

November 2, 2021

Why do Opinion Polls make Wrong Predictions?

By: Atanu Biswas

Polling is based on sound statistical theory, but pollsters seldom divulge their polling methodologies for us to assess quality. There is enough room to question the study designs, possible selection biases, and randomness in selection of samples.

The inexact science of predicting elections has been in crisis during the last few decades. In India, opinion polls conducted to project the 2004 Lok Sabha were monumental blunders. Neither the Lok Sabha elections of 2009 and 2014 nor several state elections in recent times have given much comfort to pollsters: most surveys either failed to project the winner or to foresee the margin of victory. In the UK, the performance of surveys in predicting the Brexit referendum would haunt pollsters for decades to come. Predictions for some UK general elections in recent years have exhibited serious errors, as have those in the US for presidential elections in 2016 and 2020.

The science of polling is sound, based on solid statistical theories. It can be statistically shown that, in a homogeneous socio-economic-political condition, a sample size of 1,004 is enough to estimate the vote percentages with an error margin of 3%. Thus, all one needs to do is divide the country or a state into several such ‘homogeneous’ regions and sample roughly a thousand respondents from each region.

How much of the projections depends on the judgments of pollsters, on their choices of the models and methodologies?

But the execution of these polls is often problematic. If pollsters ask questions to the wrong group of people, they can get the wrong answers. Once the raw data is obtained from questionnaires, how are they converted into vote shares? How are they turned into seat numbers? And, most importantly, how much of the projections depends on the judgments of pollsters, on their choices of the models and methodologies? There is enough room to question the study designs, possible selection biases, and randomness, and of course the methodologies of many surveys on different socio-economic-political issues.

Flawed samples

To understand the ways how opinion surveys can go wrong, let us consider the disastrous failure of opinion polls ahead of the US presidential elections of 1936 and 1948, which are now part of lessons on how not to conduct surveys.

In 1936, the reputed magazine *Literary Digest* conducted a poll with a sample size of 2.4 million, a massive number in any standard. *Literary Digest* predicted that the Republican candidate Alfred Landon would get a 57% vote share and incumbent President Franklin Roosevelt, who was a Democrat, 43%. The actual result was worse than the opposite – Roosevelt managed 62% support, compared to 38% votes for Landon.

Why did an opinion poll with such a huge sample size fail so miserably? Retrospective analysis revealed that the main reason for the disaster was the way the samples were selected. In fact, a severe selection bias was introduced in the sample. *Literary Digest* had selected the individuals from telephone directories, club memberships, and magazine subscribers’ lists – each of which was an indicator of affluence in those days. Roosevelt had less support among such people. From a statistical perspective, the lesson was clear: samples should maintain the same proportions of different socio-economic parameters as in the population. There was also a severe non-response bias in the sample of *Literary Digest’s* survey, as only 24% of the 10 million individuals approached by the magazine responded. It is known that the response pattern of the non-responders could be markedly different from that of the responders.

We still do not know how much of these basic statistical rules are followed in practice in contemporary polling.

The lesson from Gallop’s opinion poll debacle in the 1948 US presidential elections was that in a survey, the sample should also be random in the true sense of the term. Each individual (within each stratum) should have the same probability of being included in the sample. Gallup used quota sampling, in which stratified samples were taken by maintaining the socio-economic proportions as in the population. But the samples were not ‘random’ within each stratum. Gallup’s opinion poll predicted 50% votes for Republican candidate

Thomas Dewey and 44% for Democrat Harry Truman. The election result was exactly the opposite: 50% for Truman and 45% for Dewey.

The pollsters had an opportunity there to learn the importance of ‘random’ samples. But we still do not know how much of these basic statistical rules are followed in practice in contemporary polling. The 2016 and 2020 US elections were also debacles for opinion polls. In 2016, supporters of the Republican candidate, Donald Trump, were under-counted in most opinion poll surveys. Pollsters failed to correctly translate popular vote predictions into the composition of the all-important US electoral college. They also ended up with a polling error of about 2.5% in most close-contested states and traditional Blue states. Pollsters still struggle to understand how they missed identifying white male voters without a college degree who could overwhelmingly support Trump. Later, in retrospective analyses, people attempted to label this as the “shy Trump factor” or “hidden Trump vote”, a term similar to the “shy Tory factor” that was widely used in the UK after the victory of the Conservative Party under John Major in the early 1990s.

In 2020, when Democrat Joe Biden defeated Donald Trump in the US presidential election, the overall prediction of Biden’s victory came true. But not many people bothered to look back to the predicted huge margins in nationwide popular vote shares in different popular opinion polls, which never materialised (The predictions were mostly between 8-12% while the actual margin was about 3%). In most of the US states, the difference between the predicted percentage of votes for Biden and Trump and Biden’s final lead was more than 3 percentage points. The vote margin went to Trump by a median of 2.6 additional percentage points in Democrat states, and 6.4 additional percentage points in Republican states, pointing to the possibility that the Hispanic support factor favouring Trump was missed in the surveys. This was certainly a huge failure to gauge the pulse of the electorate.

The black box of models

Poll predictions vary significantly depending on the underlying methodologies. For example, the predicted vote share may not just be the simple proportion of respondents favouring a party in the survey. Pollsters need to adjust their respondents statistically to match the demographic composition of adults in the census – more weight should be given to the underrepresented groups.

