average still puts just over 90% of the weight on the expectation-based forecast. While that number may not seem remarkable in the context of this paper, it is remarkable in the context of society. If you take step back

from this paper, consider that the average weight placed on expectation polls by pollsters, the press, and campaigns is 0%, as it is generally ignored.

Table 4: Comparing the Accuracy of Efficient Forecasts of Vote Share

	Efficient Voter Intention: $E[v_r \mid \widehat{v_r}]$	Efficient Voter Expectation: $E[v_r \mid \widehat{x_r}]$	Test of Equality
Root Mean Squared Error	0.076 (0.005)	0.060 (0.006)	t_{310}=5.75 (p<0.0001)
Mean Absolute Error	0.056 (0.003)	0.042 (0.002)	t_{310}=6.09 (p<0.0001)
How often is forecast closer?	37.0% (2.6)	63.0% (2.6)	t_{310}=4.75 (p<0.0001)
Correlation	0.593	0.768	
Encompassing regression: $v_r = \alpha + \beta_v Intention_r + \beta_x Expectation_r$	0.184** (0.089)	0.913*** (0.067)	
Optimal weights: $v_r = \beta Intention_r + (1-\beta) Expectation_r$	9.5% (6.7)	90.5%*** (6.7)	

Notes: ***, **, and * denote statistically significant coefficients at the 1%, 5%, and 10%, respectively. (Standard errors in parentheses). These are assessments of forecasts of the Democrat's share of the two-party vote in n=311 elections. Comparisons in the third column test the equality of the measures in the first two columns. In the encompassing regression, the constant $\hat{\alpha} = -0.046$ (se=0.030).

V. Efficient Poll-Based Probabilistic Forecasts

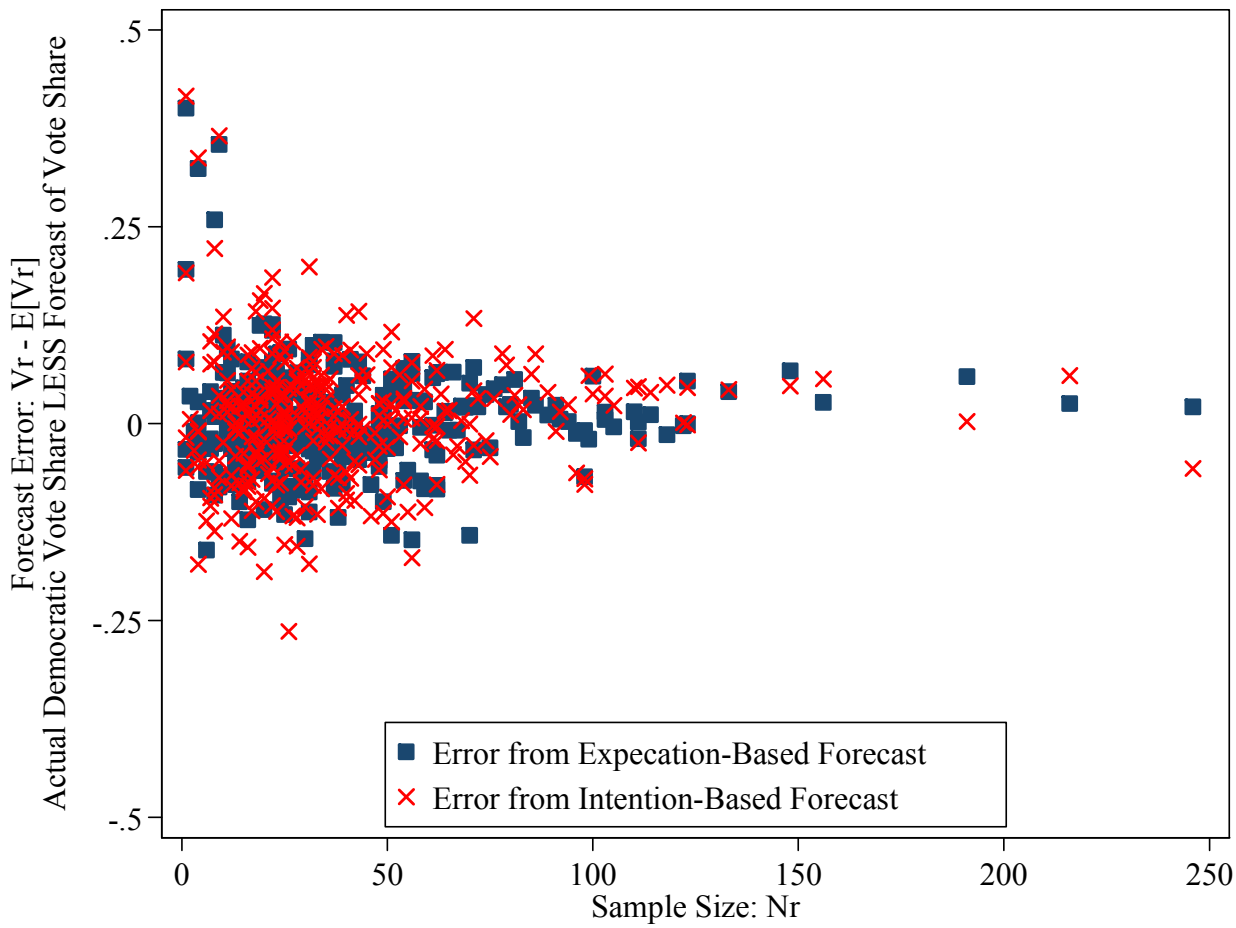
So far we have introduced two different outcome variables: the prediction of the winner and the level of the outcome, but frequently the most desirable outcome metric for stakeholders of an event is the probabilistic forecast. In this section we translate our raw polling data, from both voter intention and expectation, into efficient poll-based probabilistic forecasts. We start with our forecast of the vote shares generated from both types of data: $E[v_r \mid \widehat{v_r}]$ for intention data and $E[v_r \mid \widehat{x_r}]$ for expectation data. The probability that the Democrat wins the race is the probability that the actual vote share is greater than 50%.[12] We assume that the errors of our expectations are normally distributed and the probability of victory for the Democratic candidate becomes:

