

## Forecasting models using Facebook data can be more accurate at predicting election outcomes than polling.

*The aftermath of the last presidential election saw election forecasting models called into question when only 7 out of 12 national models predicted that Obama would remain in the White House.*

**Matthew C. MacWilliams** *proposes a new method of election prediction – using Facebook data in combination with electoral fundamentals. Applying this method to 16 of the 2014 Senate’s most close-run elections, the new model’s predictions were more accurate than five out of six other major forecasting models.*

Forecasting models designed to predict the outcome of elections in the United States suffer from a common malady – the effectiveness of campaigns or, as Lynn Vavreck noted in an [earlier post](#), the quality of a campaign is “rarely” taken into account by most calculations. Forecasting models typically relegate the billions of dollars and countless hours spent by campaigns communicating with voters to a model’s error term, presuming the effectiveness of different campaigns efforts are similar. The shortcomings of this approach were on full display in 2012 when only 7 of 12 national models predicted an Obama victory and post-election analyses of the Romney and Obama campaigns’ online efforts [documented](#) glaring disparities in their effectiveness.

Estimating campaign effectiveness is a tricky business. Quantifiable data that is transparent, replicable, and comparable across campaigns has eluded political scientists for decades. But the explosive adoption of social media and online communications by campaigns and the relentless tracking of online metrics by web platforms appear to offer a solution to election forecasting’s data dilemma.

For example, Facebook constantly tracks the growth of each congressional candidate’s fan base and the number of people engaging with candidates online. It has been doing so since 2006 when candidates launched their first campaign pages. As Facebook morphed into a ubiquitous social utility used by [seventy-one percent](#) of Americans and “popular across a diverse mix of demographic groups,” my colleague, Ed Erickson, and I speculated that Facebook data might very well provide a reliable measure of campaign effectiveness and enable the development of a dependable model for predicting the outcomes of individual congressional contests.

In the 2012 and 2014 election cycles, we set out to test our theory in the hurly-burly world of contested campaigns for US Senate. Using publicly-available Facebook data, leavened with simple electoral fundamentals similar to those that are the backbone of national forecasting models, we developed a simple model to forecast the percentage of the vote Republican and Democratic Senate candidates would win on Election Day weeks before voting began. The model’s accuracy is quite surprising.

In 2012, eight weeks before the election, our Facebook Forecasting Model predicted the winner in all seven contested elections studied. The model also produced more accurate predictions of the percentage each candidate would win in five out of the eight weeks studied than polling.

To compare our forecast with polling data, we collected the results of all polls in the seven Senate campaigns studied from September 8, 2012 to November 3, 2012. The results of these 212 polls were averaged and converted into two-party candidates’ totals to produce weekly, poll-of-polls estimates of the percent each candidate would win. The weekly predictions of the Facebook model and poll-of-polls averages for each race were then regressed against the actual percentage results in each election to compare the accuracy of the two approaches (Figure 1).

**Figure 1 – Facebook model predictions vs. Poll of Polls predictions for 2012 US Senate races**

In 2014, we predicted the results of the 16 most competitive Senate campaigns – the only major forecasting model to do so other than the Washington Post’s Election Lab. Eight weeks before the election our Facebook Forecasting Model accurately predicted the outcome of 13 of these 16 races. (During the 12 weeks leading up to the election, model predictions were published at [www.hashtageducracy](http://www.hashtageducracy) to insure transparency. They remain there today.)

By comparison, the Facebook Model accurately predicted the outcome of more Senate races than the New York Times and Washington Post models, the Cook Political and Roth & Gonzales Reports, and the University of Virginia’s Center for Politics “crystal ball” forecasts. Only FiveThirtyEight’s model predicted more Senate winners (one more) than the Facebook Model (Figure 2).

## Figure 2 – Facebook model vs. six leading public forecasts

While our Facebook Forecasting Model did a good job of predicting winners in a very competitive election year, the accuracy of the vote forecast for each candidate was much weaker in 2014 than 2012. Polling results showed a similar drop in accuracy. The root cause of this drop appears to be the electoral fundamentals in the model and not the Facebook data. In short, the incumbency variable in the model, based on the 2008 election in which there were strong Obama coattails exaggerated the strength of some Democratic incumbents and the weaknesses of some Republican incumbents.

Developing a transparent and parsimonious procedure to account for this phenomenon, while staying true to the theoretical tenets of the model, is the next practical challenge to resolve as we train the Facebook Model one more time on the 2016 elections for US Senate. The results of this third test of the model complete the six-year Senate cycle test we began in 2011. They will either add to the model’s impressive track record or raise questions about its long term viability. In either case, the forecasts will once again be published on [www.hashtageducracy.com](http://www.hashtageducracy.com) for all to see.

*This article is based on the paper, ‘Forecasting Congressional Elections Using Facebook Data’ in PS: Political Science and Politics.*

*Featured image credit: West McGowan (Flickr, CC-BY-NC-2.0)*

*Please read our comments policy before commenting.*

*Note: This article gives the views of the author, and not the position of USAPP – American Politics and Policy, nor the London School of Economics.*

*Shortened URL for this post: <http://bit.ly/1J4bYCg>*

---

## About the author

**Matthew C. MacWilliams** – *University of Massachusetts, Amherst*

Matthew C. MacWilliams is a PhD Candidate at the University of Massachusetts, Amherst. He published *The Polling Report*, wrote the Campaign Manager’s Manual for the Democratic National Committee, and is President of MacWilliams Sanders Erikson a political communications firm. His research interests including campaign communications, social media, political behavior, authoritarianism in American politics, constitutional law and judicial behavior, the Supreme Court, the political of climate change, interest group lobbying of the judiciary, and election forecasting.

- CC BY-NC 3.0 2015 LSE USAPP