FEATURES

Accuracy of Combined Forecasts for the 2012 Presidential Election: The PollyVote

Andreas Graefe, *LMU Munich*

J. Scott Armstrong, *University of Pennsylvania and University of South Australia*

Randall J. Jones, Jr., *University of Central Oklahoma*

Alfred G. Cuzán, *University of West Florida*

ABSTRACT   We review the performance of the PollyVote, which combined forecasts from polls, prediction markets, experts' judgment, political economy models, and index models to predict the two-party popular vote in the 2012 US presidential election. Throughout the election year the PollyVote provided highly accurate forecasts, outperforming each of its component methods, as well as the forecasts from FiveThirtyEight.com. Gains in accuracy were particularly large early in the campaign, when uncertainty about the election outcome is typically high. The results confirm prior research showing that combining is one of the most effective approaches to generating accurate forecasts.

Ever since elections for office have been held, people have tried to predict their results. One of the oldest approaches to predicting election results is to rely on experts' expectations of who will win. Expert surveys and betting markets were regularly conducted in the 1800s (Erikson and Wlezien 2012; Kernell 2000) and are still popular today. In the 1930s, polls that asked respondents about their intention to vote arose as another alternative, and the combination of such polls has now become an increasingly common means for forecasting presidential election results. Finally, since the 1970s, scholars have developed statistical models to forecast the popular vote based data on fundamentals such as the state of the economy, the incumbent's popularity, and the length of time the incumbent and his party were in the White House. For more information about these models, see the articles in the special symposia published in *PS: Political Science & Politics* 37(4), 41(4), and 45(4), prior to the 2004, 2008, and 2012 presidential elections. For an overview of methods that are commonly used to predict the outcomes of presidential elections see Jones (2002, 2008).

In sum, a wealth of different methods use different information to achieve the same goal: predicting election outcomes. In most situations, it is difficult to determine *a priori* which method will provide the best forecast at a given time in an election cycle. Every election is held in a different context and has its idiosyncrasies. As a result, methods that worked well in the past might not work well in the future.

In such situations, an effective way to generate accurate forecasts is to combine the various available forecasts. Combining is beneficial because it allows for incorporating different information sets provided by the respective methods. As a result, the combined forecast includes more information. In addition, combining usually increases accuracy because the systematic and random errors of individual forecasts tend to cancel out in the aggregate, particularly if the individual forecasts draw on different information and are thus likely uncorrelated (Armstrong 2001).

Since 2004, we have tested the principle of combining forecasts for predicting US presidential election outcomes and have posted the forecasts at PollyVote.com. The PollyVote project is important for at least two reasons. First, the PollyVote demonstrates the usefulness of combining forecasts to a broad audience that follows the high-profile American presidential elections. This is significant because combining can be applied to practically all forecasting and decision-making problems. For example, in two other areas of political forecasting, combining has improved accuracy in predicting outbreaks of civil wars and court decisions (Montgomery et al. 2012). Second, the PollyVote tracks the performance of individual forecasting methods over time. This enables us to learn about the

Andreas Graefe is a research fellow in the Department of Communication Science and Media Research at LMU Munich, Germany. He can be reached at a.graefe@lmu.de.

J. Scott Armstrong is a professor of marketing at the University of Pennsylvania's Wharton School and adjunct researcher at the Ehrenberg-Bass Institute at the University of South Australia. He can be reached at armstrong@wharton.upenn.edu.

Randall J. Jones, Jr. is a professor of political science in the Department of Political Science at the University of Central Oklahoma. He can be reached at ranjones@uco.edu.

Alfred G. Cuzán is a professor of political science in the Department of Government at the University of West Florida. He can be reached at acuzan@uwf.edu.

© American Political Science Association, 2014

relative accuracy of election forecasting methods under different conditions, such as the length of time to Election Day or the specific electoral context.

This article recaps the performance of the PollyVote and its components in predicting the 2012 US presidential election.