$$Prob(v_r > 0.5) = \Phi\left(\frac{E[v_r] - 0.5}{\sigma_{\varepsilon, E[v_r]}}\right) \qquad [15]$$

[12] This is not true in many types of elections, but it is true in almost all Electoral College races in our dataset.

To create probabilistic forecasts, we are going to need to determine the variance of the error. Figure 7 revisits Figure 4, but instead of showing the errors from the raw intention poll, it shows the errors from the efficient intention and expectation-based forecasts. The figure illustrates that there is still noise and that it varies with sample size.

Figure 7: Sample Size and Forecast Errors in the Forecasts of Vote Share



To determine the variance of the error, we start with the standard error of a forecast: $se = \frac{RMSE}{\sqrt{n}}\left(1+\frac{(X_r-\mu_x)^2}{\sigma_x^2}\right)^{1/2}$ (Barreto and Howland, 2006). We use equation [11], $v_r = \widehat{\mu_v} + \frac{\widehat{\sigma_v^2}}{\widehat{\sigma_v^2}+\widehat{\sigma_\tau^2}+\frac{1+(n_r-1)\rho}{4n_r}}\left(\widehat{v_r}-\hat{b}-\widehat{\mu_v}\right)$ and we think of the $\left(\widehat{v_r}-\hat{b}-\widehat{\mu_v}\right)$ as the raw data X_r, where μ_x is the average raw data. Thus, the variance of the forecasts:

$$\widehat{\sigma^2_{\varepsilon,E[v_r]}} = RMSE^2 + \frac{RMSE^2}{\sigma^2_{poll}}\left(poll-\mu_{poll}\right)^2 = RMSE^2 + \sigma^2_{Shrinkage}\left(poll-\mu_{poll}\right)^2 \qquad [16]$$

The root mean square error (squared) is of the two equations we use to determine the shrinkage estimate; equation [11] for intention and equation [14] for expectation. It captures the accuracy of the equation in estimating the within sample data used to calibrate the forecast model. The other part of the variance is the standard deviation of the shrinkage estimator (squared) interacted with the distance between the raw data and the mean data. The intuition behind this is that we should assume less accurate forecasts, the less accurate the estimation of the coefficient we use to transform our raw data is and how unusual the raw data is compared to the average raw data. For the expectation data, we use the "naïve" forecast of vote share as the raw data. This term is going to pick up the noise correlated with sample size, as low sample sizes have more polls that post very far from the average poll.

The RMSE for intention is 0.076 and expectation 0.060; as we know from Section IV, the forecasts of vote share from the expectation data is more accurate. $\frac{RMSE^2}{\sigma^2_{poll}} = \sigma^2_{Shrinkage}$ is 0.177 for intention and 0.186 for expectation. On average, about half of the $\sigma_{\varepsilon,E[v_r]}$ comes from the each side of the equation and the estimated standard deviation does decrease with sample size. Putting together equation [15] and [16]:

$$Prob(v_r > 0.5) = \Phi\left(\frac{E[v_r] - 0.5}{\sqrt{RMSE^2 + \frac{RMSE^2}{\sigma^2_{poll}}\left(poll - \mu_{poll}\right)^2}}\right) \quad [17]$$

In Table 5 we show that probabilistic forecasts based on the expectation data are more accurate than those based on the intention data. The first row shows that the expectation-based probabilities yield a smaller root mean squared error than the intention-based probabilities by a statically significant amount. The second row shows that the expectation-based probabilities are also the more accurate forecast in 80% of these elections, a statistically significant difference. In the Fair-Shiller regression, the expectation-based probabilities have a much larger weight than the intention-based forecast; the intention-based data is not statistically significant, so we cannot discount the possibility that the intention-based probability is providing no unique information. Finally, an optimally weighted average puts most of the weight on the expectation-based probability.

