# 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

March 12, 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 strong presumption in the literature on professional forecasting in other contexts (such as interest rate forecasting), which do not rely on sampling per se, 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.

To illustrate the possible problem, consider the 42 "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

* 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).

either a tie or Gore ahead in the national race. The pollsters themselves appear to have felt that they did "well" in projecting this election. 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.[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 39 or more predictions for Bush out of the 42 is less than $5 \times 10^{-7}$ percent.[4]

## I. Background

Our chief argument is that pre-election presidential polling is an activity more akin to forecasting next year's GDP or the winner of a sporting match than to 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 a primary reason why pre–election polls (particularly those close to an actual election) 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, "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

[3]In the 1996 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 national review of the pollsters. 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 assume that Gore (the Democratic candidate) and Bush (the Republican candidate) received exactly the same number of votes, and the polls were independent samples.

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*

Non–response in polls is a critical concern for pollsters. 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[5] 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?"
> 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%?"

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 $\min(100, \mu + 66)$, where $\mu$ is the population mean.[6] 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

[5] Transcribed from a televised interview with John Stewart on The Daily Show (Zogby, 2004).
[6] Thus if the population mean is between 44 and 66 percent, then the poll is utterly uninformative.

conduct than the best of the pre–election presidential polls we study.[7] Add the uncertainty involved in estimating (not sampling) voter participation to the above worst-case bound, and almost any estimate can be obtained.

### *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, estimating the election outcome 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 intention. 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 have 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 percent of the vote for candidate $A$, but the election result would be 51 percent.[8]

[7] The two groups differed in the amount of effort that was spent in trying to procure a response. The additional effort included providing a small cash incentive, unlimited and staggered attempts to initiate an interview, and up to three additional attempts to complete unfinished or refused interviews (Pew Research Center, 1998).

[8] Indeed, it is simple to construct examples where, over time, the poll and the underlying preferences of the electorate go in separate directions. See the Web appendix for such an example.

## III. Polling Data

In Web appendix Table 2 we present descriptive information on the polling results we collected from `pollingreport.com`.[9] 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 1,761 polls with an average of about 12 polls per race, although some races had as few as one poll and some as many as 80. 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 is 702.

There are several problems with the data that deserve mention and some of these are summarized in Web appendix Table 3 and Table 4. Of particular importance is the fact that 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.[10] 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.[11] Web appendix Table 4 displays a tabulation of such cases. Nearly all of the polls in our sample require this adjustment. In all of the analysis that follows we focus on the adjusted shares.[12]

### *A. Results from Analyzing Pre-Election Polls*

Tables 1 and 2 summarize several key aspects of the polls we analyze as forecasts of election results. We consider separately all polls, polls which restrict themselves to "likely voters" only, and polls conducted within two weeks of the election.[13]

Taken as a whole, the polls, on the most favorable terms we can devise, do not behave as would be

[9]See the Web appendix for details on the data and sample selection criteria.

[10]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 $p_i^{\mathrm{Adj}}$ is given by $p_i^{\mathrm{Adj}} = r_c/(\sum_{i=1}^{C} r_i)$.

[11]In a related problem, the poll results are virtually always reported rounded to the nearest percentage point. This creates rounding error so that in some cases, the poll results do not sum to exactly 100 percentage points. We handle this symmetrically to the undecided problem. A summary of this "adding up" problem is provided in Web appendix Table 3.

[12]Though the adjusted results present a more "optimistic" assessment of poll accuracy, we present some results using the unadjusted data in the Web appendix.

[13]See Web appendix Table 5 and Table 6 for a complete analysis. Web appendix Table 6 presents results for the three elections separately and the patterns roughly apply. We also conducted several other analyses available in the web appendices. Of particular note is the fact that 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.)

Table 1: Election Results For Poll Sample

| | One Observation Per Poll | | |
|---|---|---|---|
| | All Polls | "Likely Voters" | < 2 Weeks before Election |
| | N = 1,857 | N = 1,554 | N = 704 |
| 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} |
| Republican victory | 38.40 | 38.93 | 40.77 |
| Democratic victory | 61.60 | 61.07 | 59.23 |
| 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 |

*Notes:* This table presents election results corresponding to each poll in our sample. In the top panel, each observation represents one poll; in the bottom panel, each observation represents one election. Prediction errors and shares are in units of percentage points. Standard deviations in braces.

