Accuracy gains of adding vote expectation surveys to a combined forecast of US presidential election outcomes

Andreas Graefe

Abstract

In averaging forecasts within and across four-component methods (i.e. polls, prediction markets, expert judgment and quantitative models), the combined PollyVote provided highly accurate predictions for the US presidential elections from 1992 to 2012. This research note shows that the PollyVote would have also outperformed vote expectation surveys, which prior research identified as the most accurate individual forecasting method during that time period. Adding vote expectations to the PollyVote would have further increased the accuracy of the combined forecast. Across the last 90 days prior to the six elections, a five-component PollyVote (i.e. including vote expectations) would have yielded a mean absolute error of 1.08 percentage points, which is 7% lower than the corresponding error of the original four-component PollyVote. This study thus provides empirical evidence in support of two major findings from forecasting research. First, combining forecasts provides highly accurate predictions, which are difficult to beat for even the most accurate individual forecasting method available. Second, the accuracy of a combined forecast can be improved by adding component forecasts that rely on different data and different methods than the forecasts already included in the combination.

Keywords

Combining forecasts, election forecasting, vote expectations, citizen forecasts, presidential research

Introduction

Combining forecasts is a well-established and powerful method to increase forecast accuracy (Armstrong, 2001; Clemen, 1989). The reason is that a combined forecast includes more information than forecasts from any single component method. In addition, the systematic and random errors associated with individual component forecasts are likely to cancel out in the combined forecast.

As has been demonstrated with the PollyVote for predicting US presidential elections, combining forecasts is particularly beneficial if one can draw on component forecasts that use different methods and data. The PollyVote averages forecasts within and across four different component methods: polls, prediction markets, quantitative models, and expert judgment. Across the six elections from 1992 to 2012, the resulting combined forecast reduced the error of a typical poll, model, and expert judgment by more than half. Compared with prediction markets, the most accurate component method, error was reduced by 16%

(Graefe et al., 2014b). Forecasts made on Election Eve prior to the three elections from 2004 to 2012 missed the final vote share on average by 0.6 percentage points. To put this in perspective, the average error of the final Gallup poll was more than three times higher (Graefe et al., 2014a).

This performance was achieved even though the PollyVote did *not* include forecasts from vote expectation surveys, also known as 'citizen forecasts', which were recently shown to provide highly accurate forecasts of US presidential election outcomes (Graefe, 2014). These surveys simply ask respondents whom they *expect* to win.[1]

LMU Research Fellow, Department of Communication Studies and Media Research, LMU Munich, Germany
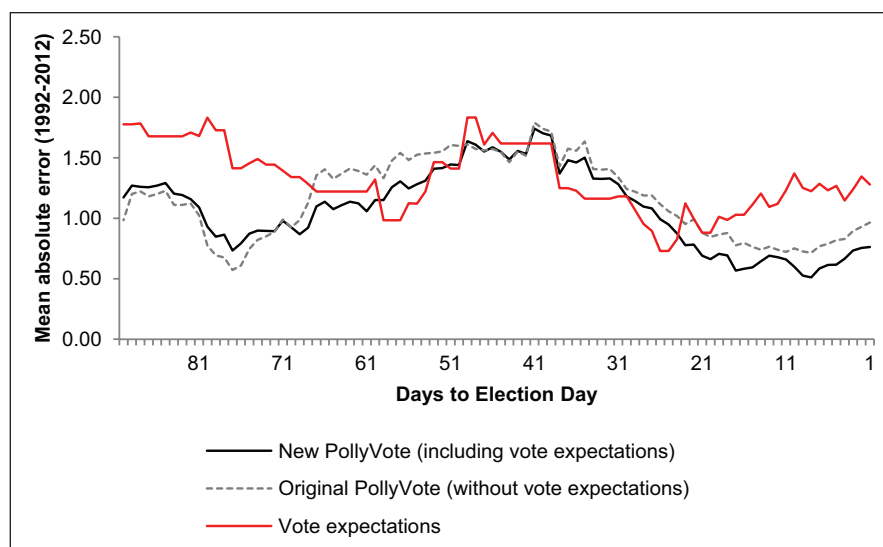
Corresponding author:
Andreas Graefe, LMU Research Fellow, Department of Communication Studies and Media Research, Oettingenstr. 67, LMU Munich 80538, Germany.
Email: a.graefe@lmu.de

Figure 1. Mean absolute errors of forecasts from vote expectations, the original PollyVote (without vote expectations) and the new PollyVote (with vote expectations), 1992–2012.

