

Predicting elections: Experts, polls, and fundamentals

Andreas Graefe*

Abstract

This study analyzes the relative accuracy of experts, polls, and the so-called ‘fundamentals’ in predicting the popular vote in the four U.S. presidential elections from 2004 to 2016. Although the majority (62%) of 452 expert forecasts correctly predicted the directional error of polls, the typical expert’s vote share forecast was 7% (of the error) less accurate than a simple polling average from the same day. The results further suggest that experts follow the polls and do not sufficiently harness information incorporated in the fundamentals. Combining expert forecasts and polls with a fundamentals-based reference class forecast reduced the error of experts and polls by 24% and 19%, respectively. The findings demonstrate the benefits of combining forecasts and the effectiveness of taking the outside view for debiasing expert judgment.

Keywords: election forecasting, expert judgment, polls, bias, reference-class forecasting

1 Introduction

Looking at election eve forecasts for the three high-profile elections in the UK (Brexit) and US (Trump) in 2016, and France (Le Pen) in 2017, FiveThirtyEight’s Nate Silver observed that, in each case, polling errors occurred in the opposite direction of what experts or betting markets had expected. This led him to conclude that: “When the conventional wisdom tries to outguess the polls, it almost always guesses in the wrong direction” (Silver, 2017). It is difficult to find evidence for or against this claim. Although the use of expert judgment in forecasting elections goes back long before the emergence of scientific polling (Kernell, 2000), we know surprisingly little about the relative accuracy of experts and polls.

Research on expert forecasting in different fields shows that the value of expertise is indeed limited when forecasting complex problems. In such situations, expert forecasts are

little more – and sometimes even less – accurate than those from novices and naïve statistical models, such as a random walk (Armstrong, 1980; Tetlock, 2005). One can however expect experts to make useful predictions for problems for which they get good feedback about the accuracy of their forecasts, and if they know the situation well (Green, Graefe & Armstrong, 2011).

Election forecasting appears to meet these conditions. First, elections have clear outcomes upon which forecast accuracy can be judged. Such feedback can help forecasters to learn about judgment errors and biases. Second, political experts can draw on a vast amount of theory and empirical evidence about electoral behavior, particularly for U.S. presidential elections, which should help them read and interpret polls. For example, research has shown that polls tend to tighten (Erikson & Wlezien, 2012), and the shares of both third-party support and undecideds decrease, as the election nears (Riker, 1982). We also know that certain campaign events such as party conventions (Campbell, Cherry & Wink, 1992) and candidate debates (Benoit, Hansen & Verser, 2003) can yield predictable shifts in the candidates’ polling numbers, not necessarily by affecting people’s vote preference but rather their willingness to participate in a poll (Gelman, Goel, Rivers & Rothschild, 2016). Furthermore, structural factors, the so-called ‘fundamentals’ (e.g., the state of the economy or the time the incumbent party has been in the White House), quite accurately predict election outcomes, even months in advance. In sum, when forecasting elections, experts receive immediate and accurate feedback about their forecasts and have access to domain knowledge, some of which may not be accounted for in the polls.

Polls, on the other hand, are far from being perfect predictors of election outcomes themselves, and are subject to various types of error (Biemer, 2010; Groves & Lyberg, 2010). Prior research found that the empirical error of polls is about

Thanks to Scott Armstrong, Alfred Cuzán, Randall Jones, and Elliott Morris for providing comments and suggestions. Thanks to the experts who participated in at least one expert survey over the years, namely Randall Adkins, Lonna Rae Atkeson, Scott Blinder, John Coleman, Omaha George Edwards, Keith Gaddie, John Geer, Ange-Marie Hancock, Seth Hill, Sunshine Hillygus, Jan Leighley, Sandy Maisel, Michael Martinez, Thomas Patterson, Gerald Pomper, David Redlawsk, Larry Sabato, Michael Tesler, Charles Walcott, and 17 experts who preferred to remain anonymous. The PollyVote project was founded in 2004 by J. Scott Armstrong, Alfred G. Cuzán, and Randall J. Jones, Jr. Andreas Graefe joined the group in 2008. Responsibilities for the administration of the expert surveys changed across the four elections. In 2004, Armstrong sent out the original invitations, collected the responses, and analyzed the data. For the elections in 2008 and 2012, Jones collected and processed the responses, and analyzed the data together with Cuzán. For the 2016 election, Graefe collected and analyzed the data.

Copyright: © 2018. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.

*Macromedia University, Munich, Germany. Email: graefe.andreas@gmail.com.

twice as large as the estimated sampling error (Buchanan, 1986; Shirani-Mehr, Rothschild, Goel & Gelman, 2018). Furthermore, polls were found to be among the least accurate methods available to forecast U.S. presidential elections, especially if conducted weeks or even months before an election (Graefe, Armstrong, Jones & Cuzán, 2017).

It thus seems reasonable to expect that experts are able to tell the direction in which the polls err. The present study provides empirical evidence to answer this question by analyzing the relative accuracy of expert judgment, a simple polling average, and fundamentals-based forecasts for predicting the popular vote in the four U.S. presidential elections from 2004 to 2016.