Table 5: Comparing the Accuracy of Efficient Probabilistic Forecasts

	Efficient Voter Intention: $Prob(v_r > 0.5 \mid \widehat{v_r})$	Efficient Voter Expectation: $Prob(v_r > 0.5 \mid \widehat{x_r})$	Test of Equality
Root Mean Squared Error	0.442 (0.007)	0.345 (0.012)	t_{344}=10.4 (p<0.0001)
How often is forecast closer?	20.0% (2.3)	80.0% (2.3)	t_{344}=13.2 (p<0.0001)
Encompassing regression: $I(DemWin)_r = \Phi\left(\alpha + \beta_v \Phi^{-1}(Prob_I) + \beta_x \Phi^{-1}(Prob_x)\right)$	0.216 (0.155)	1.776*** (0.201)	
Optimal weights: $I(DemWin)_r = \Phi\left(\beta \Phi^{-1}(Prob_I) + (1-\beta)\Phi^{-1}(Prob_x)\right)$	-0.784*** (0.185)	1.784*** (0.185)	

Notes: ***, **, and * denote statistically significant coefficients at the 1%, 5%, and 10%, respectively. (Standard errors in parentheses). These are assessments of forecasts of the Democrat's probability of victory in n=311 elections. Comparisons in the third column test the equality of the measures in the first two columns. In the encompassing regression, the constant $\hat{\alpha} = -0.035$ (se=0.114).

VI. Efficient Poll-Based Forecasts of 2008

We now test how the coefficients, developed with data from 1952-2004, forecast the 2008 Electoral College races. Since this is an "out-of-sample" comparison, the forecast of vote share use the coefficients derived in Section IV and the probabilistic forecast use the coefficients derived in Section V. In Table 2 we show that 2008 is just one of two years in which the expectation poll is not more correct in predicting the winner of the races than the intention poll; they were tied in 2008. Yet, Table 6 shows that the expectation question provides a much more accurate and informative forecast of 2008 than the intention question. The expectation-based forecasts have smaller errors and higher correlation in all of the forecast for vote share and probability of victory comparisons, and carry more weight in all of the joint estimations for forecast of vote share and partially in probability of victory. Since the expectation-based forecasts are statistically significant in encompassing regression and the intention-based forecast is not, we cannot rule out that intention data provides no useful information beyond expectation data in 2008.

Table 6: Comparing the Accuracy of Efficient Forecasts for the 2008 Electoral College

Forecast of Vote Share:	Efficient Voter Intention: $E[v_r \mid \widehat{v_r}]$	Efficient Voter Expectation: $E[v_r \mid \widehat{x_r}]$	Test of Equality
Root Mean Squared Error	0.093 (0.021)	0.085 (0.022)	t_{33}=1.28 (p<0.2105)
Mean Absolute Error	0.063 (0.012)	0.056 (0.011)	t_{33}=0.92 (p<0.3656)
How often is forecast closer?	47.1% (8.7)	52.9% (8.7)	t_{33}=0.34 (p<0.7371)
Correlation	61.6%	69.2%	
Encompassing regression: $v_r = \alpha + \beta_v Intention_r + \beta_x Expectation_r$	0.330 (0.291)	0.684*** (0.250)	
Optimal weights: $v_r = \beta Intention_r + (1-\beta) Expectation_r$	24.7% (26.7)	75.3%*** (26.7)	
Probabilistic Forecasts:	$Prob$ $(v_r > 0.5 \mid \widehat{v_r})$	$Prob$ $(v_r > 0.5 \mid \widehat{x_r})$	
Root Mean Squared Error	0.458 (0.022)	0.403 (0.048)	t_{344}=1.55 (p<0.1295)
How often is forecast closer?	23.5% (7.4)	76.5% (7.4)	t_{344}=3.58 (p<0.0011)
Encompassing regression: $I(DemWin)_r = \Phi\left(\alpha + \beta_v \Phi^{-1}(Prob_I) + \beta_x \Phi^{-1}(Prob_x)\right)$	1.618 (1.289)	1.224** (0.520)	
Optimal weights: $I(DemWin)_r = \Phi\left(\beta \Phi^{-1}(Prob_I) + (1-\beta) \Phi^{-1}(Prob_x)\right)$	2.4% (39.1)	97.6%** (39.1)	

Notes: ***, **, and * denote statistically significant coefficients at the 1%, 5%, and 10%, respectively. (Standard errors in parentheses). These are assessments of forecasts of the Democrat's forecast of voter share and probability of victory in n=34 elections (the 34 Electoral College races chosen by the ANES). Comparisons in the third column test the equality of the measures in the first two columns. In the encompassing regression, the constant $\hat{\alpha} = 0.032$ (se=0.097) for vote share and $\hat{\alpha} = 1.242$ (se=0.504) for probability of victory.

VII. Efficient Poll-Based Forecasts from Non-Random Samples

Under many circumstances we are left making forecasts based off polls with overtly non-random samples. An extreme example of that would be a poll where all of the respondents support just one of the candidates. In this section we test the accuracy of forecasts based off of the expectation question data originating from respondents who are exclusively voting for one major party or the other, by comparing

those forecasts to forecasts created from the full sample of intentions. For the non-random samples, the sample size drops from an average of 38 to about 19 respondents.

