

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/237050409>

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

Article in Political Science and Politics · April 2014

DOI: 10.1017/S1049096514000341

CITATIONS

23

READS

1,978

4 authors, including:



**Andreas Graefe**

MHMK, Macromedia Hochschule für Medien und Kommunikation

101 PUBLICATIONS 1,555 CITATIONS

[SEE PROFILE](#)



**J. Scott Armstrong**

University of Pennsylvania

482 PUBLICATIONS 28,691 CITATIONS

[SEE PROFILE](#)



**Alfred G. Cuzán**

University of West Florida

91 PUBLICATIONS 443 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Forecasting Methods and Principles [View project](#)



Guidelnes for Science [View project](#)

**Accuracy of combined forecasts for the 2012 Presidential Elections:  
The PollyVote**

*Forthcoming (with changes) in PS: Political Science and Politics*

January 28<sup>th</sup>, 2013

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

J. Scott Armstrong  
Wharton School  
University of Pennsylvania, Philadelphia, PA, USA  
armstrong@wharton.upenn.edu

Randall J. Jones  
Department of Political Science  
University of Central Oklahoma, Edmond, OK, USA  
ranjones@uco.edu

Alfred G. Cuzan  
Department of Government  
University of West Florida, Pensacola, FL, USA  
acuzan@uwf.edu

We review the performance of the PollyVote, which combined forecasts from polls, prediction markets, experts' judgment, and quantitative models, for forecasting the two-party popular vote shares in the 2012 U.S. Presidential Elections. Throughout the election year, the PollyVote provided highly accurate forecasts, outperforming each of its components 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 high. The results confirm prior research showing that combining is one of the most effective approaches to generating accurate forecasts.

## **Accuracy of combined forecasts for the 2012 Presidential Elections: The PollyVote**

Since there have been elections, people have tried to predict their results. One of the oldest approaches is to rely on experts' expectations of who will win. Expert surveys and betting markets were regularly conducted already in the 1800s, and are still popular today (Erikson & Wlezien, 2013; Kernell, 2000). In the 1930s, polls that ask respondents about their intention to vote arose as another alternative and the combination of such polls is undoubtedly the most popular means for forecasting the election results these days (Hillygus, 2011). Finally, since the 1970s, scholars developed statistical models to predict the election outcome based on fundamental data such as the state of the economy, the incumbent's popularity, and the time the incumbent and his party were in the White House (Holbrook, 2010; Jones, 2002). In sum, there is a wealth of different methods that use different information to achieve the same goal: predicting election outcomes. In most situations, it is a priori difficult to determine which method will provide the best forecast at different times in an election cycle. Every election is held in a different context and has its own 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 forecasts (Armstrong 2001). Combining is beneficial because it allows for incorporating the different information sets provided by the various methods. As a result, the combined forecast includes more information. In addition, combining usually increases accuracy, as the idiosyncratic errors of individual forecasts tend to cancel out in the aggregate, in particular if the individual forecasts draw upon different information and are thus likely uncorrelated.

Since 2004, we tested the principle of combining forecasts for predicting U.S. presidential election outcomes and posted the forecasts at PollyVote.com (Graefe et al., 2012). The goal of the project has been twofold. First, the PollyVote uses the high-profile American presidential elections to demonstrate the usefulness of combining forecasts to a broad audience. This is important because combining can be applied to all forecasting and decision-making problems. For example, for political

forecasting, combining can improve the accuracy forecasts for 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 relative accuracy of political forecasting methods under different conditions, such as the time to Election Day or the specific electoral context.

The present study recaps the performance of the combined PollyVote and its components for predicting the 2012 U.S. presidential election.

## **Method**

In 2012, PollyVote averaged forecasts of President Obama's popular two-party vote shares across five component methods, each of which represents already combined forecasts: poll 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.

### ***Polls***

Aggregating and combining polls has become common for U.S. presidential elections. In 2004, the PollyVote calculated its own rolling average but switched to external polling aggregators in 2008. For the 2012 election, we averaged the results from five polling aggregators, namely Election Projection, Pollster.com, Princeton Election Consortium, RealClearPolitics.com, and Talking Points Memo. Data from RealClearPolitics.com were collected as of January 2011. Data from the remaining four poll aggregators were added as of September 2012.