2 Materials and methods

2.1 Forecast data

Expert forecasts. Expert forecasts for the four U.S. presidential elections from 2004 to 2016 were collected over a 12-year period within the PollyVote.com research project. Since 2004, the PollyVote team has periodically asked experts to predict the national vote in U.S. presidential elections, starting many months before the election. The total number of surveys conducted across the four elections is 36 (Appendix I). The survey results were published prior to each election at pollyvote.com.

The expert panel consisted of American political scientists and, in 2004 and 2008, some practitioners. The panel composition changed across elections, with the number of panelists varying from 15 to 17. Some experts participated in only one election, others participated in all four elections. The total number of experts who participated in at least one survey round was 36, with the number of available forecasts per individual expert ranging from 1 to 36. The average number of experts for a single survey round was 13, and ranged from 8 to 17 (Appendix II).

From 2004 to 2012, the experts predicted the two-party vote for the candidate of the incumbent party (“*Considering only the major party candidates, what share of the 2-party vote do you expect the nominee of the incumbent [Democratic/Republican] party to receive in the [YEAR] general election?*”). In 2016, the experts predicted the vote for each party, including third-parties and others (“*What share of the national vote (in %) do you expect the nominees to receive in the 2016 presidential election?*”).¹

Polling average. The RealClearPolitics poll average (RCP) was used as the benchmark for the performance of polls. The

¹In addition, experts were asked how confident they are in their forecasts. In 2004, experts were asked to reveal an upper and lower bound for their given forecast. For forecasts of elections since 2008, experts were asked to give a probability that the actual vote will fall within +/-5 percentage points of the forecast they have given.

RCP is an established and widely-known polling average, and one of the few that was active as early as 2004. The RCP average does not weight by sample size or recency, and it does not correct for house effects (partisan lean of pollsters). The RCP is thus a very raw representation of publicly available opinion polls.

Fundamentals-based forecast. For nearly four decades, political scientists and economists have developed quantitative models for forecasting U.S. presidential elections. Most of these models are based on the theory of retrospective voting, which assumes that voters reward or punish the incumbent (party) based on its performance. Thereby, different models measure performance in different ways. While most models include at least one measure of economic performance, some models include military fatalities (e.g., Hibbs, 2012), the incumbent president’s job approval rating (e.g., Abramowitz, 2016), or the candidates’ performance in primary elections (e.g., Norpoth, 2016). In addition, several models measure the time that the incumbent party (or president) has been in office to account for the electorate’s periodic desire for change (e.g., Abramowitz (2009), Cuzán (2012), Fair (2009)).²

For the present study, I created a fundamentals-based forecast by calculating rolling averages of forecasts from five established models (equally weighted) that were published prior to each of the four elections from 2004 to 2016.³ The five models were those by Abramowitz (2016), Cuzán (2012), Fair (2009), Hibbs (2012) and Norpoth (2016). I deliberately decided to select these models and combine their forecasts for several reasons. First and foremost, these models are ‘pure’ in that they rely only on fundamental data. That is, they ignore trial-heat polls that measure support for the party nominees.⁴ Hence, these models provide a base rate prediction (or reference class forecast) of what one would expect to happen under ‘normal’ circumstances (i.e., with generic candidates). Second, ex ante forecasts published prior to each election were available for all five models. Third, each model uses different variables and thus includes different information. A combined forecast based on those models thus captures more information than any single model and minimizes the danger that the results are due to cherry-picking a single model.

2.2 Comparison of forecasts

The individual expert forecasts were compared to the respective forecasts from polls and fundamentals from the day an

²See, for example, the special symposiums in *PS: Political Science & Politics* published before each of the U.S. presidential elections from 2004 to 2016 (Campbell, 2004; 2008; 2012; 2016).

³Appendix III provides details on the publication date and source of each forecast.

⁴Appendix IV shows the variables used in each of the five models (based on their 2016 specification).

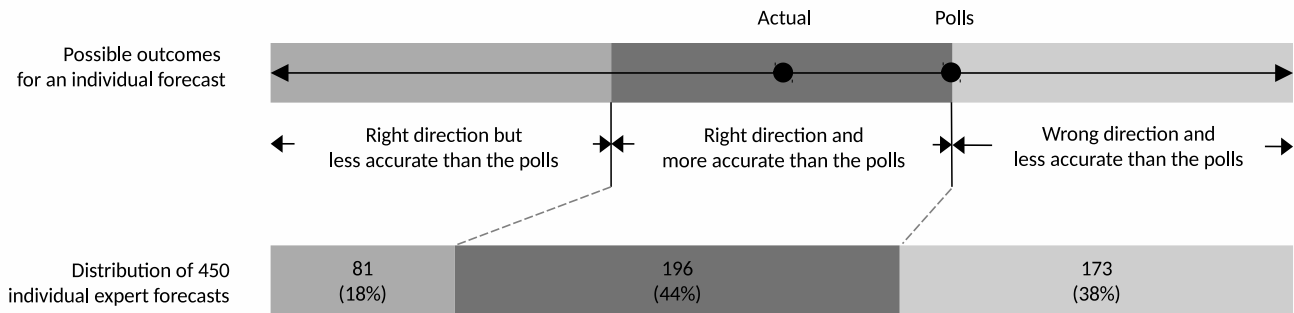


FIGURE 1: Distribution of expert forecasts relative to the polls, 2004 to 2016.

expert forecast was made. If the exact date of the expert forecast was unavailable, the last day of the expert survey was used as a reference. For the two expert surveys conducted in May and July of 2011, more than a year before the 2012 election, no RCP data were available. These surveys were thus excluded from the analysis.