We start by revisiting equation [12] from Section IV: $E[x_r|\widehat{x_r}] = \mu_x + \frac{\sigma_x^2}{\sigma_{\hat{x}}^2}(\widehat{x_r} - b - \mu_x)$, where μ_x is the mean across all elections of the proportion of the population who expect the Democrat to win and σ_x^2 is the corresponding variance. These values are the same in this biased sample as they are in Section IV.[13] $\sigma_{\hat{x}}^2$ is the variance of the sample estimator, and our bias parameter, b, will attempt to account for the oversample people who expect Democrats to win (which should be heavily positive in a Democratic only sample and heavily negative in a Republican only sample).

The bias term is $\hat{b} = \sum(\widehat{x_r} - \Phi\left(\frac{v_r - 0.5}{\widehat{\sigma_\epsilon}}\right))/R$. This is 0.203 (0.016) for the Democratic sample and -0.112 (0.012) for the Republican sample. In Section IV, we conclude that the full dataset has a bias of 0.042, approximately half the difference in the absolute bias of the Democratic and Republican supporters.

We make the same assumptions as in Section IV in regard to $\widehat{\sigma_{\hat{x}}^2}$, which, combining equations [13] and [14], is $\widehat{\sigma_x^2} + \frac{(1+(n_r-1)\rho_x)}{4n_r}$. We are going to have to re-derive ρ for these two biased samples, as the clustering is going to have a totally different affect within party. Both samples have the same $\widehat{\sigma_x^2}$ and about half the observations as the full samples. This yields shrinkage estimators (i.e., $\frac{\sigma_x^2}{\sigma_{\hat{x}}^2}$) for the Democratic sample with an average of 0.59, ranges from 0.14 to 0.74. Not surprisingly, the full dataset has shrinkage estimators that are on average 7.5 percentage points larger, ranging from a low of the same to a higher of 54 percentage points larger. For the Republican sample, the differences are not as stark, with the Republican sample having slightly smaller $\widehat{\rho_x}$ than the full sample; the full sample has shrinkage estimators that are on average 2.5 percentage points larger, but they range from a low of -7 percentage points smaller to 42 percentage points larger.

In Table 7 we show that forecasts based on the expectations of the Democratic voters only or the Republican voters only, a non-random selection of approximately half of the sample, yield a lower root mean squared error, mean absolute error, and higher correlation than forecasts based on the full sample of voter intention. The first two rows show that the expectation-based forecasts yield both a root mean squared error and mean absolute error that is less than the intention-based forecasts by a statically

[13] There is a slight variation in the variables because 5 of the races in our dataset have no Democratic supporters and 4 no Republican supporters. Thus, when we drop those races from the dataset when we explore forecast accuracy of the respective non-random samples.

significant amount for the Republican sample and a weakly significant amount for the Democratic sample. The third row shows that the expectation-based forecasts are also the more accurate forecast in the majority of elections. The expectation-based forecasts are still more highly correlated with actual vote shares than are the intention-based forecasts by a sizable margin. In the Fair-Shiller regression, the expectation-based forecasts have a much larger weight than the intention-based forecast; both are statistically significant, so the intention-based data may be providing some unique information. Finally, an optimally weighted average still puts nearly 60% of the weight on the expectation-based forecast in the Democratic sample and just over 70% in the Republican sample.

Table 7: Comparing the Accuracy of Efficient Forecasts of Vote Shares from Biased Samples

	Democratic Sample		Republican Sample	
Forecast of Vote Share:	Intention: $\boldsymbol{E[v_r \mid \widehat{v_r}]}$ Full Sample	Expectation: $\boldsymbol{E[v_r \mid \widehat{x_r}]}$ Democratic Supporters	Intention: $\boldsymbol{E[v_r \mid \widehat{v_r}]}$ Full Sample	Expectation: $\boldsymbol{E[v_r \mid \widehat{x_r}]}$ Republican Supporters
Root Mean Squared Error	0.075 (0.005)	0.070 (0.006)	0.071 (0.004)	0.062 (0.004)
Mean Absolute Error	0.056 (0.003)	0.050 (0.003)	0.054 (0.003)	0.048 (0.002)
How often is forecast closer?	46.7% (2.9)	53.3% (2.9)	44.0% (2.8)	56.0% (2.8)
Correlation	0.592	0.664	0.604	0.718
Encompassing regression: $\boldsymbol{v_r = \alpha + \beta_v Intention_r + \beta_x Expectation_r}$	0.625*** (0.078)	0.790*** (0.071)	0.489*** (0.077)	0.786*** (0.065)
Optimal weights: $\boldsymbol{v_r = \beta Intention_r + (1-\beta) Expectation_r}$	38.5%*** (6.6)	61.5%*** (6.6)	29.8%*** (6.3)	70.2%*** (6.3)
	306 Elections		307 Elections	

Notes: ***, **, and * denote statistically significant coefficients at the 1%, 5%, and 10%, respectively. (Standard errors in parentheses). These are assessments of forecasts of the Democrat's share of the two-party vote. In the encompassing regression with the Democratic sample, the constant $\hat{\alpha} = -0.196$ (se=0.036)*** and in the Republican sample, the constant $\hat{\alpha} = -0.129$ (se=0.031)***. A few of the 311 observations are dropped in each sample, because there is no intention for the given party.

