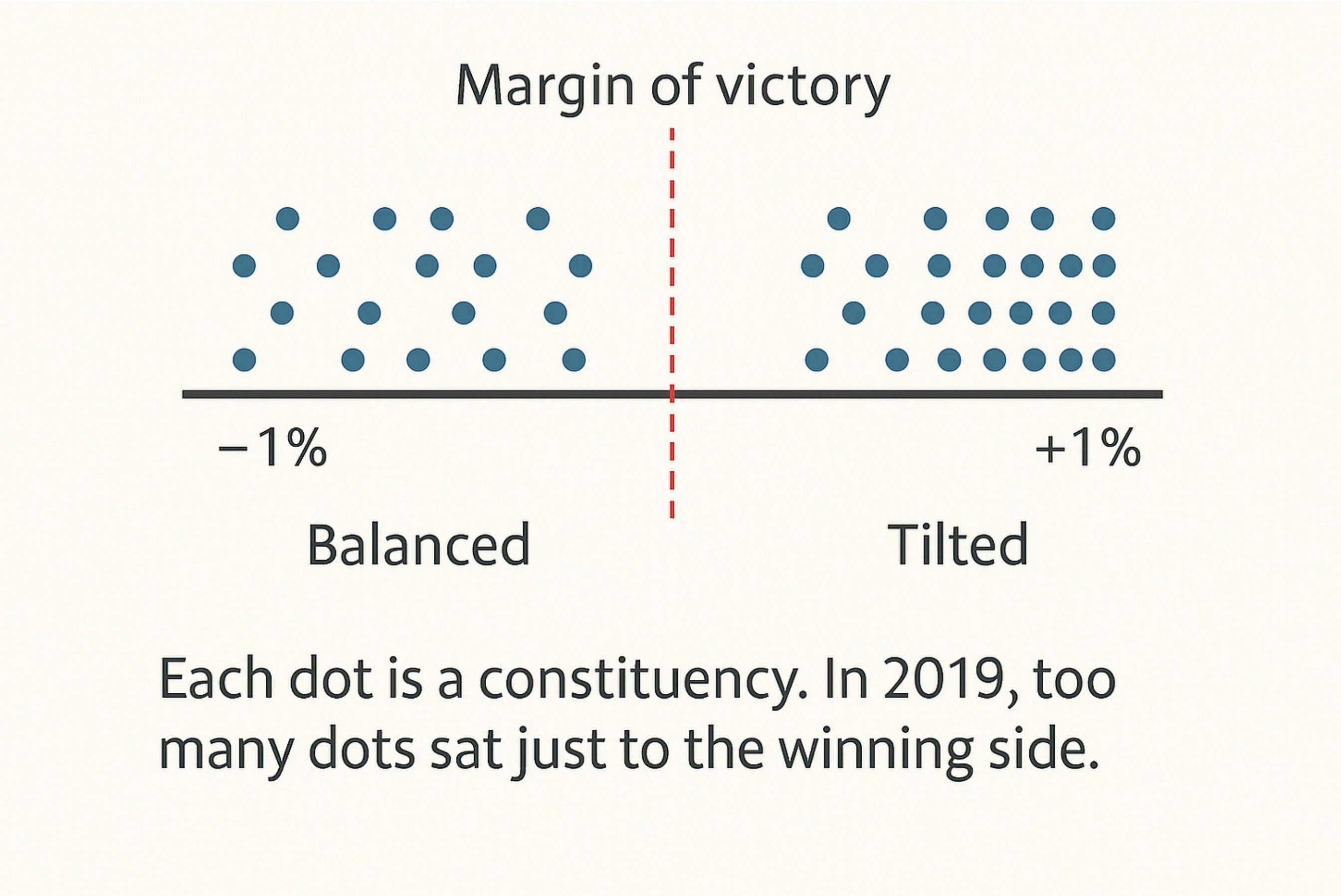
Too Many Close Wins? Reading a 2019 Election Study in 2025 (Part I)

thirdworldecon.substack.com/p/too-many-close-wins-reading-a-2019

Rohith Jyothish

November 15, 2025



This essay revisits the statistical study by economist Sabyasachi Das that examined patterns in the 2019 Lok Sabha results—now newly relevant as the Election Commission undertakes its nationwide voter-roll "Special Intensive Revision".

