Response to Critical Comments on
Electoral Irregularities in the World's Largest Democracy

Sabyasachi Das

October, 2026

Various researchers made critical observations about the analysis and arguments made in the paper when it first came out. I had prepared a response document at that time responding to each one of those claims, but chose not to release it because of various reasons. Here is an updated version of the document.

Comment 1: The data discrepancy in ECI's EVM turnout numbers reflects revision of its interim numbers to final vote counts arising out of late surge in turnout, i.e., surge in turnout towards the end of the election day. Higher data discrepancy in closely contested elections that BJP won can be due to greater late surge in BJP's favor in those constituencies.

Response: The description of the data revision as a simple change from interim to final figures is erroneous. The 2019 general election was conducted over 7 phases, spanning over a month. The discrepancy examined in the paper is for the first four phases of elections (373 PCs) for which ECI had already released the final numbers of EVM votes polled. For the last three phases it had released 'estimated' turnout numbers by that time. One can reasonably claim the 'estimated' numbers to be "interim". However, I do not examine those constituencies in my analysis of turnout manipulation. Moreover, my analysis shows that, in the PCs that BJP won, assignment of State Civil Service (SCS) officers from BJP ruled states as counting observers positively predicts large mismatches in turnout data; in PCs that BJP lost, no such pattern exists. This is consistent with manipulation, but inconsistent with the "late surge" hypothesis, since counting observers' responsibility begins after voting is completed.

Comment 2: Last minute campaigning are not captured in the data on campaigning since they are ignored by exit polls that occur early.

Response: There are two related errors in the claim. I use CSDS National Election Survey for my analysis on campaigning. NES is not an exit poll, but a post-poll voter survey carried out in a randomly selected sample of PCs. In 2019, it was carried out in 208 PCs. The survey randomly selects names of voters from electoral rolls and visits the respondents at their homes to carry out the survey, each lasting about 40-50 minutes.[1] This is in stark contrast to exit polls, which are typically very short, and conducted on the day of the election with voters who have cast their votes, immediately after they come out of polling stations. The purpose of exit polls is to predict vote shares of major parties and winner in each PC. The purpose of NES is to gain a deeper understanding of voter behavior, knowledge, and attitude across all voters, irrespective of whether they voted or not. Importantly, the NES is conducted *after* the election is over in the PC (and before the declaration of results). Hence, it *will* capture home visits by party workers, irrespective of whether the campaigning was close to the election day or otherwise.

[1] See the survey documentation: https://www.lokniti.org/media/PDF-upload/1565073104_34386100_method_pdf_file.pdf.

Comment 3: Registration manipulation requires precise ex-ante prediction about results, and therefore, is an unlikely mechanism.

Response: For any manipulation (registration or otherwise) or campaigning to be effective in winning disproportionate share of close elections, the party needs to have a good idea about the PCs where it expects a close election and it might be lagging behind. I explain this point in the paper when I discuss precise control. Not all PCs are equally amenable to registration manipulation and it is costly for the ruling party to engage in such activity, due to heightened media attention.[2] Hence, it is a noisy instrument. The PCs where the party managed to delete sufficient names, selected themselves to the right of the win margin threshold, resulting in lower electorate growth rates for them.
This leads to the next criticism.

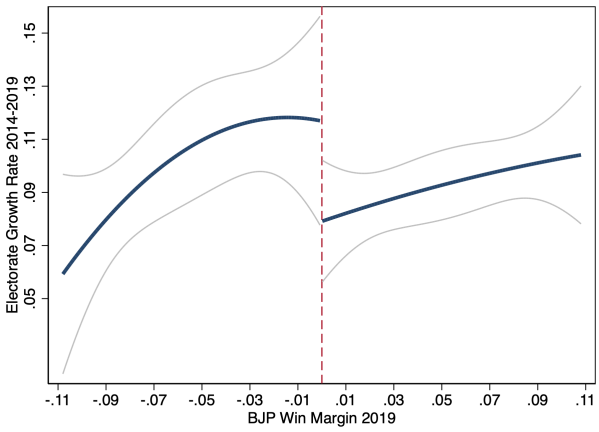


Figure 1—Electorate Growth Rate: RDD Estimation

Comment 4: The differentially lower electorate growth rate in the PCs that BJP barely won is not because of low growth rate in the PCs that BJP won, but because of high growth rate in PCs that BJP lost.

