Accuracy of Vote Expectation Surveys in Forecasting Elections

Andreas Graefe
Department of Communication Science and Media Research
LMU Munich, Germany
a.graefe@lmu.de

March 30th, 2013

Word count: 5,914

Abstract. Simple surveys that ask people who they expect to win are among the most accurate methods for forecasting U.S. presidential elections. Vote expectation surveys failed to pick the winner in only 18 (8%) of 214 surveys conducted from 1932 to 2012. Across the seven elections from 1988 to 2012, vote expectation surveys outperformed four established methods (trial-heat polls, prediction markets, econometric models, and experts' judgment) in predicting election winners and vote shares. Vote expectation surveys are accurate, inexpensive, and easy to conduct. They should be more strongly utilized by election observers as well as researchers.

Keywords: combining forecasts, Iowa Electronic Markets, election forecasting, forecast accuracy, econometric models; FiveThirtyEight

Acknowledgments: J. Scott Armstrong, Alfred Cuzán, and Christopher Wlezien provided helpful suggestions.

"Who do you think will win the U.S. presidential election?"

Pollsters have regularly asked variations of this question since the 1940s. Yet, the results of such *vote expectation surveys* have been virtually ignored by election observers and are rarely reported in the media. Instead, the trial-heat question, which asks respondents for whom they *intend* to vote, dominates the horse race coverage of elections (Patterson 2005). Likewise, researchers have widely overlooked vote expectation surveys as a method for election forecasting. Little is known about their accuracy compared to established methods (Hillygus 2011). The present study provides empirical evidence on the accuracy of vote expectation surveys for forecasting U.S. presidential elections by comparing their results to predictions from four established methods. The results suggest that vote expectation surveys are among the most accurate methods for forecasting U.S. presidential elections available to date.

Established methods for forecasting U.S. presidential elections

As long as there have been elections, people have tried to predict their results. Nowadays, the most common methods for election forecasting are trial-heat polls, prediction markets, experts' judgment, and quantitative models (Graefe et al. 2013).

Experts' judgment

Judgment of political insiders and experienced election observers were used to forecast elections long before the emergence of scientific polling (Kernell 2000); and they still are. The common assumption is that expert analysts have experience in reading and interpreting polls, assessing their significance during campaigns, and estimating the effects of recent or expected events on the aggregate vote. Given their omnipresence, we know surprisingly little about the relative accuracy of experts' judgment for election forecasting, although some evidence exists that experts are more accurate than non-experts (Sjöberg 2009, Lemert 1986).

Polls

Trial-heat polls, hereafter simply referred to as polls, ask respondents for whom they *intend* to vote if the election were held today. That is, polls do not provide predictions; they provide snapshots of public opinion at a certain point in time. However, this is not how the media commonly treat polls. Polling results are routinely interpreted as forecasts of what will happen on Election Day (Hillygus 2011). This can result in poor predictions, in particular if the election is still far away, because public opinion can be difficult to measure and fragile over the

course of a campaign. However, researchers found ways to deal with these problems and to increase the accuracy of poll-based predictions.

Combining polls to reduce measurement error

There is often high variance in the results of polls by different survey organizations, even if these polls were conducted at around the same time. Such variance can be caused by sampling problems, non-responses, inaccurate measurement, and faulty processing (Erikson and Wlezien 1999). Therefore, one should not look at the results from single polls. Rather, one should combine polls that were conducted at around the same time. The reason is that the systematic (and random) errors that are associated with individual forecasts tend to cancel out in the aggregate. The power of combining for generating accurate predictions is one of the major findings from forecasting research conducted since the 1970s, which impacted many fields such as weather forecasting and economic forecasting (Graefe et al. 2013).

Projecting poll results to Election Day

Polls conducted by the same survey organization, and by the polling industry as a whole, can fluctuate widely across the course of the campaign. The reason is that a large share of the electorate has not spent much time thinking about the election (e.g., the important issues and the candidates' positions) if the election is still far away. As a result, people's response behavior in early polls is strongly influenced by campaign events such as conventions (Campbell, Cherry, and Wink 1992) and debates (Benoit, Hansen, and Verser 2003).

The effects of such events on the outcome of high-visibility elections such as U.S. presidential elections are limited, however. As the election nears, people are less influenced by the latest campaign events and have formed stable vote intentions based on a combination of information they have learnt during the campaign, such as the state of the economy, and their basic predispositions, such as ideology and party identification (Gelman and King 1993). Therefore, polls provide accurate forecasts not until shortly before Election Day.

However, researchers found ways for how to harness early polls for forecasting by calculating *poll projections*, as they are termed hereafter. Poll projections take into account the historical record of polls for forecasting. For example, assume that the incumbent leads the polls by fifteen points in August. In analyzing historical polls conducted around the same time along with the respective election outcomes, one can derive a formula for translating August polling figures into an estimate of the incumbent's final vote share in November. This is commonly done

by regressing the incumbent's share of the vote on his polling results during certain time periods before the election. Prior research found that such *poll projections* are much more accurate than treating raw polls as forecasts (e.g., Erikson and Wlezien 2008).

Calculating combined poll projections

Graefe et al. (2013) combine both strategies (i.e., aggregating polls and calculating poll projections) to generate poll-based forecasts. The authors first calculate rolling averages of all polls that were published in a one-week period. Then, they use these results to calculate poll projections. This procedure resulted in large gains in accuracy. Across the last 100 days prior to each of the six elections from 1992 to 2012, such *combined poll projections* reduced the error of a randomly picked poll that was published the same day by 39%.

Prediction markets.

Prediction (or betting) markets allow people to bet on the election outcome. The resulting betting odds can then be interpreted as forecasts of the election results. Such markets were already highly popular in the late 19th and early 20th century, when newspapers such as the *New York Times* regularly reported the latest predictions. However, around the time of World War II, prediction markets began to disappear, likely due to a combination of factors such as the rise of the polling industry in the 1930s, the introduction of laws to eliminate organized election betting, and the emergence of alternative betting opportunities such as horse-racing (Rhode and Strumpf 2004). It took almost half a century, and the rise of the Internet, for the method to be rediscovered. In 1988, researchers at the *University of Iowa* launched the online *Iowa Electronic Markets* (IEM) to predict the U.S. presidential elections held in the same year. Since then, interest in prediction markets resurged (Arrow et al. 2008).