### ***Prediction markets***

While already popular in the late 1800s (Erikson & Wlezien, 2013), prediction markets have regained attention with the launch of the Internet-based *Iowa Electronic Markets (IEM)* by the University of Iowa in 1988. Today, there are many popular real-money markets such as intrade.com and betfair.com. These markets draw a large number of self-selected "experts" and are thus much more liquid than the relatively small IEM. However, the IEM vote-share market remains the only market at which participants predict the two-party vote shares. Therefore, this market provides the prediction market component of the PollyVote.

The IEM for the 2012 election was launched on July 1<sup>st</sup> of 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 due to herd behavior (Graefe et al., 2012).

### ***Experts***

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

### ***Econometric models***

We collected forecasts from a total of fourteen econometric models. Most of these models were so-called political economy models. That is, they include at least one economic variable, along with one or more political variables. The basic idea underlying most of these models is that U.S. presidential elections can 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 became available. Forecasts from most of these models were published in *PS: Political Science & Politics* 45(4).

### ***Index models***

An important difference of the PollyVote 2012 compared to its earlier versions is the addition of index models as a fifth component. In comparison, the earlier versions of the PollyVote combined all quantitative models within one component. The decision to treat index models separately was driven by the idea to "create" conditions that are ideal for combining forecasts (Graefe et al., 2012). In particular, the goal was to generate a situation in which component forecasts are likely to contain different biases. Index models use a different method and different information than econometric models. Thus, they were expected to contribute different information to the combined forecast, such as the impact of candidates' biographies (Armstrong & Graefe, 2011) or their perceived issue-handling competence (Graefe & Armstrong, 2012a, 2012b).

## Results

With its first forecast released on January 1<sup>st</sup> of 2011, almost two years prior to Election Day, the PollyVote predicted President Obama to win the popular vote. And 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 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 is 0.6 percentage points. In comparison, the corresponding error of the Gallup pre-election poll is nearly three times higher, at 1.7 percentage points.

Forecasts published the day before the election are generally of limited value. 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 last 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 remained stable even as other approaches such as prediction markets or polls briefly pointed to a Republican victory. This is similar to the performance in the two previous elections, when the PollyVote consistently predicted George W. Bush (eight months in advance) and Barack Obama (fourteen months) to win. That is, the PollyVote 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, which already reflect combined forecasts, 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 error of PollyVote was about 3.2 percentage points lower than the error of the combined index model forecasts. In ten of the eleven months, the PollyVote provided more accurate forecasts than each of its components. Often, the error reductions were above one percentage point, in particular compared to econometric and index methods early in the year. 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 1: Error reduction of the PollyVote compared to its components for 2012**

