Lies, Damn Lies, and Pre–Election Polling

Elias Walsh
University of Michigan, Ann Arbor,
fgelias@umich.edu,
Sarah Dolfin
Mathematica Research Inc.,
SDolfin@mathematica-mpr.com,
and John DiNardo,
University of Michigan, Ann Arbor and NBER
jdinardo@umich.edu

Corresponding Author: John DiNardo

Session Title: Polls, Votes, and Elections

Session Chair: JOHN DINARDO, University of Michigan

Discussants: MIREILLE JACOBSON, University of California-Irvine

HEATHER ROYER, University of California, Santa Barbara

Web Appendices:

1. Ten (10) web appendix tables.
2. Eleven (11) web appendix figures.
3. Five page discussion of intentions problem.

January 8, 2009

Lies, Damn Lies, and Pre-Election Polling

By Elias Walsh, Sarah Dolfin and John DiNardo*

In this paper we ask the question: how well do pre–election polls forecast the actual results of elections in the U.S.? The question is interesting for a number of reasons. First, even polling data suggests about 1/3 of polling respondents do not believe that polls work in "the best interests of the general public."[1] The situation is such that even many national governments have undertaken to restrict some aspect of pre–election polling. A 1997 international survey of governments, for example, found 30 of 78 surveyed nations had some kind of ban on publication of poll results (Røhme, 1992). Second, there is a a strong presumption in the literature on professional forecasting in other contexts which do not rely on sampling *per se*, (such as interest rate forecasting) that forecasts will be biased.[2] There are a variety of explanations for why forecasts will be biased; one "honest" motivation is that pollsters may avoid reporting results from the unavoidable "atypical" polls. Third, in the literature in economics it is sometimes assumed that polls are unbiased forecasts (of potentially time–varying) underlying preferences for candidates. For a recent example, see Keppo et al. (2008) who characterize pre–election polling as a "noisy observation of the actual election outcome that would have obtained that day." Fourth, unlike much "opinion" polling, it is possible (albeit imperfectly) to verify the accuracy of the poll. It is therefore possible, with certain caveats, to compare the behavior of polls to what might be expected from probability sampling.

Although the art of polling has become considerably more sophisticated in some respects, the *practice* of polling is a far cry from a textbook description of the power of random sampling and the central limit theorem. Indeed, our analysis of pre–election polling in presidential races suggests some reason for skepticism. Our view is that presidential pre–election polling should be considered to be an activity more akin to predicting next year's GDP or the winner of a sporting match than to something resembling scientific random sampling.

To illustrate the possible problem, consider the 43 "last–minute" national horse race polls from `pollingreport.com` (see Web-appendix Table 1) for the 2000 U.S. Presidential Election. This election is particularly well-suited for illustration of the problem since the actual vote was a virtual "tie" (with Gore actually winning the popular vote) and the predictions were generally for a close election. Only 3 of the 42 polls predicted either a tie or Gore ahead in the national race.

While such an analysis might be considered unfair to pollsters, in actual fact, the pollsters themselves appear to have felt that they did "well." Traugott (2001), for example, observes that the performance of the 2000 pre-election presidential polls stands in stark (favorable) contrast to the their performance in the 1996 Presidential election. In that election, the well–respected director of the Roper Center argued that poll performance was so bad that it represented an "American Waterloo" (Ladd, 1996) despite the fact that the polls were virtually unanimous in picking Clinton the winner of the election. Ladd (1996) argued that the systematic overprediction of Clinton's vote share required a

* DiNardo: University of Michigan, School of Public policy, `jdinardo@umich.edu`. Dolfin: Mathematica, Mathematica Policy Research Inc., NY, NY, `SDolfin@mathematica-mpr.com`. Walsh: University of Michigan, Department of Economics and School of Public Policy, `fgelias@umich.edu`. We would like to thank Carolina Krawiec, Amy Kandilov, and Il Myoung Hwang for excellent research assistance.

[1] More than two thirds of the respondents to the same poll doubted that a random sample of 1,500 people can "accurately reflect the views" of the American public (Pew Research Center, 1998). This, of course, could reflect skepticism about the central limit theorem as well as issues such as non–response!

[2] See, for example, Ehrbeck and Waldmann (1996) or Ottaviani and Norman (2006).

national review of the pollsters.[3]

For our purpose, what is of immediate import is how *unlikely* it is that these polls – conducted by well-regarded polling agencies – are generated by an unbiased procedure. Consultation of the tables for the binomial distribution reveals that the probability of 42 or more predictions for Bush out of the 45 displayed above is less than 5×10^{-7} percent.[4]

I. Background

Our chief argument is that pre-election presidential polls are more akin to forecasting models of economic activity or gambling than to the results of "scientific probability sampling."

Unlike forecasts of economic outcomes which routinely point to a "model" that is generally expected to be different for different forecasters, pre–election polls (and opinion polls in general) routinely characterize themselves as involved in *sampling*. Reports from polls are routinely accompanied by a "margin of error" which is a variant of the confidence interval.

One problem for our analysis which we can not evade is that it is possible that the intent of pollsters is not to forecast an election result, but to correctly sample the current "state of opinion". Since the current state of opinion can't be observed, maintaining this view requires maintaining a view that can't be rejected or accepted by any research design of which we are aware.

Nonetheless, it seems clear to us that for pre–election polls (at least close to an actual election) a primary reason why they are interesting to many is because they are viewed as forecasts of election results. This is also the view of some analysts as well: Crespi (1988) observes that "concluding that even if a poll were conducted immediately before an election, one cannot hope to measure voter preferences accurately enough to approximate election results closely is to impugn the meaningfulness of all polls. If polls cannot achieve such accurate predictability, why should we accept any poll results as having meaning relevant to real life? In fact, using the deviation of pre-election polls conducted close to election day from election results as a measure of accuracy does provide an objective criterion when evaluating alternative methodologies for measuring voting preferences."

Our approach to assessing bias in pre–election polls is to treat polls as reporting the sample means resulting from random sampling of voters. We find that polls do not fare well by this standard. We also observe that it is impossible to explain "why" polls are biased: there are too many different reasons.

II. Some Basic Problems With Polls

The polls we analyze are largely conducted by profit-making private firms who do not disclose key details of how they arrive at their estimates. Nonetheless, the most reputable pollsters readily acknowledge potential departures from probability sampling.

A. Non–response

A possible "deal–breaker" that makes pre–election sampling difficult or impossible is non–response. The 2004 National Elections Study had a non–response rate of 24 percent which varied with the time of year and level of media coverage (Stroud and Kenski, 2007). Non-response in telephone surveys can be more than 10 percentage points higher (Brehm, 1993). The case for pre–election horse race polls, is probably much worse. For example, take this snippet from an interview with the highly respected pollster John Zogby:

> Stewart: "How many people do you have to call... to get 1,300 [responses]?"
> Zogby: "Oh boy, figure about 10,000 telephone numbers."
> Stewart: "Really?"

[3]See also Panagakis (1999) and Mitofsky (1998) who, despite disagreeing on how "bad" the 1996 polling was, both document substantial statistical bias. See Moon (1999) for similar evidence from England. See Traugott (2001) for evidence from the 2000 U.S. Presidential Election and Butler and Kavanagh (1992) for the 1992 British Elections.

[4]In making this calculation we use the assumption that Gore (the Democratic candidate) and Bush (the Republican candidate) received exactly the same number of votes, and the polls were independent samples.

> Zogby: "Yeah, really. A lot of people are not home, and about 2 out of 3 people refuse."
> Stewart: "So why isn't the margin of error 70%?"[5]

In fact, ignoring sampling error and assessing the worst-case bounds (Horowitz and Manski, 1998) arising *only* from non–response bias produces an interval that ranges from $\max(0, \mu - 66)$ to $\min(100, \mu + 66)$. In one study which performed an informal version of the analysis suggested in DiNardo et al. (2005), Pew Research Center (1998) found significant differences between "amenable respondents" and "reluctant respondents" in a poll that was likely far more rigorous and expensive to conduct than the best of the pre–election presidential polls we study.[6] Adding the uncertainty involved in estimating (not sampling) voter participation to the above worst-case bound, almost any estimate could be obtained.

[5]Transcribed from televised interview with John Stewart (Zogby, 2004).

[6]The two groups differed in the amount of effort that was spent in trying to procure a response:

> Households in the Rigorous sample with listed telephone numbers – for whom a mailing address could be obtained – were sent an advance letter asking for their participation in the survey. A $2 bill was enclosed with this letter as an additional incentive. There was no limit on the number of attempts to complete an interview at every sampled telephone number – numbers were called throughout the survey period until an interview was completed. The calls were staggered over times of day and days of the week to maximize the chances of making a contact with a potential respondent. A random selection procedure was used to select the respondent to be interviewed in each household. In addition, all interview breakoffs and refusals were contacted up to two additional times in order to attempt to convert them to completed interviews. For households with a known mailing address, respondents who refused to be interviewed after two calls were sent a conversion letter by priority mail before they were called a third time. (Pew Research Center, 1998)

B. Uncertain Turnout, Uncertain Preferences

In the simplest case, where all voters are certain of their intentions and whether or not they will vote, a suitable probability sample would be sufficient to get an accurate prediction of an election outcome. With certain intentions but uncertainty about whether someone will actually vote or not, requires, at a minimum, an estimator of the form:

$$\overline{Y} = \sum_{i=1}^{N} P_i X_i$$

where P_i is the probability a person will vote and X_i is their certain outcome. To the extent that P_i is not 1 or zero, an estimate of the election outcome requires a *model* of participation since mere sampling cannot produce a valid estimate of participation even if it could produce a valid estimate of "opinion."

The problem is exacerbated by the possibility that some important fraction of voters are uncertain about which candidate they support. (Manski, 1990) Since pollsters generally ask respondents to express their intentions of voting for one candidate or the other as a binary variable, the poll could be biased as a forecast of the election result even if there was ready information on P_i and a proper probability sample was possible.

A simple example will make this clear. Imagine that people can express their preference as a probability from 0 to 1, and that no "surprises" or new information occurs between the time of a poll and the election. Furthermore, for simplicity, imagine voters are identical, are all (correctly) certain that they will vote and can express their views as having a 51 percent probability of voting for candidate A. Suppose further that they respond to the pollster by saying they would vote for candidate A if their underlying probability is greater than 0.5. In this simple example, the poll would record 100% of the vote for candidate A, but the election result would be 51%. Indeed, it is simple to construct examples where, over time, the poll and the underlying preferences of the electorate go in separate directions.[7]

III. Polling Data

In Web Appendix Table 2 we present descriptive information on the polling results we collected from pollingreport.com.[8] We focus on state level presidential polls completed on or after the first day of June in the relevant election year because these tend to be the most consistently well-reported and conducted. Our sample from the 2000, 2004, and 2008 elections contains 1761 polls with an average of about 12 polls per statewide race. Polling organizations sometimes distinguish between polls of "likely voters" and "all voters" and roughly 83 percent of our polls are from "likely voters." The mean reported size of a poll in our sample was $N = 702$.

As might be expected, there is considerable heterogeneity in the amount of polling activity by state reflecting "interest" in the outcome. The mean number of polls per race was about 12, although some races had as few as 1 poll and some as many as 80.

There are several problems with the data that deserve mention and some of these are summarized in web appendix Table 3 and Table 4.

First, some polls report "undecided" voters, and other polls simply drop some fraction of respondents. For virtually all of the analysis we assume that the missing data are "strongly ignorable" – that is, we assume that the "missing" or "undecided" individuals share preferences in the same proportion as those who announce a preference. If a poll reports 40 percent for candidate A, 40 percent for candidate B, 20 percent undecided, and no other candidates, our "adjusted" measure would assign both candidates 50 percent.[9] Web appendix Table 4 displays a tabulation of such cases. Nearly all of the polls in our sample require this adjustment. In our analysis, we also look at the "raw" (unadjusted) shares but focus on adjusted shares, leading to a more "optimistic" assessment of poll accuracy.

Second, the percentages are virtually always rounded to the nearest percentage point. This means that in some cases, the poll results do not sum to exactly 100 percentage points. A summary of this "adding up" problem is provided in web appendix Table 3. We handled this symmetrically to the undecided problem and used the share of the total reported poll as the prediction.

A. *Results from Analyzing Pre-Election Polls*

Table 1 summarizes several key aspects of the polls we analyze as forecasts of election results

[7]See the web appendix for such an example.

[8]As discussed in the text, we include all general election polls including at minimum both of the major party candidates and completed after June 1 of the election year. We identify and drop polls reported multiple times. When a single poll reports responses to the question phrased to allow third party candidates and another question phrased to force a choice between the Democratic and Republican candidates we use only the poll that allows the respondent more options. When a poll reports the results of the full sample in addition to some number of subsamples we use only the sample that limits respondents to "likely voters." Finally, we drop 39 polls with no reported sample size.

We obtained official 1996, 2000 and 2004 presidential election results from the Federal Election Commission website: accessed http://www.fec.gov/pubrec/fe1996/presge.htm on February 11, 2008 accessed http://www.fec.gov/pubrec/2000presgeresults.htm on February 11, 2008 accessed http://www.fec.gov/pubrec/fe2004/federalelections2004.pdf on February 11, 2008 According to the FEC these results are "the official, certified federal election results obtained from each state's election office and other official sources." http://www.fec.gov/pubrec/electionresults.shtml.

Official results of the 2008 presidential election were not yet available at the time of this writing. For this election we obtain results from the most up-to-date tallies from media websites or from the state Secretary of State office when available. These results are conveniently available with sources from Wikipedia.com (accessed from http://en.wikipedia.org/wiki/2008_presidential_election on November 19, 2008).

[9]Slightly more formally, if we let r_c denote the percentage point reported in the poll for candidate c among the C candidates reported, our adjusted measure p_i^{Adj} is given by

$$p_i^{\text{Adj}} = \frac{r_c}{\sum_{i=1}^{C} r_i} \qquad (1)$$

(see web appendix Table 5 and Table 6 for a complete analysis).

There are several points to be made:

- Taken as a whole the polls, on the most favorable terms we can devise, do not behave as would be suggested by simple random (probability) sampling and are biased.

  We consider all polls, polls which restrict themselves to "likely voters" only, and polls conducted within two weeks of the election. The first column reports results based on the raw data, unadjusted for undecided and missing respondents. Given the problems of rounding, non–reporting of third-party candidates, undecided, and others, these unadjusted numbers underpredict both the Democrat and Republican pollsters. Thus, for all our subsequent analysis we consider only adjusted numbers.

  As to the departures from what might be expected under random sampling (with certain and unchanging intentions, and certainty about participation) they are easiest to see from the table by considering our "standardized" prediction errors:

$$\frac{\widehat{p}_i - \mu_i}{\sqrt{\frac{\mu(1-\mu)}{N}}}$$

  Under the null of random sampling, the usual Central Limit Theorem argument suggests that these standardized prediction errors should have a variance of 1.

  As is evident from the estimates in Table 1, corrected or not, the actual variance of the prediction errors is much larger in magnitude than implied by sampling theory.

  Another view of bias and dispersion of the standardized poll errors is provided by a simple kernel density estimate of the standardized prediction errors in Figure 1.[10] The estimated densities are too disperse, are not centered at 0, and generally do not share the shape of the standard normal density.

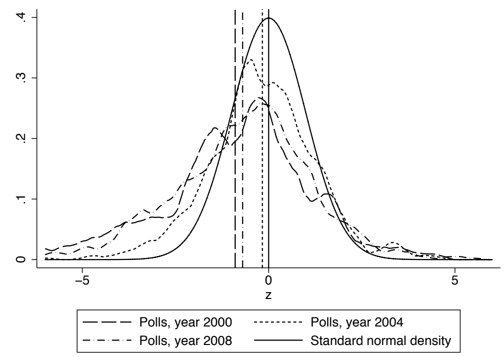


FIGURE 1: DENSITY ESTIMATES OF STANDARDIZED PREDICTION ERRORS OF DEMOCRATIC CANDIDATES

The figure displays a kernel density of the standardized prediction errors for presidential state races for the Democratic candidate. The vertical lines are the estimated mean associated with the appropriate density.

  In a subsequent section, we further demonstrate that the difference between the polls and the election outcomes does not appear to be pure noise, but rather is correlated with information available to pollsters (and everyone) at the time the poll is taken.

- The table also makes clear that the polls predict the winner more often than not, but the polls guess the winner incorrectly about 18-20 percent of the time.

- A very crude "benchmark" model uses the outcome from the previous election as a prediction for the subsequent presidential race. Perhaps surprisingly, by this benchmark pre–election polls do not fare too well. If we compute one prediction per race (as opposed to one prediction per poll) the crude model generally outperforms the polls and is competitive with polls conducted two weeks before the election campaign.

  As can be seen in Table 1 by comparing the fraction of "mispredicted victors" using one prior race outcome per poll (top panel), the

[10]See the web appendix for density estimates of the prediction errors for Republicans; the appendix also includes density estimates for subsamples of the polls we analyze.