Studies of prediction market accuracy for election forecasting commonly compare the daily market forecasts to results from polls published the same day. These studies generally find that prediction markets yield more accurate forecasts than single polls. For example, Berg, Nelson, and Rietz (2008) compare the accuracy of IEM forecasts to results from nearly one thousand polls across the five U.S. presidential elections from 1988 to 2004. The IEM forecasts were more accurate than single polls 74% of the time. However, as outlined above, single polls provide rather poor predictions and thus only serve as a weak benchmark. Erikson and Wlezien (2008) account for this problem and compare the IEM forecasts to poll projections during the same time period analyzed by Berg, Nelson, and Rietz (2008). The authors find that poll

projections were more accurate than the IEM. I extended these analyses and compared the IEM forecasts to combined poll projections, using the approach suggested by Graefe et al. (2013). Across the last 100 days prior to each of the seven elections from 1988 to 2012, the IEM forecasts yielded an average error of 1.7 percentage points, which is 17% below the corresponding error of combined poll projections of 2.0 percentage points (cf. Table 2).

Quantitative models.

A common theory of electoral behavior is that elections are referenda on the incumbent's performance. That is, voters are expected to reward the government for good performance and punish the incumbent party otherwise. Since the late 1970s, economists and political scientists tested this theory by developing quantitative models to predict election results. That is, an important advantage of these models is that they do not only predict but also explain election outcomes. Most models are based on multiple regression analysis of two to five predictor variables, which typically capture economic conditions, the incumbent's popularity, and how long the President or his party have controlled the White House. The development and testing of these models has become a well-established sub-discipline of political science and the models' forecasts are regularly published about two months prior to Election Day in scientific journals.

These models predict the correct election winner most of the time. Across the six elections from 1992 to 2012, 34 of 39 forecasts of seven well-known models correctly predicted the winner. However, the models' performance in predicting vote shares is mixed. Their mean absolute error (MAE) was three percentage points, and ranged from zero to ten points (cf. Appendix I).

Vote expectation surveys

But why not simply ask voters whom they *expect* to win and then use the aggregate result as forecast? Such *vote expectation surveys* were conducted even before the emergence of trial-heat polls in the late 1930s. And, researchers have long pointed to the surveys' accuracy. Hayes (1936) reports results from a 1932 survey of 8,419 men and women, of which the majority correctly predicted Roosevelt to defeat Hoover in that year's presidential election. Lewis-Beck and Skalaban (1989) show that the vote expectation question from the pre-election ANES surveys correctly predicted the election winner in six of the eight U.S. presidential elections from 1956 to 1984. When including the 1952 election and the elections from 1988 to 2008, the ANES

vote expectation question correctly predicted the election winner in twelve of the fifteen elections (Rothschild and Wolfers 2011). Given their accuracy, it is surprising that these surveys have been widely ignored in election forecasting. Although pollsters at times include the vote expectation question in their surveys, the responses are rarely reported in the media. In addition, few researchers studied the accuracy of vote expectation surveys compared to other methods.

v. polls

Rothschild and Wolfers (2011) study the relative accuracy of the vote expectation question and the trial-heat (i.e., vote intention) question when both are asked in the same survey. Based on an analysis of ANES data from the fifteen U.S. presidential elections from 1952 to 2008, the authors find that expectations were more accurate than intentions when predicting election winners, vote shares, and probabilities of victory. Given these results, it is surprising that trial-heat polls dominate media coverage and virtually no attention is paid to expectation surveys. In particular, since the reason for the superior performance of expectations seems obvious: expectations capture more information than intentions. Expectations incorporate information about one's own and other people's vote intentions, as well as information from other sources. For example, people might form expectations of the election outcome from following general media coverage of the campaign, reading the latest polls, and listening to expert predictions. Therefore, vote expectations *should* provide more accurate predictions than vote intentions.[1] In fact, as shown by Rothschild and Wolfers (2011), each person's expectation is equivalent to a multi-person intention poll. This is a major advantage of expectation surveys since sample size and composition are less critical. Rothschild and Wolfers (2011) demonstrate this by calculating forecasts based on expectations from biased subsamples (i.e., only Democrats and only Republicans). In both cases, the expectation-based forecasts provided more accurate forecasts then the complete sample of vote intentions.

These results leave little doubt about the superior performance of the vote expectation question compared to the trial-heat question. However, as outlined earlier, single polls are a poor benchmark of forecast accuracy, in particular early in the campaign. Further research is

[1] The only exception would be a situation in which each respondent is subject to wishful thinking. Then, vote expectations should perform at least as well as vote intentions, since each vote expectation is identical to one's vote intention.

necessary on the relative accuracy of vote expectations and sophisticated poll-based forecasts such as combined poll projections.

v. prediction markets and experts

In eliciting expectations, forecasts from experts and prediction markets are closely related to vote expectation surveys. One major difference between the three approaches lies in the composition of the sample. While vote expectation surveys sample respondents randomly, the other two approaches rely on selected (as in the case of expert surveys) or self-selected (as in the case of prediction markets) experts.[2] Given the similarities of these methods, it is surprising that few researchers studied their relative accuracy. I found only two studies that provide empirical evidence, each of which analyzes only a single election. Miller et al. (2012) find that an online expectation survey of nearly 20,000 respondents was more accurate than the *Intrade* prediction market when forecasting winning probabilities for the 2008 U.S. presidential elections. Sjöberg (2009) finds that the average forecasts of non-experts were more accurate than the average forecasts of experts when forecasting the 2006 Swedish parliament elections.