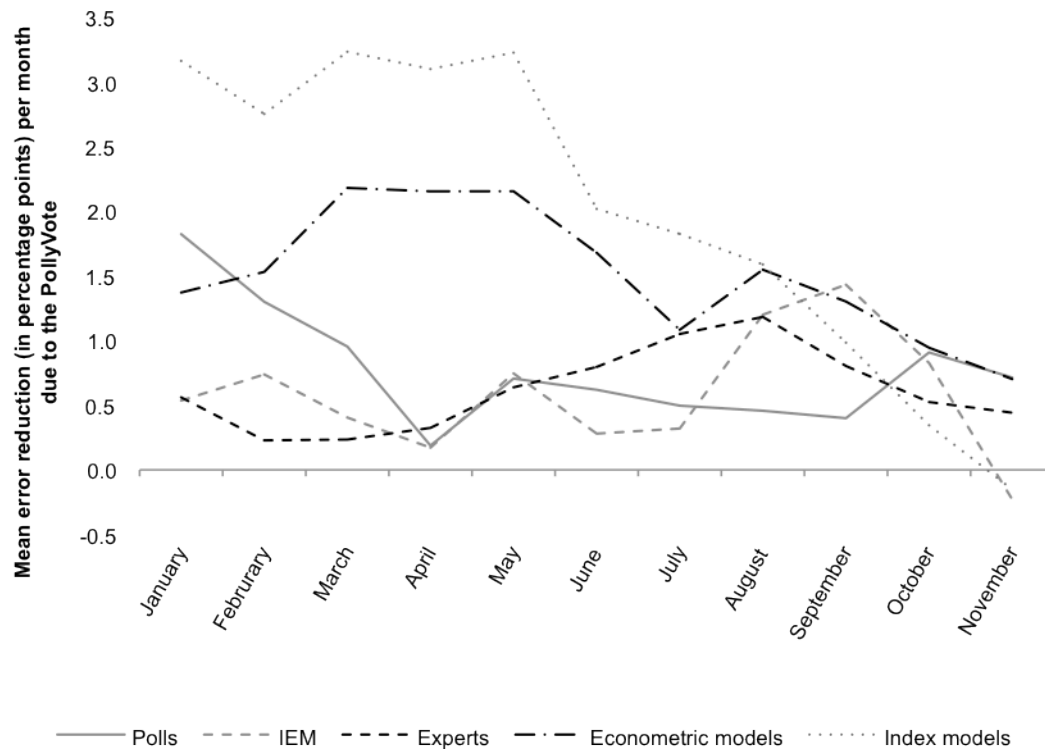
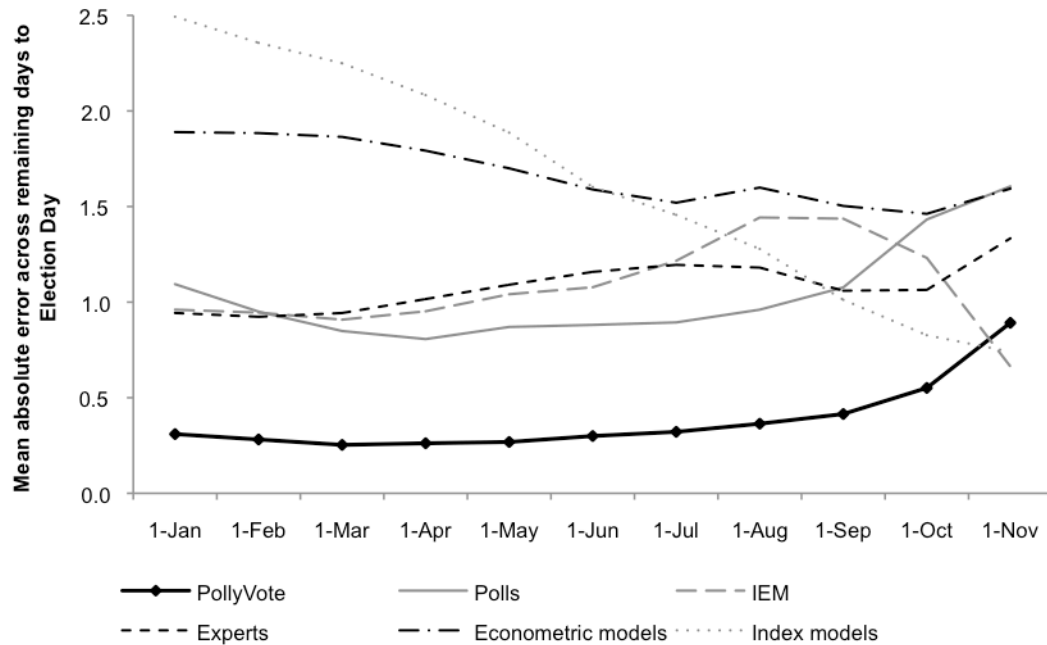


Figure 2 presents the data in a different way, by showing the mean absolute errors (MAE) of the PollyVote and its components *for the remaining days in the forecast horizon*, calculated at the beginning of each month. That is, each data point in the chart shows the average error that one would have achieved when relying on the method for the remaining days prior to Election Day. For example, from January 1<sup>st</sup> of 2012 to Election Eve, the MAE of the PollyVote was 0.31 percentage points. That is, if one had relied on the PollyVote forecast on each single day in 2012, one would have yielded an average error of 0.31 percentage points. In comparison, the respective errors were 0.94 for experts, 0.96 for the IEM, 1.09 for polls, 1.89 for econometric models, and 2.50 for index models. That is, the error of the PollyVote was only one third the error of the combined experts, which provided the most accurate forecasts of all components. From October 1<sup>st</sup> to Election Eve, the MAE of the PollyVote was 0.55 percentage points, compared to 1.06 for experts, 1.46 for econometric models, 1.43 for polls, 0.83 for index models, and 1.23 for the IEM. The results demonstrate the high accuracy 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.

