The Upshot

THE 2016 RACE

To Know Who's Leading in the Voting, Just Ask Google

Justin Wolfers @JustinWolfers   MARCH 1, 2016

I'm always looking for new methods for predicting elections.

And in this election cycle, Google might be on to something.

It now releases sufficiently fine-grained data that it is possible to track in real time the number of searches for each candidate, zeroing in only on those searches that come from a primary state while the polls are open.

A simple count of the total number of searches has nailed the early Republican results with uncanny accuracy.

In New Hampshire, Donald Trump beat his fellow Republicans in both Google searches and actual votes. John Kasich surprised many with his strong second-place showing, but his popularity had already become evident in Google searches earlier on the day of the primary.

Interestingly, the Kasich search surge was evident only within New Hampshire. Marco Rubio's disappointment in New Hampshire paralleled lackluster search interest on Primary Day. So, too, Chris Christie, whose debate theatrics resulted in little interest either on computer screens or in the

voting booth. Perhaps Google's biggest miss was Jeb Bush — remember him? — who was ranked fifth among the Republicans on Google, but surprised many by edging ahead of Mr. Rubio to finish fourth.

The Republican South Carolina primary also yielded a very close match between Primary Day Googling and final vote tallies. Mr. Trump held a big lead in online interest, and he did even better in the final results. Mr. Rubio and Ted Cruz were left to duke it out for runner-up, and while Mr. Cruz held a slight lead in online searches, Mr. Rubio ended up getting slightly more votes. And Google correctly predicted that neither Mr. Bush nor Mr. Kasich would make much of an impression in South Carolina.

The glass-half-empty version of this story would emphasize that Google picked Mr. Cruz rather than Mr. Rubio to be the runner-up. The glass-half-full version notes that their figures got the broad contours of the race right, picking Mr. Trump to win with a 30-plus-point share, with Mr. Cruz and Mr. Rubio locked in a close race for second with around 20 percent each. It may be too much to ask for more than this.

The predictive power of these Google search data was brought to my attention by Seth Stephens-Davidowitz, an economist and former quantitative analyst at Google, who has spent much of his career crunching search data. (Mr. Stephens-Davidowitz is also a contributing opinion writer for The New York Times.)

Even Mr. Stephens-Davidowitz was surprised by the forecasting performance of these search data. When he first started trawling Google data before the 2012 election, he said that he "was really, really skeptical because there are so many reasons to search for a candidate."

But these early primaries have caused him to revisit his pessimism. It is not clear why online searches have proved so predictive, Mr. Stephens-Davidowitz said, but "my interpretation is that people just searched for their candidate either before or after they voted for them."

It's not a crazy idea. After all, who among us hasn't searched for reviews of a car, a stereo or a phone on the day of buying it? And if we do this when we're shopping, who's to say that people don't do the same on Election Day?

It's a bit harder to figure out how to use search data to predict a caucus. After all, caucuses are a bit like my university lectures: Once they start, you're meant to be deeply engaged by the speaker, rather than staring at your phone and calling up Google.

Even so, Nikhil Madadi, an election enthusiast, shared with me his data on Google searches made within Nevada over the four hours before the caucus. He reckons that, like my students, perhaps the voters were doing their reading ahead of class.

The big question leading into the Nevada race was whether the Trump campaign could turn out voters for a time-consuming caucus. Google search data suggested that Mr. Trump should be optimistic. They were right. Mr. Trump's win roughly paralleled his lead in online search activity. As in South Carolina, Mr. Cruz and Mr. Rubio were left to fight for second place, with each winning a bit more than a fifth of the vote. Again, Mr. Cruz was slightly more popular online, but Mr. Rubio edged ahead of him among voters.

I have been unable to find directly comparable data for the Iowa caucuses, although Simon Rogers, a data journalist at Google, tracked online searches within Iowa of the form "How to caucus for…" Mr. Cruz topped this measure, and he also surprised many analysts by outperforming his pre-election polls to win in Iowa.

But the news from Iowa was not all good for Google. Mr. Trump, who came in second, was not even in the top five candidates for this type of search. Given the large number of newcomers that Mr. Trump was drawing to the caucus, it's surprising that a larger number didn't ask Google how it's done. Mr. Trump ended up winning nearly a quarter of the votes, and so I count this incorrect forecast as a big miss.

Putting the evidence together, it looks as if online searches give a pretty useful first indication about how people are voting, although I would be more confident about primaries than caucuses, which involve a far smaller slice of the population. My sense is that a key part of the magic here comes from zeroing in on searches made within the local area and focusing only on Election Day. Indeed, past attempts at predicting congressional races from total search traffic were basically a bust.

The idea here is that search data yield what economists call a nowcast — that they're useful for "predicting the present." They're not pre-election forecasts like polls, but they're available in real time throughout Election Day, and you can check them before exit polls are released.

To a polling purist, there are dozens of reasons this shouldn't work: Internet users are not a representative slice of the electorate — they skew young, educated and professional. Those searching for information about the candidates are likely even less representative. And what sort of voter waits until the day of voting to bother doing research? Campaigns themselves probably also affect the extent to which their supporters need to search for more information. Finally, Google's data omits those who use Bing (more blue-collar, older) or Yahoo (yes, it still exists).

As Mr. Stephens-Davidowitz told me: "One of the reasons I'm still a bit skeptical of searches predicting votes is my own experience. I Google Trump every 25 minutes. But, I'm a Bernie Sanders supporter. Plus, I don't vote."

Mr. Stephens-Davidowitz's behavior might explain why Google's data isn't as predictive for the Democratic contest. After all, this is a two-horse race in which one of the candidates — Mr. Sanders — draws much of his enthusiastic support from younger and tech-savvy folks. By contrast, Hillary Clinton is popular with older voters, who spend less time online, but have very high turnout rates.

Indeed, Google's record so far in forecasting Democratic races has been more mixed. In both the Iowa and Nevada caucuses, there were more searches asking specifically how to caucus for Mrs. Clinton than how to caucus for Mr. Sanders, but in terms of broader search interest, there were more searches for Mr. Sanders. Mrs. Clinton won both caucuses narrowly; in Iowa it was a virtual tie. In New Hampshire, Mr. Sanders's 22-point win was larger than most pollsters had anticipated, but Google had actually predicted him to do five points better.

But South Carolina was a big miss. There were fewer Google searches for Mrs. Clinton than for Mr. Sanders in South Carolina while the polls were open, even though Mrs. Clinton walloped him among voters by nearly 50 points. It's a result consistent with Mr. Stephens-Davidowitz's observation that "people Google Sanders more than Clinton, consistently."

At this point, the sample of elections is sufficiently small that it is hard to know whether Google Trends is truly prescient, or just a bit lucky. I suspect a bit of each. But there is mounting evidence that suggests the search numbers are worth watching. There were more Greek searches for "no" than for "yes" during last year's failed Greek referendum on whether to accept the terms of an international bailout. Related search data also helped forecast David Cameron's surprise majority in the most recent British election, as well as Justin Trudeau's come-from-behind win in Canada.

Super Tuesday presents the next big test. I don't know what it's going to show. But I do know that rather than biting my nails waiting to hear exit poll rumors, I'm going to be checking Google Trends instead.

If I see anything interesting, I'll report back.

Justin Wolfers is a professor of economics and public policy at the University of Michigan. Follow him on Twitter at @justinwolfers

The Upshot provides news, analysis and graphics about politics, policy and

everyday life. Follow us on Facebook and Twitter. Sign up for our newsletter.

© 2016 The New York Times Company