The aggregate responses are then used as a forecast of who will win the election. If data on historical elections are available, the aggregate responses can also be translated to popular vote-share forecasts using simple linear regression (Lewis-Beck and Stegmaier, 2011; Lewis-Beck and Tien, 1999).

Vote expectation surveys have been around at least as long as scientific polling (Hayes, 1936), but have long been overlooked as a method for forecasting election outcomes. Although early work pointed to the accuracy of vote expectations, these studies focused on identifying factors that explain *why* most citizens are able to accurately predict election outcomes (e.g. Lewis-Beck and Skalaban, 1989; Lewis-Beck and Tien, 1999). Only recently have researchers begun to specifically study vote expectation surveys as a method for forecasting elections (Lewis-Beck and Stegmaier, 2011; Murr, 2011, 2014).

In a previous study, I compared the accuracy of vote expectations to forecasts from polls, prediction markets, quantitative models, and expert judgment (Graefe, 2014). Across the last 100 days prior to the seven US presidential elections from 1988 to 2012, vote expectations provided more accurate forecasts of election winners and vote shares than each of the four established methods. Compared with polls, vote expectations reduced the error of vote-share predictions by 51%. Compared with prediction markets, error was reduced by 6%. In other words, vote expectation surveys appear to be the most accurate individual method for forecasting US presidential elections available to date.

The present research note builds on this work and contributes to knowledge on combining forecasts by analysing (1) the relative accuracy of vote expectations and the PollyVote and (2) the accuracy gains from adding vote expectations to the PollyVote.

Method and data

Accuracy is analysed for forecasts of the national two-party popular vote in the six US presidential elections from 1992 to 2012, the time period for which forecast data on both the PollyVote and vote expectation surveys are available. The absolute error, calculated as the absolute difference of the predicted and actual national two-party popular vote of the incumbent party's candidate, was used as the measure of accuracy.

Forecasts from the original four-component PollyVote and vote expectations were obtained from publicly available datasets at the *Harvard Dataverse Network*. These datasets provide daily forecasts of the national two-party popular vote for each of the six US presidential elections from 1992 to 2012.[2] From these data, a new set of daily forecasts was calculated by adding vote expectations as a fifth component method to the original PollyVote. That is, this new (five-component) PollyVote was computed by averaging forecasts across five (instead of four) component methods: (1) polls, (2) prediction markets, (3) quantitative models, (4) expert judgment, and (5) vote expectations. For more information on the calculation of the original PollyVote see Graefe et al. (2014b). All data and calculations are available at the *Harvard Dataverse Network*.[3]

Results

Figure 1 shows the mean absolute errors (MAE) of forecasts from vote expectations, the original PollyVote (without vote expectations), and the new PollyVote (including vote expectations) across the six elections from 1992 to 2012. Vote expectations were less accurate than the original PollyVote for both long-term (90–70 days prior to Election
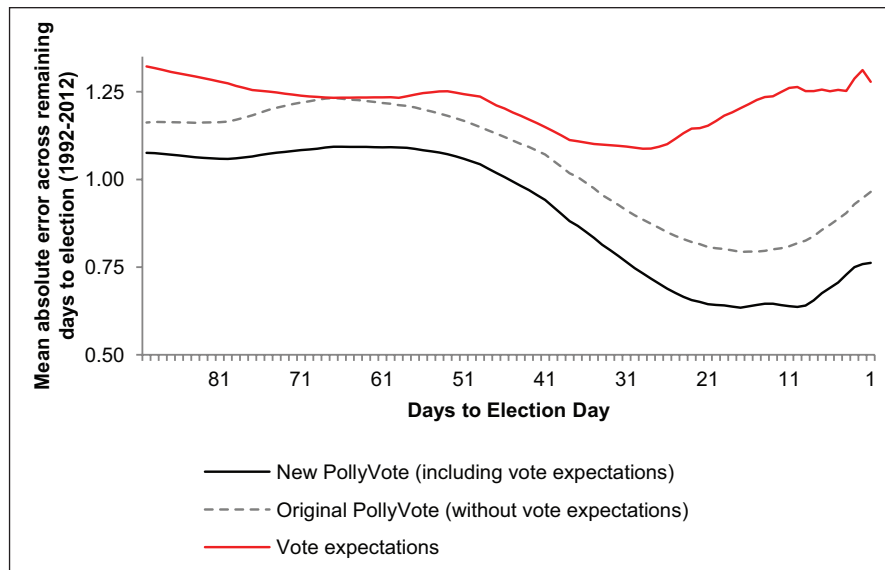
Figure 2. Mean absolute error of forecasts from vote expectations, the original PollyVote (without vote expectations) and the new PollyVote (including vote expectations), calculated across the remaining day to election, 1992–2012.