Note: This essay was drafted on 9 November 2025, before the Bihar election results were announced. It is not a commentary on that outcome. The analysis concerns a 2019 statistical study and the Election Commission's current voter-roll revision, both of which raise institutional questions independent of any single election.

1. Why Revisit This Paper Now

In June 2025 the Election Commission of India (ECI) launched a renewed version of its electoral-roll cleaning exercise in Bihar. Officially called the *Special Intensive Revision* (SIR) of electoral rolls, this door-to-door enumeration and verification drive required millions of voters to reconfirm their details or risk removal from the rolls.

What makes this important:

- The draft rolls released show that about 6.5 million names were flagged for deletion (out of about 79 million electors) in Bihar.
- Over 5.2 million electors could not be traced to their addresses during the revision.
- Reports suggest the same pattern is emerging across states. Many missing names, delays in online publication of roll changes, and burdens placed on migrant, poor or minority voters. This video by The Hindu Data Team is a good explainer of the unusual patterns in the deleted voter lists in Bihar.
- Given that the ECI is extending SIR nationwide, this is no longer a one-off exercise but a structural intervention in how voter rolls will be maintained for future elections.

Because voter rolls decide who counts as a citizen at the moment of voting, such large-scale revisions raise key questions: who is being removed, on what grounds, with what notice and audit, and with what effect on electoral competition?

These questions can be read alongside a controversial paper from a couple of years ago by economist Sabyasachi Das, then an Assistant Professor at Ashoka University, who "resigned" after the university opened an investigation into his academic work. Written two years after the 2019 Lok Sabha election, the paper asked a simple but politically charged question: *could patterns in official data reveal subtle forms of electoral manipulation?*

2. The Question: Too Many Close Wins

Elections are uncertain by design. In a healthy democracy, some races are landslides, some are comfortable wins, and a few are knife-edge close. If the process is fair, those razor-thin contests should split roughly half-and-half between winners and losers. Sometimes the incumbent scrapes through; sometimes they fall just short.

Imagine flipping a coin a hundred times. You'd expect about fifty heads and fifty tails. If you got ninety heads, you wouldn't praise the coin-flipper as masterful, you'd suspect the coin.

Das began with a similar intuition. In the 2019 Lok Sabha election the BJP was involved in dozens of extremely tight contests—constituencies decided by margins under five percent. In a random, fair system, about half of those would have gone either way. But the BJP won far more of them than it lost—so many more that, statistically, luck is an implausible explanation.

That imbalance is the puzzle at the heart of the paper. Das asks the question: what could plausibly create such a pattern?

Could the BJP's celebrated ground organisation and data-driven campaigning have allowed it to predict and capture marginal seats more effectively? Or do the data themselves point to structural irregularities—in how voters were registered, how turnout was recorded, or how counts were supervised?

3. Looking for the Kink: How Statisticians Spot Manipulation

The first analytical step in Das's paper is based on a method developed by Justin McCrary, in a 2008 article published in the *Journal of Econometrics*. Let me break down the intuition behind this method.

3.1. The Logic of Thresholds

Many public rules create sharp cut-offs.

In our university, a student scoring less than 40 percent fails an exam, while one scoring 41 percent passes.

Under the new tax regime, someone earning ₹50 lakh or less pays no tax, while another earning ₹50 lakh and one rupee pays 10 percent.

And in an election, a candidate who trails by one percentage point loses, while one who leads by one wins.

That dividing line between loss and victory is what statisticians call a **threshold**.