**Figure 2: MAE of PollyVote and its components across the remaining days to Election Day**



In addition, we compared the PollyVote to Nate Silver’s popular *New York Times* blog FiveThirtyEight.com. Starting with June 1st, 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 0.30 percentage points, compared to 0.52 for FiveThirtyEight. After October 1<sup>st</sup>, the MAE of the PollyVote was 0.55 percentage points versus 0.76 for FiveThirtyEight, and so on. The results show that the PollyVote outperformed FiveThirtyEight for longer time horizons. In comparison, FiveThirtyEight was more accurate shortly before Election Day.

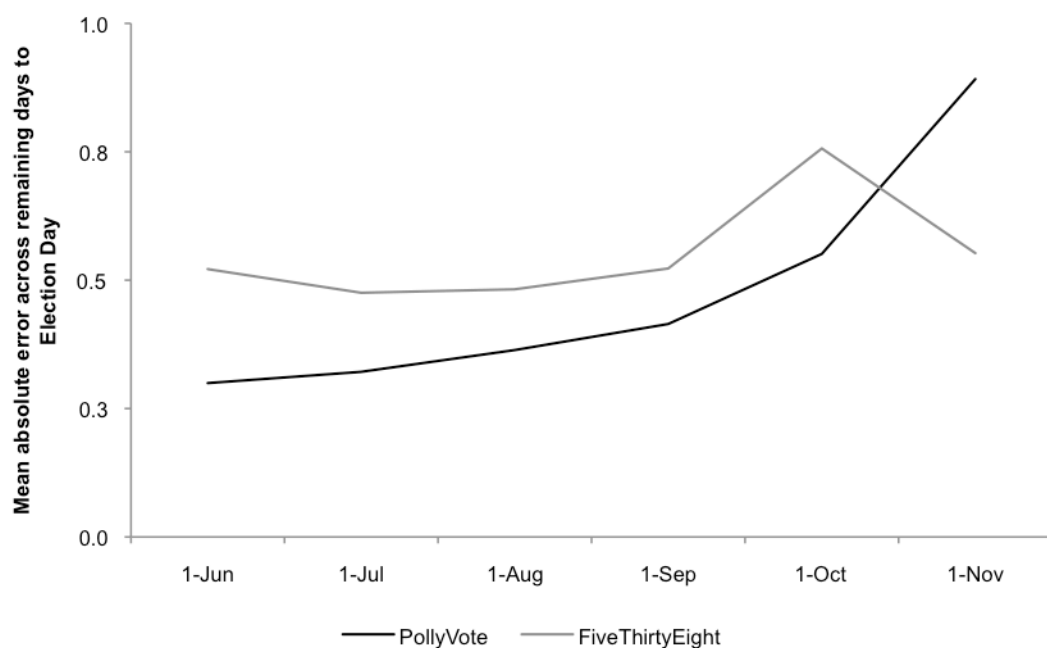
## Discussion

The results add further evidence that combining is most effective under ideal conditions. That is, if (i) multiple valid forecasts are available, (ii) the forecasts are based on different methods and data, and (iii) it is difficult to determine *ex ante* which forecast is most accurate (Graefe et al. 2012). This 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 tend to become very accurate, as much is known about how voters will decide. To adjust the PollyVote to further increase its short-term accuracy, it would thus be necessary to assign higher weights to such component methods. Polly will work on such an approach for the next appearance in 2016.