These results may surprise. How is it possible that the combined predictions of regular citizens perform as well as – or even better than – combined predictions of (self-)selected experts? In particular, since at the individual level, Sjöberg (2009) finds what one would expect: experts were consistently more accurate than the less interested, less informed, and less educated non-experts (members of the public). Sjöberg (2009) explains this result with the heterogeneity of groups. The non-experts varied in demographics and their party preferences were in line with the general public. In comparison, experts (three groups consisting of political scientists, journalists, and editors) were mostly male and well educated. In addition, experts showed a particularly low preference for the Conservative party (the second strongest party in the polls) and high preferences for parties that were less popular among the general public (e.g., the Liberals and the Center Party). As a result, the less diverse experts group was likely biased in the same direction. In such a situation, combining is of limited value, since the individual biases do not cancel out in the aggregate (Graefe et al. 2013).

---

[2] In addition, prediction markets differ from regular surveys in how they aggregate information. Prediction market participants buy and sell shares, whose prices reflect the combined expectations of all participants. Since participants can win and lose money depending on their performance, they have an incentive to be accurate. Since participants should only become active if they think they know better than the market as a whole, they are often referred to as self-selected "experts".

The same could be true for prediction markets. A study of the 1988 IEM finds that market participants exhibited substantial judgment biases and did not form a heterogeneous group. Participants were predominantly white, male, well educated, and belonged to the middle and upper income categories. In addition, participants tended to be more Republican and less independent in their partisan leanings, and were more politically active than the general public (Forsythe et al. 1992). Given these findings, there is reason to believe that vote expectation surveys can provide forecasts that are competitive with expert surveys and prediction markets.

v. quantitative models

I was unable to find prior research on the relative accuracy of vote expectations and quantitative models. The advantage of models is that they follow a structured approach to forecasting and include much information about historical elections, such as the influence of the state of the economy, the popularity of the incumbent, and the time the incumbent was in the White House. While the accuracy of single models can vary widely across elections, one usually gets accurate forecasts when combining forecasts from different models (Graefe et al. 2013). However, a disadvantage of quantitative models is their limited ability to incorporate information about the specific context of a particular election such as an economic crisis, threat of terrorism, or some scandal.

Accuracy of vote expectation surveys in forecasting U.S. presidential elections

The following analysis provides empirical evidence on the relative accuracy of vote expectation surveys for forecasting U.S. presidential elections.

Method

All data and calculations are available upon request.

Time horizon and error measures

The methods' forecast accuracy is analyzed across the last 100 days prior to Election Day. The hit rate and the absolute error were used as measures of accuracy. The hit rate is the percentage of forecasts that correctly predict the winner. For methods that provide forecasts of two-party popular vote shares, the candidate with a vote share of more than 50% is predicted to win the election. In the case that each candidate is predicted to gain 50% of the popular vote, a tie is recorded. Ties score as half of a correct prediction. The absolute error is the absolute

deviation of the predicted and the actual two-party popular vote for the incumbent party's candidate.

Data and forecast calculations

To allow for fair comparisons, all forecasts are calculated as if they were made *ex ante*. That is, the forecasts only use data that would have been available at the time of the election. In addition, comparisons include only forecasts made around the same time.

Vote expectation surveys

A total of 215 vote expectation surveys were collected across twenty elections from 1932 to 2012. For example: *„Regardless of whom you support, and trying to be objective as possible, who do you think will win the presidential election in November (2008)--Barack Obama or John McCain?"* (Gallup Poll, October 23-26, 2008). All polls were conducted within 150 days prior to Election Day. The data set includes fifteen ANES surveys, one for each of the elections from 1952 to 2008, and the 1932 survey reported by Hayes (1936). The remaining 199 surveys were derived from the *iPoll Databank* of the *Roper Center for Public Opinion Research*. For the 1936 election, no survey was found.

For each survey that was published during the past 100 days prior to Election Day, the two-party percentage of respondents that expected the candidate of the incumbent party to win was recorded. If more than one survey was published on the same day, the results of all surveys from that date were averaged. On days without any surveys, the most recent survey from preceding days was used.

Vote expectation surveys provide direct forecasts of who will win; the candidate that the majority of respondents expect to win is predicted as the election winner. However, the results of vote expectation surveys cannot be directly interpreted as vote share forecasts. For example, a survey that reveals that 60% of respondents expect the incumbent candidate to win does not mean that the incumbent can be expected to gain 60% of the vote. In order to translate vote expectation survey results into vote share forecasts it is necessary to use data from historical surveys. That is, one estimates how a hypothetical incumbent lead of 60-40 translated to the incumbent's final vote share in past elections. A simple approach for estimating this relationship is linear regression analysis. Thereby, the incumbent party's actual two-party popular vote share is regressed on the results from the vote expectation surveys. The generic equation for the vote expectation prediction can be read as

$$V_y = \alpha + \beta E_{yt} + \varepsilon_y,$$

where V_y is the actual two-party popular vote share of the candidate of the incumbent party in year y, E_{yt} is the corresponding vote expectation on day t, and ε_y is the error term. Ex ante vote share forecasts were calculated by successive updating. That is, only data that would have been available for the election under observation were used. For example, to estimate the 1988 equation, only the 40 polls available from 1932 to 1984 were used, while the 2004 equation is based on the 124 polls available through 2000.[3]

Polls

Polls that were conducted within 100 days prior to each of the sixteen elections from 1952 to 2012 were obtained from *iPoll*. For each poll, the two-party percentage of respondents that intended to vote for the candidate of the incumbent party was recorded. If more than one poll ended on the same day, the results of all polls ending that date were averaged. On days without any polls ending, the most recent poll from preceding days was used.

Three different poll-based forecasts were used as benchmarks: (1) single polls, (2) combined polls, and (3) combined poll projections. The *single polls* benchmark simply interprets the result of a single poll published on a particular day as forecast of the election outcome. The *combined polls benchmark* calculates rolling averages of all polls released over a 7-day period. The third benchmark, combined poll projections, was adopted from Graefe et al. (2013). That is, for each of the 100 days prior to an election, starting with 1952, the incumbent's actual two-party share of the popular vote was regressed on the combined polls value for that day. This process produced 100 vote equations (and thus poll projections) per election year. Again, successive updating was used to calculate *ex ante* poll projections.