2.3 Forecast combination

I also formed two combined measures. A combined forecast of polls and experts was calculated as the equally-weighted average of an individual expert forecast and the polling average that day. A combined forecast of polls, experts, and fundamentals was calculated as the equally weighted average of an individual expert forecast, the polling average that day, and the fundamentals-based forecast.

2.4 Error measure

Forecast errors were calculated as the difference between the predicted and actual Democratic vote share. The analysis is based on the two-party vote. Where necessary (as for the RCP and the 2016 expert forecasts), two party vote shares were calculated using the following formula: $(\text{Democratic vote}) / (\text{Democratic vote} + \text{Republican vote})$.

3 Results

3.1 Directional error

If experts are able to identify the directional error of polls, we would expect their own forecasts to be in the direction of the actual election outcome. Figure 1 suggests that this was in fact the case. When comparing experts' forecasts with the polling average of the same day across the four elections from 2004 to 2016, 277 (62%) of 450 expert forecasts were in the direction of the actual election outcome.⁵

⁵The total number of expert forecasts was 452. However, two forecasts were similar to the poll average and are thus excluded from the visualization in Figure 1.

However, the simple fact that the majority of expert forecasts pointed in the right direction does not imply that these forecasts are necessarily more accurate than the polls. The reason is that experts may adjust the polling numbers too far in the right direction and overshoot the actual outcome. This would result in an error larger than that of the polls. This happened for 81 (18%) of the expert forecasts. Together with the 173 (38%) cases in which experts moved the forecast in the wrong direction, the majority (56%) of expert forecasts were thus in fact less accurate (farther from the actual outcome) than the polls.

3.2 Vote share forecast error

Expert inaccuracy is also apparent from comparing the errors of experts' vote share forecasts with those from polls. Figure 2 shows the results as the mean absolute error (MAE) for each election, and across the four elections. The results were mixed. In 2008, experts outperformed the polls, whereas in 2004 the polls were more accurate than the experts. In 2012 and 2016, differences in accuracy were small. Across the four elections, the weighted (by the number of available forecasts in each election) MAE of a typical expert forecast (1.6 percentage points) was 7% higher than the respective error of the polling average (1.5 percentage points).

3.3 Bias

Figure 3 addresses the question of potential biases in depicting the mean difference between the predicted and actual Democratic two-party vote for each method. Hence, values above the horizontal axis indicate that a method overpredicted the Democratic vote, while values below the horizontal axis suggest that the method overpredicted the Republican vote. For example, in 2016, the typical expert forecast overpredicted the Democratic vote by 1.7 percentage points, while the polling average overpredicted Democrats by 1.5 percentage points.

Experts overpredicted the Democratic vote (in 2004 and 2016) and the Republican vote (in 2008 and 2012) twice each. Interestingly, experts and polls erred in the same direction in