METHOD

For forecasting the 2012 election, the PollyVote averaged forecasts of President Obama's share of the two-party popular vote within and across five component methods: trial-heat polls (as reported by polling aggregators), prediction market prices, expert judgment, econometric models, and index models. As of January 2011, forecasts were published daily at PollyVote.com and were updated whenever new data became available. All data and calculations are publicly available (Graefe 2013a).

Polls

In recent years aggregating, or combining, polls has become more common for US presidential elections. In 2004, rolling averages of polls were calculated specifically for the PollyVote. In 2008, we switched to external polling aggregators. For the 2012 election, figures were averaged from five polling aggregators, namely Election Projection, Pollster.com, Princeton Election Consortium, RealClear Politics.com, and Talking Points Memo. Data from RealClearPolitics.com were collected starting in January 2011. Data from the remaining four poll aggregators were added in September 2012.

Prediction Markets

Although already popular in the late 1800s (Erikson and Wlezien 2012), prediction markets have regained attention with the launch of the Internet-based *Iowa Electronic Markets (IEM)* by the University of Iowa in 1988. In contrast to well-known commercial markets such as betfair.com, at which participants can bet only on the election winner, IEM vote-share market participants can also wager on the candidates' vote shares, thereby making point forecasts. The IEM vote share market, therefore, provides the prediction market component of the PollyVote.

The IEM for the 2012 election was launched on July 1, 2011. As in the two previous elections, the IEM prices were combined by calculating one-week rolling averages of the last traded price on each day. This procedure was expected to protect against short-term manipulation and cascades because of herd behavior (Graefe et al. 2014).

Experts

As of December 2011, we conducted monthly surveys of 16 experts on American politics. Experts were asked to provide their best estimate of Obama's two-party vote share, along with a measure of their confidence in the estimate. On average across the 11 surveys, 14 experts participated.

Political Economy Models

We collected forecasts from 14 econometric models. Most of these were so-called political economy models. That is, they include at least one economic variable, along with one or more political variables. The idea underlying most of these models is that US presidential elections can, in part, be regarded as referenda on the incumbent's performance in handling the economy. As of January 2011, forecasts from three models were available; new or updated model forecasts were added as they were released. Forecasts from most of these models were published in the October 2012 issue of *PS: Political Science & Politics* (Campbell 2012).

Index Models

An important difference of the PollyVote 2012 compared to its earlier versions in 2004 and 2008 is the addition of index models as a fifth component. In comparison, earlier versions of the PollyVote combined all quantitative models within one component. The decision to treat index models separately was driven by the desire to create conditions that are most conducive to combining forecasts, which is when component forecasts contain different biases (Armstrong 2001; Graefe et al. 2014).

Index models use a different method and different information than econometric models. Thus, they were expected to contribute different bits of knowledge to the combined forecast, such as data from candidates' biographies (Armstrong and Graefe 2011) and candidates' issue-handling and leadership competence (Graefe 2013b; Graefe and Armstrong 2012, 2013).

RESULTS

With its first forecast released on January 1, 2011, almost two years prior to Election Day, the PollyVote predicted President Obama to win the popular vote. This forecast never changed. On Election Eve the PollyVote predicted Obama to gain 51.0% of the two-party vote and thus missed the final result by 0.9% percentage points. The corresponding figures in 2004 and