Prediction markets

Daily prediction market data from the IEM vote-share markets were obtained from the IEM website. On these markets, people buy and sell future contracts according to their own expectations of the candidates' final vote shares. The market price represents the combined forecast of all market participants. To put the prediction market forecasts on equal footing with vote expectation surveys and polls, two-party forecasts were calculated by ignoring any third-party options. The last traded prices per day were used as the market forecasts.

---

[3] Data from the ANES data of a particular election are not available until months after the election. Therefore, their results were only used when calculating forecasts of succeeding elections.

Experts

One expert survey was available for each of the elections in 1992 and 2000, four surveys were available for each of the elections in 2004 and 2008, and five surveys were available for the 2012 election. The average number of experts per survey ranged from ten to fifteen. In each survey, experts were asked to provide forecasts of the popular vote.[4] Individual and average expert forecasts are compared to the average of all vote expectation surveys conducted during the seven days prior to the publication of the expert survey.

Quantitative models

The present study uses forecasts from seven established quantitative models.[5] These models, along with their forecasts, were published in *Political Methodologist* 5(2), *American Politics Research* 24(4) and *PS: Political Science and Politics* 34(1), 37(4), 41(4), and 45(4). Most model forecasts are published about two to three months prior to the election. Therefore, the model forecasts are compared to the average of all vote expectation surveys that were published from 90 to 60 days prior to Election Day.

Results

Table 1 summarizes the performance of vote expectation surveys for forecasting election winners. The vote expectation survey result correctly predicted the winner in 191 (89%) of 215 surveys across twenty elections from 1932 to 2012. Only 18 (8%) of the surveys predicted the wrong winner, half of which were conducted during the very close 2000 election. The remaining six surveys predicted a tie.

v. polls and prediction markets

The accuracy of daily forecasts from vote expectations surveys, polls, and the IEM was analyzed across the full 100-day period prior to Election Day. Table 2 shows the methods' hit rate and mean absolute error (MAE) across and for each of the seven elections from 1988 to 2012. The vote expectation surveys were most accurate. If one had simply relied on the most recent vote expectation survey available on a particular day, one would have achieved a hit rate

---

[4] The 1992 survey was published in *The Washington Post. Pundits' brew: How it looks; Who'll win? Our fearless oracles speak, November 1, 1992, p. C1, by David S. Broder*. The 2000 survey was published in *The Hotline. Predictions: Potpourri of picks from pundits to professors, November 6, 2000*. The surveys conducted for the three elections from 2004 to 2012 were derived from Graefe, Armstrong et al. (2013).

[5] These models were developed by Abramowitz, Campbell, Erikson & Wlezien, Holbook, Lewis-Beck & Tien, Lockerbie, and Norpoth.

of 92%. That is, one would have predicted the correct winner more than nine out of ten times. In comparison, if one had relied on the most recent single poll at the same day, one would have predicted the correct winner only 79% of the time. As expected, combining polls (86% correct predictions) and calculating combined poll projections (88%) increased the accuracy of single polls. Surprisingly, with a hit rate of 79%, the IEM vote-share markets were no more accurate than single polls.[6]

Vote expectation surveys also provided accurate vote-share forecasts. On average, the vote expectation surveys missed the election results by 1.6 percentage points and were thus more accurate than each of its competitors. While gains in accuracy compared to the IEM vote-share markets and combined poll projections were small, vote expectation surveys clearly outperformed combined and individual polls. Compared to single polls, vote expectation surveys reduced error by 50%.

v. experts

Table 3 compares the accuracy of vote expectation surveys and experts' judgment. In every single comparison, the vote expectation surveys predicted the correct winner. In comparison, the combined experts predicted the correct winner 70% of the time and were thus slightly more accurate than the typical expert (66% correct).[7] In terms of MAE, vote expectation surveys (1.2 percentage points) were more accurate than the typical (1.8) and the combined experts (1.4).

v. quantitative models

Table 4 shows the hit rates and MAE of vote expectation surveys and the typical and average forecast of seven quantitative models for the six elections from 1992 to 2012. The vote expectation surveys as well as the average forecast of all available models correctly predicted the winner in each election. The track record of the individual models is not perfect with two models missing the winner in 2012, and one each in 1992, 2004, and 2008. This results in an average hit rate of 86% for the typical model. That is, if one had randomly picked a model in each election,

---

[6] I also analyzed the accuracy of the IEM winner-takes-all markets, which were first launched in 1992. The winner-take-all markets were specifically designed to predict popular vote winners and thus provide the better benchmark for this type of task. These markets achieved a hit rate of 88% across the six elections from 1992 to 2012. The corresponding hit rate of vote expectation surveys for the same time period was 93%.

[7] The performance of the typical expert is the performance that one would achieve if one would randomly pick an expert.

one would have correctly picked the winner 86% of the time. In terms of MAE, the vote expectation surveys yielded an error of 1.5 percentage points and thus were more accurate than typical (2.9) and combined (2.3) model forecast.

Discussion

There has been much progress in our ability to forecast elections over the last three decades. Aggregating polls and projecting their results to Election Day yielded substantial improvements in accuracy compared to single polls. Researchers have developed econometric models that can quite accurately predict election outcomes from structural information that is available months before Election Day. Finally, prediction markets reappeared as a powerful tool.

One simple method, which has existed at least since the advent of scientific polling, has been largely overlooked in this development: surveying people on who they expect to win. Given their accuracy, the disregard of vote expectation surveys is puzzling. Across the past seven U.S. presidential elections, vote expectation surveys provided more accurate forecasts of election winners and vote shares than any other established method.[8] Gains in accuracy were particularly large compared to single polls. The error of vote-share forecasts derived from vote expectation surveys was 50% lower than the corresponding error of a single poll. Nevertheless, the widespread belief that trial-heat polls provide accurate forecasts remains.