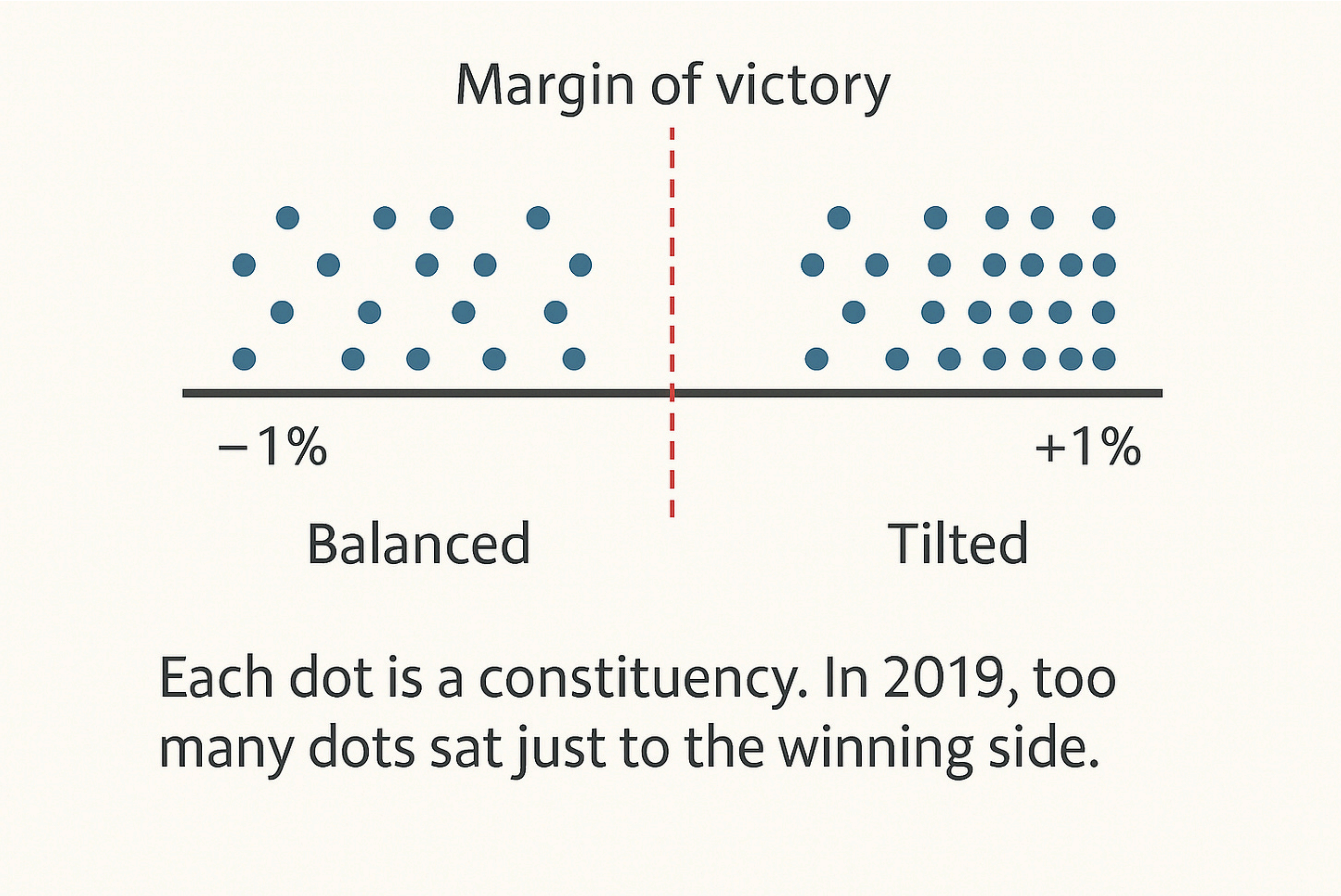
These pairs of people—just below and just above the line—are almost identical in every other way. They live in the same world and face the same rules; they differ only because the threshold happened to place them on opposite sides of a decision. For researchers, such lines are goldmines because they let us watch how outcomes change *only* because of the rule itself.