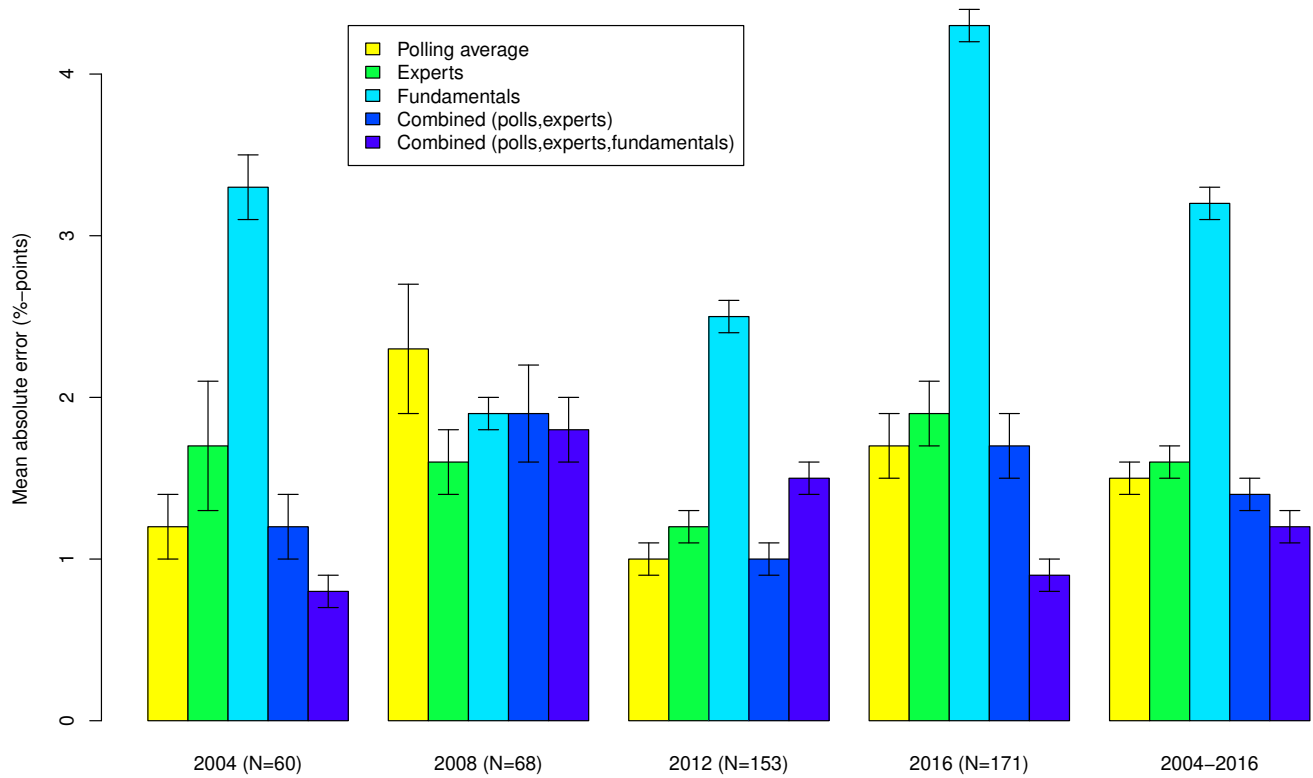


FIGURE 2: Mean absolute error (in %-points) of experts, the polling average, fundamentals-based forecasts, and two combined forecasts, 2004 to 2016 (error bars show 95%-confidence intervals).

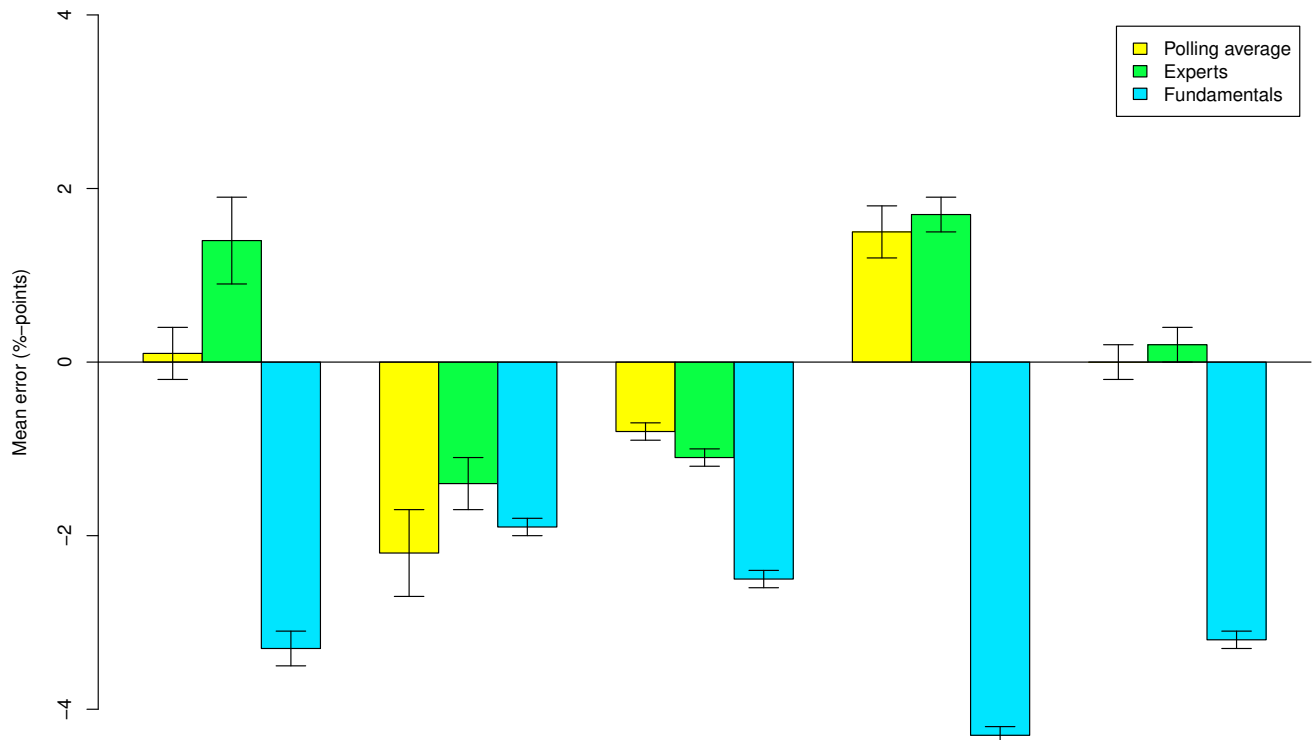


FIGURE 3: Mean error (in %-points) of experts, the polling average, and fundamentals-based forecasts, 2004 to 2016 (error bars show 95%-confidence intervals). Positive numbers favor Democrats.

each election. This may indicate that experts draw heavily on polls when making their forecasts. In three of the four elections (except for 2012), experts were more favorable for the Democrats than the polls. Across all forecasts, the polls showed virtually no bias, while the typical expert slightly overshot the Democratic vote by 0.2 percentage points.