Day) and short-term forecasts (from 20 days prior to Election Day). For medium-term forecasts (60–20 days), however, vote expectations performed similar to – and sometimes better than – the original PollyVote.

Figure 1 further shows that adding vote expectations to the original PollyVote increases accuracy. Except for long-term forecasts, the new five-component PollyVote provides at least as – and usually more – accurate forecasts as the original four-component PollyVote.

Figure 2 presents the same data in a different way by showing the MAE of vote expectations and both PollyVote versions *across the remaining days* in the forecast horizon. That is, at any given day, the chart depicts the average error that one would have achieved by picking one of the three methods and relying on its forecast until Election Day. For example, if one had relied on the vote expectation forecasts starting 90 days before the election, an average error of 1.32 percentage points would have resulted. In comparison, the corresponding error of the original PollyVote would have been 12% lower (1.16 percentage points). In general, the gains in accuracy by relying on the PollyVote rather than vote expectations tend to increase as the election comes closer. Furthermore, Figure 2 demonstrates the benefit of adding vote expectations, as the error of the new (five-component) PollyVote was consistently lower than the error of the original (four-component) PollyVote. For example, starting 90 days prior to Election Day, the MAE of the new PollyVote was 1.08, which is 7% lower than the corresponding error of the original PollyVote.

Discussion

This research note provides empirical evidence in support of two major findings from the forecasting literature. First,

combining forecasts from different methods that use different data provides highly accurate forecasts, which are difficult to beat by even the most accurate individual method available. Across the past 90 days prior to each of the six elections from 1992 to 2012, the original PollyVote – which averages forecasts within and across polls, prediction markets, quantitative models, and expert judgment, but does not include vote expectations – missed the incumbent party's final vote share on average by 1.16 percentage points. This error is 12% lower than the corresponding error of vote expectation surveys, which prior research found to the most accurate method for the examined time period (Graefe, 2014).

Second, and more importantly, the accuracy of a combined forecast can be further improved by adding component forecasts that rely on a different method and different data than the forecasts already included in the combination. After adding vote expectations as a fifth component method, the new PollyVote reduced the error of the original four-component version by 7%, a substantial improvement given the already very low forecast error. On average across the 90 days prior to the six elections, the new five-component PollyVote missed the final election result by little more than one (i.e. 1.08) percentage point.

This performance was achieved by calculating simple unweighted averages within and across forecasts of five component methods. Calculating unweighted averages may appear as a naïve approach to combining forecasts, as it does not account for the component methods' relative accuracy. However, an early review of more than 200 papers showed that the simple average provides a good starting point for combining forecasts, and is difficult to beat by more complex approaches (Clemen, 1989). These results still hold today, despite many efforts in search of

sophisticated methods for combining. The problem with complex statistical procedures that aim to estimate component weights from historical data is that they tend to perform poorly in situations with limited and messy data, which are common in the social sciences. A recent example is Ensemble Bayesian Model Averaging (EBMA), a method that has been shown to perform well for combining forecasts in the data-heavy domain of weather forecasting. However, when applied to problems with scarce and noisy data, such as in economic and election forecasting, EBMA provided less accurate forecasts than the simple equal-weights average (Graefe et al., 2015).

When pre-specifying equal weights to component forecasts, analysts ignore the components' relative accuracy. Instead, they deliberately introduce a bias that reduces variance and thereby limits a model's ability to *explain* given data. At the same time, however, lower variance avoids the danger of overfitting a model to historical data. Thus, low variance can be beneficial when *predicting* new data, in particular, in situations that involve much uncertainty. In statistical theory, this relationship is known as the bias–variance tradeoff (Hastie et al., 2001).

When forecasting presidential elections, for example, uncertainty occurs due to ambiguity about the component methods' relative accuracy, external shocks (e.g. campaign events), or the existence of noisy data. Calculating unweighted averages across forecasts is a simple way to account for such uncertainties or, in other words, to incorporate prior knowledge that prediction in the situation at hand is difficult.

Differential weights can be useful if there is strong prior knowledge about the methods' relative accuracy. For example, polls are well known to have little predictive value until shortly before the election (Erikson and Wlezien, 2012). Thus, it might be useful to assign lower weights to polls early in the campaign and then gradually increase their weight as the election comes closer, an approach that is becoming standard practice in models that combine structural (fundamental) data and updated polls over time. For a review of existing models see Lewis-Beck and Dassonneville (2015), who also incorporate this prior knowledge about the relative accuracy of polls over time to develop forecasting models for French, German, and UK elections.