Figure 1

**Error Reduction of the PollyVote Compared to its Components for 2012**
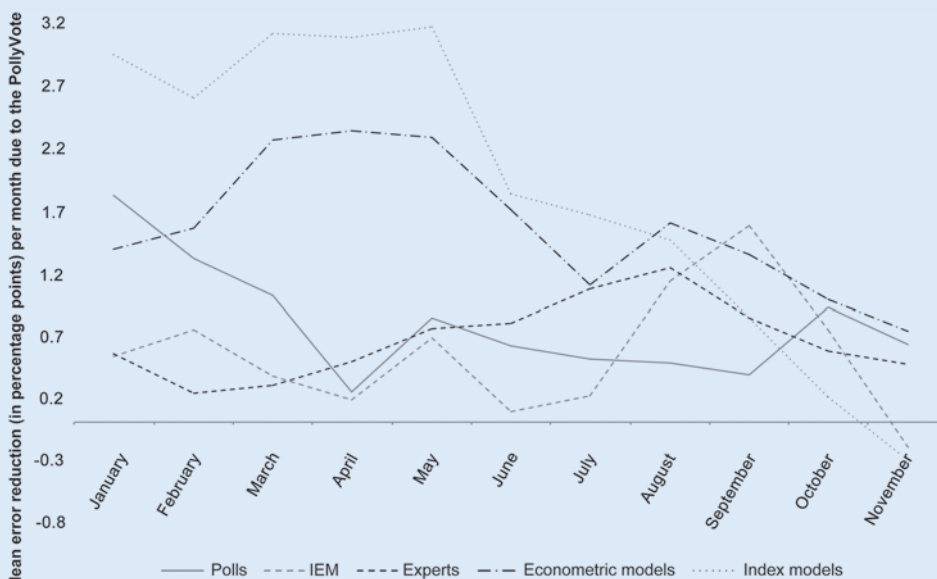
2008 were 0.3 and 0.7 percentage points, respectively. Thus, the mean absolute error for the PollyVote's final forecast across the past three elections was 0.6 percentage points. In comparison, the corresponding error of the final Gallup preelection polls was nearly three times higher, at 1.7 percentage points.

Forecasts published the day before the election are generally of limited value, however. The time for action has passed. Furthermore, in most cases, one will obtain quite accurate predictions by simply looking at the mean of the polls that were published in the week prior to Election Day. The more interesting question is how accurate forecasts are over longer time horizons. The PollyVote consistently predicted that President Obama would be reelected, and its forecasts remained stable even as other approaches, such as prediction markets or polls, at times pointed to a Republican victory. This is similar to the performance in the two previous elections, when the PollyVote also consistently predicted wins by George W. Bush (eight months in advance) and Barack Obama (14 months ahead). PollyVote, therefore now has a track record of more than 44 months of correct daily forecasts of the election winner across its three appearances.

Figure 1 shows the mean error reduction of the PollyVote compared to its five components for each month in 2012. Positive values above the x-axis mean that the PollyVote was more accurate than the particular component. Negative values mean that the component was more accurate. For example, in January, the PollyVote error was about 2.9 percentage points lower than the error of combined forecasts of the index models. In 10 of the 11 months, the PollyVote provided more accurate forecasts than

Figure 2

MAE of PollyVote and its Components across the Remaining Days to Election Day 2012
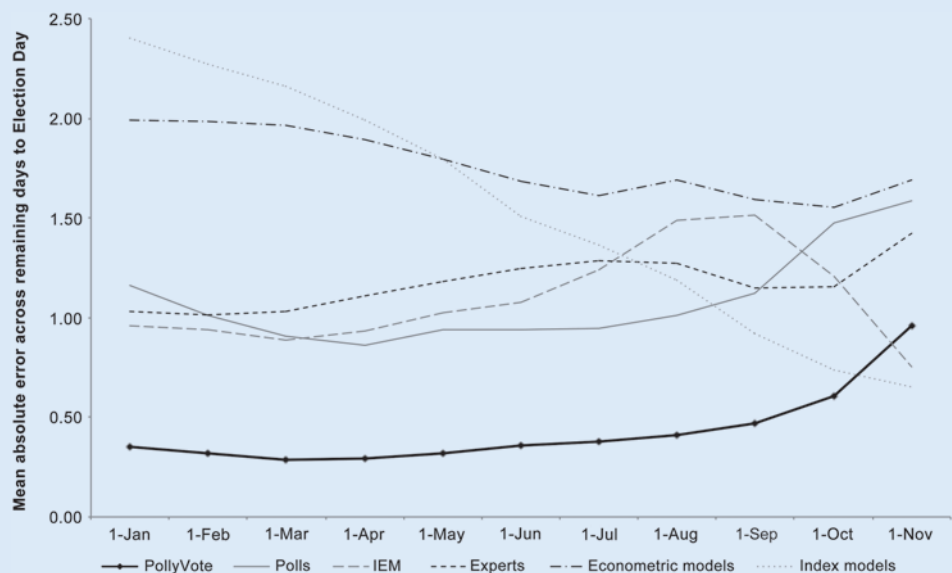
the average error of a given method for the remaining days until Election Day.