Relative performance of polls and vote expectation surveys in the 2012 election

The *National Council on Public Polls* (NCPP) analyzed 25 national polls that were conducted within the final week of the 2012 campaign and concluded that, with an average error of 1.46 percentage points, the polls "came close to the election outcome".[9] However, the NCPP failed to compare the polls' performance to an adequate benchmark. I calculated the corresponding error of vote expectation surveys and the IEM for the same time period. Both benchmarks provided more accurate predictions than the NCPP poll sample. The vote

[8] The performance of vote expectation surveys in predicting vote shares is particularly noteworthy, since the analysis is constrained by limited historical data. For early elections, the regression analysis that was used to estimate the vote equation is based on only few polls. This makes it impossible to account for the uncertainty that occurs over the course of a campaign. As with projecting the results of trial-heat polls to Election Day, one might be able to further increase the accuracy of vote expectation surveys by taking into account that the accuracy of people's expectations decreases for long time horizons (Lewis-Beck and Skalaban 1989, Miller, Wang et al. 2012).

[9] The NCPP's "Analysis of Final 2012 Pre-Election Polls" is available at www.ncpp.org.

expectation surveys yielded an error of 0.71 and thus reduced the error of polls by more than 50%. With an error of 0.99 percentage points, the IEM were 30% more accurate than the polls.

Figure 1 extends this analysis and shows the relative accuracy of vote share predictions of 110 polls and 20 vote expectation surveys that were published during the last 100 days prior to the 2012 election. The vertical axis shows Obama's predicted lead in the two-party popular vote and the dotted grey line depicts the final election outcome; Obama won the election with a four-point advantage over Romney. Single polls varied wildly and predicted anything from a three-point lead for Romney to an eighteen-point lead for Obama (standard deviation: 4.3). In comparison, vote expectation surveys were much more stable, less extreme, and closer to the election result throughout the 100-day time horizon. The vote expectation survey forecasts ranged from a one-point to a six-point lead for Obama (standard deviation: 1.3).

Relative performance of FiveThirtyEight.com and vote expectation surveys in the 2012 election

One cannot discuss forecasts of the 2012 election without mentioning Nate Silver's *FiveThirtyEight.com*, a polling aggregation website that was launched in 2008 and has become part of the *The New York Times* online in 2010. Silver uses sophisticated statistical analyses to analyze the type and extent of biases of single pollsters, and to demonstrate the value of polling aggregation for forecasting. His forecast model aggregates information from state-level polls by accounting for the relative performance of different pollsters and considering relationships between states. In addition, the model incorporates an index of economic indicators, whose weight decreases as the election nears. Simply put, Silver's model is an enhanced and much more sophisticated version of the traditional quantitative models, which also combine polls and economic fundamentals.

FiveThirtyEight has become extremely popular. In the week prior to the 2012 election, almost three out of four politics visits at the *New York Times* website included a stop at FiveThirtyEight. The day before the election, one in five nytimes.com visitors looked at Silver's site (Tracy 2012). FiveThirtyEight had become a synonym for election forecasting, which becomes evident when looking at the volumes of Google searches for variants of "Fivethirtyeight" and "election forecast" (cf., Appendix II).

I compared the accuracy of vote expectation surveys to Silver's popular vote forecast. Figure 3 reports the error of both approaches for the last 100 days prior to the election. Any point

on the lines in the chart shows the average error for the remaining days in the forecast horizon. For example, if one had relied on the FiveThirtyEight forecast on each of the 100 days prior to Election Day (i.e., starting from July 29th), one would have achieved a MAE of 0.65 percentage points. If one had relied on FiveThirtyEight on each day from October 11th, one would have achieved a MAE of 1.2 percentage points, and so on. The corresponding values for the MAE of the vote expectation surveys are 0.40 (from July 29th) and 0.76 (from October 11th). As shown in Figure 3, at any point in the campaign, one would have fared better by relying on the most recent vote expectation survey than on the forecasts at FiveThirtyEight.com. Across the full 100-day forecast horizon, vote expectation surveys reduced the error of FiveThirtyEight on average by 38%.[10]

Barriers to the adoption of vote expectation surveys

Vote expectation surveys provide highly accurate forecasts, are easy to conduct, and the results are easy to understand. So why have they been virtually ignored by election observers? I see at least three reasons: people do not know about the benefits of combining, are not interested in accuracy, and have no faith in simple methods.

Lack of knowledge about the benefits of combining

When asked to reveal expectations, people are influenced by their preferences. In the case of elections, this means that people tend to predict their preferred candidate to win. This bias, which is known as *wishful thinking,* is long known to be common in the context of elections. Hayes (1936) finds that 72% of Hoover supporters predicted Hoover to win, whereas 91% of Roosevelt supporters predicted Roosevelt to win the 1932 election; Roosevelt won by a landslide. In their seminal study of voting behavior, Lazarsfeld, Berelson, and Gaudet (1948) report a strong relationship between people's vote intentions and their expectations of who will win. Granberg and Brent (1983) study wishful thinking across the eight U.S. presidential elections from 1952 to 1980. Their analysis revealed a strong positive relationship between expectations and candidate preferences (r=.8). Since these early studies, evidence has accumulated. Wishful thinking occurs in all types of elections, from local referenda to national

[10] Silver's model also provides forecasts at the state level and correctly predicted the winner in all 50 states. The present study did not collect state-level vote expectation surveys and thus cannot compare the relative performance of both approaches for this task.

elections, and across various countries. See Miller et al. (2012) for a useful overview of recent research.

Wishful thinking limits people's ability to provide accurate predictions. However, the mere existence of wishful thinking at the individual level should not prevent one from using aggregated expectations. The reason is that wishful thinking can be expected to cancel out in the aggregate. Although the benefits of combining forecasts have long been demonstrated in many fields (Graefe et al. 2013), they are not intuitively obvious. In a series of experiments, a majority of highly qualified participants thought that an average of estimates would yield only average accuracy (Larrick and Soll 2006). But there is some good news. Combining has impacted how people nowadays consume polls and online polling aggregators such as realclearpolitics.com, pollster.com, and FiveThirtyEight.com, in particular, have become increasingly popular.