Table 1: Pre-Election Polls

	All Polls		"Likely Voters"		< 2 Weeks before Election	
	N = 1857		N = 1554		N = 704	
	Raw	Adj	Raw	Adj	Raw	Adj
Republican share	48.17		48.21		48.31	
	{6.12}		{5.90}		{5.36}	
Democratic share	49.99		49.98		49.75	
	{5.93}		{5.66}		{5.15}	
Predicted Republican	44.70	48.20	45.03	48.31	45.14	47.84
	{5.99}	{6.31}	{5.71}	{6.00}	{5.24}	{5.48}
Predicted Democratic	45.42	48.95	45.71	49.01	46.55	49.31
	{5.87}	{5.91}	{5.59}	{5.61}	{5.19}	{5.22}
Republican error	-3.48	0.03	-3.18	0.10	-3.17	-0.47
	{3.48}	{3.36}	{3.31}	{3.21}	{2.67}	{2.49}
Democratic error	-4.57	-1.04	-4.27	-0.96	-3.19	-0.43
	{4.00}	{3.45}	{3.79}	{3.29}	{3.02}	{2.70}
Standardized	-1.80	0.02	-1.63	0.07	-1.59	-0.22
Republican error	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Variance of stand'd	3.32	3.07	2.82	2.69	1.86	1.58
Republican error	(0.16)	(0.16)	(0.12)	(0.14)	(0.12)	(0.10)
Standardized	-2.38	-0.55	-2.22	-0.51	-1.63	-0.23
Democratic error	(0.05)	(0.04)	(0.05)	(0.04)	(0.06)	(0.05)
Variance of stand'd	4.38	3.20	3.91	2.84	2.37	1.89
Democratic error	(0.19)	(0.14)	(0.20)	(0.13)	(0.15)	(0.12)
Republican victory	38.40		38.93		40.77	
Democratic victory	61.60		61.07		59.23	
Republican victory predicted	40.01		40.22		38.64	
Democratic victory predicted	55.57		55.15		56.53	
Mispredicted victor	20.73		20.46		19.18	
Mispredicted victor using prior race	24.23		24.26		28.41	
	One Observation Per Race					
	N = 143		N = 136		N = 117	
Republican share	50.01		49.68		50.11	
	{8.97}		{8.72}		{8.02}	
Democratic share	47.69		48.09		47.65	
	{8.92}		{8.53}		{7.85}	
Republican victory	53.15		52.21		53.85	
Democratic victory	46.85		47.79		46.15	
Mispredicted victor using prior race	16.08		16.18		19.66	

"Adj" means treating undecided respondents as strongly ignorable. The standardized prediction errors are calculated using the equation in the text. Under the null that the poll results are i.i.d. draws from the true distribution, the mean of the standardized prediction error is 0 and the variance is 1. Prediction errors and shares are in units of percentage points. Standard deviations in braces. Standard errors in parentheses. Standard errors on variance estimates are bootstrapped with 1000 repetitions.

success of the crude benchmark forecast is only partly explained by the fact that more polls are conducted for "hard to predict" races. (See web appendix Figure 7 for a visual description of where polls are most likely to be conducted).

Web appendix Table 6 repeats the same analysis, except this time we analyze the three elections separately and the same patterns described roughly apply to each of the three presidential elections we analyze. We also conducted several other analyses (see the web appendices) from which we summarize two important points:

First, in the 2000 elections, for example, polls that included any third party candidate provided forecasts with more bias for the Democratic candidate, less bias for the Republican candidate, and much less disperse forecasts for both. However, in 2004 we see precisely the opposite pattern. (See web appendix Table 7).

Second, although there is some *slight* improvement in the poll forecasts closer to the election date, the key features of the errors – bias and over-dispersion – are unchanged. Figure 2 displays the median, and the 10th and 90th quantile regression lines of the prediction errors for all three presidential elections we analyze (Democratic candidates only), demonstrating some decline in the amount of over-dispersion as election day approaches.

The point estimates from the quantile regression of the forecast error for the Democratic candidate on a constant and the number of days confirms the impression from the figure. If a simple linear trend is correct for all three quantiles, the estimates suggest that 100 days closer to the election moves the 90th quantile by 2 standardized units (quite a large amount), and the 10th quantile by about 0.6. Both move in the expected direction – dispersion in the polls diminishes over time. The constant term in the quantile regressions can be interpreted as the hypothetical distribution of poll errors on the day of the election.

As the following tabulation makes clear, there is significant over-dispersion. The 95 percent confidence interval for the constant term for 10th quantile regression does not cover the value suggested by standard normality (-1.28). Likewise the 95 percent confidence interval for the constant term in the median regression does not

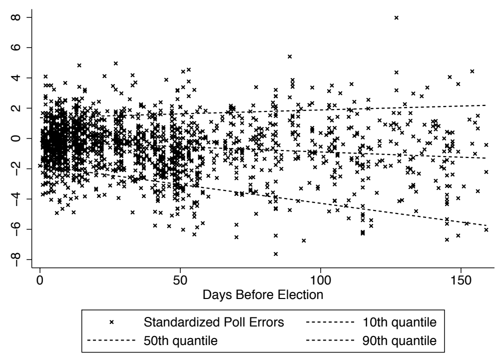


FIGURE 2: SCATTER PLOT OF DEMOCRATIC PREDICTION ERRORS FOR 2000, 2004, 2008 ELECTIONS

The figure displays a scatter plot of standardized prediction errors for presidential state races for the democratic candidate and quantile regressions at the 10th, 50th, and 90th quantile.

cover its "theoretical" value of zero. For the 90th quantile, the theoretical value suggested by standard normality (1.28) just lies inside the upper part of the estimated 95 percent confidence interval.

Quantile/ N(0,1)	Estimate	95 Percent CI for estimate	
10th/ -1.28	-1.80	-1.97	-1.63
50th/ 0	-0.20	-0.32	-0.08
90th/ 1.28	1.37	1.22	1.52

IV. How "Informative" are the Polls

Ottaviani and Norman (2006) argue that there are many reasons that polls should be biased. A simple reason is because pollsters may act as "honest Bayesians" and report their posterior distribution instead of the actual poll result.

For instance, imagine a pollster response to a "rogue poll" – a polling result that is wildly inconsistent with other reliable information (such as previous polls). This will happen infrequently of course, but it will happen. Faced with an "unrepresentative" or "unusual" sample, the pollster may "honestly" decide not to report the result of the polling, but massage the answer with his/her prior information to be more consistent with what s/he knows.

Table 2: The Relationship Between Forecast Errors and Prior Information

	(1)	(2)	(3)	(4)
Dependent Variable = 2008 Polls				
2008 Outcome	0.821		0.507	0.492
	(0.041)		(0.085)	(0.099)
2004 Outcome		0.855	0.360	0.500
		(0.045)	(0.090)	(0.154)
2000 Outcome				-0.144
				(0.106)
1996 Outcome				0.023
				(0.135)
Constant	7.967	10.108	7.222	7.007
	(2.098)	(2.250)	(1.756)	(2.591)
R-squared	0.715	0.692	0.733	0.736
N = 677				
Dependent Variable = 2004 Polls	(1)	(2)	(3)	(4)
2004 Outcome	0.915		0.886	0.881
	(0.032)		(0.099)	(0.104)
2000 Outcome		0.828	0.033	0.006
		(0.111)	(0.103)	(0.128)
1996 Outcome				0.043
				(0.137)
Constant	3.851	8.480	3.666	3.095
	(1.643)	(5.447)	(1.700)	(2.472)
R-squared	0.729	0.582	0.730	0.730
N = 705				
Dependent Variable = 2000 Polls	(1)	(2)	(3)	
2000 Outcome	0.764		0.594	
	(0.047)		(0.143)	
1996 Outcome		0.932	0.228	
		(0.059)	(0.159)	
Constant	9.920	1.090	6.889	
	(2.399)	(3.067)	(2.467)	
R-squared	0.598	0.558	0.602	
N = 475				

Each column is an OLS regression clustered by state. The dependent variable is the adjusted Democratic poll prediction treating undecideds as strongly ignorable. Standard errors clustered by state in parentheses.

The canonical Bayesian approach to this procedure is sometimes referred to as the "Beta–binomial model" which takes the usual binomial distribution likelihood and combines it with a (conjugate) prior of the Beta distribution.

Suppose the likelihood of seeing x votes for candidate A from a poll of size N is binomial and the true fraction supporting A is θ. Taking the prior and likelihood together generates the following posterior distribution for the "honest" Bayesian:

$$\text{Posterior} = \frac{\theta^{\alpha+x-1}(1-\theta)^{\delta-1+N}}{B((\alpha+x),(\delta+(N-x)))}$$

Letting $\alpha' \equiv \alpha - 1$, $\delta' \equiv \delta - 1$, and $\mathcal{P} \equiv \frac{\alpha'}{\alpha'+\delta'}$ the mode of the posterior occurs at:

$$\frac{\alpha'+x}{\alpha'+\delta'+N} = \left(\frac{\alpha'+\delta'}{\alpha'+\delta'+N}\right)\mathcal{P} + \left(\frac{N}{\alpha'+\delta'+N}\right)\frac{x}{N}$$

Thus the mode of the posterior is merely the weighted average of the prior and the actual sample, where the weights reflect the strength of the prior. This suggests an OLS regression

$$\text{(2)} \quad \text{poll}_i = \text{constant} + a * \text{Prior}_i + b * \text{Actual}_i$$

where the parameters a and b are respectively the weights that the typical pollster puts on his prior and the actual polling result. If the pollster was merely reporting the results obtained from sampling, then on average the polls would provide the true result, and both a and the constant would be equal to zero.

The "model" as described is easily rejected by the data (although it does remarkably well considering how tightly parameterized the model is) so we instead consider a "just identified" version of equation 2 where we allow an additional parameter that allows the identical priors to vary from the previous election result by a constant μ (that is identical across all state races) and assume that the prior can be summarized by a linear combination of previous election results (E):

$$\begin{aligned}\text{poll}_i &= a*(\sum_{j=1}^{J}\phi_j E_i^{(t-1)} + \mu) + b*\text{Actual}_i \\ &= a*\text{constant} + \sum_{j=1}^{J}\phi'_j E_i^{(t-j)} + b*\text{Actual}_i\end{aligned}$$

where the constant term (up to scale) identifies a shift from the previous election result, ϕ_j is the weight on the previous election result, and J is as large as two previous election results. These are reported in Table 2 (see web appendix Table 8 for a complete analysis). Our main result is that the coefficient on the actual outcome is always below 1 (what would be predicted by a pure sampling error model.) When we include two previous races in the regression, the coefficient on the actual outcome is about 0.5 for the 2008 election. This suggests that for "honest Bayesians" reported poll results are "one part sample, one equal part prior information."

This finding helps explain a puzzle: if there are so many reasons for the poll to be biased (non-response, participation model error, the difference between intentions the pollsters questions) why do the polls seem to perform "o.k."? The simplest answer is that they are very easy to predict. Indeed, it is in 2004, when the polls seem to perform the best, that the crude benchmark model most outperforms the pollsters: the 2004 election was, to a large extent, a "replay" of the 2000 election. (See web appendix table 6). Indeed, use of the 2000 election result as a prediction would have correctly guessed the winner 94% of the time: the polls we analyzed guessed the victor less than 74% of the time.

V. A Poll that Allows for Uncertain Preferences

While a large literature (see Crespi (1988) for a nice summary) suggests that horse race polls – those that ask respondents about who they intend to vote for in an election – should, if conducted properly and under the right conditions, reflect actual outcomes, an old statistical literature, most recently Manski (1990) suggests the opposite. Manski (1990) observes that if a potential voter is uncertain about for whom s/he will vote then a simple "intention" question: "who are you likely to vote for" will be biased in general for the outcome even if agents are perfectly rational, etc. The only hope for generating an unbiased prediction of an outcome from intentions data requires asking the question in such a way that allows the voter to express his or her uncertainty. (See the web appendix for a further discussion of the intentions problem.)

Instead of asking: *If the election were held today, would you*:

- Vote for John Kerry, the Democratic nominee for president.
- Vote for George Bush, the Republican nominee for president.
- Vote for another candidate.

one should ask the question in terms of *probabilities* for voting for each of the candidates.

It seems worthwhile to ask whether this "'theoretical" source of bias can explain much of the bias we observe in actual polls. In a sense, we would like to see the extent to which this purely statistical problem addresses the question posed by Gelman and King (1993) – are polls variable only because the questions are posed as intentions instead of probabilities?

A. Our Poll

Our purpose in designing the questions for our poll was to evaluate the extent to which bias in the polls as forecasts of the outcome are generated by not allowing respondents to characterize their preferences as probabilities. Although described as an attempt to generate a "representative" sample[11] the sampling process appears to be a variant of quota sampling, where (conditional on participation) an attempt is made to make the distribution of a few key demographic characteristics similar to a representative sample.[12] Thus, we had little reasonable expectation of the poll as a reliable measure of electorate

[11] See TESS (2005a), for example.

[12] The data and documentation for our survey is available at `http://www.experimentcentral.org/data/data.php?pid=298`. The poll was conducted by TESS (2005b). We had originally planned and were encouraged to use TESS for a second survey in 2008. Unfortunately, they decided against running the poll at a point too late in the process to

opinion, but find it of limited use in assessing the extent to which allowing for probabilistic intentions influences the estimate for whatever (non-representative) population it achieves (i.e. those willing to participate).

To that end, there were two sets of questions. One was administered to half the sample; the other set of questions to the (demographically balanced) other half. We call the first set of questions "the Probabilistic way" and the second, "the Usual way."

Our study design consisted of the following two pairs of questions:

1) *Are you a registered voter? If yes:*

   - *Given your other obligations, on a scale of 0 to 100 what is the chance that you will actually cast a vote for president? If you are certain you will vote, state "100." If you are certain you will not vote, state "0". If there is a 40 in 100 chance you will vote, state 40, and so on.*

   If no,

   - *Given your other obligations, what is the chance that you will register to vote* **and** *vote for president in November 2004. Use a scale of 0 to 100. If you are certain you will register and you will vote, state "100." If you are certain you will not register, or you will register and not vote, state "0". is a 40 in 100 chance you will both register and vote, state 40, and so on.*

2) *Regardless of whether or not you are likely to vote in the presidential election, given what is likely to happen during the course of the campaign, on a scale of 0 to 100 what is the likelihood that you would vote for John Kerry, George Bush, or some other candidate for president?*

   The sum of your answers should be 100. For instance, if there is a 40% chance you would vote for John Kerry and a 40% chance you would vote for George Bush, and a 20% chance you would vote for someone else, your response should be:

John Kerry	40
George Bush	40
Other Candidate	20

   If you are certain that you would vote for Ralph Nader (or a candidate other than Bush or Kerry), your response should be:

John Kerry	0
George Bush	0
Other Candidate	100

For the other demographically balanced half-sample, the two questions are designed to mimic typical poll practice.

1) *Are you registered to vote?*

   If yes:

   - *Are you likely to cast a vote for a presidential candidate in the 2004 election?*

   If no,

   - *Are you likely to register in time for the election* **and** *cast a vote for a presidential candidate in the 2004 election?*

2) *Regardless of whether or not you are likely to vote in the presidential election, and given what is likely to happen during the course of the campaign, for whom would you vote:*

   - *Vote for John Kerry, the Democratic nominee for president.*
   - *Vote for George Bush, the Republican nominee for president.*
   - *Vote for another candidate.*

find an alternative means to conduct it. The first wave was conducted between October 19th and October 24th, 2004. The second wave was conducted between October 26th and November 1st, 2004. We drop four observations from the Manski group with no response for probability of voting (three of these also have missing poll results). We also drop a combined 58 observations from both groups with missing poll results. The survey completion rate is 68% for the first wave and 71% for the second wave.

The foregoing questions were intended to mimic how questions are actually asked in presidential horse race polls[13].

B. Results

Web appendix Table 9 presents descriptive statistics of the experimental (Probabilistic) and control (Usual) samples. In both waves we fail to reject differences in mean demographics. As Table 3 demonstrates, neither version of the poll does particularly well and, echoing earlier results, use of Probabilistic style questions does not significantly alter the result (see web appendix Table 10 for the complete analysis). Of course, as is true for any poll results, there are several explanations including non–representative sampling, selection bias and considerable problems with the implementation of the polling by *TESS* and *Knowledge Networks*. In addition, over 3/4 of the Probabilistic group reported that they were virtually certain of going to the polls, and a similar fraction expressed certainty about their choice of candidate. With such a high degree of certainty among respondents it might have been surprising to see important differences in the preferences of the two groups.[14]

VI. Conclusion

Voter "uncertainty" and sample selection bias are only two possible problems that might render pre–election polls as unreliable and biased forecasts of the election outcome even when conducted close to the election. There is an enormous literature that proposes other possible reasons which, because of limitations of space, we do not discuss here. Nonetheless, it remains the case that either problem would be *sufficient* to render pre–election polls as unreliable and biased estimates of *trends* – even for the narrowest construct pollsters might care to estimate, i.e. if the election were held *today* ....

Given the relative ease with which one can arrive a good guess of the outcome of a presidential race at the state level by using the previous election's result, it is clear that the fact that the polls can often predict the winner is little reason to be sanguine about the "value added" they provide. Our analysis suggests that until a more severe test (Mayo, 1996; DiNardo, 2009) is proposed there is considerable reason for skepticism.

[13]See McDermott and Frankovic (2003) for a description of how different pollsters ask the question.

[14]Indeed, the possibility that the the 2004 race was unusual for the high degree of "certainty" most voters had about their intentions, was our primary motivation for attempting to undertake a second poll for the 2008 campaign.

REFERENCES

Brehm, John, *The Phantom Respondents: Opinion Surveys and Political Representation*, Ann Arbor, MI: University of Michigan Press, 1993.

Butler, David and Dennis Kavanagh, "The Waterloo of the Polls," in "The British General Election of 1992," New York: Pallgrave Macmillan, 1992.

Crespi, Irving, *Pre–Election Polling: Sources of Accuracy and Error*, New York: Russell Sage Foundation, 1988.

DiNardo, John E., "Introductory Remarks on Metastatistics for the Practically Minded Non-Bayesian Regression Runner," in Terence C. Mills and Kerry Patterson, eds., *Applied Econometrics*, Vol. 2 of *Palgrave Handbook of Econometrics*, Palgrave Macmillan, 2009. *Forthcoming.*

DiNardo, John, Justin McCrary, and Lisa Sanbonmatsu, "Randomizing Attrition and Simple Estimators for Treatment Effects and Bounds: An Empirical Example," Unpublished Manuscript, University of Michigan, Ann Arbor, MI November 21 2005.

Ehrbeck, Tilman and Robert Waldmann, "Why are Professional Forecasters Biased? Agency Versus Behavioral Explanations," *Quarterly Journal of Economics*, February 1996, *111* (1), 21–40.

Gelman, Andrew and Gary King, "Why are American Presidential Election Campaign Polls So Variable When Votes are So Predictable?," *British Journal of Political Science*, October 1993, *23* (4), 409–451.