For example, from January 1, 2012 to Election Eve, the MAE of the PollyVote was 0.35 percentage points. That is, if one had relied on the PollyVote forecast on each single day in 2012, an average error of 0.35 percentage points would have resulted. In comparison, the respective errors for individual component methods were 0.96 for the IEM, 1.03 for experts, 1.16 for polls, 1.99 for econometric models, and 2.40 for index models. That is, the error of the PollyVote was 65% lower than the error of the IEM, which provided the most accurate forecasts among all components.

From October 1 to Election Eve, the MAE of the PollyVote was 0.60 percentage points, compared to 0.74 for index models, 1.15 for experts, 1.21 for the IEM, 1.48 for polls, and 1.55 for econometric models. The results demonstrate the high accuracy

> *PollyVote feeds off the work of others without taking anything away from them. In so doing, this simple technique of combining through averaging has emerged as one of the most effective approaches for generating greater accuracy in forecasting.*

any of its components. Often, the error reduction achieved through the PollyVote was greater than one percentage point. The only exceptions were five days in November, when the IEM and the index models slightly outperformed the PollyVote. In general, the relative performance of the individual methods varied across the election year.

Figure 2 presents the same data in a different way by showing the mean absolute errors (MAE) of the PollyVote and its components *for the remaining days in the forecast horizon*, calculated at the beginning of each month. Each data point in the chart shows

of the PollyVote, in particular for longer time horizons. Except for the last days prior to the election, when index models and the IEM provided the most accurate forecasts, the PollyVote was the best choice.

We also compared the PollyVote to forecasts from Nate Silver's popular *New York Times* blog FiveThirtyEight. Starting with June 1, which is the day after Silver published his first forecast, figure 3 shows the mean absolute errors of both approaches for the remaining days in the forecast horizon. Across the full 159-day period, the MAE of the PollyVote was 0.36 percentage points,