Lack of incentives for accuracy

Journalists and political commentators need to meet the demands of the news cycle and constantly look for interesting stories and analyses. In this endeavor, they often select newsworthiness over accuracy and relevance. In particular, journalists increasingly generate news by focusing on who is ahead in the polls or linking the latest poll results to campaign events (Patterson 2005, Rosenstiel 2005). As shown in Figure 1, polls from different survey organizations often vary wildly, even if they are conducted at around the same time. In such a situation, journalists can cherry-pick on polls that support their story. Thereby, they might be little concerned about the accuracy of a poll, in particular if the election is still some time away.

In contrast, vote expectation surveys are much more robust and less extreme. Since their forecast rarely changes, vote expectation surveys are less suited for *generating* news. It would be desirable if journalists would focus more on vote expectation surveys. First, voters would be much better informed about who is really ahead. Second, the stability of vote expectation surveys makes it difficult to frame the election as a horse race. Rather, journalists could concentrate on providing explanations for the relative performance of candidates and their proposed policies.

Lack of complexity

Occam's Razor advises researchers to prefer simple models unless simplicity is offset by more explanatory power. Since Occam, many famous researchers advocated the use of simple models. Albert Einstein is reputed to have said that "everything should be made as simple as possible but not simpler". Zellner (2004), who coined the phrase "keep it sophisticatedly

simple", named several Nobel laureates as proponents of simplicity. Vote expectation surveys adhere to Occam's Razor; they are easy to conduct, the results are easy to understand, and they provide accurate forecasts.

Unfortunately, simple models often face resistance, because people tend to *wrongly* believe that complex solutions are necessary to solve complex problems. Hogarth (2012) reported results from four studies, which showed that simple models often perform better than more complex ones. In each case, however, people resisted the findings regarding the performance of simple models. The same appears to be true for election forecasting. People are impressed by sophistication and complexity (e.g., FiveThirtyEight.com) and overlook obvious approaches, such as simply asking people whom they expect to win.

Directions for future research

The present study demonstrates the high accuracy of vote expectation surveys for forecasting U.S. presidential elections. For this type of task, respondents are asked to make a binary forecast between two candidates. In comparison, it should be more difficult to predict the outcome of elections in which numerous parties and candidates compete for voters, and in which citizens' might engage in tactical voting. Future research should investigate whether the results from the present study generalize to other electoral systems and other countries.

Conclusion

The general election observer is probably most interested in who will win. When it comes to U.S. presidential elections, vote expectation surveys are likely to provide the best answer to this question. In addition, the results of such surveys can be translated into highly accurate vote share forecasts.[11]

Vote expectation surveys are inexpensive and easy to conduct, and the results are easy to understand. They should be more strongly utilized by election observers as well as researchers.

References

Arrow, Kenneth J., Robert Forsythe, Michael Gorham, Robert Hahn, Robin Hanson, John O. Ledyard, Saul Levmore, Robert Litan, Paul Milgrom, Forrest D. Nelson, George R. Neumann, Marco Ottaviani, Thomas

[11] For predicting U.S. presidential elections, respondents are asked to make a binary forecast between two candidates. In comparison, it should be more difficult to predict the outcome of elections in which numerous parties and candidates compete for voters, and in which citizens' might engage in tactical voting. Future research should investigate whether the results from the present study generalize to other electoral systems and other countries.

C. Schelling, Robert J. Shiller, Vernon L. Smith, Eric Snowberg, Cass R. Sunstein, Philip C. Tetlock, Hal R. Varian, Justin Wolfers, and Eric Zitzewitz. 2008. "The promise of prediction markets." *Science*, 320, 877-878.

Benoit, William L., Glenn J. Hansen, and Rebecca M. Verser. 2003. "A meta-analysis of the effects of viewing US presidential debates." *Communication Monographs*, 70, 335-350.

Berg, Joyce E., Forrest D. Nelson, and Thomas A. Rietz. 2008. "Prediction market accuracy in the long run." *International Journal of Forecasting*, 24, 285-300.

Campbell, James E., Lynna L. Cherry, and Kenneth A. Wink. 1992. "The convention bump." *American Politics Research*, 20, 287-307.

Erikson, Robert S., and Christopher Wlezien. 1999. "Presidential polls as a time series: the case of 1996." *Public Opinion Quarterly*, 63, 163-177.

Erikson, Robert S., and Christopher Wlezien. 2008. "Are political markets really superior to polls as election predictors?" *Public Opinion Quarterly*, 72, 190-215.

Forsythe, Robert, Forrest Nelson, George R. Neumann, and Jack Wright. 1992. "Anatomy of an experimental political stock market." *The American Economic Review*, 82, 1142-1161.

Gelman, Andrew, and Gary King. 1993. "Why are American presidential election campaign polls so variable when votes are so predictable?" *British Journal of Political Science*, 23, 409-451.

Graefe, Andreas, J. Scott Armstrong, Randall J. Jones Jr., and Alfred G. Cuzán. 2013. "Combining forecasts: An application to elections." *Forthcoming in the International Journal of Forecasting*, ssrn.com/abstract=1902850.

Granberg, Donald, and Edward Brent. 1983. "When prophecy bends: The preference-expectation link in US presidential elections, 1952–1980." *Journal of Personality and Social Psychology*, 45, 477-491.

Hayes, Samuel P. Jr. 1936. "The predictive ability of voters." *Journal of Social Psychology*, 7,

Hillygus, D. Sunshine. 2011. "The evolution of election polling in the United States." *Public Opinion Quarterly*, 75, 962-981.

Hogarth, Robin M. 2012. "When simple is hard to accept." In *Ecological Rationality: Intelligence in the World*, edited by Peter M. Todd, Gerd Gigerenzer and ABC Research Group, 61-79. Oxford: Oxford University Press.

Kernell, Samuel. 2000. "Life before polls: Ohio politicians predict the 1828 presidential vote." *PS: Political Science and Politics*, 33, 569-574.

Larrick, Richard P., and Jack B. Soll. 2006. "Intuitions about combining opinions: Misappreciation of the averaging principle." *Management Science*, 52, 111-127.