Table 2: Pre-Election Polls

| | All Polls | "Likely Voters" | < 2 Weeks before Election |
|---|---|---|---|
| | N = 1,857 | N = 1,554 | N = 704 |
| Predicted Republican | 48.20 | 48.31 | 47.84 |
| | {6.31} | {6.00} | {5.48} |
| Predicted Democratic | 48.95 | 49.01 | 49.31 |
| | {5.91} | {5.61} | {5.22} |
| Republican error | 0.03 | 0.10 | -0.47 |
| | {3.36} | {3.21} | {2.49} |
| Democratic error | -1.04 | -0.96 | -0.43 |
| | {3.45} | {3.29} | {2.70} |
| Standardized Republican error | 0.02 | 0.07 | -0.22 |
| | (0.04) | (0.04) | (0.05) |
| Variance of standardized Republican error | 3.07 | 2.69 | 1.58 |
| | (0.16) | (0.14) | (0.10) |
| Standardized Democratic error | -0.55 | -0.51 | -0.23 |
| | (0.04) | (0.04) | (0.05) |
| Variance of standardized Democratic error | 3.20 | 2.84 | 1.89 |
| | (0.14) | (0.13) | (0.12) |
| 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 |

*Notes:* 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.

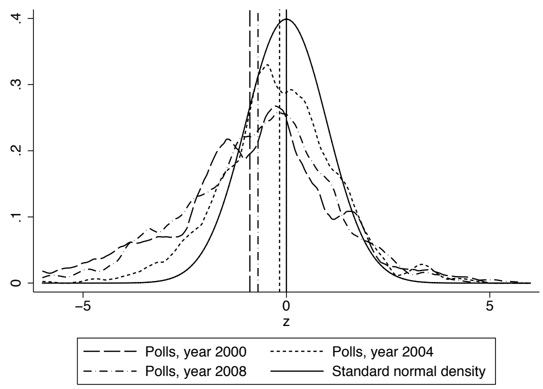


FIGURE 1: DENSITY ESTIMATES OF STANDARDIZED PREDICTION ERRORS OF DEMOCRATIC CANDIDATES

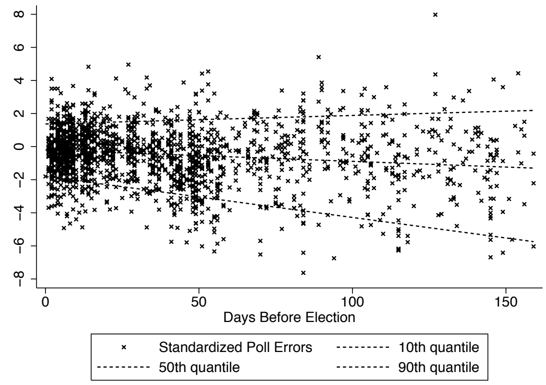


FIGURE 2: SCATTER PLOT OF DEMOCRATIC PREDICTION ERRORS FOR 2000, 2004, 2008 ELECTIONS

suggested by simple random (probability) sampling and are biased. To assess departures from what might be expected under random sampling (with certain and unchanging intentions, and certainty about participation), we construct "standardized" prediction errors by subtracting the "true" state election result $\mu$ from the corresponding poll prediction and dividing by $\sqrt{\frac{\mu(1-\mu)}{N}}$, where $N$ is the sample size of the poll. 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 2, 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.[14] The estimated densities are too disperse, are not centered at 0, and generally do not share the shape of the standard normal 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.

Table 2 also makes clear that the polls predict the winner more often than not, but the polls guess the winner incorrectly about 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. Table 1 documents the fraction of mispredicted races using the prior race outcome. 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

[14]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.

crude model generally outperforms the polls and is competitive with polls conducted two weeks before the election campaign. Focusing instead on the the fraction of "mispredicted victors" using one prediction per poll (top panel) demonstrates that the success of the crude benchmark forecast is only partly explained by the fact that more polls are conducted for "hard to predict" races.[15]

Although Table 2 demonstrates that 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.

Even at this hypothetical best–case there is significant over-dispersion. The 95 percent confidence interval for the constant term for 10th quantile regression is -1.97 to -1.63 and does not cover its theoretical value of -1.28. Likewise the 95 percent confidence interval for the constant term in the median regression is -0.32 to -0.08 and does not 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, -1.22 to 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

[15]See Web appendix Figure 7 for a visual description of where polls are most likely to be conducted.

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.