4 Discussion

The MAE across all 452 experts' vote share forecasts was 7% higher than a simple polling average from the same day. This is a small difference in accuracy, which certainly does not imply that one should ignore expert judgment when forecasting elections.

Trying to find the one best forecasting method is generally not a useful strategy for forecasting. Rather, the literature suggests combining forecasts from different methods. The reason is that a combined forecast includes more information than any single forecast, and the systematic and random errors associated with single forecasts tend to cancel out in the aggregate. This improves accuracy. If one uses the simple average as the means to combining, the combined forecast will at least be as accurate as the average error of the individual component forecasts, and often much more accurate (Armstrong, 2001).

Experimental studies have shown that many people do not understand, and thus do not appreciate, the power of combining forecasts (Larrick & Soll, 2006). One reason is that people commonly think that they know which forecast is the best one and decide to go with it. But this is not a good approach to forecasting for several reasons. First, in picking a particular forecast, one may select a forecast that suits one's biases (Soll & Larrick, 2009). Second, it is extremely difficult, if not impossible, in most practical situations to know in advance which forecast will turn out to be most accurate. Past accuracy, for example, is not a good indicator for future accuracy. Two studies have found a negative relationship between the historical accuracy of a method (Graefe et al., 2017) or model (Graefe, Küchenhoff, Stierle & Riedl, 2015) and its accuracy in predicting future elections. Third, even if one would know in advance which forecast will be most accurate, combining that forecast with less accurate forecasts can be useful. Herzog and Hertwig (2009) illustrate this – perhaps counterintuitive – finding for the simple case of combining two forecasts. The authors showed that the simple average of two forecasts is more accurate than the best single forecast if the two component forecasts bracket the outcome — i.e., the outcome is between the two forecasts — and if the error of the less accurate forecast does not exceed *three times* the error of the more accurate one. In other words, as long as adding a new forecast to the combination is likely to increase the chance that the

range of forecasts bracket the true value, that forecast's error can be quite large.

4.1 Combining forecasts from experts and polls

In the present study, the majority of expert forecasts (62%) were in the direction of the final election result. That is, there is a high chance that the expert forecasts and the polls bracket the true value. In such a situation, combining forecasts is likely to be useful. Figure 1 shows the MAE of a combined forecast of polls and individual expert forecasts. Across the four elections, the error of the combined forecast was 1.4 percentage points, which is 5% lower than the corresponding error of the polling average (1.5 percentage points), the best of the two methods. Compared to the expert forecasts (1.6 percentage points), the combined forecast reduced error by 12%. Also, note that even when the combined forecast did not provide the most accurate predictions, it helped avoiding large errors, such as the relatively large polling error in 2008.

As pointed out above, the results on the relative accuracy of polls and experts suggested that experts closely follow the polls (Figure 3). In other words, the two methods likely incorporate similar information. Combining, however, works particularly well if one combines forecasts that incorporate different information (Graefe, Armstrong, Jones & Cuzán, 2014). Hence, in order to improve upon the accuracy of forecasts from polls and experts, one should look for information that these methods may overlook, and incorporate that into the forecast.

4.2 Ignorance of election fundamentals

Figure 3 shows the mean errors of the fundamentals-based forecasts. The results reveal an interesting pattern. The fundamentals consistently overpredicted the Republican vote by substantial margins. In other words, in each of the past four elections, the Republicans underperformed relative to the fundamentals, and achieved less votes than what could be expected from historical data.

The results reveal mixed results for the performance of fundamentals relative to polls and experts. While in two of the four elections (2008 and 2012), all three methods erred in the same direction, the fundamentals pointed in the opposite direction from experts and polls in both 2004 and 2016. What is more, in both 2004 and 2016, the experts thought that the polls would underestimate the Democratic vote, even though the fundamentals pointed the other way. The results thus suggest that the experts did not sufficiently account for their fellow political scientists work on election fundamentals when making forecasts.

I can only speculate on the reasons for this behavior. For example, it may be that experts have little trust in these models, since they often incur large errors. Figure 2 shows the

MAE of the fundamentals-based forecast in each election. In three of the four elections, the fundamentals were by far the least accurate method. Only in 2008 did the polls perform even worse. Across the four elections, the fundamentals-based forecast missed by 3.2 percentage points. This error is more than double the corresponding errors of polls and experts. The large error may lead experts to think that fundamental models are generally of limited value and thus to ignore them altogether. Such neglect would be unfortunate, however, if these models provide valuable information regarding the direction of the forecast error.

4.3 Combining polls, experts, and fundamentals

The last columns in Figure 2 show the results of a combined forecast of polls, individual experts, and the fundamentals-based forecast. The combined forecast was more accurate than both the typical expert forecast and the polling average in two of the four elections (2004 and 2016). Across the four elections, the combined forecast (MAE: 1.2 percentage points) reduced the errors of the polling average (1.5) and the typical expert (1.6) forecast by 19% and 24%, respectively.

