Forecasting Challenges in India: The Case of Indian Parliamentary Elections

Gurumurthy Kalyanaram

Indian national elections are here. India chooses members of parliament – *Lok Sabha*. And then the parliamentarians choose the Prime Minister. The process is similar to other parliamentary democracies.

The Indian elections will be conducted in the months of April and May, and the electoral outcome will be known about May 23rd. There are 543 parliamentary seats. A government must enjoy the majority support in the parliament.

There are many polls predicting the outcomes of elections in India. But I have discomfort in accepting polling results in India. For most part, it appears that they get the numbers wrong.

The error-prone forecasting of Indian elections extends to forecasting in other domains in India, including business and economic forecasting.

In general, forecasting is difficult and requires careful attention to data collection, surveys, modeling and statistical analyses. Probabilistic/Stochastic modeling is critical. This Scholarly article presents many of the challenges in forecasting. These challenges are even more amplified in the Indian context.

Here, as an illustration, we examine political forecasting. The same polling companies and pollsters – for most part – also provide business and economic forecasting in India.

The data – actual outcomes and predicted outcomes -- for the last five national elections is presented in Table 1.

Table 1

Election Year	Non-Congress Grouping NDA Actual (Predicted Average)	Absolute Error for NDA	Congress Grouping UPA Actual (Predicted Average)	Absolute Error for UPA	Total Absolute Error NDA+UPA
1998	252 (233)	19	166 (155)	11	30
1999	321 (296)	25	134 (142)	8	33
2004	189 (260)	71	222 (181)	41	112
2009	159 (195)	36	262 (201)	61	97
2014	336 (275)	61	58 (111)	53	114

NDA: National Democratic Alliance
UPA: United Progressive Alliance

The outcomes were the closest to mean projections for major alliances in 1998 and 1999 elections. Since then the forecasts have produced large errors. The mean absolute error has been about 75 parliamentary seats. Setting aside the 2014 national elections, which were a perfect storm in favor of non-Congress grouping, the mean absolute error is still about 66 seats.

Considering only the 1998, 1999 and 2009 elections when the pollsters got the direction of the results right (and setting aside 2014), the mean absolute error is 50 parliamentary seats.

Why is it that the Indian electoral polls don't perform that well? The answer is simple: It is a very challenging task -- devilishly difficult.

There are many challenges. Including intimidating size and diversity of the electorate, complexity of sample size, design, and weightings, correlated error structures, the bias of the pollster, and lack of transparent data.

Columnist Mihir Sharma, here, on the complexity and diversity of the electorate: "An election as complex as India's is fiendishly difficult to predict. India is the most diverse country in the world, and each state has its own hot-button issues; many of them have their own regional leaders and political parties. Voters will examine not just who might become prime minister but what the local dynamics are, as well as the caste and connections of whoever might represent them in Parliament. Some areas have multiple equally strong parties, and a tiny swing from one to the other might translate into big advantage in terms of parliamentary seats."

Surveys done in India and subsequent modeling and analyses quite often lack the statistical and probabilistic rigor to place confidence in the empirical results. For instance, Essay by Professor Rajeeva Karandikar of Chennai Mathematical Institute presents the fundamental element that is critical but is quite often missing in surveys (economic, business or political): "Randomness should be seen as a property of the process that selects the sample and not the sample itself. Failure to select a random sample can lead to wrong conclusions."

Surveys often mistake randomization an outcome not as a process. Sometimes, they also often confound arbitrariness and randomization.

The challenges with Indian polling are succinctly summarized in this Essay by Jonah Force Hill, a former Fellow of the Belfer Center. "An additional source of poll unreliability stems from the relationship between India's political parties and the polling agencies and media outlets. Indian news outlets – which ultimately sponsor the polls – tend to be ideologically slanted. Many have long-standing historical ties to political parties. …the U.S. has trade bodies such as the American Association for Public Opinion Research and the National Council on Public Polls that issue

guidelines on the dissemination of opinion poll results. And as a result, pollsters in the U.S. are usually forced to reveal their sample size, methodology and margin of error. No such institutional oversight exists in India."

The lack of transparency, and lack of professional oversight and requirements are serious impediments to good data and credible forecasting not only in the political domain but in all domains, including economic and social.

But, for a moment, let us assume all this away.

The methodology for political forecasting itself is inherently problematic. It is a two-stage process. The poll first measures a party's vote share in a State, and then tries to convert that vote-share into number of parliament seats in that State. Essay by Professor Rajeeva Karandikar of Chennai Mathematical Institute explains this: "…in order to predict the number of seats for parties, we need to estimate not only the percentage of votes for each party, but also the distribution of votes of each of the parties across constituencies."

Columnist Mihir Sharma alerts us that this approach is ridden with errors. "Pollsters might get the overall numbers right — the number of voters who prefer the BJP to the Congress, say. But translating that into a final tally of seats for the competing alliances is almost impossible. History bears out this unpredictability. Few outside Modi's own circle believed that he would win a majority in 2014. In 1999, the BJP won fewer seats — after a border skirmish with Pakistan — than predicted. In 2004, the BJP government was unexpectedly voted out. And in 2009, the Congress increase in seat strength startled pretty much every observer."

Let me illustrate this challenge. Here is a recent poll sponsored in part by TIMES NOW – a TV channel. This survey has been reported widely in Indian public media: https://is.gd/VhRPVf

Let us look at the State of Uttar Pradesh. The vote shares are roughly 8 percent for Indian National Congress, 43 percent for Bharatiya Janata Party (BJP) and 40 percent for a formidable regional coalition of BSP-SP-RLD. (Bahujan Samaj Party-Samajwadi Party-Rashtriya Lok Dal.)

Once the vote share preference is measured, then the vote share is converted into number of potential parliamentary seats: 2 for Congress, 42 for BJP, and 36 for the local combine.

This is an impossible task: converting vote shares to seats. Multiple candidates/parties and the fact that largest vote-getter (not majority) is the winner make it even more error-prone.

For instance, assume the standard error -- not reported in this Indian poll, and never reported by most Indian pollsters -- is 3 percent in the measurement of shares. That's a very reasonable assumption.[1] Now let us get back to Uttar Pradesh. To get a reasonably confident estimate (95 percent confidence), we need to consider at least 2 standard error measures. So, in Uttar Pradesh BJP's vote share may vary between 36 and 48 percent. If it is 36 percent, BJP will be wiped out -- zero seats. If it is 48 percent, BJP will win it all -- 80 seats. You see the problem.

So the estimates of number of seats can be dangerously incorrect for each one of the states. And then when it is aggregated for the entire country it can be devastatingly misleading.

It is precisely why United States pollsters don't even attempt this. This would be akin to a pollster in US measuring the preference for Democratic party and the Republican party in -- for instance -- the State of New York, and from that preference measure trying to forecast the exact number of House of Representative seats that the parties will win -- there are 27 of them in New York. Pollsters don't do this. There are just too many measurement and modeling errors.

What the US pollsters provide is a general guidance at the national level about the number of House of Representative seats that a party may win depending on the preferences on the generic ballot. For instance, if the Democratic party is preferred 7 percentage points more than the Republican party then the Democratic party may gain 25-30 net seats in the House. But there is no way to adopt this methodology to India because there are too many viable political parties, and the number and the identity of such viable parties vary from State to State.

[1] This is an example of lack clarity and transparency in reporting of the data and results.