2014 Lok Pre-Election Surveys

Center for the Advanced Study of India at the University of Pennsylvania
& The Lok Foundation

Updated: March 12, 2014

Methodological Details

The goal of this survey is to make nuanced social science claims about the Indian voter. As such, this survey was not designed to explicitly predict electoral outcomes. Nonetheless, given the volume of data in conjunction with sophisticated statistical tools, we believe we can provide reasonable predictions of the electorate down to the state level. Moreover, we believe one of the biggest virtues our approach is that we provide full transparency (including weighting) in vote predictions.

Predicting Elections

Pollsters typically selected electoral constituencies randomly, and then select voters off the voter list randomly. Furthermore, many pollsters go the extra step of asking voters to deliver their vote preference through a simulated "secret ballot." This is an optimal procedure, but voter surveys in India typically receive a significant amount of non-response and, more problematically, respondents often strategically answer the question rather than providing their true voter preferences.

The reality is that all polling data requires some re-weighting to account for this strategic response bias. As we will show, this weighting has serious consequences for the "effective sample size" in the survey and the margin of error. Pollsters in India have typically viewed this re-weighting as a trade secret and have not divulged its details. Unfortunately, weighting has serious consequences for both the explicit prediction and the confidence in the prediction. It is our view that full transparency in opinion polls can never be achieved unless the details of the weighting are divulged. We provide these details here.

The Larger Sample

In order to conduct the survey, we turned to Centre for Monitoring Indian Economy (CMIE). First, each state was broken into collections of similar districts, called homogeneous regions. The survey has oversampled urban India. Every city with at least 2 lakh individuals as of the 2001 census was sampled in the survey. For the states reported here, the largest city in the homogeneous region was always surveyed, but often more than one city was surveyed in a homogeneous region. In addition to this urban sample, a set of smaller "census towns" were sampled to augment the urban sample.

Once cities were selected, wards of the city were randomly selected according to stratification upon the average asset wealth of the ward. Within each ward, a random set of census enumeration blocks were selected, within which households were selected randomly. Villages were selected randomly within blocks stratified upon average asset wealth within each homogeneous region.

Weighting, Effective Sample Size, Margin of Error, and Elections Sample

Since this was not an electoral poll, the survey was not drawn to be random across voter lists. Furthermore, vote preferences were elicited through face-to-face interviews as opposed to secret ballot, which provided greater strategic response bias.

In order to generate weights for strategic response bias, we asked voters to state who they voted for in the previous Lok Sabha election. We then generated the strategic response weight (SRW) for each party in each state by taking the ratio of the actual vote share in 2009 to the probability of support observed in the sample for the party in 2009. In particular:

$$SRW = \frac{(\text{Actual \% support in state for party in 2009})}{(\text{\% support reported for party in 2009 in state sample})}$$

The logic of this weighting scheme follows from ratio estimation and the belief that response was given under "similar contexts." In our survey, the respondent answered an interviewer in back-to-back questions. More precise details of the logic of this weighting will be provided in a separate note.

The weights were applied at the coalition level (e.g., BJP and SAD in Punjab). At times, there were difficulties in applying this weight, e.g., when a coalition from 2009 disintegrated or when new large parties (e.g., YSR Congress in Andhra Pradesh) emerged. If we did not believe that the assumptions for weighting were valid for a party, we reported the data unadjusted. Finally, if we did not have confidence in the numbers due to irregularities in the survey, we did not report results; this is relevant to JD(U) in Bihar and INLD in Haryana. We also observed some irregularities in the survey for Congress in Orissa, but a close look at the data suggested little change in vote share from 2009. Thus, we simply imputed the 2009 vote share for Congress in Orissa as our prediction.

A second source of bias in the data is that respondents are not randomly sampled from the voter list. The existing survey sample, since it is created for different purposes, generates a biased sample of voters. The data over the 15 states, removing non-response, yielded a total sample of 51,130, of which 31,530, or 61%, were urban respondents, while India's population is actually about one-third urban. In order to address this problem, we provided weighted correction for the rural/urban composition at the level of the city, district, homogeneous region, and state using the 2011 Census. This yielded a unique weight, w_i, for each individual. The effective sample size ($Eff\ N$) corrects for increases in uncertainty due to variation in weights at the state level, for which we used an approximation provided by Kish (1965). For a given sample size N, the

effective sample size was calculated as:

$$Eff\ N = \frac{\left(\sum_{i=1}^{N} w_i\right)^2}{\sum_{i=1}^{N} w_i^2}$$

The margin of error (MoE) is approximately a two standard deviation bound (95% confidence interval) on the provided estimate. Since predictions and SRW values differ by party, we had to construct an MoE value for each party in each state. Let p be the the predicted vote prediction for a party in a state. Using standard calculations, we define MoE as:

$$MoE = 1.96 * SRW * \sqrt{\frac{p(1-p)}{Eff\ N}}$$

The resulting calculations are shown below. Like any estimation strategy, this one will have its pitfalls, but based on our knowledge of Indian politics, we have reasonable confidence in the estimates provided.

For more information about the project visit:

CASI UPenn

India in Transition Blog