In Table 8 we show that probabilistic forecasts based on the expectation data are more accurate than those based on the intention data. The first row shows that the expectation-based probabilities yield a smaller root mean squared error than the intention-based probabilities by a statically significant amount. The second row shows that the expectation-based probabilities are also the more accurate in slightly more elections. In the Fair-Shiller regression, the expectation-based probabilities have a much larger

weight than the intention-based forecast; both are statistically significant, so the intention-based data may be providing some unique information. Finally, an optimally weighted average still puts most of the weight on the expectation-based forecast.

Table 8: Comparing the Accuracy of Efficient Probabilistic Forecasts

	Democratic Sample		Republican Sample	
Probabilistic Forecasts:	Voter Intention:	Voter Expectation:	Voter Intention:	Voter Expectation:
Root Mean Squared Error	0.444 (0.006)	0.388 (0.010)	0.442 (0.006)	0.357 (0.013)
How often is forecast closer?	28.4% (2.6)	71.5% (2.6)	19.9% (2.3)	80.1% (2.3)
Encompassing regression: $\boldsymbol{I(DemWin)_r = \Phi\left(\alpha + \beta_v \Phi^{-1}(Prob_I) + \beta_x \Phi^{-1}(Prob_x)\right)}$	1.73*** (0.40)	1.62*** (0.20)	1.29*** (0.41)	1.53*** (0.17)
Optimal weights: $\boldsymbol{I(DemWin)_r = \Phi\left(\beta \Phi^{-1}(Prob_I) + (1-\beta)\Phi^{-1}(Prob_x)\right)}$	-0.336** (0.170)	1.336*** (0.170)	-0.435*** (0.152)	1.435*** (0.152)
	306 Elections		307 Elections	

Notes: ***, **, and * denote statistically significant coefficients at the 1%, 5%, and 10%, respectively. (Standard errors in parentheses). These are assessments of forecasts of the Democrat's probability of victory. In the encompassing regression with the Democratic sample, the constant $\hat{\alpha} = -0.009$ (se=0.093). In the encompassing regression with the Republican sample, the constant $\hat{\alpha} = 0.235$(se=0.103). A few of the 311 observations are dropped in each sample, because there is no intention for the given party.

VIII. Efficient Poll-Based Forecasts from Secondary Dataset

Our secondary dataset consists of any poll from around the world that we could find that asks both an intent and expectation question. The polls are a hodgepodge of USA and non-USA elections, executive and legislative offices, and the full range of national elections to smaller districts.

Table 9 summarizes the forecast performance of our two questions in forecasting the winning candidate, mimicking the input in Table 2. Again, we use a very coarse performance metric, simply scoring the proportion of races in which the candidate who won a majority in the relevant poll ultimately won the election. The expectation question is correct more often than the intent question in both 0 to 90 and 90 to 180 days before the election; this difference is statistically significant. Indeed this difference is largest in the 90 to 180 day segment. For reasons we are not quite sure of, the expectation question is essentially random after 180 days, in the data in our dataset, but many of those polls are actually a full year and beyond before the election.

Table 9: Comparing the Accuracy from Secondary Dataset

	Days Before the Election ≤ 90				90 < Days Before the Election ≤ 180				Days Before the Election > 180			
	Proportion of observations where the winning candidate was correctly predicted by a majority of respondents to:											
	Expect	Intent	# Obs	# Elections	Expect	Intent	# Obs	# Elections	Expect	Intent	# Obs	# Elections
President	89.4%	80.7%	161	19	69.2%	61.5%	39	12	59.6%	57.7%	52	11
1936 Electoral College	72.3%	80.9%	47	47	-	-	0	0	-	-	0	0
Governor	78.9%	78.9%	19	9	83.3%	50.0%	6	6	100.0%	100.0%	2	1
Senator	81.8%	90.9%	11	7	-	-	0	0	-	-	0	0
Mayor	100.0%	100.0%	4	2	100.0%	66.7%	3	1	-	-	0	0
Other	80.0%	80.0%	10	9	100.0%	66.7%	3	2	50.0%	50.0%	2	2
USA Total	84.9%	81.3%	252	93	74.5%	60.8%	51	21	60.7%	58.9%	56	14
AUS (Parliament)	88.9%	41.7%	36	3	66.7%	33.3%	21	3	24.4%	66.3%	86	2
GBR (Parliament)	85.0%	90.0%	20	9	100.0%	92.3%	13	7	69.4%	62.9%	62	9
FRA (President)	60.9%	56.5%	23	4	40.0%	20.0%	5	3	-	-	0	0
Other	71.4%	71.4%	7	6	0.0%	0.0%	1	1	0.0%	0.0%	1	1
non-USA Total	79.1%	59.3%	86	22	72.5%	50.0%	40	14	43.0%	64.4%	149	12
Total	83.4%	75.7%	338	115	73.6%	56.0%	91	35	47.8%	62.9%	205	26
Diff (standard error)	7.7%* (2.2)				17.7%* (4.8)				-15.1%* (4.4)			

IX. A Structural Interpretation

At this point, it's worth reflecting on just why it is that the expectation-based forecasts are more accurate. The basic intuition is simply that each of us possesses substantial information that is relevant to forecasting the election outcome, but the voter intention question only elicits part of this information set. By asking about each respondent's expectation, a poll can effectively aggregate this broader information set.