Table 3: Probabilistic vs. Usual Style Questions

	Probabilistic Group			Control Group		
	Bush	Kerry	Other	Bush	Kerry	Other
		N = 1322			N = 1393	
Survey weighted	46.037	50.429	3.534	46.364	49.333	4.303
	(1.485)	(1.490)	(0.501)	(1.563)	(1.573)	(0.684)
		N = 1190			N = 1181	
Survey weighted, and	46.973	50.102	2.925	48.119	49.204	2.676
P(vote)>0	(1.582)	(1.589)	(0.495)	(1.705)	(1.709)	(0.627)
Above, and	46.886	50.445	2.669			
participation weighted	(1.633)	(1.636)	(0.459)			
Above, and	46.655	50.687	2.657	48.084	49.250	2.666
missing data weighted	(1.646)	(1.652)	(0.454)	(1.706)	(1.711)	(0.624)
p-values						
Bush(T=1) = Bush(T=0)		0.5467				
Kerry(T=1) = Kerry(T=0)		0.5457				
Joint		0.8295				

Results pool both survey waves, employing DFL weights to account for differences in observed sample demographics between waves. In addition, we employ the survey weights provided by TESS designed to match the demographics of the surveyed sample to the U.S. Census and the Knowledge Networks Panel. The likely voter weights use the reported probability of voting (for the Probabilistic group only) to adjust results. The missing data weights use DFL weights to account for 58 dropped observations with missing poll results on observed dimensions of demographics. See web appendix for details on construction of weights. Actual national 2004 election results were Bush 50.733%, Kerry 48.270%, and Other 0.996%. Heteroskedasticity robust standard errors in parentheses.

Horowitz, Joel L. and Charles F. Manski, "Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," *Journal of Econometrics*, May 1998, *84* (1), 37–58.

Keppo, Jussi, Lones Smith, and Dmitry Davydov, "Optimal Electoral Timing: Exercise Wisely and You May Live Longer," *Review of Economic Studies*, Forthcoming 2008.

Ladd, Carll Everett, "The Election Polls: An American Waterloo," *Chronicle of Higher Education*, November 22 1996, *43* (11), A52.

Manski, Charles F., "The Use of Intentions Data to Predict Behavior: A Best-Case Analysis," *Journal of the American Statistical Association*, December 1990, *85* (412), 934–940.

Mayo, Deborah G., *Error and the Growth of Experimental Knowledge* Science and Its Conceptual Foundations, Chicago: University of Chicago Press, 1996.

McDermott, Monika L. and Kathleen A. Frankovic, "Horserace Polling and Survey Method Effects: An Analysis of the 2000 Campaign," *Public Opinion Quarterly*, Summer 2003, *67* (2), 244–264.

Mitofsky, Warren J., "Was 1996 A Worse Year for Polls than 1948?," *Public Opinion Quarterly*, Summer 1998, *62* (2), 230–249.

Moon, Nick, *Opinion Polls: History, Theory, and Practice*, Mancester and New York: Manchester University Press, 1999.

Ottaviani, Marco and Peter Norman, "The Strategy of Professional Forecasting," *Journal of Financial Economics*, August 2006, *81* (2), 441–446.

Panagakis, Nick, "Response to "Was 1996 A Worse Year for Polls than 1948"," *Public Opinion Quarterly*, Summer 1999, *63* (2), 278–281.

Pew Research Center, "Possible Consequences of Non-Response for Pre-Election Surveys: Race and Reluctant Respondents," May 16 1998.

Røhme, N., "The state of the art of public opinion polling worldwide: an international study based on information collected from national market and opinion research institutes in April 1992.," *Marketing Research Today*, 1992, pp. 264–273.

Stroud, Natalie Jomini and Kate Kenski, "From Agenda Setting to Refusal Setting: Survey Nonresponse as a Function of Media Coverage Across the 2004 Election Cycle," *Public Opin Q*, 2007, *71* (4), 539–559.

TESS, "Get to Know Tess," 2005. Accessed November 1, 2008 at `http://www.experimentcentral.org/tess/`.

_, "Time-sharing Experiments for the Social Sciences," 2005. Data collected by Time-sharing Experiments for the Social Sciences, NSF Grant 0094964, Diana C. Mutz and Arthur Lupia, Principal Investigators.".

Traugott, Michael W., "Assessing Poll Performance in the 2000 Campaign," *Public Opinion Quarterly*, 2001, *65*, 389–419.

Zogby, John, "Interview of John Zogby on the *Daily Show* with Jon Stewart," October 28 2004. `http://www.thedailyshow.com/video/index.jhtml?videoId=127045&title=john-zogby`.

Web-Appendix

1) Appendix 1. Ten (10) Web tables.
2) Appendix 2. Eleven (11) Web tables.
3) Appendix 3. Short Discussion of Intensions.

Web Appendix Table 1: November "Trial Heats" for 2000 U.S. Presidential Election
Of these 43 "last minute" national horse race polls from the 2000 U.S. Presidential Election only 3 of the 42 polls predicted either a "tie" or Gore ahead in the national race, despite the fact that the actual vote was a virtual "tie" (with Gore actually winning the popular vote). Consultation of the tables for the binomial distribution reveals that the probability of 42 or more "Bush" predictions out of the 45 displayed above is less than 5×10^{-7} percent. In making this calculation we use the assumption that Gore (the Democratic candidate) and Bush (the Republican candidate) received exactly the same number of votes, and the polls were independent samples.

Web Appendix Table 2: Descriptive Statistics of Pre-Election Poll Sample, 2000-2008
The implied sample size is calculated from the reported margin of error and a mean of 0.50. Similarly, the implied margin of error is calculated from the reported sample size and mean of 0.50. The differences between the reported and implied values can be attributed to rounding error in most (but not all) cases. The sample includes all available statewide pre-election polls completed on or after the first day of June in the election year. We drop 39 polls with missing sample size from all analyses. See text for a further discussion of the sample inclusion criteria. Over a third of all polls in our sample are conducted within two weeks of election day, and approximately 85% of polls are reported as polls of "likely voters" (as opposed to registered voters, adults, or no qualification at all). The intensity of polling by state tends to increase across the three election years, with a median (mean) of 9 (13.5) polls per state in 2008 and a median (mean) of 5 (10.1) polls per state in 2000.

Web Appendix Table 3: Total Percentage Reported in Polls
The poll totals in this table include all reported categories including undecided and other candidate respondents. The sum of the predicted shares in many polls do not add up to exactly 100 percentage points. Since nearly all polls report figures rounded to two digits, many of these sums can be explained by rounding error. We do observe a small fraction of polls that sum to an amount below that which can be explained by rounding error, although over 95% of the polls in our sample do add up to 99 percentage points or higher. In these cases, as in the case of rounding error, we handle the problem symmetrically to the undecided problem and use the share of the total reported poll as the prediction (see text for details).

Web Appendix Table 4: Descriptive Statistics for Undecideds and Other Candidates in Polls
"Conditional" shares are conditional on being having any undecided or ambiguous respondents (or third party, other or none in bottom panel). "Ambiguous" shares include categories that are lumped together, such as "Other/Undecided" as well as shares left unaccounted. The vote shares are the unweighted means across polls. Only about 1% of polls have no undecided or ambiguous respondents. In polls with undecided or ambiguous respondents these respondents account for approximately 7% of the total, most of whom are classified as undecided. The fraction of polls with third party candidates varies with the election year. In the 2000 election 3.7% of the electorate voted for a third party candidate, while only about 1% did so in 2004 or 2008. As might be expected, the 2000 polls included third party candidates (or other/none) over 90% of the time, while 2008 polls included these only about 70% of the time. The composition of the third party candidate components varies by election year.

Web Appendix Table 5: Pre-Election Polls
"Adj" means treating undecided respondents as strongly ignorable. The standardized prediction errors are calculated using the equation in the text. Under the null that the poll results are i.i.d.

draws from the true distribution, the mean of the standardized prediction error is 0 and the variance is 1. Prediction errors and shares are in units of percentage points. See text for a discussion of this table.

Web Appendix Table 6: Pre-Election Polls, by Year
"Adj" means treating undecided respondents as strongly ignorable. The standardized prediction errors are calculated using the equation in the text. Under the null that the poll results are i.i.d. draws from the true distribution, the mean of the standardized prediction error is 0 and the variance is 1. Prediction errors and shares are in units of percentage points. The pattern of over-dispersion and bias is consistent across election years. The polls in 2004 are slightly less disperse and display the least bias of the three years. As noted in the text, the 2004 race was, to a large extent, a "replay" of the 2000 election, possibly making the 2004 election easier to predict. Indeed, use of the 2000 election result as a prediction would have correctly guessed the winner 94% of the time: the polls we analyzed guessed the victor less than 74% of the time. The fact that most polls are conducted for "hard to predict races" only partly explains this fact, since even accounting for where polls are conducted, the 2000 election result will correctly guess the winner 83% of the time. In the 2000 and 2008 races the polls outperform this crude benchmark, but not by a large margin.

Web Appendix Table 7: Error in Pre-Election Polls, By Inclusion of Third Party Candidates
All columns treat undecided respondents as strongly ignorable. Under the null that the poll results are i.i.d. draws from the true distribution, the mean of the standardized prediction error is 0 and the variance is 1. Prediction errors and shares are in units of percentage points. Third party candidates received 1.3% of the popular vote in 2008, 1.0% in 2004 and 3.7% in 2000. In the 2000 elections polls that *included* any third party candidate provided forecasts with more bias for the Democratic candidate, less bias for the Republican candidate, and much less disperse forecasts for both. However, in 2004 we see precisely the opposite pattern.

Web Appendix Table 8: The Relation Between Forecast Errors and Prior Information
Each column is an OLS regression clustered by state. The dependent variable is the adjusted poll result, treating undecided respondents as strongly ignorable. See text for a discussion of the Democratic candidate results. The results are qualitatively similar for the Republican candidate results with somewhat less weight placed on the prior than for the Democratic candidate, though this difference is not statistically significant.

Web Appendix Table 9: Descriptive Statistics of Manski Poll
See text for a discussion of the TESS poll. The table demonstrates that the means of observed individual characteristics do not differ significantly within wave across the treatment and control groups with a p-value of 0.34 in wave 1 and 0.90 in wave 2. Approximately 85% of the control group sample responded that they intended to vote in the election. This fraction is statistically indistinguishable from the mean of the reported probability of voting in the Manski group sample. Over 75% of the Manski sample reported that they were virtually certain of going to the polls. A similar fraction also expressed certainty about their choice of candidate. With so few respondents expressing uncertainty about their voting behavior one might be surprised to see important differences in the estimated preferences of the experimental groups.

Web Appendix Table 10: Probabilistic vs. Usual Style Questions
Results pool both survey waves, employing DFL weights to account for differences in observed sample demographics between waves. In addition, we employ the survey weights provided by TESS designed to match the demographics of the surveyed sample to the U.S. Census and the Knowledge Networks Panel. The likely voter weights use the reported probability of voting (for the Manski group only) to adjust results. The missing data weights use DFL weights to account for 58 dropped observations with missing poll results on observed dimensions of demographics. Actual national 2004 election results were Bush 50.733%, Kerry 48.270%, and Other 0.996%. See text for discussion of pooled results. The results tabulated separately by wave do not demonstrate any significant differences between the Manski and the control group respondents.

Web Appendix Table 1: November "Trial Heats" for 2000 U.S. Presidential Election

Date	Size		Gore	Bush	Prediction	Polling Agency
11/3 - 5	1801		45.9	49.0	False	ABC News
11/2 - 4	1741		45.9	50.0	False	Poll
11/1 - 3	1495		46.9	50.0	False	
10/31 - 11/2	1280		45.9	49.0	False	
10/30 - 11/1	1032		46.4	50.5	False	
11/5 - 6	2350		46.0	48.0	False	Gallup/CNN
11/5 - 6	2350	†[a]	46.4	48.5	False	USA Today
11/4 - 5	2386		46.4	48.5	False	Poll
11/2 - 4	2733		44.8	50.0	False	
11/1 - 3	2222		45.3	49.5	False	
10/31 - 11/2	2128		44.2	50.5	False	
10/30 - 11/1	2123		45.3	49.5	False	
11/1 - 2	623		45.8	51.0	False	Marist College
11/3 - 5	1026		45.8	49.0	False	NBC News/Wall
11/2 - 3	751		45.4	48.5	False	Street Journal
10/31 - 11/2	808		46.2	48.4	False	Newsweek Poll
11/2 - 5	1301		47.0	49.0	False	Pew Research Center
11/2 - 5	1301	†	46.7	48.9	False	for the People & the
11/1 - 4	1307		46.2	49.5	False	Press Survey
11/4 - 6	1091		47.9	46.8	True	CBS News Poll
11/2 - 5	1273		44.7	48.9	False	
11/1 - 3	825		46.3	48.4	False	
11/1 - 4	1158		44.2	49.5	False	CBS News/New York Times Poll
11/1 - 2	1000		47.8	47.8	False	Fox News/Opinion Dynamics Poll
11/3 - 5	1348		47.0	47.0	True	The Harris Poll
11/1 - 5	§[b]		44.4	46.5	False	ICR
11/5 - 6	1000		45.0	50.0	False	Tarrance Group-d-/
11/5 - 6	1000	†	45.6	51.1	False	Lake Snell Perry & Assoc.-R-
11/1-2,5	1000		41.6	51.7	False	Voter.com/
10/30 - 11/2	1000		41.6	51.7	False	Battleground Survey
10/29-31, 11/1	1000		43.3	51.1	False	
11/4 - 6	1292		46.0	47.9	False	Christian Science Monitor/
11/3 - 5	989		44.7	51.1	False	Investor's Business Daily/
11/2 - 4	718		42.4	52.2	False	TIPP Poll
11/1 - 3	838		41.4	48.5	False	
10/31 - 11/2	1070		42.4	47.5	False	
10/30 - 11/1	1186		45.3	50.5	False	
11/3 - 5	1253		45.2	48.4	False	Hotline Bullseye Poll
10/31 - 11/2	1000		43.0	50.5	False	
11/4 - 6	1200	‡[c]	48.0	46.0	True	Reuters/MSNBC
11/3 - 5	1200	‡	46.0	47.0	False	Tracking Poll
11/2 - 4	1200	‡	44.4	46.5	False	
11/1 - 3	1200	‡	44.2	48.4	False	
10/30 - 11/2	1200	‡	45.2	48.4	False	
10/29 - 11/1	1200	‡	42.4	45.5	False	

[a]This poll is a duplicate of the one immediately above but applies allocation algorithm as if true allocated had not been reported. In principle, they should differ only because of rounding error.

[b]No sample size reported.

[c]Only "approximate" sample size reported

[c]Source `pollingreport.com`.

Web Appendix Table 2: Descriptive Statistics of Pre-Election Poll Sample, 2000-2008

	All Polls	2000 Polls	2004 Polls	2008 Polls
Days before election	40.23	38.02	41.48	40.47
	{39.01}	{41.71}	{40.08}	{35.77}
< two weeks before election	0.38	0.48	0.37	0.32
Poll of "likely voters"	0.84	0.85	0.83	0.83
Reported sample size	697.07	626.82	733.20	708.72
	{280.32}	{213.26}	{276.62}	{314.96}
Reported margin of error	3.86	4.07	3.73	3.85
	{0.61}	{0.57}	{0.56}	{0.64}
Implied sample size	703.76	620.31	743.96	715.54
	{281.71}	{226.98}	{266.46}	{316.47}
Implied margin of error	3.88	4.04	3.74	3.90
	{0.65}	{0.56}	{0.50}	{0.81}
Number of polls	1857	475	705	677
Number of races	143	47	46	50
Mean polls per race	12.99	10.11	15.33	13.54
Median polls per race	7	5	7.5	9
Minimum polls per race	1	1	1	1
Maximum polls per race	80	37	64	80

The implied sample size is calculated from the reported margin of error and a mean of 0.50. Similarly, the implied margin of error is calculated from the reported sample size and mean of 0.50. The differences between the reported and implied values can be attributed to rounding error in most (but not all) cases. The sample includes all available state-level pre-election polls completed on or after the first day of June in the election year. We drop 39 polls with missing sample size from all analyses. See text for a further discussion of the sample inclusion criteria. The source for all polls is pollingreport.com. Standard deviations in braces.

Web Appendix Table 3: Total Percentage Reported in Polls

	All Polls	2000 Polls	2004 Polls	2008 Polls
Mean	99.82	99.85	100.02	99.58
Standard Deviation	1.39	0.81	0.58	2.11
Minimum	81	89	92	81
5th percentile	99	99	99	98
10th percentile	99	99	100	99
25th percentile	100	100	100	100
90th percentile	101	100	101	101
95th percentile	101	101	101	101
Maximum	102	102	102	102
Number of polls	1857	475	705	677

Poll totals include all reported categories including undecided and other candidate respondents.

Web Appendix Table 4: Descriptive Statistics for Undecideds and Other Candidates in Polls

	All Polls	2000 Polls	2004 Polls	2008 Polls
Fraction of polls with any undecided or ambiguous	0.989	0.981	0.996	0.987
Share of poll (conditional)	0.074	0.092	0.064	0.073
Vote shares (conditional)				
Undecided	0.057	0.069	0.053	0.054
Ambiguous	0.014	0.021	0.010	0.013
Unaccounted	0.003	0.003	0.001	0.005
Fraction of polls with any third party, other or none	0.793	0.914	0.804	0.697
Share of poll (conditional)	0.033	0.053	0.023	0.028
Vote shares (conditional)				
Green	0.012	0.039	0.000	0.000
Independent	0.008	0.000	0.015	0.006
Libertarian	0.002	0.001	0.000	0.004
Reform	0.003	0.010	0.000	0.000
Constitution	0.000	0.000	0.000	0.000
Other	0.009	0.004	0.007	0.016
None	0.001	0.000	0.000	0.001

"Conditional" shares are conditional on having any undecided or ambiguous respondents (or third party, other or none in bottom panel). "Ambiguous" shares include categories that are lumped together, such as "Other/Undecided" as well as shares left unaccounted. Vote shares are the unweighted means across polls.