In most of the social world, we can't build laboratories or randomly assign people to "treatment" and "control" groups (unless we are Abhijit Banerjee, Esther Duflo and Michael Kremer). But when a rule creates a clean dividing line, t**he world does the random assignment for us**.

Social scientists call these moments **natural experiments**—situations in which the economy, polity, or society produces the conditions of an experiment without anyone having to design one.

Over time, statisticians developed formal tools to study such cases. One of the most influential is the **regression discontinuity design**. As technical as it sounds, it simply means comparing what happens immediately on either side of a threshold. If there is a sharp *jump* in the outcome right at that line, the rule itself probably caused it.

Figure 1. A balanced election



Note: Each dot represents a constituency arranged by its margin of victory. The vertical line marks zero—the fine border between losing and winning. In a fair contest, close wins and close losses scatter evenly on both sides of that line. No side consistently benefits from luck.

Building on this logic, Justin McCrary proposed a test that asks whether the *number of cases* itself piles up unevenly at the threshold. If too many results cluster just on one side, it may signal that someone—or some process—has nudged the numbers.

The beauty of this approach is that it needs no insider knowledge about who might be responsible. It only asks whether the data behave the way chance would normally arrange them.

Thanks for reading Third World Econ! Subscribe for free to receive new posts and support my work.

3.2. Why Manipulation Breaks the Pattern

This logic depends on one crucial assumption, that people can’t arbitrarily choose which side of the line they fall on.

Think again of the examples above. If a teacher quietly adds a few marks to push favourite students over the pass mark, or if taxpayers doctor their income reports to stay just below a tax slab, the pattern of results around the cutoff stops looking balanced.

The same thing can happen in elections. If something systematically helps one side in the closest races, the overall pattern of wins and losses will start to lean unnaturally to one side.

Imagine collecting data on every parliamentary constituency from a given election and calculating the difference between the winner's and the runner-up's vote shares—that's the **margin of victory**.

Some of these margins are negative, where the party lost, and others are positive, where it won.

Now group these constituencies according to how big or small their margins were. That is, how many were decided by <1 percent, 1-2 percent, 2-3 percent, and so on.

When you plot those counts on a graph—with the margin of victory running from losses on the left to wins on the right—you get a picture of how close or one-sided the contests were.