Figure 3

MAE of the PollyVote and FiveThirtyEight across the Remaining Days to Election Day 2012
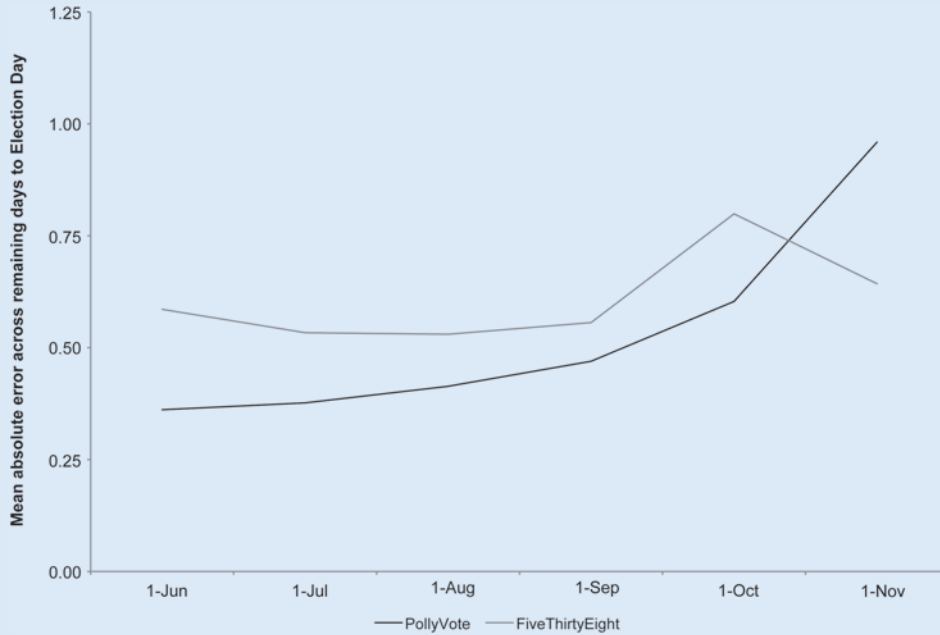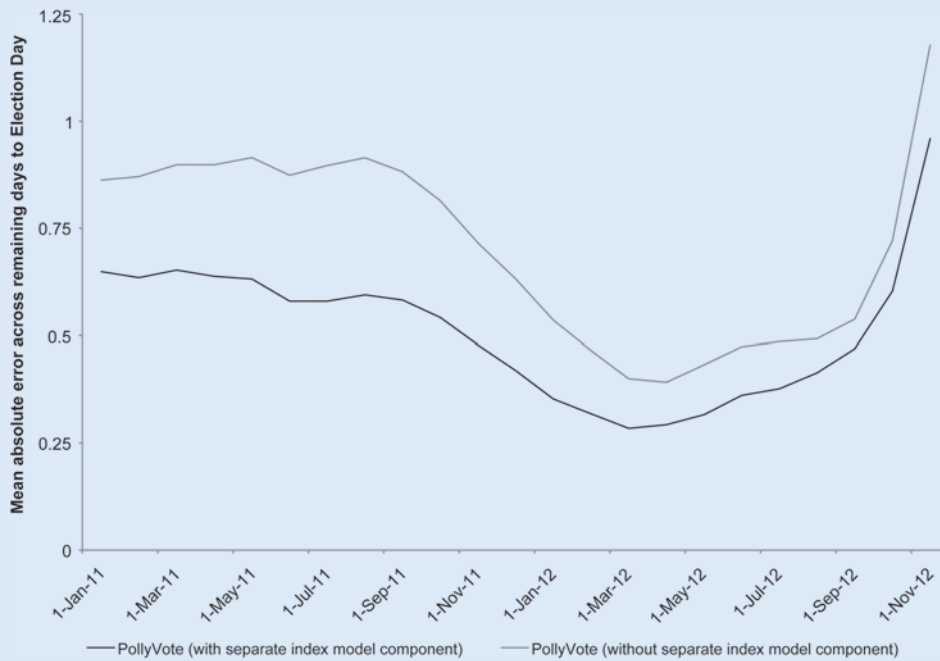

Figure 4

MAE of the PollyVote 2012 with and Without a Separate Index Model Component

DISCUSSION

The results add further evidence that combining is most effective (1) if multiple valid forecasts are available, (2) if the forecasts are based on different methods and data, and (3) if it is difficult to determine *ex ante* which forecast is most accurate (Graefe et al. 2014). This result conforms to what one would expect from the literature on combining forecasts. Combining is particularly valuable in situations that involve high uncertainty, which is usually the case with long time horizons. The PollyVote was designed to provide accurate long-term forecasts. For very short-term predictions, individual methods such as polls and prediction markets are usually accurate, as more information becomes known about how voters will decide. To increase the PollyVote's short-term accuracy it would be necessary to assign higher weights to these component methods. PollyVote will work on such an approach for the next appearance in 2016.