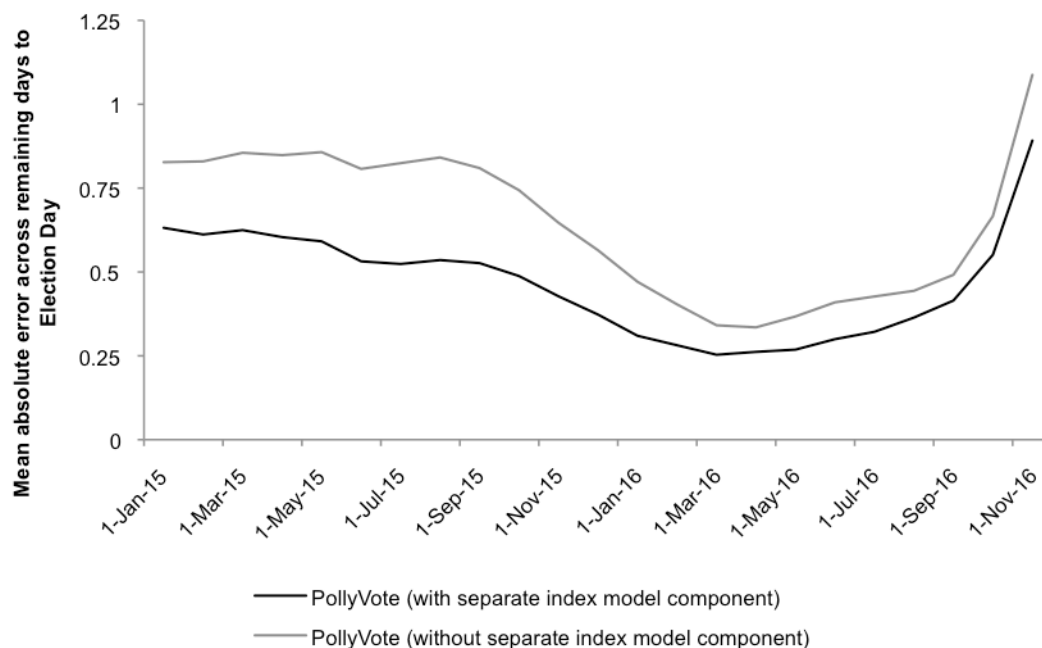
**Figure 3: MAE of the PollyVote and FiveThirtyEight across the remaining days to Election Day**



One important difference of the PollyVote 2012 compared to its earlier versions was the addition of index models as a fifth component. A look at Figures 1 and 2 might suggest that the special treatment of index method was a misguided decision, as 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 “old” version of the PollyVote, in which the econometric and index models 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 one would have achieved with a four component version. In addition, the 2012 version had a perfect 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.

**Figure 4: MAE of the PollyVote 2012 with and without a separate index model component**



## Concluding remarks

The accuracy problem in forecasting U.S. presidential elections has been solved in the past decade. For the past three elections, the combined PollyVote has provided highly accurate forecasts of the election outcome, starting months before Election Day, and outperformed each individual component method. Combining is one of the most effective approaches to generating accurate forecasts.

## References

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., & Graefe, A. (2011). Predicting elections from biographical information about candidates: A test of the index method. *Journal of Business Research*, 64, 699-706.

Erikson, R. S., & Wlezien, C. (2013). Markets vs. polls as election predictors: An historical assessment. *Electoral Studies*, 31, 532-539.

Graefe, A., & Armstrong, J. S. (2012). Forecasting elections from voters' perceptions of candidates' ability to handle issues. *Journal of Behavioral Decision Making* (forthcoming), doi: 10.1002/bdm.1764.

Graefe, A., & Armstrong, J. S. (2012). Predicting elections from the most important issue: A test of the take-the-best heuristic. *Journal of Behavioral Decision Making*, 25, 41-48.

Graefe, A., Armstrong, J. S., Jones Jr., R. J., & Cuzán, A. G. (2012). Combining forecasts: An application to elections. *2011 APSA Annual Meeting Paper*, Available at: [ssrn.com/abstract=1902850](http://ssrn.com/abstract=1902850).

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

Holbrook, T. M. (2010). Forecasting US presidential elections. In J. E. Leighley (Ed.), *The Oxford Handbook of American Elections and Political Behavior* (pp. 346-371). Oxford: Oxford University Press.

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

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

Montgomery, J. M., Hollenbach, F., & Ward, M. D. (2012). Improving predictions using ensemble Bayesian model averaging. *Political Analysis*, 20, 271-291.