Response: Unfortunately, the claim does not hold ground in close scrutiny. The average growth rate in electorate in the close election sample (PCs with absolute win margin of 0.1 or less, as per the optimal bandwidth) is 9.6%. The estimated growth rate to the right of threshold is 7.5%, and that to the left of threshold is 12.5%.[3] The sample mean, therefore, is approximately in the middle of the two numbers. Hence, while the left estimate is higher than average, the right estimate is also lower than the average. Figure 1 above shows this clearly. As can be observed, the sample mean of 9.6% lies in the middle of the left and right estimated values. Moreover, on the both sides of the cut-off, the relationship is positively sloped. The RDD estimate, therefore, appears more as a shift in the intercept.

The RD estimate of electorate growth rate for previous years show no such discontinuity, except in 1991, where the discontinuity is in the opposite direction (Table 1).[4] Additionally, Figure 2 shows the RD estimates for high and low Muslim share PCs over a range of bandwidths. The RD estimate is

[2] Several media outlets did report on targeted deletion of voter names from electoral rolls, as I point out in the paper.
[3] These estimates come out of the RD estimation exercise, and stored as output of the STATA command `rdrobust`.
[4] Electorate growth rate for 2009 can not be computed because of delimitation in 2008.

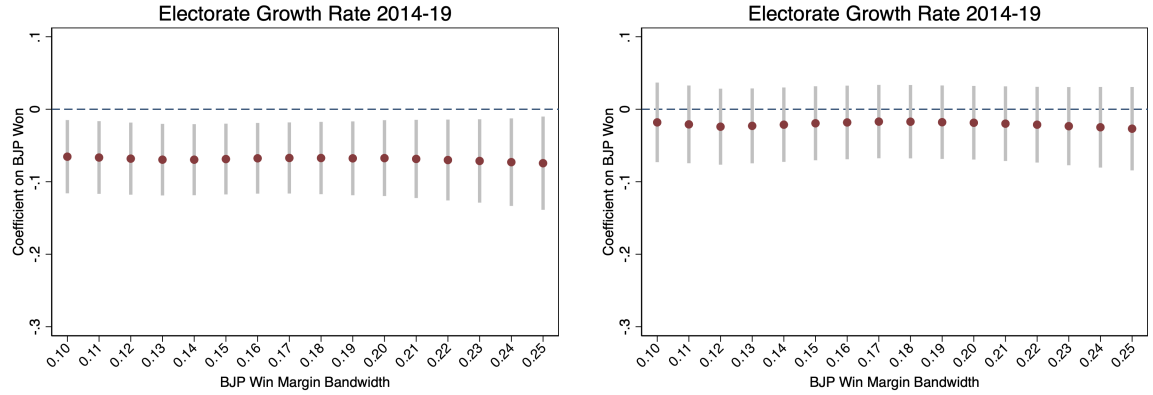
negative and statistically significant for high Muslim share PCs across all bandwidths, while it is always insignificant for low Muslim share PCs.

Table 1—Electorate Growth Rate Not Different in PCs Barely Won by BJP in Previous Years Except 1991

	Electorate Growth Rate (G_p) 2009-2014					
	2014	2004	1999	1998	1996	1991
	(1)	(2)	(3)	(4)	(5)	(6)
BJP Won	-0.00	-0.01	0.00	0.00	-0.04	0.02**
	(0.03)	(0.03)	(0.01)	(0.01)	(0.04)	(0.01)
Mean Dep. Var.	0.16	0.12	0.02	0.02	0.17	0.02
Observations	183	209	247	253	213	191
Polynomial order	1	1	1	1	1	1
Bandwidth (h^*)	0.16	0.16	0.16	0.16	0.16	0.16