Web Appendix Table 5: Pre-Election Polls

	All Polls N = 1857		"Likely Voters" N = 1554		< 2 Weeks before Election N = 704	
	Raw	Adj	Raw	Adj	Raw	Adj
Republican share	48.17		48.21		48.31	
	{6.12}		{5.90}		{5.36}	
Democratic share	49.99		49.98		49.75	
	{5.93}		{5.66}		{5.15}	
Predicted Republican	44.70	48.20	45.03	48.31	45.14	47.84
	{5.99}	{6.31}	{5.71}	{6.00}	{5.24}	{5.48}
Predicted Democratic	45.42	48.95	45.71	49.01	46.55	49.31
	{5.87}	{5.91}	{5.59}	{5.61}	{5.19}	{5.22}
Republican error	-3.48	0.03	-3.18	0.10	-3.17	-0.47
	{3.48}	{3.36}	{3.31}	{3.21}	{2.67}	{2.49}
Democratic error	-4.57	-1.04	-4.27	-0.96	-3.19	-0.43
	{4.00}	{3.45}	{3.79}	{3.29}	{3.02}	{2.70}
Standardized	-1.80	0.02	-1.63	0.07	-1.59	-0.22
Republican error	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Variance of stand'd	3.32	3.07	2.82	2.69	1.86	1.58
Republican error	(0.16)	(0.16)	(0.12)	(0.14)	(0.12)	(0.10)
Standardized	-2.38	-0.55	-2.22	-0.51	-1.63	-0.23
Democratic error	(0.05)	(0.04)	(0.05)	(0.04)	(0.06)	(0.05)
Variance of stand'd	4.38	3.20	3.91	2.84	2.37	1.89
Democratic error	(0.19)	(0.14)	(0.20)	(0.13)	(0.15)	(0.12)
Republican victory	38.40		38.93		40.77	
Democratic victory	61.60		61.07		59.23	
Republican victory predicted	40.01		40.22		38.64	
Democratic victory predicted	55.57		55.15		56.53	
Mispredicted victor	20.73		20.46		19.18	
Mispredicted victor using prior race	24.23		24.26		28.41	
	One Observation Per Race					
	N = 143		N = 136		N = 117	
Republican share	50.01		49.68		50.11	
	{8.97}		{8.72}		{8.02}	
Democratic share	47.69		48.09		47.65	
	{8.92}		{8.53}		{7.85}	
Republican victory	53.15		52.21		53.85	
Democratic victory	46.85		47.79		46.15	
Mispredicted victor using prior race	16.08		16.18		19.66	

"Adj" means treating undecided respondents as strongly ignorable. The standardized prediction errors are calculated using the equation in the text. Under the null that the poll results are i.i.d. draws from the true distribution, the mean of the standardized prediction error is 0 and the variance is 1. Prediction errors and shares are in units of percentage points. Standard deviations in braces. Standard errors in parentheses. Standard errors on variance estimates are bootstrapped with 1000 repetitions.

Web Appendix Table 6: Pre-Election Polls, by Year

	2000 Polls N = 475		2004 Polls N = 705		2008 Polls N = 677	
	Raw	Adj	Raw	Adj	Raw	Adj
Republican share	46.88		50.37		46.80	
	{6.43}		{4.90}		{6.39}	
Democratic share	49.37		48.69		51.78	
	{6.08}		{4.80}		{6.44}	
Predicted Republican	42.86	47.12	46.51	49.63	44.10	47.47
	{6.15}	{6.70}	{5.23}	{5.59}	{6.13}	{6.47}
Predicted Democratic	43.38	47.62	45.37	48.38	46.91	50.47
	{6.00}	{6.01}	{5.08}	{5.14}	{6.10}	{6.25}
Republican error	-4.02	0.24	-3.86	-0.74	-2.70	0.67
	{3.74}	{3.64}	{3.02}	{2.81}	{3.60}	{3.52}
Democratic error	-5.99	-1.75	-3.32	-0.31	-4.87	-1.31
	{4.54}	{4.07}	{3.02}	{2.71}	{4.11}	{3.53}
Standardized	-1.98	0.14	-2.07	-0.40	-1.39	0.37
Republican error	(0.09)	(0.08)	(0.06)	(0.06)	(0.07)	(0.08)
Variance of stand'd	3.47	3.34	2.86	2.45	3.43	3.23
Republican error	(0.27)	(0.33)	(0.27)	(0.26)	(0.23)	(0.19)
Standardized	-3.01	-0.90	-1.78	-0.17	-2.55	-0.70
Democratic error	(0.11)	(0.09)	(0.06)	(0.05)	(0.07)	(0.07)
Variance of stand'd	5.55	4.17	2.69	2.13	4.64	3.38
Democratic error	(0.49)	(0.31)	(0.17)	(0.16)	(0.28)	(0.21)
Republican victory	43.58		49.93		22.75	
Democratic victory	56.42		50.07		77.25	
Republican victory predicted	43.58		45.53		31.76	
Democratic victory predicted	52.84		47.52		65.88	
Mispredicted victor	19.58		26.95		15.07	
Mispredicted victor using prior race	26.95		12.91		34.12	
	One Observation Per Race					
	N = 47		N = 46		N = 50	
Republican share	49.90		52.36		47.97	
	{8.71}		{8.28}		{9.48}	
Democratic share	45.94		46.47		50.46	
	{8.32}		{8.28}		{9.50}	
Republican victory	57.45		58.70		44.00	
Democratic victory	42.55		41.30		56.00	
Mispredicted victor using prior race	23.40		6.52		18.00	

"Adj" means treating undecided respondents as strongly ignorable. The standardized prediction errors are calculated using the equation in the text. Under the null that the poll results are i.i.d. draws from the true distribution, the mean of the standardized prediction error is 0 and the variance is 1. Prediction errors and shares are in units of percentage points. Standard deviations in braces. Standard errors in parentheses. Standard errors on variance estimates are bootstrapped with 1000 repetitions.

Web Appendix Table 7: Error in Pre-Election Polls, by Inclusion of Third Party Candidates

	Republican Prediction Error			Democratic Prediction Error			
	Adj	Stand'd	Stand'd Var.	Adj	Stand'd	Stand'd Var.	Number of Polls
All 2000 polls	0.24	0.14	3.34	-1.75	-0.90	4.17	475
	(0.17)	(0.19)	(0.33)	(0.08)	(0.09)	(0.32)	
Buchanan included	-0.15	-0.07	2.77	-2.08	-1.06	4.03	292
	(0.20)	(0.24)	(0.27)	(0.10)	(0.12)	(0.34)	
Buchanan not included	0.87	0.47	4.08	-1.22	-0.64	4.32	183
	(0.30)	(0.30)	(0.68)	(0.15)	(0.15)	(0.58)	
Nader included	0.00	0.01	2.63	-2.03	-1.05	3.87	393
	(0.16)	(0.20)	(0.26)	(0.08)	(0.10)	(0.32)	
Nader not included	1.41	0.75	6.35	-0.40	-0.20	5.10	82
	(0.56)	(0.50)	(1.27)	(0.28)	(0.25)	(0.94)	
Both Buchanan	-0.13	-0.06	2.74	-2.15	-1.10	3.96	277
and Nader included	(0.20)	(0.24)	(0.27)	(0.10)	(0.12)	(0.35)	
Any third party	-0.09	-0.03	2.70	-1.88	-0.97	3.97	434
candidate included	(0.16)	(0.19)	(0.25)	(0.08)	(0.10)	(0.32)	
No third party	3.74	1.92	6.80	-0.38	-0.16	5.87	41
candidate included	(0.80)	(0.74)	(1.66)	(0.41)	(0.38)	(1.62)	
All 2004 polls	-0.74	-0.40	2.45	-0.31	-0.17	2.13	705
	(0.11)	(0.10)	(0.26)	(0.06)	(0.05)	(0.15)	
Nader included	-0.76	-0.42	2.54	-0.79	-0.44	2.08	391
	(0.14)	(0.13)	(0.34)	(0.08)	(0.07)	(0.18)	
Nader not included	-0.72	-0.38	2.34	0.29	0.16	2.00	314
	(0.16)	(0.15)	(0.37)	(0.09)	(0.08)	(0.25)	
Any third party	-0.92	-0.51	2.57	-0.57	-0.30	2.16	567
candidate included	(0.12)	(0.11)	(0.31)	(0.07)	(0.06)	(0.19)	
No third party	-0.03	0.02	1.72	0.75	0.38	1.67	138
candidate included	(0.21)	(0.21)	(0.22)	(0.11)	(0.11)	(0.21)	
All 2008 polls	0.67	0.37	3.23	-1.31	-0.70	3.38	677
	(0.14)	(0.14)	(0.19)	(0.07)	(0.07)	(0.20)	
Any third party	0.04	0.07	2.93	-1.58	-0.87	3.33	472
candidate included	(0.15)	(0.16)	(0.23)	(0.08)	(0.08)	(0.27)	
No third party	2.13	1.05	3.26	-0.68	-0.31	3.30	205
candidate included	(0.25)	(0.26)	(0.32)	(0.13)	(0.13)	(0.32)	

All columns treat undecided respondents as strongly ignorable. Under the null that the poll results are i.i.d. draws from the true distribution, the mean of the standardized prediction error is 0 and the variance is 1. Prediction errors and shares are in units of percentage points. Third party candidates received 1.3% of the popular vote in 2008, 1.0% in 2004 and 3.7% in 2000. Standard errors in parentheses. Standard errors on variances are bootstrapped with 1000 repetitions.

Web Appendix Table 8: The Relation Between Forecast Errors and Prior Information

	Dependent Variable = 2008 Polls							
	Republican Candidate				Democratic Candidate			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
2008 Outcome	0.861		0.569	0.571	0.821		0.507	0.492
	(0.041)		(0.091)	(0.104)	(0.041)		(0.085)	(0.099)
2004 Outcome		0.899	0.338	0.440		0.855	0.360	0.500
		(0.053)	(0.105)	(0.215)		(0.045)	(0.090)	(0.154)
2000 Outcome				-0.205				-0.144
				(0.114)				(0.106)
1996 Outcome				0.130				0.023
				(0.076)				(0.135)
Constant	7.166	0.908	3.289	2.716	7.967	10.108	7.222	7.007
	(1.978)	(2.595)	(2.159)	(2.260)	(2.098)	(2.250)	(1.756)	(2.591)
R-squared	0.723	0.690	0.738	0.741	0.715	0.692	0.733	0.736
N = 677								

	Dependent Variable = 2004 Polls							
	Republican Candidate				Democratic Candidate			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
2004 Outcome	0.986		0.927	0.951	0.915		0.886	0.881
	(0.032)		(0.122)	(0.131)	(0.032)		(0.099)	(0.104)
2000 Outcome		0.927	0.061	0.143		0.828	0.033	0.006
		(0.089)	(0.119)	(0.093)		(0.111)	(0.103)	(0.128)
1996 Outcome				-0.139				0.043
				(0.159)				(0.137)
Constant	-0.034	5.125	-0.006	0.456	3.851	8.480	3.666	3.095
	(1.567)	(4.268)	(1.616)	(1.540)	(1.643)	(5.447)	(1.700)	(2.472)
R-squared	0.747	0.681	0.747	0.750	0.729	0.582	0.730	0.730
N = 705								

	Dependent Variable = 2000 Polls					
	Republican Candidate			Democratic Candidate		
	(1)	(2)	(3)	(1)	(2)	(3)
2000 Outcome	0.883		0.745	0.764		0.594
	(0.049)		(0.081)	(0.047)		(0.143)
1996 Outcome		1.021	0.185		0.932	0.228
		(0.072)	(0.093)		(0.059)	(0.159)
Constant	5.719	6.574	4.834	9.920	1.090	6.889
	(2.213)	(2.726)	(2.298)	(2.399)	(3.067)	(2.467)
R-squared	0.717	0.637	0.721	0.598	0.558	0.602
N = 475						

Each column is an OLS regression clustered by state. The dependent variable is the adjusted poll result, treating undecideds as strongly ignorable. Standard errors clustered by state in parentheses.

Web Appendix Table 9: Descriptive Statistics of Manski Poll

	Wave 1		Wave 2	
	Probabilistic	Control	Probabilistic	Control
Number of respondents	647	682	675	711
Fraction expressing no uncertainty in candidate preference	0.764		0.767	
Fraction expressing little (<10%) uncertainty in candidate preference	0.897		0.908	
Probability of voting				
Mean	0.841	0.839	0.857	0.857
Standard deviation	0.338	0.368	0.315	0.351
10th percentile	0	0	0.2	0
25th percentile	0.99	1	0.99	1
50th percentile	1	1	1	1
Demographics				
Age	47.209	47.443	47.108	47.498
	{16.908}	{16.744}	{16.940}	{17.701}
White	0.810	0.792	0.796	0.788
Male	0.488	0.493	0.484	0.498
Household head	0.819	0.833	0.839	0.826
Married	0.603	0.589	0.582	0.536
Metro area	0.807	0.826	0.847	0.840
Employed	0.621	0.572	0.573	0.589
Less than high school	0.130	0.166	0.166	0.166
High school graduate	0.272	0.224	0.273	0.276
Some college or associate degree	0.332	0.359	0.289	0.293
B.A. or higher	0.266	0.251	0.273	0.266
Northeast	0.176	0.188	0.188	0.173
Midwest	0.283	0.249	0.276	0.294
South	0.331	0.331	0.313	0.329
West	0.210	0.232	0.224	0.204
F-statistic from joint test of significance	1.12		0.54	
p-value from joint test of significance	0.3393		0.8987	

Standard deviations in braces.

Web Appendix Table 10: Probabilistic vs. Usual Style Questions

	Probabilistic Group			Control Group		
	Bush	Kerry	Other	Bush	Kerry	Other
Wave 1		N = 647			N = 682	
Survey weighted	46.534	49.873	3.593	47.190	49.551	3.259
	(2.137)	(2.157)	(0.751)	(2.276)	(2.290)	(0.823)
P(vote) > 0		N = 577			N = 572	
Survey weighted	48.919	48.078	3.004	48.806	48.900	2.293
	(2.318)	(2.336)	(0.750)	(2.487)	(2.496)	(0.840)
Above, and	48.915	48.410	2.675			
participation weighted	(2.403)	(2.415)	(0.637)			
Above, and	48.585	48.761	2.654	48.763	48.949	2.288
missing data weighted	(2.437)	(2.455)	(0.628)	(2.490)	(2.499)	(0.839)
p-values						
Bush(M=1) = Bush(M=0)		0.9593				
Kerry(M=1) = Kerry(M=0)		0.9573				
Joint		0.9445				
Wave 2		N = 675			N = 711	
Survey weighted	45.528	50.997	3.474	45.519	49.110	5.371
	(2.069)	(2.061)	(0.661)	(2.144)	(2.153)	(1.093)
P(vote) > 0		N = 613			N = 609	
Survey weighted	45.037	52.117	2.846	47.435	49.507	3.058
	(2.173)	(2.173)	(0.647)	(2.337)	(2.341)	(0.931)
Above, and	44.913	52.425	2.662			
participation weighted	(2.232)	(2.231)	(0.661)			
Above, and	44.772	52.567	2.661	47.408	49.551	3.042
missing data weighted	(2.237)	(2.238)	(0.656)	(2.338)	(2.342)	(0.924)
p-values						
Bush(M=1) = Bush(M=0)		0.4155				
Kerry(M=1) = Kerry(M=0)		0.3518				
Joint		0.6374				
Wave 1 & 2 Combined		N = 1322			N = 1393	
Survey weighted	46.037	50.429	3.534	46.364	49.333	4.303
	(1.485)	(1.490)	(0.501)	(1.563)	(1.573)	(0.684)
P(vote) > 0		N = 1190			N = 1181	
Survey weighted	46.973	50.102	2.925	48.119	49.204	2.676
	(1.582)	(1.589)	(0.495)	(1.705)	(1.709)	(0.627)
Above, and	46.886	50.445	2.669			
participation weighted	(1.633)	(1.636)	(0.459)			
Above, and	46.655	50.687	2.657	48.084	49.250	2.666
missing data weighted	(1.646)	(1.652)	(0.454)	(1.706)	(1.711)	(0.624)
p-values						
Bush(T=1) = Bush(T=0)		0.5467				
Kerry(T=1) = Kerry(T=0)		0.5457				
Joint		0.8295				

In all results we employ survey weights provided by TESS designed to match the demographics of the surveyed sample to the U.S. Census and the Knowledge Networks Panel. The pooled results employ DFL weights to account for differences in observed sample demographics between waves. Likely voter weights use the reported probability of voting (for the Probabilistic group only) to adjust results. The missing data weights use DFL weights to account for 58 dropped observations with missing poll results on observed dimensions of demographics. All weights (except the TESS survey weights) are estimated using probit regressions of the appropriate outcome on a flexible set of the individual demographics including age, age squared, and dummies for each of the categorical variables in web appendix Table 9. Actual national 2004 election results were Bush 50.733%, Kerry 48.270%, and Other 0.996%. Heteroskedasticity robust standard errors in parentheses.

Web Appendix: Figures

Elias Walsh, Sarah Dolfin and John DiNardo

January 8, 2009

Web Appendix Figure 1: Density Estimates of Standardized Prediction Errors, by Election Year
The figure displays a kernel density of the standardized prediction errors for presidential state races by election year. The vertical lines are the estimated mean associated with the appropriate density. In comparison to the standard normal density, the theoretical prediction under random probability sampling, the poll densities are more disperse and are not centered at 0, indicating bias.

Web Appendix Figure 2: Density Estimates of Standardized Prediction Errors, by Poll Subgroup
The figure displays a kernel density of the standardized prediction errors for presidential state races by poll subgroup. In comparison to the standard normal density, the theoretical prediction under random probability sampling, the poll densities are more disperse, though the polls within two weeks of the election do show less dispersion.