We can formalize this idea with a very simple model that also does surprisingly well at matching the key facts about voter expectations. We conceptualize a survey respondent's expectation as deriving from her knowledge of her own voting intention, as well as those of $m-1$ of her friends, family, and coworkers, effectively creating a survey of m likely voters. We denote the proportion of person i's sample who plan to vote for the Democrat as $\widehat{v_r^i}$. If each person's individual "informal poll" is drawn from an unbiased sample, then $\widehat{v_r^i}$ is a random variable with mean v_r and variance of $v_r(1-v_r)/m$. Using the normal approximation to the binomial distribution, this suggests that the probability an individual respondent expects the Democrat to win is:

$$Prob\left(\widehat{v_r^i} > 0.5\right) = \Phi\left(\frac{v_r - 0.5}{\sqrt{\frac{v_r(1-v_r)}{m}}}\right) \approx \Phi\left(2\sqrt{m}(v_r - 0.5)\right) \qquad [18]$$

where $\Phi(.)$ is the standard normal cdf. The approximation equality follows because in competitive elections $1/\sqrt{v_r(1-v_r)} \approx 2$.

Thus, equation [15] suggests that in a simple probit regression of whether an individual forecasts the Democrat to win, the coefficient on the winning (or negative losing) margin of the Democrat candidate reveals $\sqrt{m}$.[14] The intuition here can be explained by thinking about two extreme cases. If each survey respondent based their expectations on an informal poll of thousands of friends, then even in very close races, most of them will agree on which candidate they expect to win. But if each respondent polls only a couple of friends, then sampling variation will have a larger influence, and there will be much greater disagreement over the likely winner. Thus, the basic logic identifying our estimate of m can be seen in Figure 2, which shows the likelihood that the expectations of survey respondents are consistent with each other, as a function of how close the race is.

In fact, we have already run the regression described in equation [18]—it is identical to that described in equation [4], which we used to estimate $\widehat{\sigma_e}$. Comparing these equations we see that $2\sqrt{m} = \frac{1}{\sigma_\epsilon}$, or $m = \frac{1}{4\sigma_\epsilon^2}$. Thus given our estimate of $\widehat{\sigma_\epsilon} = 0.150$, we infer that $m = 11$ (the exact coefficient is 11.11, and the standard error clustering by state-year and applying the delta method is 1.12). That is, our simple model suggests that each poll respondent bases her expectation on a poll of herself, plus 10 friends.

We should add a qualifier to this interpretation which highlights that this simple model also serves as a metaphor for a slightly more realistic scenario. In particular, it is likely that the social network of any survey respondent is not a representative random sample. It may be that their social networks contain a known partisan bias or that specific clusters of friends have correlated voting intentions. Our point is simply that the voter expectation question draws upon a broader information set, whose forecasting value is similar to that of an unbiased or random sample with an effective sample size of eleven.

Table 10 shows the relationship between intention, the winning candidate, and expectation in the primary dataset. In approximately half of our individual-level observations (48.8%), the respondent has the same expectation and intention and that candidate wins the elections. In less than 10% of individual-level observations (9.2%) the respondent does not expect the candidate for whom she intends to vote to win, but the candidate does win. In the remaining 42% of observations, where the voter intention is not

[14] We can avoid the approximation noted above, and run a probit regression where the dependent variable is instead $\frac{p_s-0.5}{\sqrt{p_s(1-p_s)}}$; this yields similar results.

for the winning candidate, the respondents expect the winner 19.7% and expect the candidate they intend 22.3%. Another way to look at the data is that in 71.1% of observations the respondent expecting their intended candidate to win: 48.8% correctly and 22.3% incorrectly. And 28.9% of observations the respondent expects their intended candidate to lose: 19.7% correctly and 9.2% incorrectly.