Some readers may be puzzled by the fact that these large accuracy gains occurred despite adding a forecast to the ensemble that incurred an error that was more than twice the corresponding errors of polls and experts. The results thus provide further evidence that combining can be useful even in situations when one has strong evidence that a particular method will be most accurate. The key here is that the fundamentals-based forecast provided different information than both experts and polls, thus increasing the likelihood that the combined forecast would bracket the true value (Graefe et al., 2014).

4.4 Limitations

The analysis presented in this paper is based on a rather large sample of expert forecasts (N=452) collected over a 12-year period. That said, the generalizations that can be drawn are limited, since the data cover only the four U.S. presidential elections from 2004 to 2016. Further studies for different election types and in other countries are necessary to learn more about the relative predictive accuracy and potential biases of expert judgment in forecasting elections.

4.5 Conclusions

The present study provides evidence on the accuracy of expert judgment in forecasting elections. Although the majority of expert forecasts correctly predicted the directional error of polls, the error of a typical expert's vote share forecast was on average 7% higher than a polling average. The

results further suggest that experts ignored information captured by structural fundamental data available months before election day, which prior research found to be useful for election forecasting. Combining expert forecasts and polls with such a fundamentals-based reference class forecast reduced the error of polls and experts by 19% and 24%, respectively.

These large gains in accuracy are in line with prior research, which showed that reference class forecasts and base rates are one of the most effective tools for debiasing judgmental forecasts (Chang, Chen, Mellers & Tetlock, 2016). Experts in any field should refrain from focusing too much on the specifics of a situation ("this time is different") but also take the outside view (Lovallo & Kahneman, 2003). In addition, they should be conservative about large changes and take into account all cumulative knowledge about a situation (Armstrong, Green & Graefe, 2015). A structured approach of combining forecasts from different methods that use different information provides a valuable and simple strategy to achieve that goal.

References

- Abramowitz, A. I. (2016). Will time for change mean time for Trump? *PS: Political Science & Politics*, 49(4), 659–660.
- Armstrong, J. S. (1980). The Seer-Sucker theory: The value of experts in forecasting. *Technology Review*, 1980(83), 16–24.
- Armstrong, J. S. (2001). Combining forecasts. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 417–439). New York: Springer.
- Armstrong, J. S., Green, K. C., & Graefe, A. (2015). Golden Rule of Forecasting: Be conservative. *Journal of Business Research*, 68(8), 1717–1731.
- Benoit, W. L., Hansen, G. J., & Verser, R. M. (2003). A meta-analysis of the effects of viewing US presidential debates. *Communication Monographs*, 70(4), 335–350.
- Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, 74(5), 817–848.
- Buchanan, W. (1986). Election predictions: An empirical assessment. *Public Opinion Quarterly*, 50(2), 222–227.
- Campbell, J. E. (2004). Introduction — The 2004 Presidential Election Forecasts. *PS: Political Science & Politics*, 37(4), 733–735.
- Campbell, J. E. (2008). Editor's Introduction: Forecasting the 2008 National Elections. *PS: Political Science & Politics*, 41(4), 679–682.
- Campbell, J. E. (2012). Forecasting the 2012 American National Elections: Editor's introduction. *PS: Political Science & Politics*, 45(4), 610–613.

- Campbell, J. E. (2016). Introduction. *PS: Political Science & Politics*, 49(4), 649–654.
- Campbell, J. E., Cherry, L. L., & Wink, K. A. (1992). The convention bump. *American Politics Research*, 20(3), 287–307.
- Chang, W., Chen, E., Mellers, B., & Tetlock, P. (2016). Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments. *Judgment and Decision Making*, 11(5), 509–526.
- Cuzán, A. G. (2012). Forecasting the 2012 presidential election with the fiscal model. *PS: Political Science & Politics*, 45(4), 648–650.
- Erikson, R. S., & Wlezien, C. (2012). *The Timeline of Presidential Elections: How Campaigns Do (And Do Not) Matter*. Chicago: University of Chicago Press.
- Fair, R. C. (2009). Presidential and congressional vote-share equations. *American Journal of Political Science*, 53(1), 55–72.
- Gelman, A., Goel, S., Rivers, D., & Rothschild, D. (2016). The mythical swing voter. *Quarterly Journal of Political Science*, 11(1), 103–130.
- Graefe, A. (2015). Improving forecasts using equally weighted predictors. *Journal of Business Research*, 68(8), 1792–1799.
- Graefe, A., Armstrong, J. S., Jones, R. J. J., & Cuzán, A. G. (2014). Combining forecasts: An application to elections. *International Journal of Forecasting*, 30(1), 43–54.
- Graefe, A., Küchenhoff, H., Stierle, V. & Riedl, B. (2015). Limitations of Ensemble Bayesian Model Averaging for forecasting social science problems. *International Journal of Forecasting*, 31(3), 943–951.
- Graefe, A., Armstrong, J. S., Jones, R. J. J., & Cuzán, A. G. (2017). Assessing the 2016 U.S. presidential election popular vote forecasts. In A. Cavari, R. Powell, & K. Mayer (Eds.), *The 2016 Presidential Election: The Causes and Consequences of a Political Earthquake* (pp. 137–158). Lanham, MD: Lexington Books.
- Green, K. C., Graefe, A., & Armstrong, J. S. (2011). Forecasting principles. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 527–534). Berlin Heidelberg: Springer.
- Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly*, 74(5), 849–879.
- Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, 20(2), 231–237.
- Hibbs, D. A. (2012). Obama's reelection prospects under "Bread and Peace" voting in the 2012 US presidential election. *PS: Political Science & Politics*, 45(4), 635–639.
- Kernell, S. (2000). Life before polls: Ohio politicians predict the 1828 presidential vote. *PS: Political Science & Politics*, 33(3), 569–574.
- Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management science*, 52(1), 111–127.
- Lovallo, D., & Kahneman, D. (2003). Delusions of success. *Harvard Business Review*, 81(7), 56–63.
- Norpoth, H. (2016). Primary model predicts Trump victory. *PS: Political Science & Politics*, 49(4), 655–658.
- Riker, W. H. (1982). The two-party system and Duverger's law: An essay on the history of political science. *American Political Science Review*, 76(4), 753–766.
- Shirani-Mehr, H., Rothschild, D., Goel, S., & Gelman, A. (2018). Disentangling bias and variance in election polls. *Journal of the American Statistical Association* (Online first), <https://doi.org/10.1080/01621459.2018.1448823>.
- Silver, N. (2017). Conventional wisdom may be contaminating polls. May 9. Retrieved from <https://fivethirtyeight.com/features/conventional-wisdom-may-be-contaminating-polls/>
- Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 780–805.
- Tetlock, P. C. (2005). *Expert political judgment*. Princeton: Princeton University Press.

