The present paper reviews the performance of different methods – as well as the PollyVote, a combined forecast based on these methods – in forecasting the popular vote in the 2016 U.S. presidential election.

The PollyVote

The PollyVote is an evidence-based formula designed to forecast election outcomes that relies on the principle of combining forecasts (Armstrong, 2001). Forecast error is reduced by combining forecasts within and across different component methods using unweighted averages. Following this rule, the PollyVote has predicted the popular vote in the three U.S. presidential elections from 2004 to 2012 by as much as a year in advance of Election Day. Updated twice a week in 2004 (Cuzán et al., 2005) and in subsequent elections at least once daily, at no time has the PollyVote called the election for any other than the winner. Moreover, on average across the six elections from 1992 to 2012, the PollyVote forecast has been more accurate than any of its component methods (Graefe et al. 2014a, 2014b).

In 2016, the PollyVote averaged forecasts within and across six different component methods: (1) trial heat polls, (2) prediction markets, (3) econometric models, (4) expert judgment, (5) index models, and (6) citizen forecasts. The first four methods listed above comprised the original specification of the PollyVote used in 2004 (Cuzán et al., 2005) and 2008 (Graefe et al., 2009). In 2012, index models were added to the formula (Graefe et al. 2014a), and citizen forecasts have been
included for the first time for forecasting the 2016 election. Each component method has been shown in research findings to be an appropriate and accurate election predictor.

Since its first launch last January of 2016, the combined PollyVote forecast has consistently – and correctly – predicted Hillary Clinton to win the popular vote. However, with a MAE of 1.9 percentage points across the last 100 days before the election, the forecast error was almost twice as large as in previous elections (cf. Figure 1).

**Figure 1. Forecast error by method**
(Mean absolute error, historical vs. 2016, across last 100 days before the election)

Polls

Trial heat polls—are most prevalent and highly visible in the news media coverage. The method asks respondents a variation of this question: "If the election for President were held today, for whom would you vote: Donald Trump, the Republican, or Hillary Clinton, the Democrat?"

The PollyVote relied on several poll aggregators, each of which collected and aggregated results of individual polls. To calculate its combined poll component, the PollyVote averaged the forecasts across the different poll aggregators.

Across the last 100 days, the mean absolute error (MAE) of combined vote-intention polls was 1.6 percentage points (cf. Figure 1). That is, national polls were considerably more accurate than in previous elections.¹

**Expectations-based methods**

An alternative to polls that obtain people’s vote intentions is to ask them how the election will turn out. Despite their long history, which goes back before the rise of scientific polling, these expectations-based approaches are often overlooked as methods to forecast elections.

¹ This is of course in stark contrast with the large polling error at the state-level, particularly in certain key states such as Michigan, Wisconsin, and Pennsylvania, which ended up deciding the election in the Electoral College.
**Expert judgment**

Asking experts to predict what is going to happen is one of the oldest forecasting methods available. When it comes to forecasting elections, experts can be expected to have knowledge and expertise about how to read and interpret polls. Most importantly, experts know that polls have errors. If they can assess in which direction the polling error is, and thus adjust their forecast accordingly, their judgment should improve upon the accuracy of polls. Prior evidence shows that this is in fact the case. Jones and Cuzán (2013) found that experts provided more accurate forecasts than polls or the IEM early in the election season, when the election is still nine months to a year or more in the future.

The PollyVote includes the judgment of prominent academics (and in 2004 some practitioners, as well) knowledgeable of American politics. In 2016, a panel of 15 political scientists was polled monthly, and the mean forecast was incorporated into the PollyVote.

In contrast to previous elections, the experts tended to adjust the polling numbers in the wrong direction in 2016. That is, they overestimated Clinton’s vote share relative to polls in five of the six surveys conducted since the parties’ conventions. In other words, the expert consensus was that the polls would even underestimate Clinton’s support. Thus, the forecast error of the combined expert forecast was half a percentage point higher than the combined polls (cf. Figure 1).

**Prediction markets**

Prediction markets are another expression of people’s expectations of who will win. Yet, instead of asking a representative sample for their opinion, prediction markets are open for anyone to participate. And, people reveal their opinion by betting money on the election outcome. And the betting quotes provide a forecast of what is going to happen. Depending on the accuracy of their individual predictions, participants can either win or lose money, and thus have an incentive to be right. Hence, people should only participate if they think they have information that improves the current market forecast.

Graefe (2016) reviewed prediction market accuracy for providing vote-share forecasts for elections in different countries. He found that prediction markets tend to outperform forecasts made by experts, as well as forecasts based on quantitative models and trial-heat polls, although compared to citizen forecasts the evidence was mixed.

Most available markets provide probability forecasts for the candidates’ likelihood to win. However, the PollyVote requires a national popular vote share prediction. We know of only one prediction market, the University of Iowa’s Iowa Electronic Market (IEM), which provides such information. One limitation of the IEM is a lack of efficiency. That is, the IEM is relatively low volume and participants are not allowed to invest more than $500. The PollyVote uses the IEM’s daily market prices but calculates one-week rolling averages to diminish short-term fluctuations.

As shown in Figure 1, the IEM vote-share market was the second most accurate among the PollyVote’s components across the six elections from 1992 to 2012. In 2016, however, the IEM
performed poorly and provided by far the worst forecasts of all components throughout the campaign by wildly overestimated Clinton’s vote share.

Prior research shows that participants in these markets tend to be well educated and belong to middle and upper income groups. It may have been the case that this certain demographic group allowed their wishes to get the better of their judgment.