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 $\theta$. 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 mean from the (unreported) actual polling sample, where the weights reflect the strength of the prior. This suggests an OLS regression,

$$\text{poll}_i = \text{constant} + a * \text{Prior}_i + b * \text{Actual}_i \tag{1}$$

where the parameters $a$ and $b$ are respectively the weights that the typical pollster puts on his prior and the (unreported) 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 1 where we allow an additional parameter that allows the identical priors to vary from the previous election result by a constant $\gamma$ and assume that the prior can be summarized by a linear combination of previous election results $E$:

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

Table 3: The Relationship Between Forecast Errors and Prior Information

| | Dependent Variable = 2008 Polls | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| 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 | | | | |

where the constant term (up to scale) identifies a shift from the previous election result, $\phi_j$ is the weight on the previous election result, and $J$ is as large as two previous election results. These are reported in Table 3.[16] Our main result is that the coefficient on the actual outcome is always below 1 (the prediction from 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

[16]See Web appendix Table 8 for a complete analysis.

to perform "o.k."? The simplest answer is that the elections 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 X, Y or Z?" 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 poll was conducted by *Time–Sharing Experiments for Social Scientists* (TESS, 2005*b*) and *Knowledge Networks*.[17] Our purpose in designing the questions for the 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

[17]The data and documentation for our survey is available at `http://www.experimentcentral.org/data/data.php?pid=298`. Although described as an attempt to generate a "representative" sample (see TESS (2005*a*), for example) 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. Thus, we had little reasonable expectation of the poll as a reliable measure of electorate 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).

preferences as probabilities. To that end, there were two sets of questions. We call the first set of questions, administered to half of the sample, "the Probabilistic way"; the second set of questions, administered to the (demographically balanced) other half, we call "the Usual way." Both sets of questions are available in a Web appendix. The Usual style questions are intended to mimic how questions are actually asked in presidential horse race polls,[18] while the Probabilistic style questions allowed respondents to express their voting intentions on a scale of 0 to 100.

### *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 4 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.[19]

## 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* . . . "

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

[19]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. 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 find an alternative means to conduct it.

Table 4: Probabilistic vs. Usual Style Questions

| | Probabilistic Group (N = 1,190) | | |
|---|---|---|---|
| | Bush | Kerry | Other |
| Survey weighted | 46.97 | 50.10 | 2.93 |
| | (1.58) | (1.59) | (0.50) |
| Above, and | 46.89 | 50.45 | 2.67 |
| likely voter weights | (1.63) | (1.64) | (0.46) |
| Above, and | 46.66 | 50.69 | 2.66 |
| missing data weights | (1.65) | (1.65) | (0.45) |
| | Control Group (N = 1,181) | | |
| | Bush | Kerry | Other |
| Survey weighted | 48.12 | 49.20 | 2.68 |
| | (1.71) | (1.71) | (0.63) |
| Above, and | 48.08 | 49.25 | 2.67 |
| missing data weights | (1.71) | (1.71) | (0.62) |
| p-values | | | |
| Bush(T=1) = Bush(T=0) | | 0.5467 | |
| Kerry(T=1) = Kerry(T=0) | | 0.5457 | |
| Joint | | 0.8295 | |

*Notes:* See Web appendix Table 11 for details about weights and sample.

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.

## REFERENCES

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

**Butler, David and Kavanagh, Dennis**. (1992), The Waterloo of the Polls, *in* 'The British General Election of 1992', Pallgrave Macmillan, New York.

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

**DiNardo, John E.** (2009), 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. *Forthcoming.*

**DiNardo, John, McCrary, Justin and Sanbonmatsu, Lisa**. (2005), Randomizing Attrition and Simple Estimators for Treatment Effects and Bounds: An Empirical Example, Unpublished manuscript, University of Michigan, Ann Arbor, MI.

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

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

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

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

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

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

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

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

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

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

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

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

**Pew Research Center**. (1998), 'Possible Consequences of Non-Response for Pre-Election Surveys: Race and Reluctant Respondents'.

**Røhme, N.** (1992). '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* pp. 264–273.

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

**URL:** *http://poq.oxfordjournals.org/cgi/content/abstract/71/4/539*

**TESS**. (2005*a*). 'Get to Know Tess'. Accessed November 1, 2008 at `http://www.experimentcentral.org/tess/`.

**URL:** *http://www.experimentcentral.org/tess/*

**TESS**. (2005*b*). 'Time-sharing Experiments for the Social Sciences'. Data collected by Time-sharing Experiments for the Social Sciences, NSF Grant 0094964, Diana C. Mutz and Arthur Lupia, Principal Investigators.".

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

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