

Xbox Polling and the Future of Election Prediction


by David Rothschild

September 30, 2014

For generations, pollsters have used probability polling (think of the Gallup polls quoted on the nightly news) as their go-to method to forecast the outcomes of elections. But cost increases and concerns about accuracy have called the method into question. A new form of polling called non-probability sampling — opt-in surveys on the internet, prediction markets, and even polls on gaming systems — has emerged as an improvement, and a viable replacement.

First, let's take a look at probability polling, which works like this: ask a random sample of likely voters who they would vote for if the election were held that day, and the answer is almost as accurate as asking everyone. This method has worked relatively well in countless election cycles, but it's growing more difficult to receive accurate results. One reason: the rise of cell phones. For a period in the 1980s nearly all likely voters owned a land-line; now the catalog of likely voters is spread over landlines and cell phones, or both, which makes it hard to figure out what the sample really is. In other words, where are all of the likely voters? The next problem is non-response error. Not all likely voters are equally likely to answer the poll, if contacted. This error is due to simple things like differences between demographics (e.g., some groups are more likely to answer calls from unknown numbers) and more complex things like household size. In other words, which likely voters are responding to polls, and do they differ from likely voters who don't?

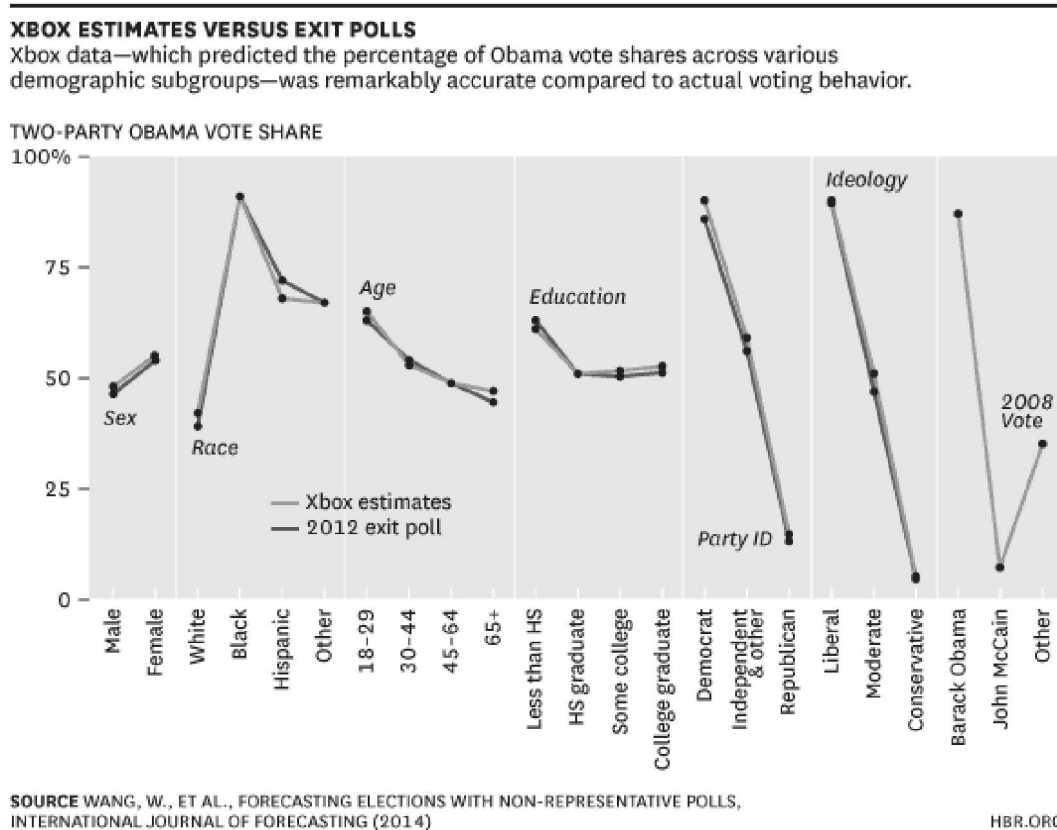
There are serious selection issues with non-probability samples as well — just like probability samples, they are prone to coverage and non-response errors — but the data is so much faster and cheaper to acquire. For example, in 2012, my colleagues and I collected opinions on the U.S. Presidential election from Xbox users by conducting a series of daily voter intention polls on the Xbox gaming platform. We pulled the sample from a non-representative group of users who had opted-in to our polls. In total, over 350,000 people answered 750,000 polls in 45 days, with 15,000 people responding each day and over 30,000 people responding 5 or more times. At a small fraction of the cost, we increased time granularity, quantity of response, and created a panel of repeated interactions.



But the raw data still needed to be turned into an accurate forecast. With our Xbox data, we first needed to create a model that incorporated the key variables of the respondents. We did this by determining the likelihood that a random person, from any given state, would poll for Obama or Romney on any given day, based on state, gender, age, education, race, party identification, ideology, and previous presidential vote. Then, we post-stratified all possible demographic combinations, thousands per state, over their percentage of the estimated voting

population; for transparency we used exit poll data from previous elections. Finally, we transformed the Xbox data into an expected vote share — by detailed demographics, and probability of victory, for all states.

In the below figure, you can see the accuracy of the Xbox pre-election estimates compared to exit-poll data.



As you can see, our forecasts were accurate — even compared with the best aggregations of the traditional polls — and they provided detailed demographic insight as well. Further, we were able to gain a new understanding of the movement (or lack thereof) of swing voters, because we had so many repeated users. The accurate forecasts, new relevant insights, and ability to easily update daily, all came at a much lower cost than traditional probability polling.

Yet there are meaningful groups of researchers that cling to the past, even as more papers confirm our findings. Their argument is that declining response rates don't affect results in a major way, so why worry and innovate?

Yes, it is possible that our Xbox polls would be slightly less accurate in other domains or with smaller samples. But within a few years, there's no doubt that traditional polls will lose their statistical power and become less accurate.

Xbox polling, and other forms of non-probability polling, will be an increasingly crucial tool for campaigns and advertisers in future elections. Campaigns have the capacity to target detailed demographic groups, or individuals, with messages specifically designed for them. And, because non-probability polling allows for continually updated forecasts for specific demographic groups, they can be even more efficient at targeting and delivering those messages.



David Rothschild is an economist at Microsoft and a fellow at the Applied Statistics Center at Columbia.



Read more on **Technology and analytics**