Table 10: Relationship between Voter Intention, the Winner, and Expectation

Intention Democratic	Winner Democratic	Expectation Democratic	Percent of Respondents
1	1	1	19.0%
1	0	1	15.8%
1	1	0	3.4%
1	0	0	12.3%
0	1	1	7.5%
0	0	1	5.8%
0	1	0	6.5%
0	0	0	29.8%

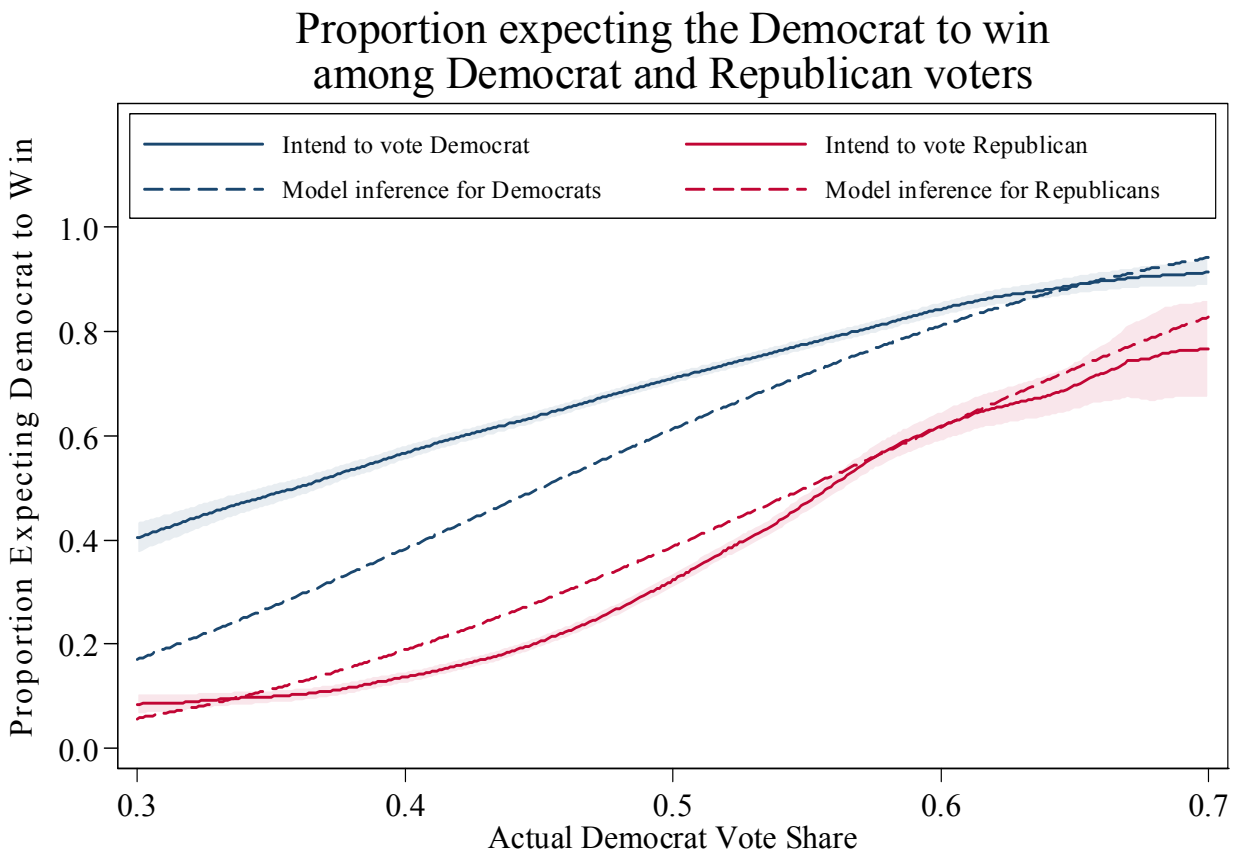
Notes: There are n=13,208 individual-level observations, including 2008 data.

Figure 8 charts the proportion of respondents that expect the Democratic candidate to win for any given actual vote share. We do so for both the full data set of 13,208 individual responses and the structural interpretation from above. From equation [18], if a random respondent has a probability of $\Phi\left(2\sqrt{11}(v_r - 0.5)\right)$ of expecting the Democratic candidate to win, a Democratic supporter has a $\Phi\left(2\sqrt{10}(v_r - 0.45)\right)$ probability of expecting the Democratic candidate to win. One of the eleven votes is guaranteed to the Democrats, so that only 5 of the remaining 10 votes need to go Democratic for the respondent to expect the Democratic candidate to win. Likewise, a Republican supporter expects the Democratic candidate to win with a probability of $\Phi\left(2\sqrt{10}(v_r - 0.55)\right)$.

Table 10 and Figure 8 confirm that there is a mix of information reaching the respondents about the election. As in Figure 2, Figure 8 shows a positive correlation between vote share and expecting to win, for both types of intended voters. If the respondents only considered their own intention, they would expect their own intended candidate 100%, but only do so 71.1% (Table 10) and there would be a straight line of circles at 0 for the Republicans and 100 for the Democrats (Figure 8). If the respondents only considered an unbiased signal, respondents of different intentions would be equally likely to expect the other candidate, at a given vote share. Figure 8 illustrates that respondents are much more likely to expect a candidate to win, at any given vote share, if they intend to vote for that candidate. We know that there

is a noisy signal; 9.2% of respondents switch their expectation away from their intended candidate to the losing candidate (Table 10). Further, the noisy signal cannot be too biased towards the voter's intention, because if all signals were completely biased towards their intention then there would be no explanation for a respondent to expect a losing candidate to win, despite a preference for that candidate. Finally, the figure shows that our simple structural interpretation is a reasonable fit for the actual data.