While combining is most powerful when aggregating many forecasts that use a different method and different data, the approach can also be used if fewer methods are available. In a recent study, combining forecasts from three methods (a quantitative model, prediction markets, and polls) yielded accurate predictions of the 2012 US Electoral College and senatorial elections, a situation in which data are scarce (Rothschild, 2014).

Finally, the benefits of combining are of course not limited to US elections. In a recent validation test, the PollyVote was used to predict vote shares of seven parties in the 2013 German federal election by averaging forecasts within and across the four-component methods that were used in the original US PollyVote. Across the 58 days for which forecasts from all four components were available, the combined PollyVote forecast was more accurate than each component's typical forecast. Error reductions ranged from 5%, compared with a typical poll, to 41%, compared with a typical prediction market (Graefe, 2015).

Conclusion

Since 2004, the PollyVote has demonstrated the benefits of combining for forecasting the national vote in US presidential elections by averaging forecasts from four-component methods. Combining is a simple and powerful strategy to generate accurate forecasts. Combining forecasts from different component methods typically yields more accurate predictions than the average (i.e. randomly selected) component, and often outperforms even the best component. Adding forecasts that use a different method and different data to the combination can be expected to further improve accuracy. Given the results of the present research note, the PollyVote will add vote expectations as a fifth component for forecasting the 2016 US presidential election.

Declaration of conflicting interest

Funding

Notes

1. 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.
2. The PollyVote data are available at http://dx.doi.org/10.7910/ DVN/23184. The vote expectations data are available at http:// dx.doi.org/10.7910/DVN/VOTEEXPECTATIONSURVEYS.
3. The data and calculations for the present research note are available at: http://dx.doi.org/10.7910/DVN/27967.

Supplementary material

The replication files are available at: http://thedata.harvard.edu/dvn/ dv/agraefe/faces/study/StudyPage.xhtml?globalId=doi:10.7910/ DVN/27967

References

Armstrong JS (2001) Combining forecasts. In: Armstrong JS (ed) *Principles of Forecasting: A Handbook for Researchers and Practitioners*. New York: Springer, pp.417–439.

Clemen RT (1989) Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5: 559–583.

Erikson RS and Wlezien C (2012) *The Timeline of Presidential Elections: How Campaigns Do (And Do Not) Matter*. Chicago: University of Chicago Press.

Graefe A (2014) Accuracy of vote expectation surveys in forecasting elections. *Public Opinion Quarterly* 78: 204–232.

Graefe A (2015) German election forecasting: Comparing and combining methods for 2013. *German Politics (forthcoming)* http://ssrn.com/abstract=2540845.

Graefe A, Armstrong JS, Jones RJJ, et al. (2014a) Accuracy of combined forecasts for the 2012 Presidential Election: The PollyVote. *PS: Political Science & Politics* 47: 427–431.

Graefe A, Armstrong JS, Jones RJJ, et al. (2014b) Combining forecasts: An application to elections. *International Journal of Forecasting* 30: 43–54.

Graefe A, Küchenhoff H, Stierle V, et al. (2015) Limitations of Ensemble Bayesian Model Averaging for forecasting social science problems. *International Journal of Forecasting* DOI: 10.2139/ssrn.2266307 (forthcoming).

Hastie T, Tibshirani R and Friedman J (2001) *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. New York: Springer.

Hayes SP (1936) The predictive ability of voters. *The Journal of Social Psychology* 7: 183–191.

Lewis-Beck MS and Dassonneville R (2015) Forecasting elections in Europe: Synthetic models. *Research & Politics* Epub ahead of print January 2015. DOI: 10.1177/2053168014565128.

Lewis-Beck MS and Skalaban A (1989) Citizen forecasting: can voters see into the future? *British Journal of Political Science* 19: 146–153.

Lewis-Beck MS and Stegmaier M (2011) Citizen forecasting: Can UK voters see the future? *Electoral Studies* 30: 264–268.

Lewis-Beck MS and Tien C (1999) Voters as forecasters: A micromodel of election prediction. *International Journal of Forecasting* 15: 175–184.

Murr AE (2011) "Wisdom of crowds"? A decentralised election forecasting model that uses citizens' local expectations. *Electoral Studies* 30: 771–783.

Murr AE (2014) The wisdom of crowds: Applying Condorcet's jury theorem to forecasting U.S. Presidential Elections. *International Journal of Forecasting* (forthcoming).

Rothschild D (2014) Combining forecasts for elections: Accurate, relevant, and timely. *International Journal of Forecasting* http://dx.doi.org/10.1016/j.ijforecast.2014.1008.1006.