**Citizen forecasts**

Vote expectation surveys—or citizen forecasts—are the newest addition to the PollyVote. Vote expectation surveys ask respondents who they expect to win the election, rather than asking people for whom they themselves intend to vote (Lewis-Beck and Skalaban 1989). A typical question might be: “Who do you think will win the U.S. presidential election, Donald Trump or Hillary Clinton?” The aggregate responses are then used to predict the election winner.

Though often overlooked, these citizen forecasts are highly accurate predictors of election outcomes (Graefe 2014). In 89% of 217 surveys administered between 1932 and 2012, a majority of respondents correctly predicted the winner. Regressing the incumbent share of the two-party vote on the percent of respondents who expect the incumbent party ticket to win accounts for two-thirds of the variance. Moreover, in the last 100 days of the previous seven presidential elections, vote expectations provided more accurate forecasts than vote intention polls, prediction markets, econometric models, and expert judgment. Compared to a typical poll, for example, vote expectations reduced the forecast error on average by about 50%.

In deriving a forecast using this component, we translate the results of vote expectation surveys into a two-party vote share prediction using the vote equation estimated by Graefe (2014). Then, we calculate the combined component forecast by exponential smoothing.

As in previous elections, a forecast derived from survey respondents’ answer to a simple question “Who will win?” was again the most accurate methods for predicting the 2016 popular vote. Across the final 100 days before the election, citizen forecasts missed on average by only 1.2 percentage points, which is the same as the average error across the six previous elections.

**Models**

The PollyVote combined forecasts from two types of models: index and econometric models.

**Econometric models**

For the past several presidential election cycles at least a dozen political scientists and economists have computed regression equations to forecast the election results (Campbell, 2013). Many of the models use economic data through the second quarter of the election year, the first official estimate of which becomes available in late July. Forecasts from those models are made shortly after that. There are exceptions, however. The predictions of some models are available well
before then, even two years ahead of the election, while at least one is delayed until the first polls after Labor Day are released (Campbell, 2013).

Most of these models are based on the idea of retrospective voting. That is, voters are expected to look back at how well the incumbent government has done its job, particularly in handling the economy. In addition, many models include some measure about how long the incumbent party has been in office to account for Americans’ desire for change, and some also include a measure of the president’s popularity.

On average, these models predicted a very close race, at least on average. The models’ average forecast missed the final election outcome by 1.5 percentage points, which makes them the second most accurate component method in this election. That said, the final forecasts from the 18 individual models differed by as much as 10-points, ranging from 44.0% to 53.9% of the two-party vote.

**Index models**

In contrast to econometric models, most index models are based on the idea of prospective voting. That is, they rely on the idea that voters evaluate the candidates and how they stand on the issues when deciding upon for whom to vote. Indexes are typically constructed based on ratings of specific characteristics of candidates or events. Ratings can be made by experts or members of the public (e.g., based on survey data) and can cover factors such as the candidates’ biographic information (Armstrong & Graefe, 2011), leadership skills, or issue-handling competences (Graefe, 2013), as well as exogenous effects, such as economic performance or the presence of a third party (Lichtman, 2005). Point forecasts of an election are provided by inserting current data into an equation specified by regressing the vote on the respective index scores.

On average, the five available index models overestimated Clinton’s support, particularly due to two models that were especially far off (the bio-index and the issue-index). Three of these models (e.g., the big-issue model, the Issues and Leaders model, and the Keys to the White House) were close the final election outcome.

**Discussion**

Prior research shows that the relative accuracy of different forecasting methods varies from one election to the next. We can see this again in 2016. Prediction markets, which were among the most accurate methods historically, were dramatically off, while econometric models, historically high in error, turned out to be more accurate this time. That is one of the reasons why combining forecasts usually works well. It’s extremely difficult to predict *ex ante* which method will end up being most accurate.

Combining works best when the various component forecasts bracket the true value (Graefe et al., 2014a). Compared to the previous six elections, there wasn’t much bracketing in 2016. Five of the combined forecasts overshot Clinton’s share of the vote while only one component, the econometric models, fell short. Thus, the PollyVote did not do as well as in previous elections and performed only
slightly better than the typical forecast. That is, it performed worse than econometric models, citizen forecasts, and polls but outperformed expert judgment, index models, and prediction markets.

But the principle of combining forecasts makes no claim that the combined forecast will always outperform its most accurate component, although that can happen, as was the case in 2004 (Cuzán et al., 2005) and 2012 (Graefe et al., 2014b). The claim that is made is that over time, as the component methods’ relative accuracy varies, the PollyVote will surpass its components. This is demonstrated in Figure 2, which displays the mean absolute error of all methods across all seven elections from 1992 to 2016. On average, the PollyVote continues to minimize error while avoiding making large errors.

Interestingly, citizen forecasts performed nearly as well as the PollyVote. So why not just use this one method in the future, you might ask? One major advantage of combining forecasts is that it’s often among the most accurate methods and, most importantly, it avoids large errors. There is not guarantee that citizen forecasts will perform as well in future elections.

**Conclusion**

At the PollyVote, we are currently reviewing what we can learn from this election. This process includes reviewing which forecasts to include, how to better combine them, and how to better communicate their surrounding uncertainty.

We do know from prior research that, first, the combined forecast will always be at least as accurate than the typical component forecasts, in any single election. So, combining forecasts prevents you from making large errors. We also know that, over time and across many elections, the combined forecast will be among the most accurate forecasts available, because the performance of individual forecasts varies wildly.
If you accept that it is extremely difficult to predict which forecast will turn out to be most accurate in a single election, there is no better way to forecasting than combining forecasts.

References


Campbell, James E. 2013 “Recap: Forecasting the 2012 Election,” PS: Political Science and Politics, 46, 1, 37.