Appendix I: Overview of expert surveys per election.

Election	Survey	Start date	End date	N of experts	RCP available
2004	1	N/A	4-Jul-04	15	yes
2004	2	N/A	29-Jul-04	12	yes
2004	3	N/A	6-Sep-04	16	yes
2004	4	N/A	25-Oct-04	17	yes
2008	1	25-May-07	2-Jul-07	15	no
2008	2	8-Jul-07	18-Aug-07	13	no
2008	3	6-Dec-07	3-Jan-08	11	yes
2008	4	28-Feb-08	13-Mar-08	9	yes
2008	5	29-May-08	5-Jun-08	10	yes
2008	6	4-Sep-08	13-Sep-08	12	yes
2008	7	13-Oct-08	17-Oct-08	13	yes
2008	8	31-Oct-08	2-Nov-08	13	yes
2012	1	14-Dec-11	2-Jan-12	15	yes
2012	2	3-Feb-12	9-Feb-12	15	yes
2012	3	9-Mar-12	13-Mar-12	12	yes
2012	4	11-Apr-12	20-Apr-12	15	yes
2012	5	19-May-12	25-May-12	13	yes
2012	6	22-Jun-12	27-Jun-12	10	yes
2012	7	23-Jul-12	28-Jul-12	16	yes
2012	8	24-Aug-12	29-Aug-12	15	yes
2012	9	24-Sep-12	28-Sep-12	15	yes
2012	10	17-Oct-12	19-Oct-12	12	yes
2012	11	1-Nov-12	4-Nov-12	15	yes
2016	1	26-Dec-15	31-Dec-15	8	yes
2016	2	26-Jan-16	31-Jan-16	17	yes
2016	3	26-Feb-16	28-Feb-16	15	yes
2016	4	31-Mar-16	28-Mar-16	14	yes
2016	5	27-Apr-16	30-Apr-16	15	yes
2016	6	29-May-16	31-May-16	13	yes
2016	7	28-Jun-16	30-Jun-16	12	yes
2016	8	29-Jul-16	31-Jul-16	12	yes
2016	9	29-Aug-16	31-Aug-16	13	yes
2016	10	28-Sep-16	30-Sep-16	13	yes
2016	11	13-Oct-16	14-Oct-16	14	yes
2016	12	29-Oct-16	1-Nov-16	13	yes
2016	13	6-Nov-16	7-Nov-16	12	yes

Appendix II: Number of forecasts per expert and election.

Expert	N of elections	N of forecasts			
		2004–2016	2004	2008	2012 2016
1	4	36	4	8	11 13
2	4	36	4	8	11 13
3	4	34	4	8	11 11
4	4	32	4	6	9 13
5	4	31	4	6	9 12
6	3	27	0	7	10 10
7	3	26	0	7	6 13
8	3	25	0	4	8 13
9	2	24	0	0	11 13
10	2	22	0	0	9 13
11	2	21	0	0	10 11
12	3	18	0	7	7 4
13	1	11	0	0	11 0
14	1	11	0	0	11 0
15	1	11	0	0	11 0
16	2	12	4	8	0 0
17	1	10	0	0	0 10
18	2	10	4	6	0 0
19	1	9	0	0	0 9
20	2	10	3	7	0 0
21	1	8	0	0	0 8
22	1	8	0	0	8 0
23	2	9	4	5	0 0
24	1	5	0	5	0 0
25	1	4	0	0	0 4
26	1	4	4	0	0 0
27	1	4	4	0	0 0
28	1	4	4	0	0 0
29	1	4	4	0	0 0
30	1	3	3	0	0 0
31	1	2	2	0	0 0
32	1	2	2	0	0 0
33	2	3	2	1	0 0
34	1	2	0	2	0 0
35	1	1	0	0	0 1
36	1	1	0	1	0 0

Appendix III: Forecasts of fundamentals-based model (continued on next page).