Prior to the 2016 US presidential elections, in partnership with Siena College, *The Upshot*, a section of The New York Times – surveyed 867 likely Florida voters, showing Hillary Clinton leading Donald Trump by a 1% margin. The raw data was shared with four well-respected pollsters, who were asked to predict the result. Three of these pollsters predicted a Clinton victory with 4%, 3%, and 1% margins, respectively, while the remaining pollster predicted a Trump victory by a 1% margin. A clear 5 percentage point difference between the five estimates was observed, even though all were based on the same data. [“Their answers illustrate just a few of the different ways that pollsters can handle the same data – and how those choices can affect the result.”](#) The pollsters made different decisions in adjusting the sample and identifying likely voters which resulted in four different electorates, and four different results.

In this example, the survey was not conducted by these pollsters. How much more would the estimates vary had the pollsters designed and conducted their own surveys? And what would have happened if, in addition, they had to estimate the number of seats in a first-past-the-post electoral system like in India?

|| Poll methodologies often remain undisclosed [...] It is almost impossible to comment on the quality of their predictions from a statistical point of view.

Predicting the number of seats various political parties will win is much more complicated than simply estimating their vote shares. The relationship between vote shares and number of seats is never a simple straight line even if there are only two contesting parties. One can try to find some thumb rule in a scenario of only two contesting parties. But the degree of complication gets amplified in a multi-party democracy like in India. Heterogeneity, induced by numerous factors such as several regional issues, regional parties, complex caste-religion-language divisions, a number of official and unofficial alliances, makes the overall prediction a daunting task. Errors are compounded in most cases, and it would be extremely difficult to get an idea of their magnitude.

Pollster should try to make the samples a representative cross-section of the population. But do they, in real life, properly follow the underlying statistical principles in designing, sampling, and analysing their data? It is not quite clear, as poll methodologies often remain undisclosed. Are their samples ‘random’? Do the samples represent the population by approximately maintaining the proportions across gender, age, income, religion, caste, and other important factors? It is almost impossible to comment on the quality of their predictions from a statistical point of view. All that we can say is that their predictions are wrong quite often.

Nate Silver, in his 2012 book, *The Signal and the Noise: The Art and Science of Prediction*, envisaged that the truth was ‘out there’, but it was becoming increasingly difficult to find. Although it is relatively easy to churn out data-driven forecasts, much of the information out there is simply noise “which distracts us from the truth.” Silver, who built an innovative system for predicting baseball performance and predicted the 2008 US election within a hair’s breadth, notes that “our bias is to think we are better at prediction than we really are.”

As early as 1954, Darrell Huff in his *How to lie with statistics* outlined the misuse and misinterpretation of statistics, and how these errors could create incorrect conclusions. Darrell Huff also pointed on the selection bias and lack of randomness in these polls. In Michael Wheeler’s *Lies, Damned Lies, and Statistics: The Manipulation of Public Opinion in America* (1978), leading pollsters like George Gallup and Louis Harris admitted that their record was marred by serious errors. Wheeler’s book showed that it was very much worse than they would have admitted.

Designing better surveys

Opinion poll predictions create huge uproar before every major election, almost everywhere in the world. The contradictory predictions and wrong results pose a serious crisis in terms of people’s faith in such surveys. Many people might end up sceptical of polls, especially when their opinion is in the ‘wrong’ direction. This could enhance the non-response error of surveys. In the US, response rates to telephone public opinion polls conducted by Pew Research Center have exhibited a steady decline to 6% in 2018 from 9% in 2016. At present, only 5% of people – or even less – respond to opinion polls. Thus, the poll predictions are bound to ignore the views of the remaining 95%.

Poll respondents, in general, score higher on social trust than the general population simply because they agree to respond to the surveys. Estimating the views of the total population based on such exceptional people might very well go the wrong direction.

These respondents might not be a representative sample at all. They are clearly different from the remaining 95% as they agreed to respond to the survey. In the American context, they are older and whiter, and more likely to be women. Poll respondents, in general, score higher on social trust than the general population simply because they agree to respond to the surveys. Estimating the views of the total population based on such exceptional people might very well go the wrong direction. (This is also true for surveys other than opinion polls.)

There is also no guarantee that some respondents are not lying. How can pollsters identify and eradicate them from the study? Or how can they deal with instances like ‘Lazy Labour’ voters, a class of electorates puzzled the pollsters in the 2015 UK general elections, who declared a Labour voting intention to pollsters but did not turn up to vote. In many cases, these are attempts to explain the errors in opinion polls in posterior analyses. Should not pollsters gauge the public mood while making their predictions? Are not ‘shyness’ or ‘laziness’ parts of the public mood which they should consider while making their predictions? Should not the respondents have been asked how likely they were to vote on the election day?

An opinion poll is a tool, not a principle. The tool can be modified to redress some of these problems. But, understandably, not much can be done about the undecided voters and non-responders. Circumventing non-response by ‘attributing’ answers to them is attempted sometimes, but that often depends on various unrealistic assumptions and the results are not always good. Do the pollsters use a ‘randomised response’ design that allows respondents to respond while maintaining confidentiality?

Nate Silver thinks that most predictions fail, often at great cost to society, because most of us have a poor understanding of probability and uncertainty. This is the “prediction paradox”. More humility about our ability to make predictions would make it more successful. Overall, opinion polls, in general, miss the proverbial pinch of salt. West Bengal or Tamil Nadu, Israel or Holland, US or UK – a sense of déjà vu persists.

References:

Huff, Darrel (1954). *How to lie with statistics*. W.W. Norton & Company.

Silver, Nate (2012). *The Signal and the Noise: The Art and Science of Prediction*, Penguin.