Web Appendix Figures 3 & 4: The Relation Between Forecasts and Election Results
Each circle represents the mean of all poll results in a statewide election. The dashed line is the estimated line from a regression of the poll prediction on the actual election outcome. The solid line is the 45-degree line. The slope of the estimated line is always less than 1 (see also web appendix Table 8). Thus bias in polls tends to work in a way that understates larger vote shares and overstates smaller vote shares. This could be explained as a result of "honest Bayesian" type behavior on the part pollsters, or simply an artifact of other problems in polling that cause bias. For Democratic candidates the point at which the regression line crosses the 45-degree line is below 50%, while for Republicans this crossing point tends to be higher. If pollsters do act like "honest Bayesians" then these crossing points may be indicative of the pollsters' prior beliefs about a candidate's vote share. If for example, pollsters are reporting the maximum posterior density, then the nonzero intercept and departure of the slope from 1 are the consequence of the standard omitted variable calculations where the omitted variable is pollsters' prior information. These general findings are not changed much if we limit the analysis to only those polls conducted within two weeks of the election.

Web Appendix Figures 5 & 6: The Relation Between Forecast Errors and Election Results
Figures 5 and 6 are analogues to Figures 3 and 4 with poll prediction errors in the place of the predicted vote shares. Each circle represents the mean of all poll prediction errors in a statewide election. The estimated line from a regression of the poll prediction errors on the actual election outcome is always negative. The main benefit to displaying the prediction errors rather than the predicted shares is that the scatter plot is more clearly presented.

Web Appendix Figure 7: Distribution of Polls Across States By Election Result and Number of Electoral Votes
Each circle represents a statewide election. The area of the circle is proportional to the number of polls in that race. Races with more polls tend to be concentrated in states with more electoral votes and in states that are more highly contested. We would expect to see very large circles in states that both have many electoral votes and are close races, however, the only state with more than 40 electoral votes is California, a state that is not particularly competitive.

Web Appendix Figure 8: Standardized Prediction Errors Over Time
The figure displays scatter plots of standardized prediction errors for presidential statewide races and quantile regressions at the 10th, 50th, and 90th quantile. The lines in panels (a) and (c) present the results of a

quantile regression of the prediction errors on the number of days before the election and a constant term. Panels (b) and (d) present the 10th and 90th quantiles, and associated confidence intervals from a design-adaptive bandwidth quantile regression, limiting the sample to only those polls within 10 weeks of the election. The two dotted horizontal lines in each of panels (b) and (d) indicate the theoretical prediction of the 90th and 10th percentiles under standard normality (1.28/-1.28). The panels demonstrate that dispersion in the poll errors diminishes over time, but even for the closest polls to the election the dispersion exceeds that of a standard normal density.

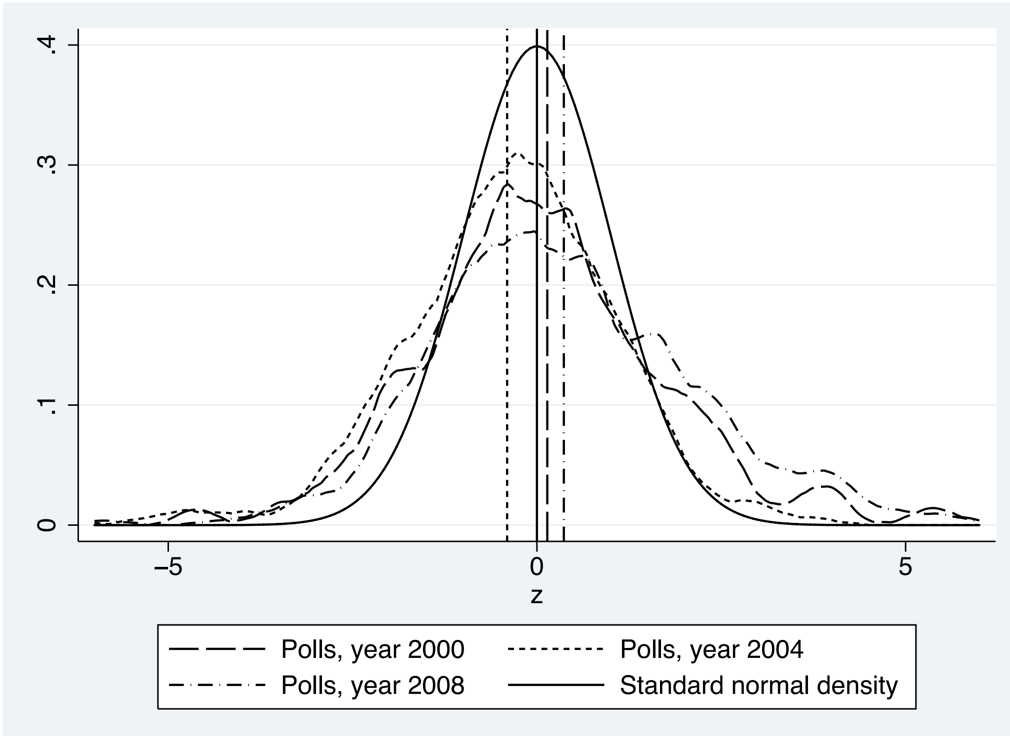
Web Appendix Figures 9 & 10: Standardized Prediction Errors Over Time, by Election Year

The figures display scatter plots of standardized prediction errors for presidential statewide races and quantile regressions at the 10th, 50th, and 90th quantile for polls separately by election year. The lines in panels (a), (c), and (e) present the results of a quantile regression of the prediction errors on the number of days before the election and a constant term. Panels (b), (d), and (f) present the 10th, and 90th quantiles and associated confidence intervals from a design-adaptive bandwidth quantile regression, limiting the sample to only those polls within 10 weeks of the election. The two dotted horizontal lines in each of panels (b), (d), and (f) indicate the theoretical prediction of the 90th and 10th percentiles under standard normality (1.28/-1.28). As in Figure 9, the panels generally demonstrate that dispersion in the poll errors diminishes over time, but even for the closest polls to the election the dispersion exceeds that of a standard normal density. We see some variation across election, with the 2004 polls for both the Republican and Democratic candidate displaying more-or-less constant dispersion over time. Also, the design-adaptive bandwidth quantile regressions do not always reject the prediction for the 10th and 90th quantiles of the standard normal density for the closest polls to the election.

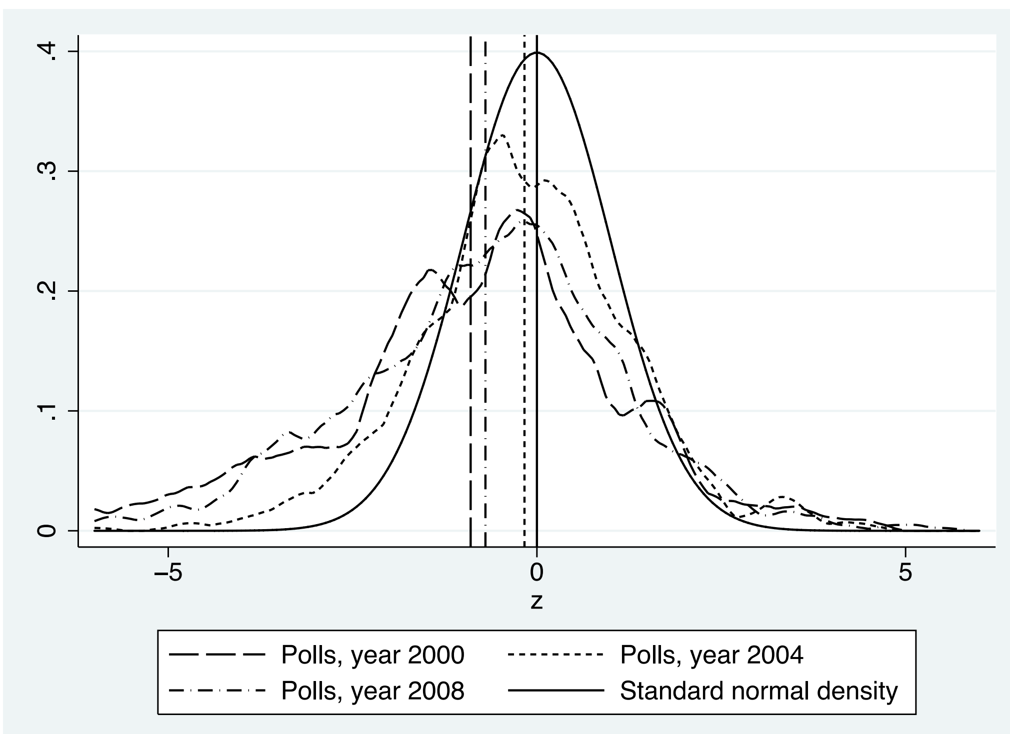
Web Appendix Figure 11: Density Estimates of Standardized Prediction Errors, by Detailed Poll Subgroup

Figure 11 is an extension of Figure 2 with two additional subgroups: polls that sum to 100-102 percentage points, and polls that do not allow third party candidates as an option for respondents. The polls that sum to 100-102 do not look much better than the density of all polls. The polls that exclude third parties show about the same amount of dispersion as polls more generally, but in the case of the Republican share the density is shifted to the right, indicating bias in the direction of over-prediction.

Web Appendix Figure 1: Density Estimates of Standardized Prediction Errors, by Election Year



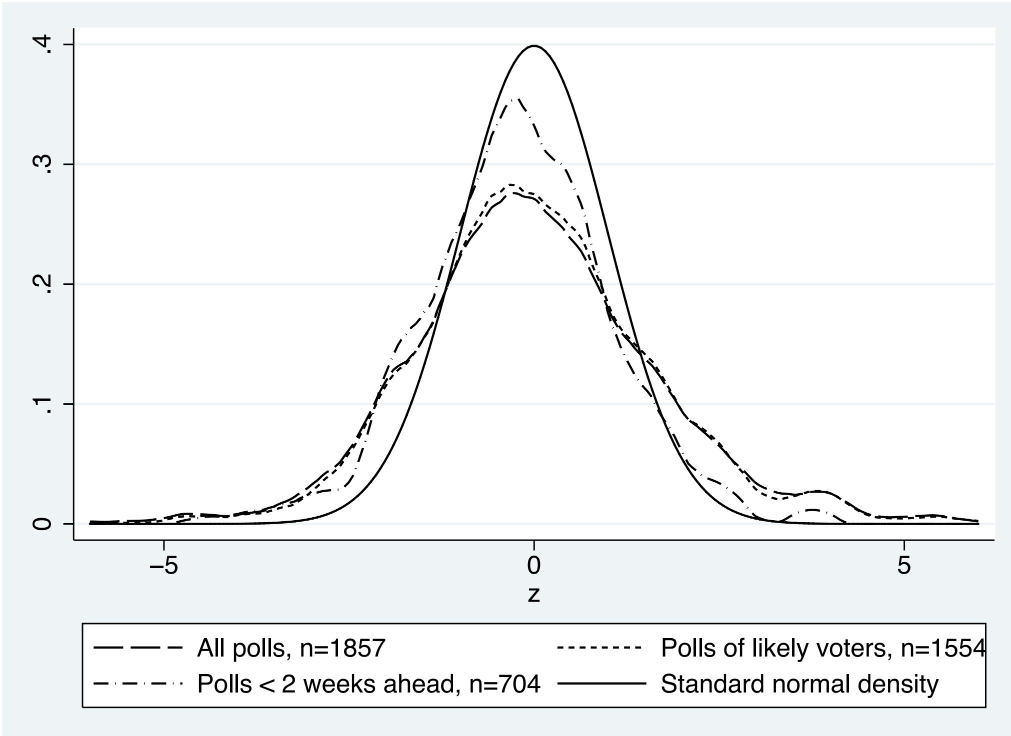
(a) Republican Prediction Error



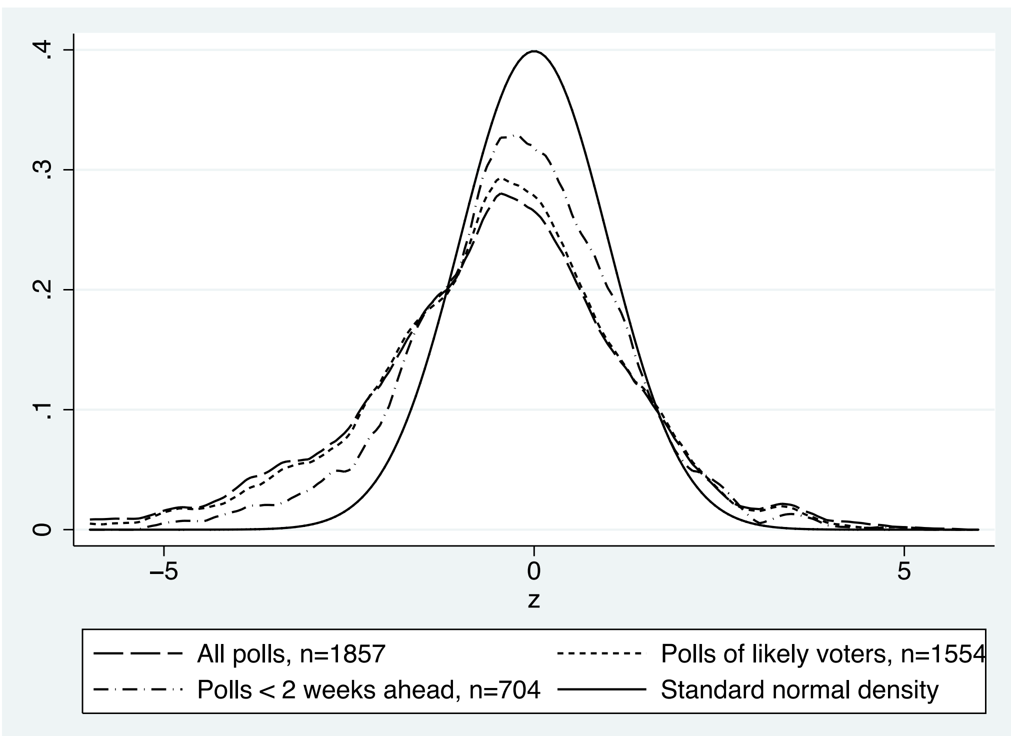
(b) Democratic Prediction Error

The figure displays a kernel density of the standardized prediction errors for presidential state races by election year. The vertical lines are the estimated mean associated with the appropriate density.

Web Appendix Figure 2: Density Estimates of Standardized Prediction Errors, by Poll Subgroup



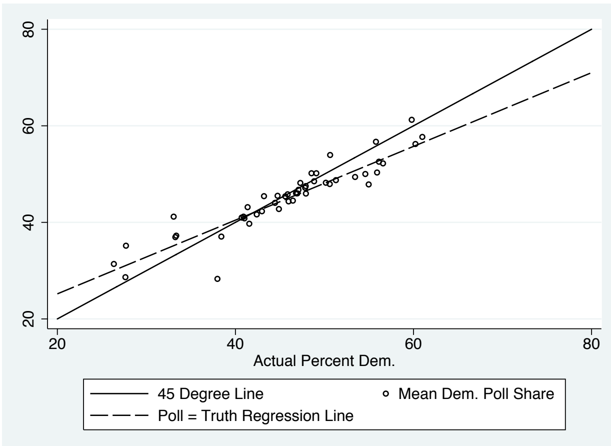
(a) Republican Prediction Error



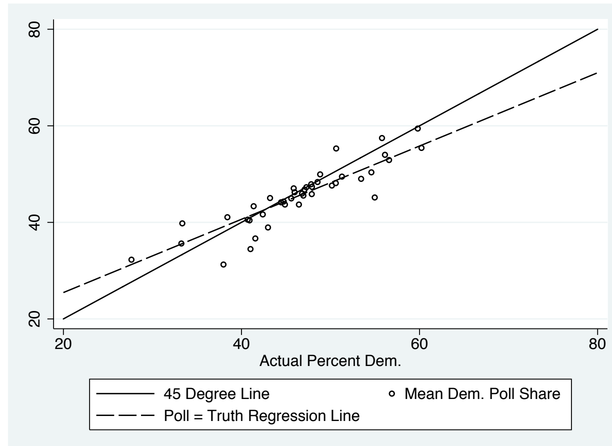
(b) Democratic Prediction Error

The figure displays a kernel density of the standardized prediction errors for presidential state races by poll subgroup.