Figure 8: Relationship between Intention, Expectation, and Actual Vote Share



Local linear regression estimates, using Epanechnikov kernal and rule-of-thumb bandwidth.
Shaded area shows 95% confidence interval.

X. Discussion

A Fox News poll, in the field on September 8 and 9 of 2008, prompted a headline trumpeting McCain's lead in the intent question, but buried Obama's continued lead in the expectation question.[15] This is an editorial choice that almost any polling or news organization would have made. Yet, while McCain and Obama occasionally traded the lead in the intent question, Obama, the eventual winner, led

[15] http://www.foxnews.com/story/0,2933,420361,00.html

in every expectation question we could find. We hope that this paper will give editors, as well as pollsters, campaigns, and researchers, pause in future elections, as we have proved the expectation question significantly useful in creating accurate forecasts of all major outcome variables.

The structural interpretation of the response to the expectation question helps illustrate why expectations are such a powerful polling tool. The answer we receive from the expectation question includes all of the information in the intention question as well as the intention of approximately ten friends, family, and coworkers who are likely voters. This multiplication of the sample size is crucial in making the expectation question remarkably valuable with small sample sizes. That phenomenon, combined with the identification of likely voters, is what enables expectation polls to provide valuable information, even with a non-random sample of respondents.

The structural interpretation offers clues into the ratio of the sources of data that people are using in creating expectations of elections and this can be applied to a wide set of disciplines. One example is questions about economic voting. There is very important question about whether people vote (or frame their political preferences in general) on their own pocketbook (egotropic) or the national economy (sociotropic). We have a tool to look at which source of information is framing their expectations, where their expectations are the main input in their preferences or revealed utility. The same example can be done for the popularity of a new style of jeans in marketing or for consumer sentiment on durable goods in coming months in economics. In any situation where individual-level stakeholders are exposed to private information and public signals the results and methods of this paper illustrate meaningful information that can be extracted by individual-level questions of expectations.

XI. References

Alford, Richard F. 1977. "Estimation of the Probability of Bryan Victory in 1896" Presented at the 1977 *Cliometrics Conference.*

Barreto, Humberto and Frank Howland. 2006. *Introductory Econometrics: Using Monte Carlo Simulation with Microsoft Excel*. New York, NY: Cambridge University Press.

Berg, Joyce E., and Thomas A. Rietz. 2006."The Iowa Electronic Market: Stylized Facts and Open Issues.". *In: Hahn Robert W., Tetlock Paul, editors. Information Markets: A New Way of Making Decisions in the Public and Private Sector*, eds. Robert W. Hahn, and Paul Tetlock. Washington, DC*:* AEI-Brookings Joint Center.

Berg, Joyce E., Forrest D. Nelson and Thomas A. Rietz. 2008. "Prediction Market Accuracy in the Long Run," *International Journal of Forecasting* 24(2):283-298.

Campbell, James E. 2000. *The American Campaign*. College Station: Texas A&M University Press.

Erikson, Robert S., and Christopher Wlezien. 2008. "Are Political Markets Really Superior to Polls as Election Predictors? " *Public Opinion Quarterly* 72:190-21.

Fair, Ray, and Robert Shiller. 1989. "The Informational Content of ex-Ante Forecasts." *Review of Economics and Statistics* 71(2):325-31.

----------. 1990. "Comparing Information in Forecasts From Econometric Models." *American Economic Review* 80(3):375-89.

Granberg, Donald and Edward Brent. 1983. "When Prophecy Bends: The Preference-Expectation Link in U.S. Presidential Elections." *Journal of Personality and Social Psychology* 45(3):477-91.

Lock, Kari and Andrew Gelman. 2010. Bayesian Combination of State Pols and Election Forecasts. *Political Analysis*, 18(3):337-348.

Imai, Masami and Cameron Shelton. 2010. "Elections and Political Risk: New Evidence from Political Prediction Markets in Taiwan." Wesleyan Economics Working Papers.

Irwin, Galen and Joop Van Holsteyn. 2002. "According to the Polls, the Influence of Opinion Polls on Expectations." *Public Opinion Quarterly* 66:92-104.

Moulton, Brent. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unites." *The Review of Economics and Statistics* 72(2):334-338.

Mutz, Diana. 1995. "Effects of Horse-Race Coverage on Campaign Coffers: Strategic Contributing in Presidential Primaries." *The Journal of Politics* 57(4):1015-1042.

Robinson, Claude E.. 1937. "Recent Developments in the Straw-Poll Field" *Public Opinion Quarterly* 1(3):45-56.

Rothschild, David. 2009. “Forecasting Elections, Comparing Prediction Markets, Polls, and Their Biases.” *Public Opinion Quarterly* 73(5):895-16.

Wolfers, Justin and Eric Zitzewitz. 2004. “Prediction Markets.” *Journal of Economic Perspectives*, 18(20 107-126.