Lazarsfeld, Paul F., Bernard Berelson, and Hazel Gaudet. 1948. *The People's Choice. How the Voter Makes up His Mind in a Presidential Campaign*. 2nd ed. New York: Columbia University Press.

Lemert, James B. 1986. "Picking the winners: Politician vs. voter predictions of two controversial ballot measures." *Public Opinion Quarterly*, 50, 208-221.

Lewis-Beck, Michael S., and Andrew Skalaban. 1989. "Citizen forecasting: can voters see into the future?" *British Journal of Political Science*, 19, 146-153.

Miller, Michael K, Guanchun Wang, Sanjeev R Kulkarni, H Vincent Poor, and Daniel N Osherson. 2012. "Citizen Forecasts of the 2008 US Presidential Election." *Politics & Policy*, 40, 1019-1052.

Patterson, Thomas E. 2005. "Of polls, mountains." *Public Opinion Quarterly*, 69, 716-724.

Rhode, Paul W., and Koleman S. Strumpf. 2004. "Historical presidential betting markets." *Journal of Economic Perspectives*, 18, 127-141.

Rosenstiel, Tom. 2005. "Political polling and the new media culture: A case of more being less." *Public Opinion Quarterly*, 69, 698-715.

Rothschild, David, and Justin Wolfers. 2011. "Forecasting elections: voter intentions versus expectations." *Working paper.*,

Sjöberg, Lennart. 2009. "Are all crowds equally wise? A comparison of political election forecasts by experts and the public." *Journal of Forecasting*, 28, 1-18.

Tracy, Marc. 2012. "Nate Silver is a one-man traffic machine for the Times." *New Republic*, November 6,

Zellner, Arnold. 2004. "Keep it sophisticatedly simple." In *Simplicity, Inference and Modelling: Keeping It Sophisticatedly Simple*, edited by Arnold Zellner, Hugo A. Keuzenkamp and Michael McAleer, 242-262. Cambridge: Cambridge University Press.

Table 1

Accuracy of vote expectation surveys in predicting the winner (1932-2012)

		Predicted winner		
Election	No. of surveys	Correct	Wrong	Tie
1932	1	1	0	0
1940	3	3	0	0
1944	6	6	0	0
1948	1	0	1	0
1952	3	3	0	0
1956	5	5	0	0
1960	1	0	1	0
1964	1	1	0	0
1968	1	1	0	0
1972	3	3	0	0
1976	3	3	0	0
1980	3	1	2	0
1984	9	9	0	0
1988	17	16	1	0
1992	23	20	3	0
1996	20	20	0	0
2000	24	13	9	2
2004	43	39	1	3
2008	22	21	0	1
2012	26	26	0	0
Total	**215**	**191**	**18**	**6**

Table 2

Hit rate and mean absolute error of vote expectation surveys, trial-heat polls, and the IEM vote share prediction markets across the last 100 days prior to Election Day (1988-2012)

	Mean	1988	1992	1996	2000	2004	2008	2012
Hit rate (in %)								
Vote expectation surveys	**92**	85	100	100	60	98	99	100
Single polls	**79**	74	100	100	48	68	90	76
Combined polls	**86**	78	100	100	56	69	96	100
Combined poll projections	**88**	84	99	100	97	100	39	100
IEM vote share market	**79**	70	67	100	30	88	100	100
Mean absolute error (in %-points)								
Vote expectation surveys	**1.6**	3.0	1.4	1.1	2.2	1.1	1.8	0.4
Single polls	**3.2**	4.0	4.9	5.0	3.1	1.9	1.9	1.8
Combined polls	**2.9**	3.9	4.7	4.7	2.7	1.5	1.7	1.3
Combined poll projections	**2.0**	1.4	1.9	2.4	1.8	1.1	4.1	1.2
IEM vote share market	**1.7**	3.0	2.8	1.0	1.4	0.7	1.2	1.5

The analysis of the 1988 election covers only the last 91 days prior to Election Day, since no vote expectation survey was available before that day.

Table 3

Hit rate and mean absolute error of vote expectation surveys and experts (1992, 2000-2012)

	Mean	1992	2000	2004	2008	2012
Hit rate (in %)						
Vote expectations surveys	**100**	100	100	100	100	100
Typical expert forecast	**66**	87	20	47	94	82
Average expert forecast	**70**	100	0	50	100	100
Mean absolute error (in %-points)						
Vote expectations surveys	**1.2**	0.3	2.4	1.1	1.8	0.5
Typical expert forecast	**1.8**	1.8	2.4	1.8	1.7	1.3
Average expert forecast	**1.4**	0.7	2.3	1.4	1.5	1.3

Table 4

Hit rate and mean absolute error of vote expectation surveys and seven established quantitative models (1992-2012)

	Mean	1992	1996	2000	2004	2008	2012
Hit rate (in %)							
Voter expectations	**100**	100	100	100	100	100	100
Typical model forecast	**86**	75	100	100	86	86	71
Average model forecast	**100**	100	100	100	100	100	100
Mean absolute error (in %-points)							
Voter expectations	**1.5**	1.8	0.4	3.9	0.8	1.7	0.2
Typical model forecast	**2.9**	1.7	2.1	5.7	2.8	3.2	1.8
Average model forecast	**2.3**	1.4	2.1	5.7	2.5	1.1	0.8

Only four model forecasts were available for the 1992 election. See Appendix I for the forecasts of each model.