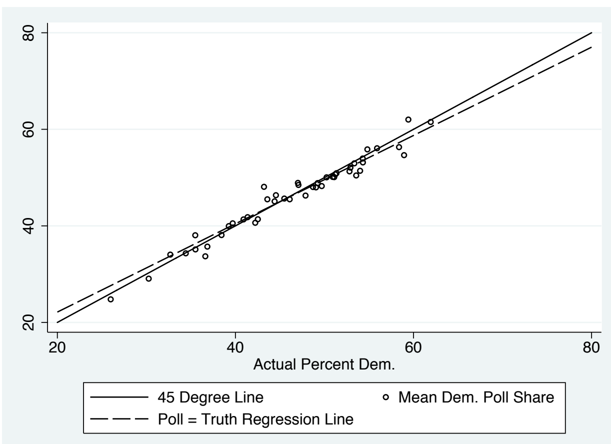
Web Appendix Figure 3: The Relation Between Forecasts and Election Results, Democratic Vote Share



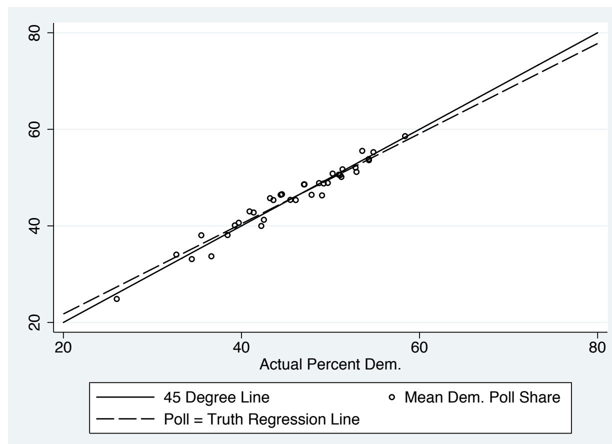
(a) All Polls, 2000



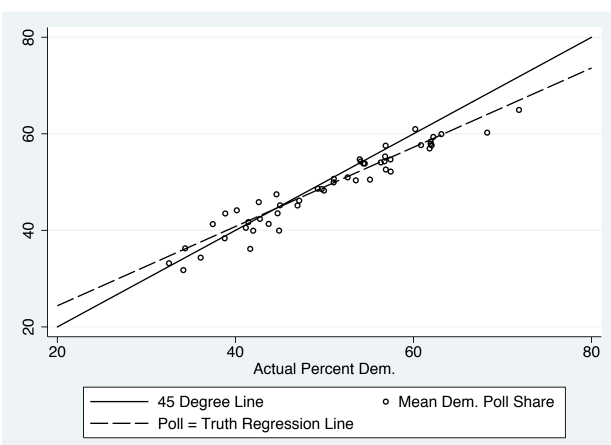
(b) Polls Within Two Weeks of Election, 2000



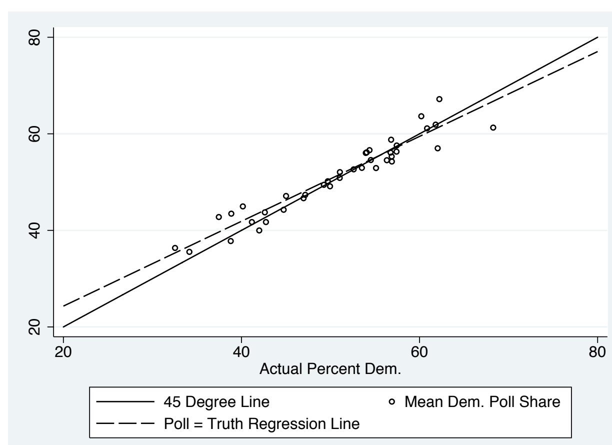
(c) All Polls, 2004



(d) Polls Within Two Weeks of Election, 2004



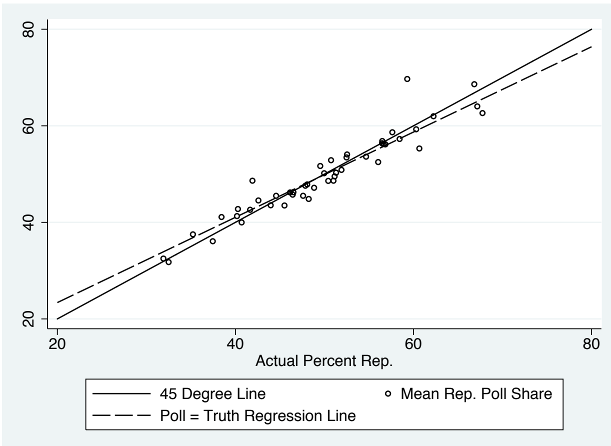
(e) All Polls, 2008



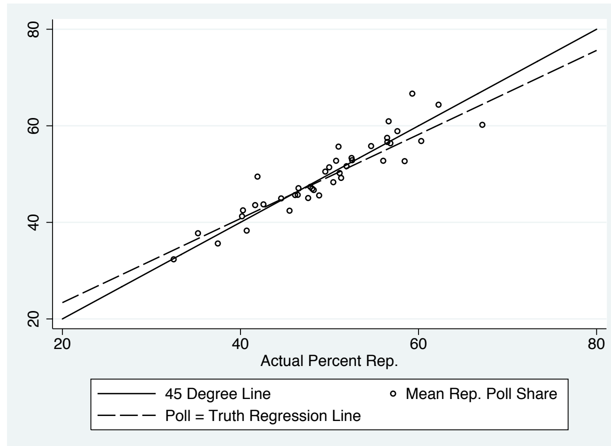
(f) Polls Within Two Weeks of Election, 2008

Each circle represents the mean of all poll results in a statewide election. The dashed line is the estimated line from a regression of the poll prediction on the actual election outcome. The solid line is the 45-degree line.

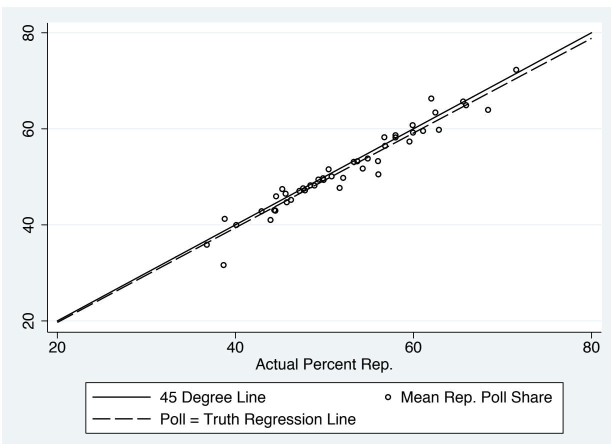
Web Appendix Figure 4: The Relation Between Forecasts and Election Results, Republican Vote Share



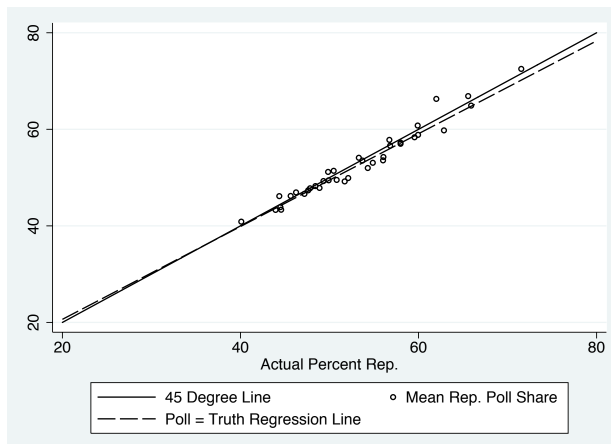
(a) All Polls, 2000



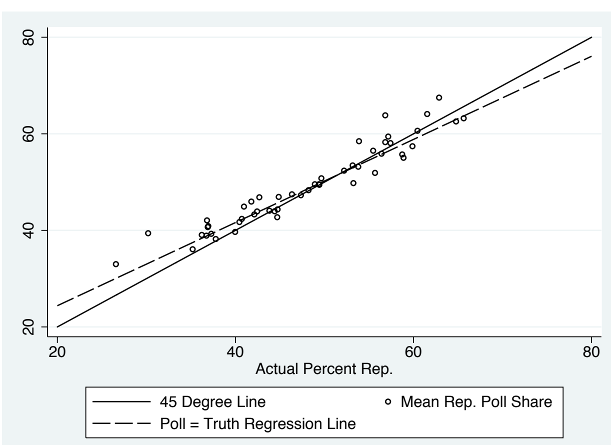
(b) Polls Within Two Weeks of Election, 2000



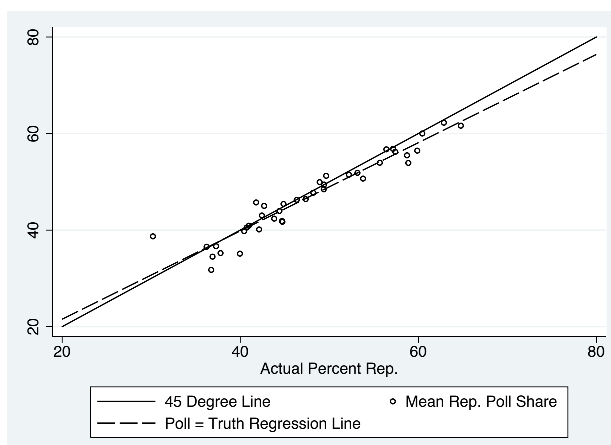
(c) All Polls, 2004



(d) Polls Within Two Weeks of Election, 2004



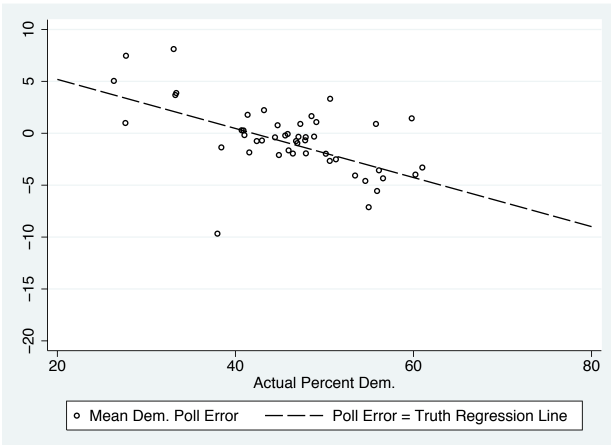
(e) All Polls, 2008



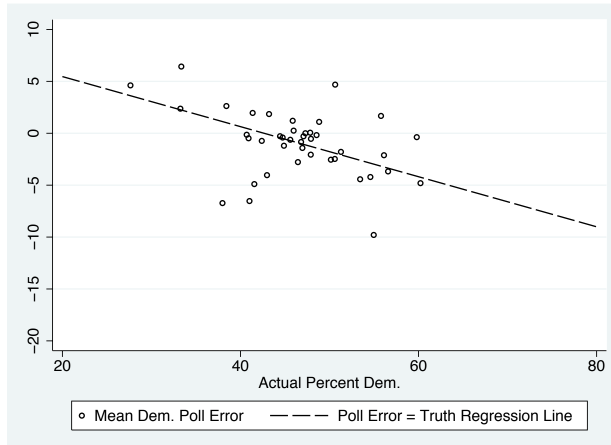
(f) Polls Within Two Weeks of Election, 2008

Each circle represents the mean of all poll results in a statewide election. The dashed line is the estimated line from a regression of the poll prediction on the actual election outcome. The solid line is the 45-degree line.

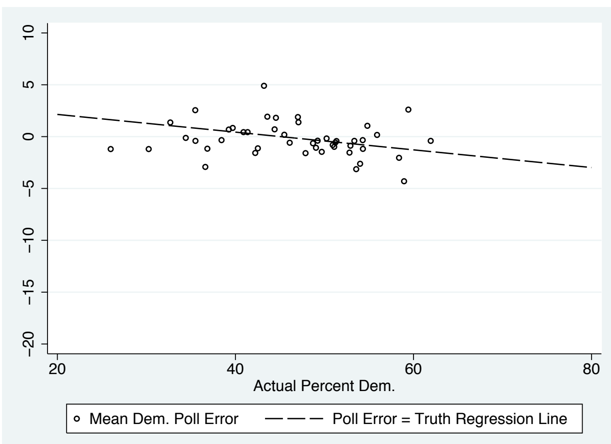
Web Appendix Figure 5: The Relation Between Forecast Errors and Election Results, Democratic Vote Share



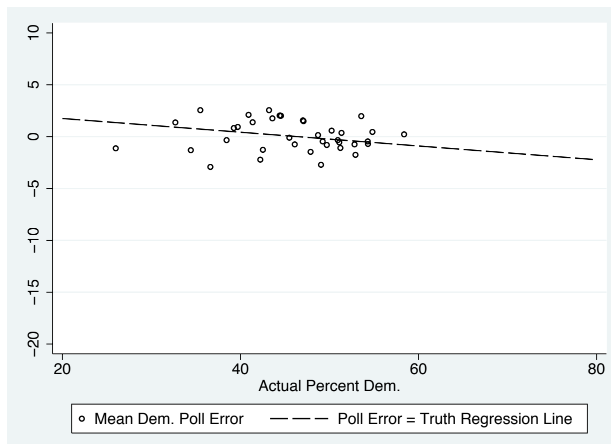
(a) All Polls, 2000



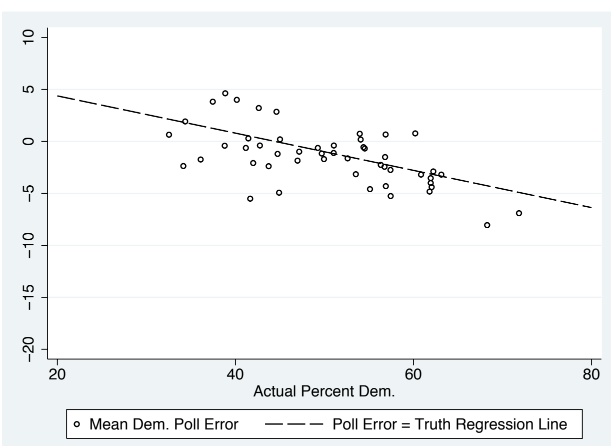
(b) Polls Within Two Weeks of Election, 2000



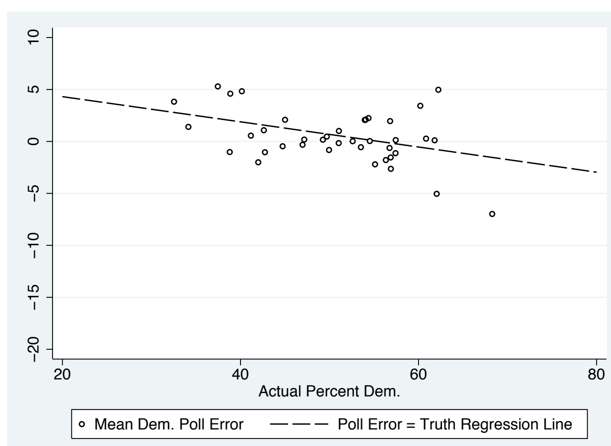
(c) All Polls, 2004



(d) Polls Within Two Weeks of Election, 2004



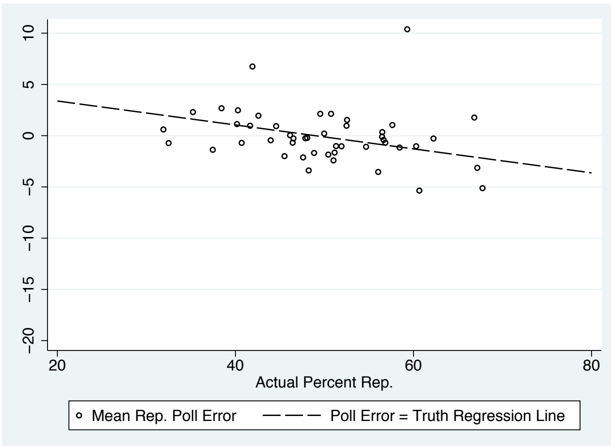
(e) All Polls, 2008



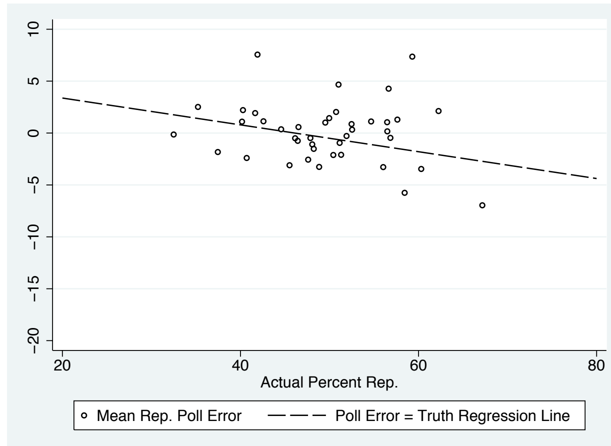
(f) Polls Within Two Weeks of Election, 2008

Each circle represents the mean of all poll prediction errors in a statewide election. The dashed line is the estimated line from a regression of the poll prediction errors on the actual election outcome.

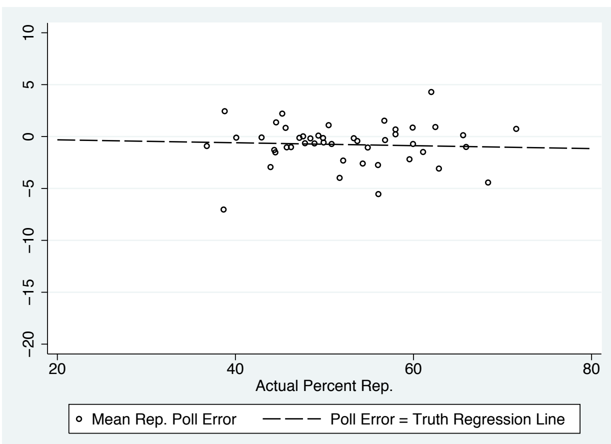
Web Appendix Figure 6: The Relation Between Forecast Errors and Election Results, Republican Vote Share



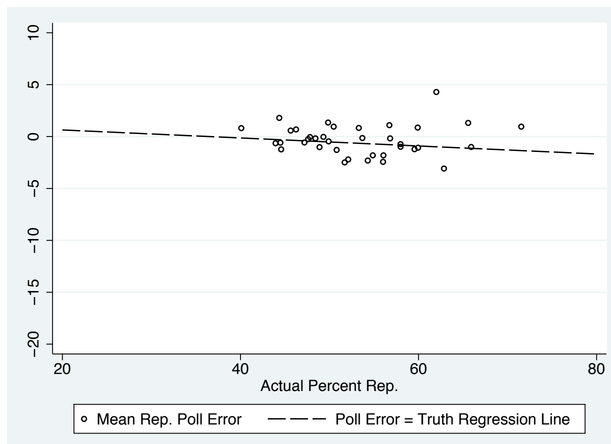
(a) All Polls, 2000



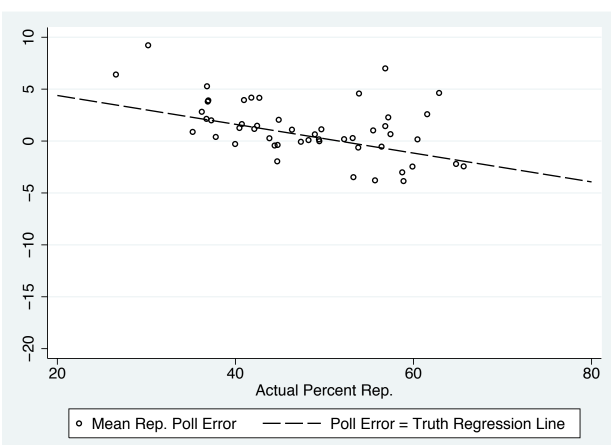
(b) Polls Within Two Weeks of Election, 2000