Notes: The data is at Parliamentary Constituency (PC) level. The table reports the regression discontinuity design estimate using BJP win margin as the running variable. The dependent variable in all the columns is the growth rate in the PC electorate between the year stated at the top of the column and the previous election year. BJP Won is a dummy indicating whether BJP won the PC in the year mentioned at the top. Robust standard errors are are reported in the parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 2—RDD Estimates for Electorate Growth Rate (2014-2019)



(a) High Muslim Share PCs

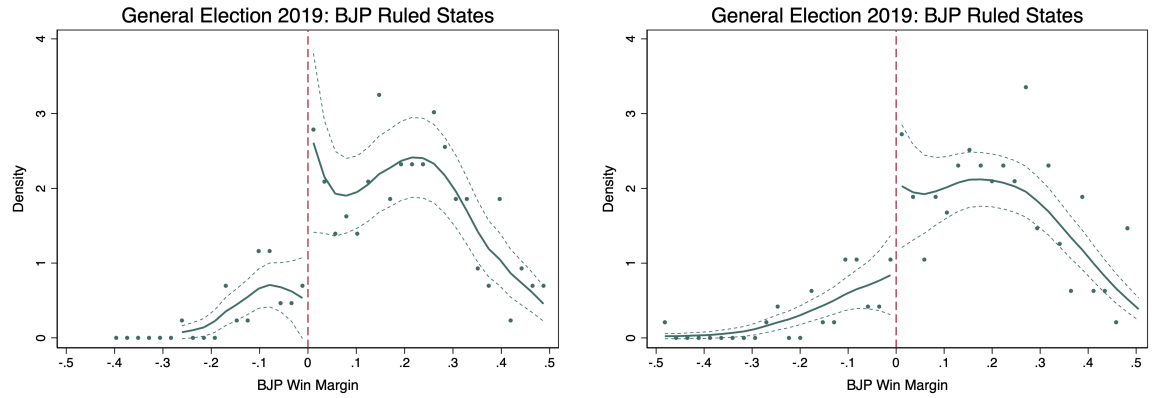
(b) Low Muslim Share PCs

Comment 5: The BJP-ruled states do not include some of the Union Territories.

Response: There are only 6 PCs in/as Union Territories in the entire country. Their classification has no bearing on the results I report in the paper. Most of the action comes from large states that were controlled by BJP in 2019, as one would expect. Figure 3 below shows density discontinuity in BJP win margin for BJP ruled states with and without UTs.

Comment 6: Appendix Table A1 shows that BJP also won disproportionate number of close elections in 2004.

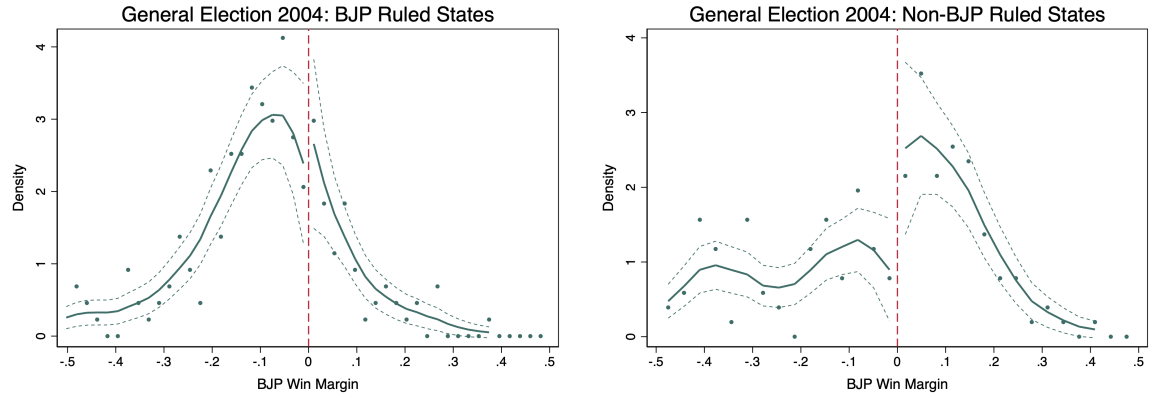
Figure 3—BJP Win Margin Density in BJP Ruled States – With and Without UTs



(a) BJP Ruled States without UT **(b)** BJP Ruled States with UT

Response: As I show in the paper, the density discontinuity in 2019 is concentrated in BJP ruled states (see Figure 3 above). The density for the 2004 BJP win margin in the same states does not exhibit discontinuity (Figure 4a). It appears to be concentrated in the non-BJP ruled states of 2019 (Figure 4b). The two states that drive this discontinuity are Karnataka and Maharashtra, where BJP won 17 out of 22 PCs with its win margin less than 0.05. Both the states were also non-BJP ruled during 2004. Therefore, the pattern of discontinuity observed in 2019 is different from the one in 2004.