Figure 1

2012 U.S. presidential election vote-share forecasts

of 20 vote expectation surveys and 110 trial-heat polls

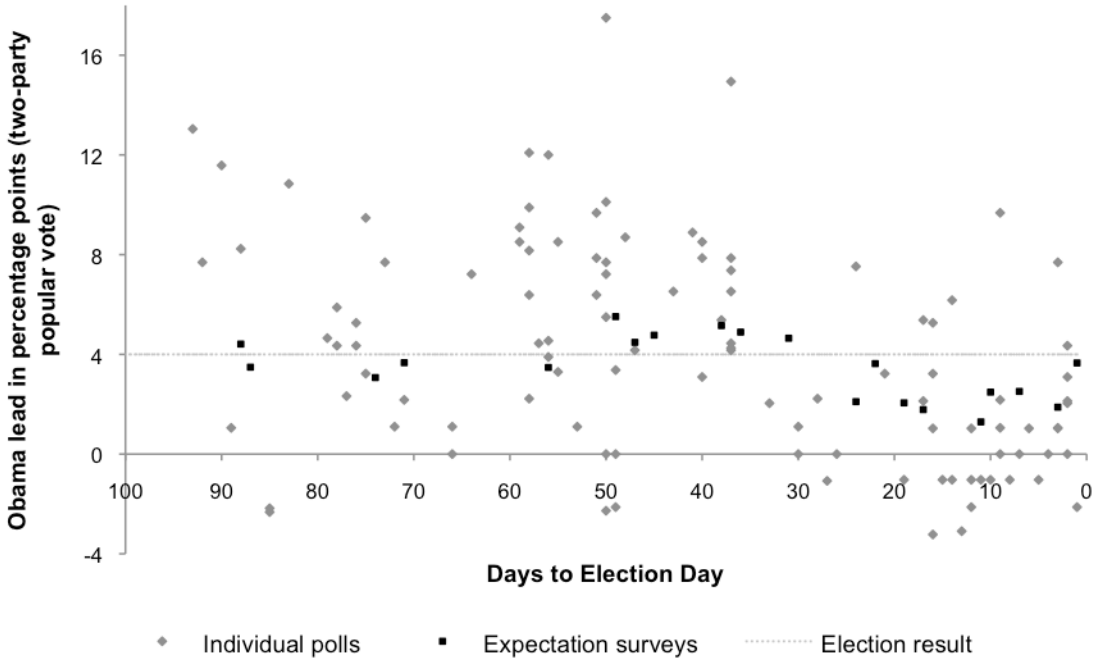
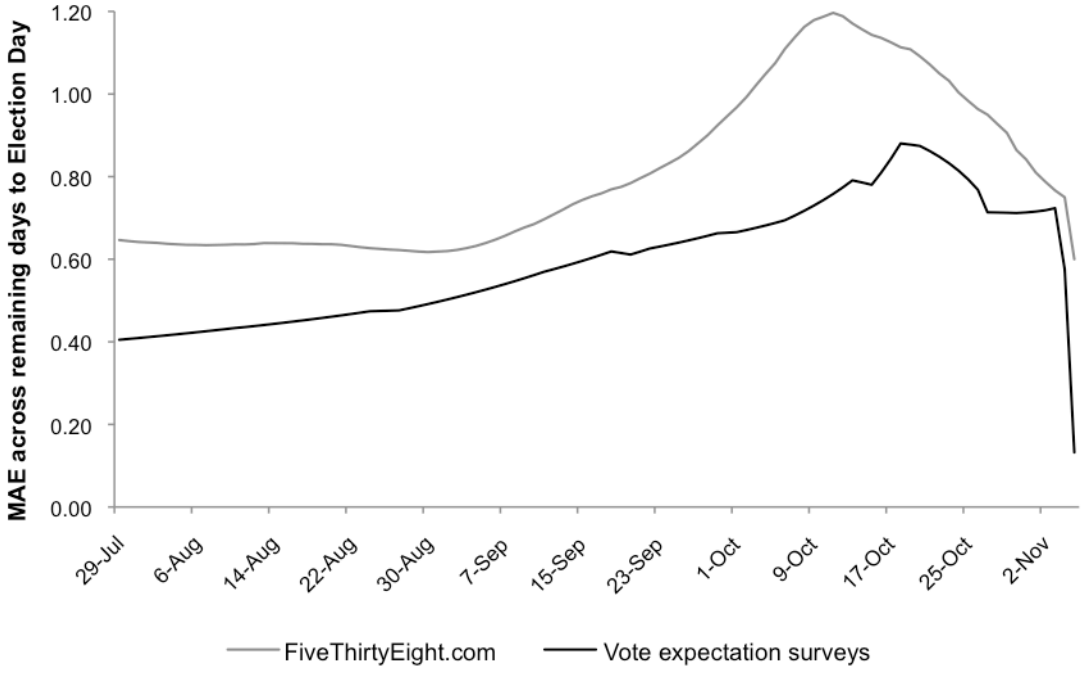


Figure 2

Mean absolute error of popular vote share forecasts

from vote expectation surveys and FiveThirtyEight.com



Appendix I

Quantitative model forecasts of U.S. presidential elections from 1992 to 2012

Election	1992		1996		2000		2004		2008		2012	
Election result	46.4		54.7		50.3		51.2		46.3		52.0	
	FC	AE	FC	AE	FC	AE	FC	AE	FC	AE	FC	AE
Abramowitz	46.7	0.3	56.8	2.1	53.2	2.9	53.7	2.5	45.7	0.6	50.6	1.4
Campbell	47.1	0.7	58.1	3.4	52.8	2.5	53.8	2.6	52.7	6.4	52.0	0.0
Holbrook			57.2	2.5	60.0	9.7	54.5	3.3	44.3	2.0	47.9	4.1
Lewis-Beck & Tien	51.5	5.1	54.8	0.1	55.4	5.1	49.9	1.3	49.9	3.6	48.2	3.8
Lockerbie			57.6	2.8	60.3	10.0	57.6	6.4	41.8	4.5	53.8	1.8
Norpoth			57.1	2.4	55.0	4.7	54.7	3.5	49.9	3.6	53.2	1.2
Wlezien & Erikson	45.9	0.5	56.0	1.3	55.2	4.9	51.7	0.5	47.8	1.5	52.6	0.6

Forecasts were published in *Political Methodologist* 5(2), *American Politics Research* 24(4) and *PS: Political Science and Politics* 34(1), 37(4), 41(4), and 45(4)

Appendix II

Google searches for FiveThirtyEight and "election forecast" prior to the 2012 election