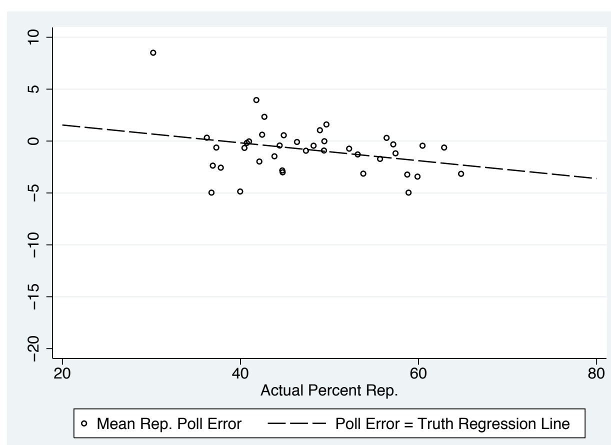
(c) All Polls, 2004



(d) Polls Within Two Weeks of Election, 2004



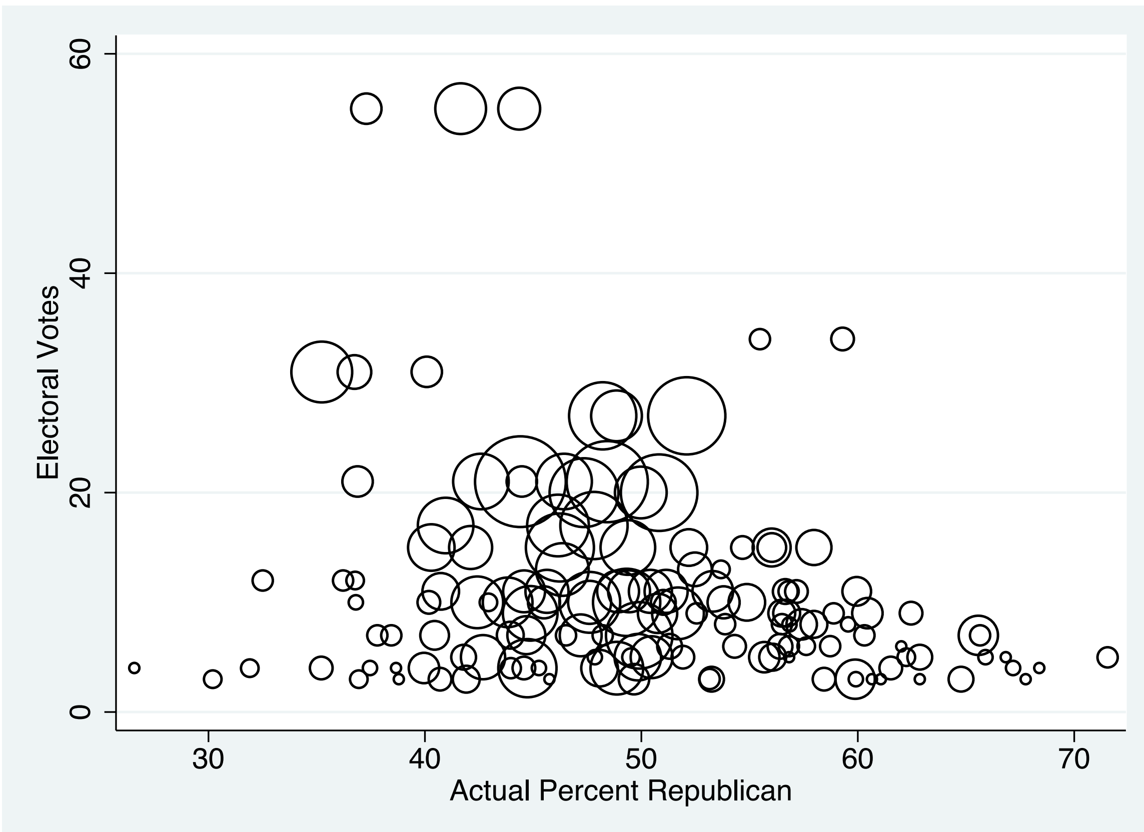
(e) All Polls, 2008



(f) Polls Within Two Weeks of Election, 2008

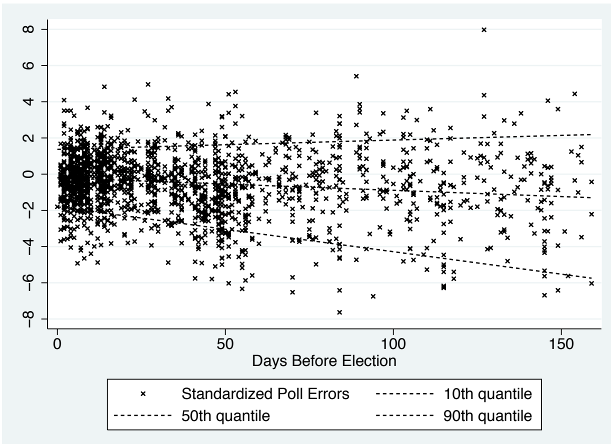
Each circle represents the mean of all poll prediction errors in a statewide election. The dashed line is the estimated line from a regression of the poll prediction errors on the actual election outcome.

Web Appendix Figure 7: Distribution of Polls Across States By Election Result and Number of Electoral Votes

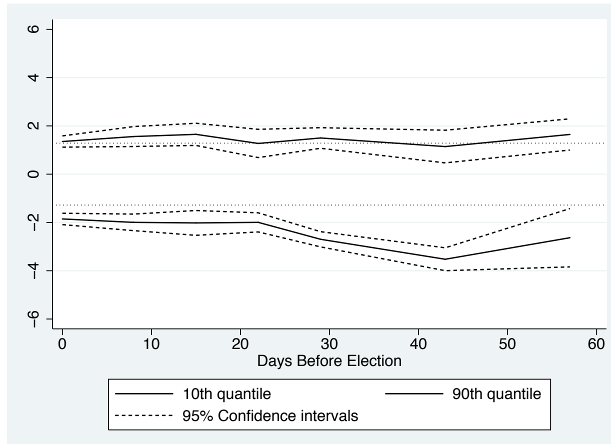


Each circle represents a statewide election. The area of the circle is proportional to the number of polls in that race.

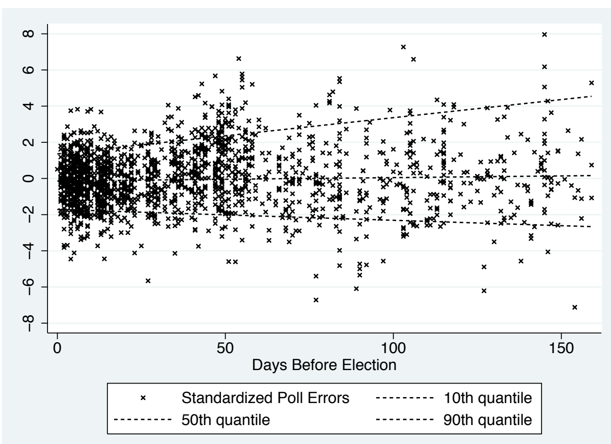
Web Appendix Figure 8: Standardized Prediction Errors Over Time



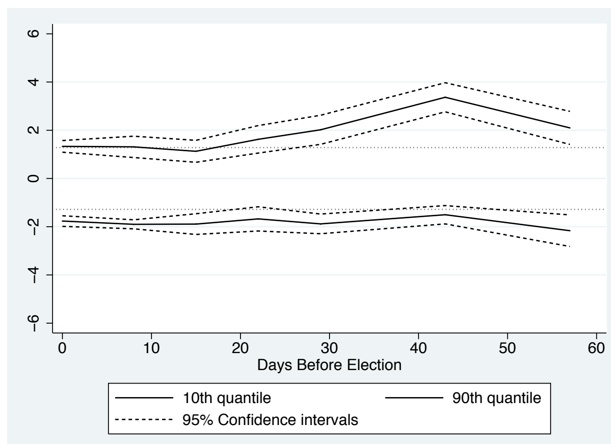
(a) All Democratic Standardized Prediction Errors



(b) All Democratic Standardized Prediction Errors



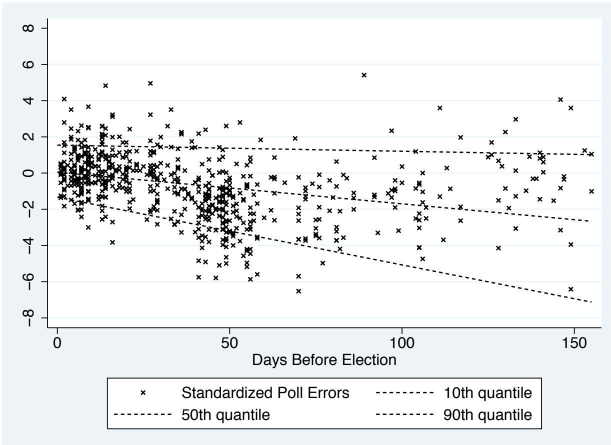
(c) All Republican Standardized Prediction Errors



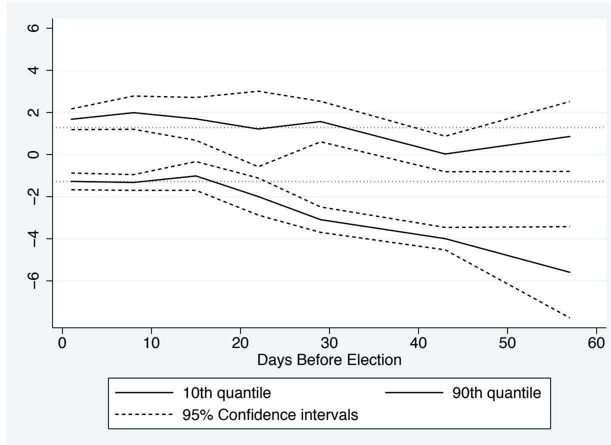
(d) All Republican Standardized Prediction Errors

The figure displays scatter plots of standardized prediction errors for presidential statewide races and quantile regressions at the 10th, 50th, and 90th quantile. The lines in panels (a) and (c) present the results of a quantile regression of the prediction errors on the number of days before the election and a constant term. Panels (b) and (d) present the 10th, and 90th quantiles and associated confidence intervals from a design-adaptive bandwidth quantile regression, limiting the sample to only those polls within 10 weeks of the election. The two dotted horizontal lines in each of panels (b) and (d) indicate the theoretical prediction of the 90th and 10th percentiles under standard normality (1.28/-1.28).

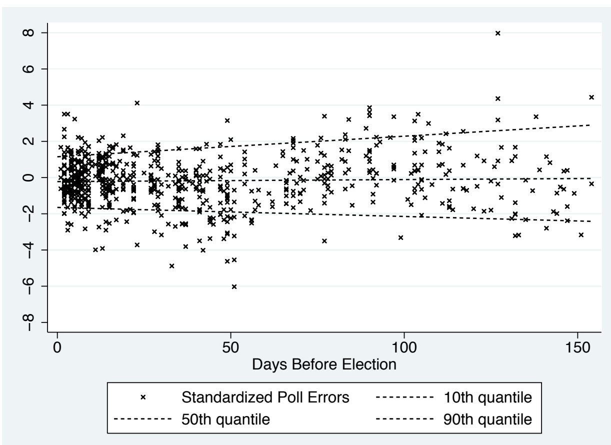
Web Appendix Figure 9: Standardized Democratic Prediction Errors Over Time, by Election Year



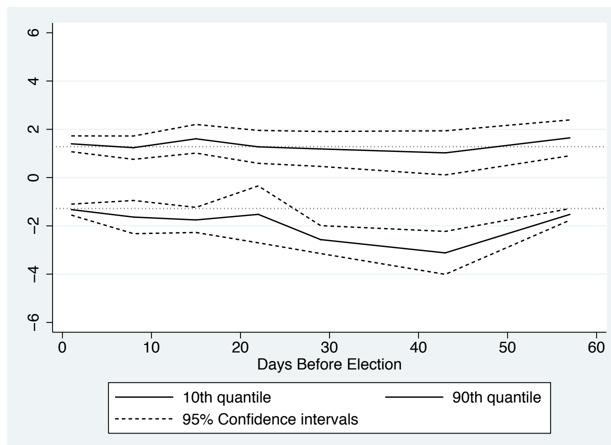
(a) 2008 Standardized Prediction Error



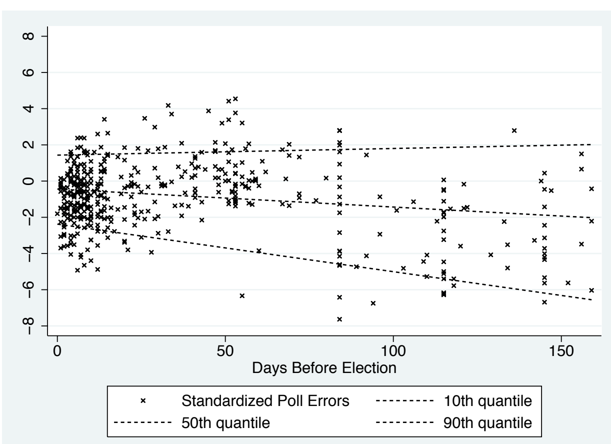
(b) 2008 Standardized Prediction Error



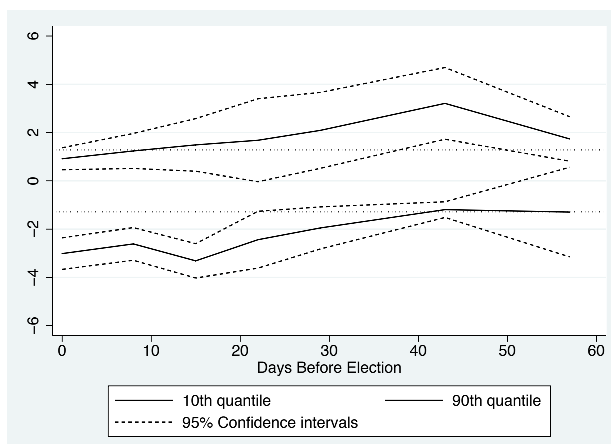
(c) 2004 Standardized Prediction Error



(d) 2004 Standardized Prediction Error



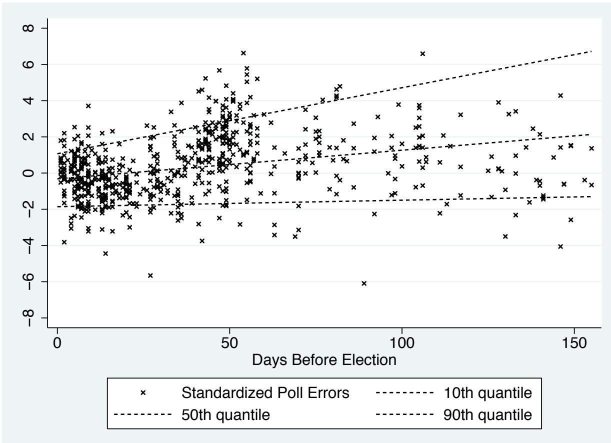
(e) 2000 Standardized Prediction Error



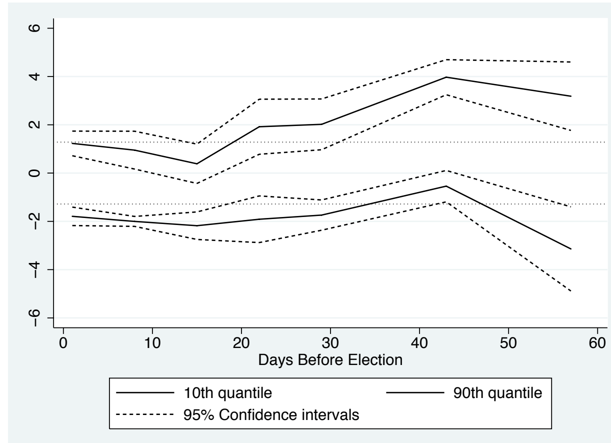
(f) 2000 Standardized Prediction Error

The figures display scatter plots of standardized prediction errors for presidential statewide races and quantile regressions at the 10th, 50th, and 90th quantile for polls separately by election year. The lines in panels (a), (c), and (e) present the results of a quantile regression of the prediction errors on the number of days before the election and a constant term. Panels (b), (d), and (f) present the 10th, and 90th quantiles and associated confidence intervals from a design-adaptive bandwidth quantile regression, limiting the sample to only those polls within 10 weeks of the election. The two dotted horizontal lines in each of panels (b), (d), and (f) indicate the theoretical prediction of the 90th and 10th percentiles under standard normality (1.28/-1.28).