Forecaster	Election	Forecast date	Dem	Rep	Source	
Abramowitz (2016)	2004	July 31, 2004	46.3	53.7	a	
		August 27, 2008	54.3	45.7	b	
		May 23, 2012	51.0	49.0	e	
	2016	August 1, 2012	50.5	49.5	e	
		August 29, 2012	50.6	49.4	c	
		June 14, 2016	48.7	51.3	k	
		July 29, 2016	48.6	51.4	d	
Cuzán (2012)	2004	April 3, 2004	48.0	52.0	i	
		May 24, 2004	47.2	52.8	i	
		August 11, 2004	48.9	51.1	i	
		October 29, 2004	48.8	51.2	e	
	2008	August 2, 2008	52.0	48.0	b	
	2012	January 1, 2011	52.7	47.3	e	
		September 8, 2011	49.9	50.1	e	
		November 3, 2011	46.6	53.4	e	
		May 24, 2012	47.6	52.4	e	
	2016	August 1, 2012	46.2	53.8	c	
		August 11, 2016	48.2	51.8	j	
	Fair (2009)	2004	January 30, 2003	43.7	56.3	h
April 25, 2003			43.7	56.3	h	
July 31, 2003			43.3	56.7	h	
October 31, 2003			41.7	58.3	h	
February 5, 2004			41.3	58.7	h	
April 29, 2004			41.3	58.7	h	
July 31, 2004			42.5	57.5	h	
October 29, 2004			42.3	57.7	h	
2008			November 1, 2006	53.5	46.5	h
			January 31, 2007	53.4	46.6	h
			April 27, 2007	53.2	46.8	h
			July 27, 2007	52.0	48.0	h
			October 31, 2007	51.9	48.1	h
			January 31, 2008	52.0	48.0	h
			April 30, 2008	52.2	47.8	h
			July 31, 2008	51.5	48.5	h
2012		October 30, 2008	51.9	48.1	h	
		November 11, 2010	55.9	44.1	h	
		January 29, 2011	52.5	47.5	h	
		April 28, 2011	52.8	47.2	h	
		July 31, 2011	53.4	46.6	h	
		October 30, 2011	50.0	50.0	h	
		January 28, 2012	50.3	49.7	h	
		April 27, 2012	50.2	49.8	h	
		July 27, 2012	49.5	50.5	h	
		October 26, 2012	49.0	51.0	h	
2016		November 11, 2014	48.7	51.3	h	
		January 31, 2015	46.0	54.0	h	
		April 29, 2015	48.6	51.4	h	
		July 31, 2015	46.4	53.6	h	
		October 31, 2015	45.8	54.2	h	
		January 30, 2016	45.7	54.3	h	
	April 28, 2016	45.0	55.0	h		
	July 29, 2016	44.0	56.0	h		
October 28, 2016	44.0	56.0	h			

Appendix III, continued.

Forecaster	Election	Forecast date	Dem	Rep	Source
Hibbs (2012)	2004	July 26, 2004	46.9	53.2	g
	2008	July 6, 2008	51.8	48.2	f
		October 31, 2008	53.8	46.3	e
		May 28, 2011	46.2	53.8	e
	2012	October 27, 2011	47.8	52.2	e
		February 29, 2012	48.2	51.8	e
July 27, 2012		47.5	52.5	c	
Hibbs (2012)	2016	October 22, 2016	53.9	46.1	l
Norpoth (2016)	2004	January 29, 2004	45.3	54.7	a
Norpoth (2016)	2008	January 15, 2008	50.1	49.9	b
	2012	January 12, 2012	53.2	46.8	c
	2016	March 7, 2016	47.5	52.5	d

Sources

a	Campbell (2004)
b	Campbell (2008)
c	Campbell (2012)
d	Campbell (2016)
e	Graefe, A. (2013). Replication data for: Combining forecasts: An application to elections. https://doi.org/10.7910/DVN/23184 , HarvardDataverse, V3
f	http://www.douglas-hibbs.com/Election2008/2008Election-MainPage.htm
g	http://www.douglas-hibbs.com/Elections2004--00--96--92/election2004.pdf
h	https://fairmodel.econ.yale.edu/
i	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2821878
j	https://uwf.edu/media/university-of-west-florida/colleges/cassh/departments/government/cdocs/I.A.10.-Cuzan-2004-fiscaleffectsprselect.pdf
k	https://www.vox.com/2016/6/14/11854512/trump-election-models-political-science
l	Personal communication (Email, October 22, 2016)

Appendix IV: Independent variables included in each of the five models.

	Abramowitz (2016)	Cuzán (2012)	Fair (2009)	Hibbs (2012)	Norpoth (2016)
Total number of variables	3	5	7	2	4
Economic growth	1	2	3	1	.
Federal spending	.	1	.	.	.
Incumbency	1	2	3	.	.
War / military fatalities	.	.	1	1	.
Primary support	2
Presidential job approval	1
Previous vote share	2