Figure 4—McCrary Test for General Elections in 2004



(a) 2004 Election: BJP Ruled States (2019) **(b)** 2004 Election: Non-BJP Ruled States (2019)

Comment 7: The density discontinuity in 2019 is not large, compared to the estimates for the previous election years. Therefore, we cannot say that 2019 exhibits abnormal pattern.

Response: This needs a bit longer discussion. To understand whether the discontinuity estimated in 2019 was particularly stark compared to other years is not straight forward. The estimates of absolute density discontinuities are directly not comparable across elections. This is because the same density discontinuity is more likely to arise if there are more close elections in a year. One way to see this is to think about density discontinuity at zero arising out of lop sided wins among the set of closely contested elections. If there are lots of close elections then some lop sided wins can happen simply due to chance,

even if the data generating process is not lop-sided. Therefore, how large a density discontinuity estimate is depends on the density height at zero. Consequently, a higher discontinuity estimate in a given year may not imply a *larger* discontinuity.

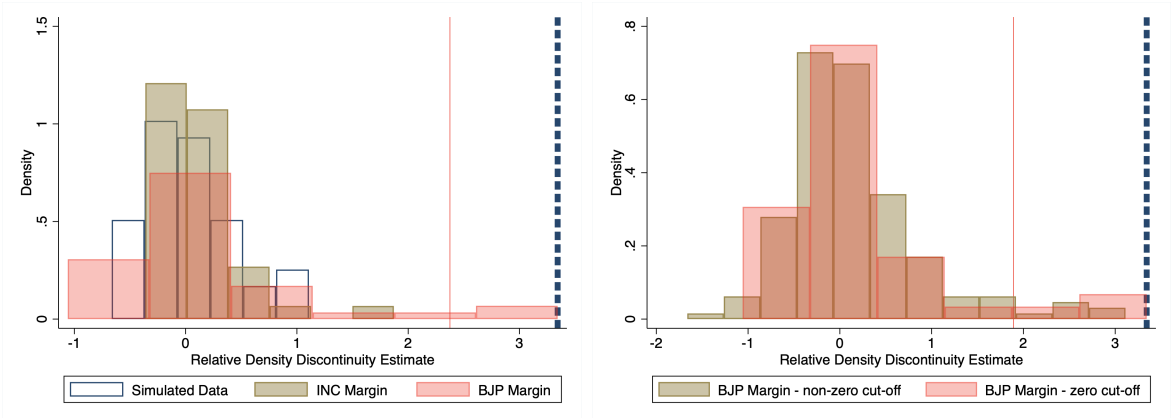
In order to make the estimates comparable, I compute for each year the *relative density discontinuity*, defined as the ratio of the density discontinuity and the left limit of the density. Denoting density discontinuity by $\Delta f(0) \equiv f_+(0) - f_-(0)$, the relative density discontinuity is defined as $\rho(0) \equiv \frac{\Delta f(0)}{f_-(0)}$. This can be interpreted as the rate of change in the right limit of the density at zero relative to its left limit. This is similar to the estimator proposed by McCrary (2008), which estimates the difference in log density heights at zero, i.e, the estimator is $\theta = (\log f_+(0) - \log f_-(0))$. Additionally, the difference between the fractions of close election constituencies that fall on the left and right of the zero margin threshold, i.e., the extent of lop-sided wins within close races follows a distribution that depends on $\rho(0)$.