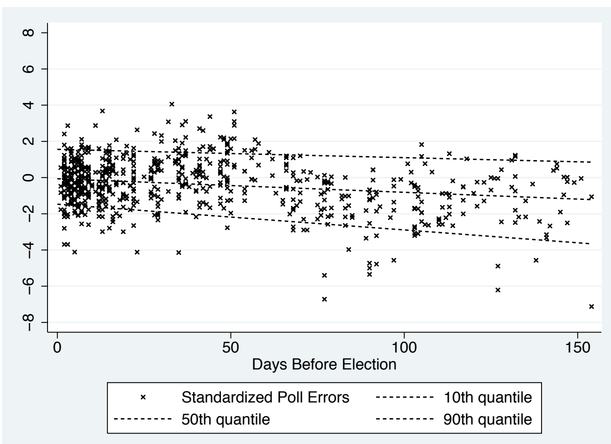
Web Appendix Figure 10: Standardized Republican Prediction Errors Over Time, by Election Year



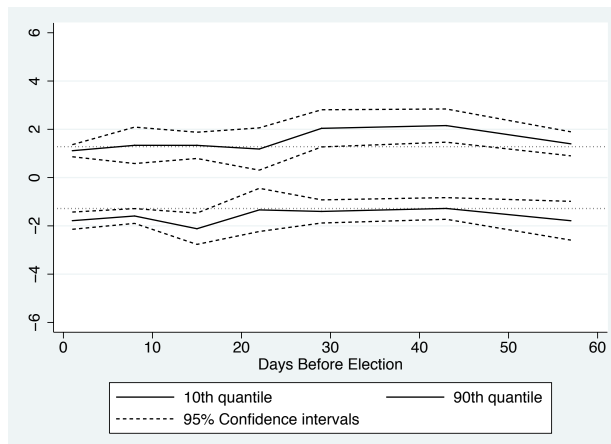
(a) 2008 Standardized Prediction Error



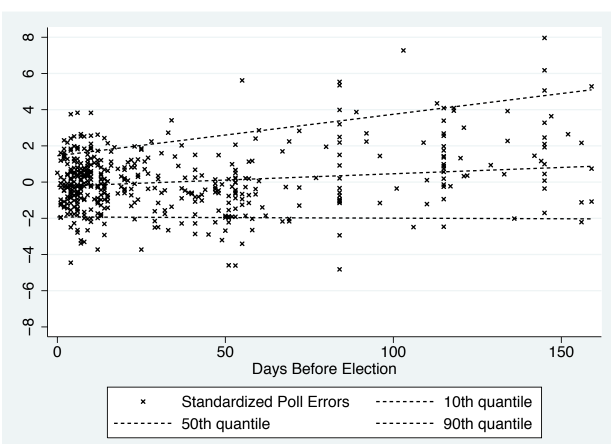
(b) 2008 Standardized Prediction Error



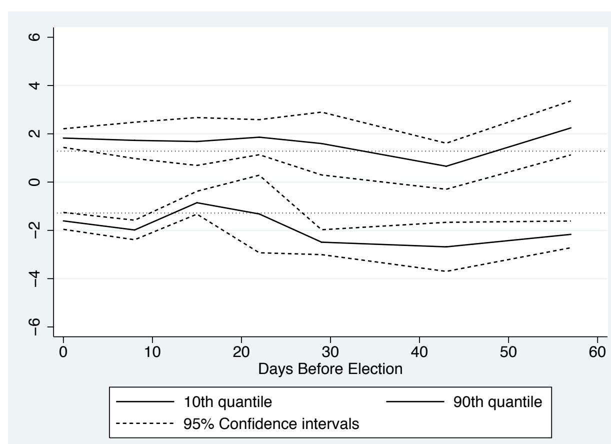
(c) 2004 Standardized Prediction Error



(d) 2004 Standardized Prediction Error



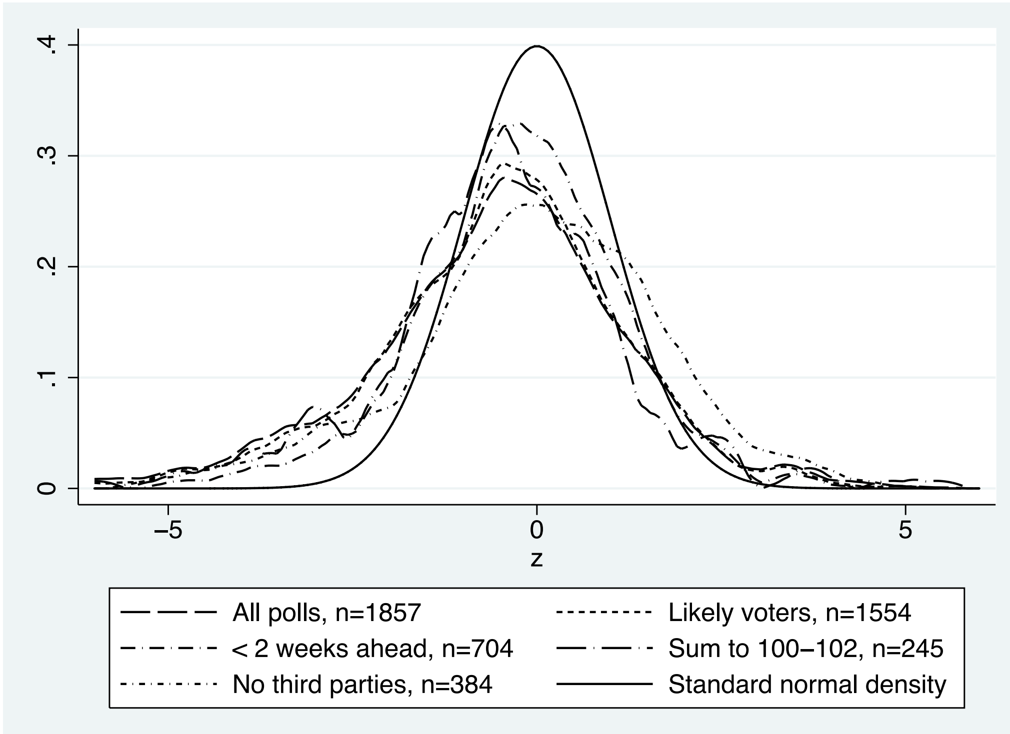
(e) 2000 Standardized Prediction Error



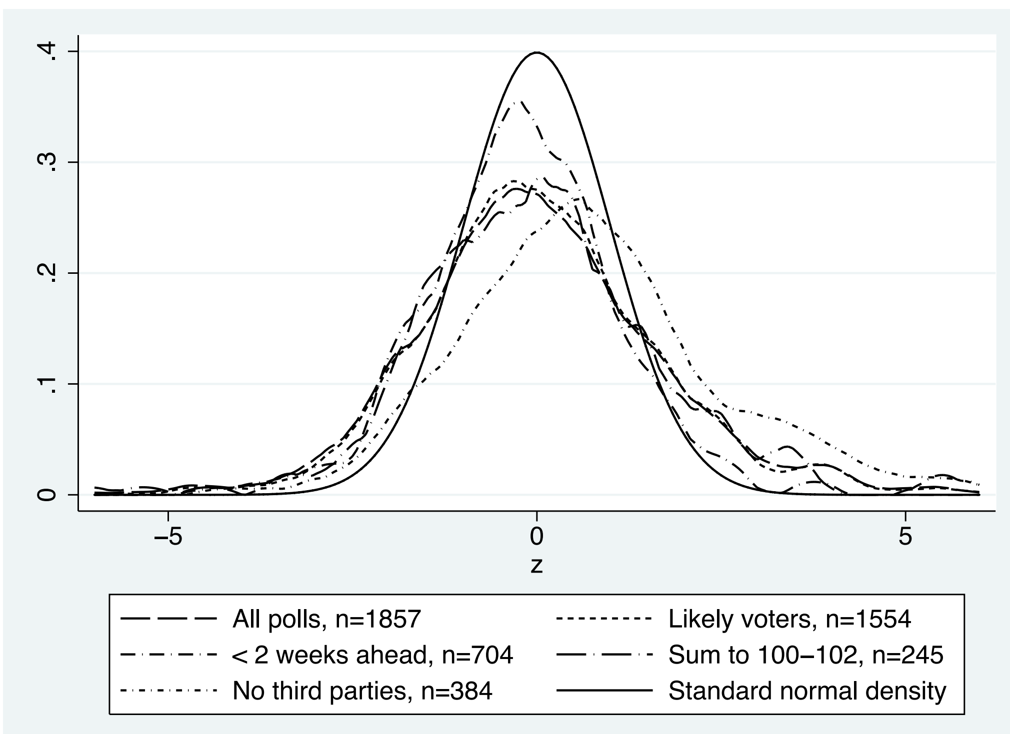
(f) 2000 Standardized Prediction Error

The figures display scatter plots of standardized prediction errors for presidential statewide races and quantile regressions at the 10th, 50th, and 90th quantile for polls separately by election year. The lines in panels (a), (c), and (e) present the results of a quantile regression of the prediction errors on the number of days before the election and a constant term. Panels (b), (d), and (f) present the 10th, and 90th quantiles and associated confidence intervals from a design-adaptive bandwidth quantile regression, limiting the sample to only those polls within 10 weeks of the election. The two dotted horizontal lines in each of panels (b), (d), and (f) indicate the theoretical prediction of the 90th and 10th percentiles under standard normality (1.28/-1.28).

Web Appendix Figure 11: Density Estimates of Standardized Prediction Errors, by Detailed Poll Subgroup



(a) Democratic Prediction Error



(b) Republican Prediction Error

The figure displays a kernel density of the standardized prediction errors for presidential state races by poll subgroup.

Web Appendix: Discussion of Intentions and Polling

Elias Walsh, Sarah Dolfin and John DiNardo

January 8, 2009

I. Probabilistic Intentions

While a large literature (see Crespi (1988) for a nice summary) suggests that "horse race" polls – those that ask respondents about who they intend to vote for in an election – should, if conducted properly and under the right conditions, reflect actual outcomes, an old statistical literature, most recently Manski (1990) suggests the opposite. Manski (1990) observes that if a potential voter is uncertain about who s/he will vote then a simple "intention" question: "who are you likely to vote for" will be biased in general for the outcome even if agents are perfectly rational, etc. The only hope for generating an unbiased prediction of an outcome from intentions data requires asking the question in such a way that allows the voter to express his or her uncertainty.

Instead of asking: *If the election were held today, would you*:

- Vote for John Kerry, the Democratic nominee for president.
- Vote for George Bush, the Republican nominee for president.
- Vote for another candidate.

one should ask the question in terms of *probabilities* for voting for each of the candidates.

It seems worthwhile to ask whether this "'theoretical" source of bias can explain much of the bias we observe in actual polls. In a sense, we would like to see the extent to which this purely "statistical" problem addresses the question posed by Gelman and King (1993) – are polls variable only because the questions are posed as intentions instead of probabilities? The purpose of this section of the paper is to investigate the importance of this question by a comparison of responses to "horse race" questions asked the usual way, and the way suggested by Manski's analysis. Both trends and the reliability of the implied forecast may be quite different for the two sets of questions and this might yield insights as to why polls tend to be biased forecasts of the outcomes.

While this source of bias has been studied extensively for continuous outcomes such as income (see Domintz and Manski (1997) for a review and example) to the best of our knowledge has not been studied in this context. This problem arises routinely in data of interest to political scientists, economists, sociologists and others and may have implications for broader issues than merely horse race election polling *per se.*

Although "horse race" polls are routinely used to forecast the likelihood that some candidate will win an election, it is well understood in the statistics literature that even in the "best case" there is no reason to suppose that "intentions" ("I am likely to vote for candidate X") should yield unbiased forecasts of actual behavior. (Manski, 1990)

We first focus on a "best case" scenario and illustrate with some simple numerical examples why

1. Polls should be biased in general.
2. Even large positive changes in poll results over time do not necessarily indicate increased support for the candidate.

In doing so, we focus only on the possibility that some individuals are uncertain about who they will vote for. We assume that all the other possible problems (sample selection biases, question ordering, etc.) that have been cited in the literature are solved.[1] As a rule, assuming something worse than the "best case" results in an even greater bias and for reasons of brevity and clarity we omit that discussion here.

A The Best Case

Following Manski (1990), let i be a binary indicator denoting an intention – "talking about the presidential elections in November, for whom are you likely to vote – George Bush?" and let y be the indicator corresponding to the actual behavior (the individual votes for Bush). Letting s denote the information available at the time of the survey to the respondent and let z denote the events that have not yet occurred but that will affect his future action.[2] Let $P_z|s$ denote the objective distribution of z conditional on s. Let $P(y|s)$ denote the objective distribution of y conditional on s. The event $y = 1$ occurs $\iff$ the *realization* of z is such that $y(s, z) = 1$.

In the *best* case, we assume rational expectations: this means the respondent *knows* how they will act depending on the possible realizations of z and that they also know $P_z|s$ – that is they know the stochastic process generating z – in words, the respondent knows the correct distribution of the behavior influencing events z and moreover uses that information optimally. To take a concrete example, suppose z is the public exposure of a scandal involving "morals" or sexual behavior of a candidate. This assumption is the requirement that I know how I would behave if my candidate were involved in a scandal and the the probability that I would learn about such a scandal before election day.

The second aspect of the "best case" scenario is that the respondent states her best point prediction of her behavior. The best prediction depends on her "loss function" associated with either $(i = 1, y = 0)$ and $(i = 0, y = 1)$. Manski observes that under these two sets of assumptions the responses satisfy:

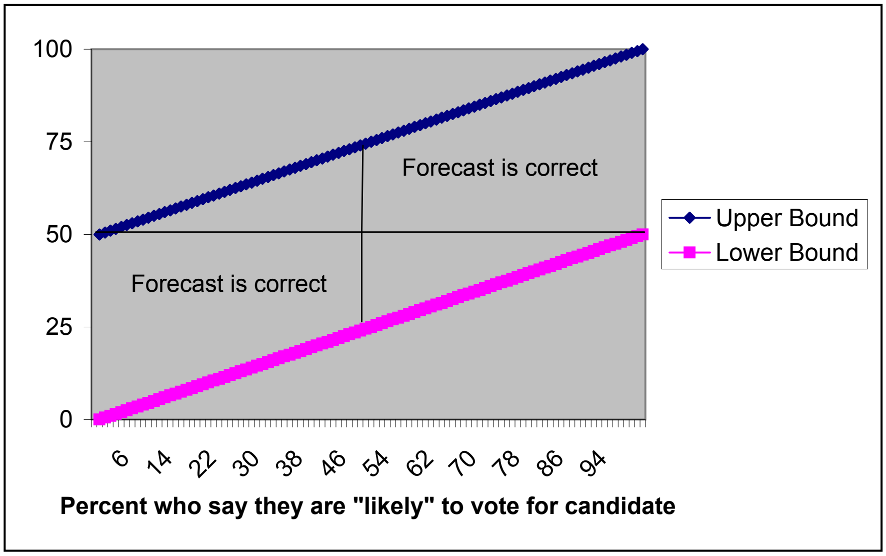
$$\begin{array}{lcl} i = 1 & \Longrightarrow & P(y = 1|s) \geq \pi \\ i = 0 & \Longrightarrow & P(y = 1|s) \leq \pi \end{array} \tag{1}$$

In words, if the action y is "voting for candidate X", then a respondent tells the interviewer that she will vote for candidate X if the probability that she will do so is greater than π. If both possible errors are equally "costly" than $\pi = .5$ Specializing to the case of horse race polls, the object of the poll is to learn the probability $P(y = 1|i, s)$. As Manski observes, however, the pollster's data on "intentions" does not identify that probability. Even in this "best case" – assuming that persons have identical loss functions – they only imply a "bound". As Manski shows:

[1] See for example, Gelman and King (1993) or Ottaviani and Norman (2006) for discussions.

[2] To make the problem even more simple, we assume that a person's participation is known with certainty. Allowing for uncertainty in participation only strengthens the negative result.

Figure 1: The Bounds Implied by "Intentions" are not tight: A comparison of intentions with outcomes



$$P(y = 1|s, i = 0) \leq \pi \leq P(y = 1|s, i = 1)$$

expresses *all* the information in intentions data.

Figure 1 displays the bounds implied by the data assuming no sampling error, that individuals have identical symmetric loss functions, and that there is no "new information" s between the time the poll is taken and the behavior occurs. The dependent variable is the actual voting outcome on election day.

The lower right and upper left triangles that lie within the polygon formed by the bounds indicate that 25 percent of the area within the bounds fail even to cover the correct binary prediction of the outcome. Note that it would be incorrect to draw the inference that the polls would get it right 75 percent of the time in this best case. Rather, the correct inference is that the correct bounds do not have to cover the correct binary prediction of the election outcome. Of course, if the sample is not a random sample, new information occurs between the poll and the event, or that there is a double uncertainty (i.e. the voter does not know for certain whether s/he votes) the bounds could easily be *completely* uninformative.

B A Rise in the Polls Doesn't Necessarily Imply Increased Support

Observe that we have gone a bit beyond even the "best case" in this simple illustration. As Manski observes (and as was observed earlier by Juster (1966), for example) it has been well known in the statistical literature that such polls will *not* be unbiased in general, even in this *best case*. As a consequence, a poll is especially unsuited to assessing "trends" in voter support for a candidate, even when the electorate is composed of Bayesian statisticians with correct rational expectations.

The following contrived example, although not altogether unreasonable, shows an example where support for candidate X is falling (measured as what would actually have happened if an election had been conducted), at the same time the polls are showing a massive increase in support for the candidate. For simplicity, we have three types of voters. Type "C" voters strongly support candidate X, type "B" voters less strongly support candidate X, and type "A" voters strongly oppose candidate X. Between the two periods, type "A" voters grow much more strongly opposed to candidate X, and type "B" voters slightly shift in favor of candidate X. As a consequence, the polling shows a large increase in support for candidate X from period 1 to period 2, even as the actual probability of X being elected fell over this time!

Table 1: Polls show increased support, when support is falling

		Time Period 1		Time Period 2	
Voter "Type"	Fraction in Population	Probability vote for X	Response to Pollster	Probability vote for X	Response to Pollster
A	0.25	0.4	0	.1	0
B	0.5	0.46	0	.51	1
C	0.25	0.8	1	.8	1
		Actual Outcome	0.53	Actual Outcome	0.48
		Poll Result	0.25	Poll Result	0.75

C Voter Participation

The above analysis applies *mutatis mutandis* to an analysis of voter participation as a separate inquiry. As far as we have been able to ascertain, polling organizations routinely use a binary measure of whether or not an individual is likely to vote. Again, if the decision to participate is uncertain, in general there is no reason to believe that restricting to the sample to "likely voters" or "registered voters" (the two most frequently used screens in practice) will yield an unbiased rate of participation.

Moreover, since – in the simplest model – the act of the voting for a specific candidate is the product of two uncertain decisions (a decision to support the candidate, and the act of going to the polling booth) it is clear that treating the corresponding sets of intentions as certain – i.e. binary – is biased as a forecast of the actual vote or the "strength" of the support for a candidate.

References

Crespi, Irving, *Pre–Election Polling: Sources of Accuracy and Error*, New York: Russell Sage Foundation, 1988.

Dominintz, Jeff and Charles F. Manski, "Using Expectations Data to Study Subjective Income Expectations," *Journal of the American Statistical Association*, September 1997, *92* (439), 855–867.

Gelman, Andrew and Gary King, "Why are American Presidential Election Campaign Polls So Variable When Votes are So Predictable?," *British Journal of Political Science*, October 1993, *23* (4), 409–451.

Juster, Thomas, "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design," *Journal of the American Statistical Association*, 1966, *61*, 658–696.

Manski, Charles F., "The Use of Intentions Data to Predict Behavior: A Best-Case Analysis," *Journal of the American Statistical Association*, December 1990, *85* (412), 934–940.

Ottaviani, Marco and Peter Norman, "The Strategy of Professional Forecasting," *Journal of Financial Economics*, August 2006, *81* (2), 441–446.