One important difference between PollyVote 2012 and its earlier versions was the addition of index models as a fifth component. One might think, by studying figures 1 and 2, that treating index models as a separate component was a misguided decision: the combined index models were among the least accurate components, particularly for long time horizons. However, less accurate components can still increase the accuracy of a combined forecast if they contribute unique information. Figure 4 shows the mean absolute errors of the PollyVote 2012 and a hypothetical "original" version of the PollyVote, in which the econometric and index models

compared to 0.59 for FiveThirtyEight. After October 1, the MAE of the PollyVote was 0.60 percentage points versus 0.80 for Five ThirtyEight, and so on. The results show that the PollyVote outperformed FiveThirtyEight for longer time horizons. However, FiveThirtyEight was more accurate shortly before Election Day.

are merged into one component. Again, each data point reflects the average error across the remaining days in the forecast horizon. The results show that the 2012 version of the PollyVote performed well. At all times, the five-component PollyVote had a lower error than what would have been achieved with a

four-component version. In addition, the 2012 version had a perfect daily record in predicting the popular vote winner (i.e., a hit rate of 100%). In comparison, the four-component version would have predicted the correct winner on 95% of the 675 days in the forecast horizon.

CONCLUDING REMARKS

In the past decade the accuracy problem in forecasting US presidential elections has largely been solved. For the last three elections, the combined PollyVote has provided highly accurate forecasts of the election outcome, starting months before Election Day, and has outperformed each individual component method. Of course, PollyVote is only as good as the underlying forecasts from various sources. All that the PollyVote does is to combine all available forecasts in the structured manner described. Thus, to borrow an analogy from biology, the relationship between the PollyVote and its components is a form of commensalism. PollyVote feeds off the work of others without taking anything away from them. In doing so, this simple technique of combining through averaging has emerged as one of the most effective approaches for generating greater accuracy in forecasting. ■

REFERENCES

Armstrong, J. Scott. 2001. "Combining Forecasts." In *Principles of Forecasting: A Handbook for Researchers and Practitioners,* ed. J. Scott Armstrong. 417–39. New York: Springer.

Armstrong, J. Scott, and Andreas Graefe. 2011. "Predicting Elections from Biographical Information about Candidates: A Test of the Index Method." *Journal of Business Research* 64(7): 699–706.

Campbell, James E. 2012. "Forecasting the 2012 American National Elections." *PS: Political Science and Politics* 45(4): 610–74.

Erikson, Robert S., and Christopher Wlezien. 2012. "Markets vs. Polls as Election Predictors: An Historical Assessment." *Electoral Studies* 31(3): 532–39.

Graefe, Andreas. 2013a. "Replication Data for: Accuracy of Combined Forecasts for the 2012 Presidential Election: The PollyVote," *Harvard Dataverse Network,* http://dx.doi.org/10.7910/DVN/POLLYVOTE2012.

Graefe, Andreas. 2013b. "Issue and Leader Voting in U.S. Presidential Elections." *Electoral Studies* 32(4): 644–57.

Graefe, Andreas, and J. Scott Armstrong. 2012. "Predicting Elections from the Most Important Issue: A Test of the Take-the-Best Heuristic." *Journal of Behavioral Decision Making* 25(1): 41–48.

Graefe, Andreas, and J. Scott Armstrong. 2013. "Forecasting Elections from Voters' Perceptions of Candidates' Ability to Handle Issues." *Journal of Behavioral Decision Making* 26(3): 295–303.

Graefe, Andreas, J. Scott Armstrong, Randall J. Jones Jr., and Alfred G. Cuzán. 2014. "Combining Forecasts: An Application to Elections." *International Journal of Forecasting* 30(1): 43–54.

Jones, Randall J. Jr. 2002. *Who Will Be in the White House? Predicting Presidential Elections.* New York: Longman.

Jones, Randall J. Jr. 2008. "The State of Presidential Election Forecasting: The 2004 Experience." *International Journal of Forecasting* 24(2): 310–21.

Kernell, Samuel. 2000. "Life before Polls: Ohio Politicians Predict the 1828 Presidential Vote." *PS: Political Science & Politics* 33(3): 569–74.

Montgomery, Jacob M., Florian M. Hollenbach, and Michael D. Ward. 2012. "Improving Predictions Using Ensemble Bayesian Model Averaging." *Political Analysis* 20(3): 271–91.