To see this, notice that conditional on N close elections (defined as win margins falling in some small neighborhood $(-\epsilon, \epsilon)$ around zero), the fraction of them falling on the left side follows a distribution that depends on $\rho(0)$. To see this, notice that the conditional probability that any of the N constituencies will fall in the range $(-\epsilon, 0)$ is given by

$$p \approx \frac{\epsilon f_-(0)}{\epsilon f_-(0) + \epsilon f_+(0)} = \frac{f_-(0)}{2f_-(0) + \Delta f(0)} = \frac{1}{2 + \rho(0)}$$

where $f_-(0)$ and $f_+(0)$ are the true left and right densities at zero. Let the number of close election constituencies falling in left (right) range be N_- (N_+). Then the observed discrepancy in the fraction of constituencies to the right vs left is $\delta \equiv \frac{(N_+ - N_-)}{N}$. The mean and variance of δ is a function of p, and hence of $\rho(0)$. Therefore $\rho(0)$ can be considered a more meaningful statistic than $\Delta f(0)$ in shaping how lop-sided the wins are within closely contested elections.

Figure 5—Histogram of relative density discontinuities in past elections



(a) Relative Density Jumps in All Elections

(b) Relative Density Jumps across Non-zero Cutoffs

Notes: Figure (a) plots three histograms. The distribution of relative density discontinuity estimates across all national and state elections from 1985-2021 for BJP win margin, the INC win margin and the same for a simulated data with no true discontinuity in the data generating process. Figure (b) plots two histograms. The relative density discontinuity distribution for BJP win margin as in (a) and the relative density discontinuities for BJP win margin at non-zero cut-off points, in the range $[-0.1, -0.02] \cup [0.02, 0.1]$. The dashed vertical line represents the estimate for 2019 BJP win margin, while the solid vertical line represents the 99.5 percentile of estimate distribution for the BJP win margin.

I estimate $\rho(0)$ for BJP and INC win margins for all national and state elections for each year from 1985-2021. Figure 5a plots the histograms of estimated $\rho(0)$ for BJP and INC win margins. It additionally plots the histogram of $\rho(0)$ using simulated data, where I generate (fictitious) BJP vote share in each PC using a truncated Normal distribution with mean 0.45 and variance 0.2.[5] The rest of the vote share is assigned to another candidate, and the simulated BJP win margin is computed for each PC. I run the simulation 40 times, to match the number of election years in the data. The true estimate for 2019 is shown with a dashed line and the 99.5 percentile of $\rho(0)$ estimates is shown with a solid line. We observe that the 2019 estimate for BJP win margin is the largest across all cases and is unlikely to arise purely due to chance given the standard deviation of the estimate across years.

Additionally, in Figure 5b I compare the histogram of $\rho(0)$ for BJP win margin with that of $\rho(k)$ for $k \in [-0.1, -0.02] \cup [0.02, 0.1]$ at intervals of 0.01 for the same BJP win margin density for all general election years sine 1989. The idea is to estimate the density discontinuities at non-zero values of BJP win margin. Since there is no reason to expect density discontinuity away from zero, the estimate distribution could be considered a placebo.[6] The solid vertical line now depicts the 99.5 percentile of the $\rho(k)$ distribution. As before, the 2019 estimate of $\rho(0)$ is the highest and lies to the right of the solid line.

Comment 8: Marshall (2024) shows that the close election RD estimate does not give us the causal effect that many researchers have been claiming in their papers.

Response: The analysis of Marshall (2024) pertains to the causal effect of electoral victory on *down-stream* outcomes. My paper, on the other hand, uses the RD method to examine endogenous sorting of PCs around the win margin threshold. Consequently, I examine discontinuity in the *pre-election* or contemporaneous outcomes (such as electorate growth rate, turnout data discrepancy, election campaigning etc. that happened prior to the election). Therefore, the criticisms of Marshall (2024) do not apply to my analysis.

References

Marshall, J. (2024). Can close election regression discontinuity designs identify effects of winning politician characteristics? *American Journal of Political Science 68*(2), 494–510.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics 142*(2), 698–714.

[5] The mean and variance in the true data is 0.25 and 0.2, due to presence of large number of candidates.

[6] The values of k are chosen to be not too far away from zero since the reduced sample size around extreme values would result in higher variance in those estimates.