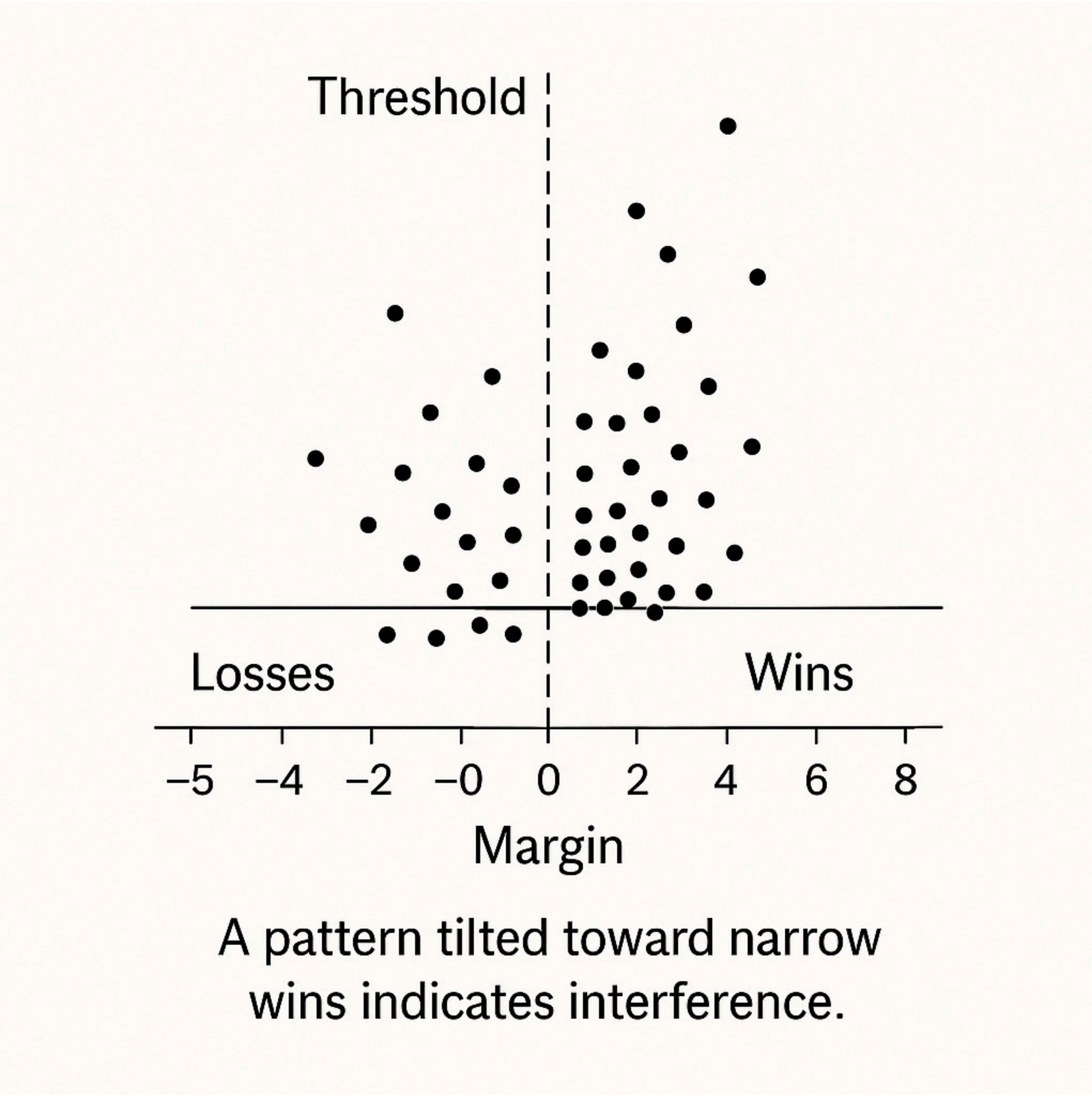
In a normal, fair election, that picture rises and falls gradually. There are plenty of comfortable victories and defeats, and a smaller, balanced number of very close races on each side of the zero line.

If no one has interfered, the pattern changes smoothly, without sudden jumps.

Statisticians call this kind of even, natural shape a **continuous distribution**—simply meaning that numbers spread out in a gentle, regular way you'd expect if they were left to chance.

But when manipulation creeps in, that smooth shape changes. You begin to see a small mound of results bunched just on one side of the zero line—too many seats that were won by a whisker and too few that were lost by the same margin.

Figure 2. When a Contest Tilts



Note: ***Each dot shows one constituency, arranged by its margin of victory. I****n a fair election, close wins and close losses should balance across the zero line. When the dots bunch just to the right of zero, it means one side has too many narrow wins and too few narrow losses, a small but systematic tilt. This is the pattern the McCrary test detects in Das's analysis of the 2019 results.*

This "aberration" is what econometricians would call a **discontinuity**. A break or kink in what should have been a smooth pattern. It's the statistical equivalent of dusting for fingerprints in forensics.

Detecting such a kink became a methodological problem of its own. When McCrary published his paper, he proposed a simple visual way of detecting it. His idea was to take all the values near a threshold, count how many cases fall into each small interval, and plot those counts as bars on a graph. If the process is clean, the bars rise and fall gently as you move across the threshold. If they show a sudden step upward or downward right at the line, that step reveals possible manipulation. McCrary called this pattern a **discontinuity in the density**—in plain language, it means that something has pushed too many cases to one side of the rule.

The power of this approach lies in its simplicity. You don't need to know *who* manipulated the numbers or *how*; you only need to see whether the pattern behaves the way chance would predict. **It's a diagnostic, not a verdict**, a way to see if the world looks slightly tilted at the very point where fairness should make it flat.

That is why Das begins his analysis of the 2019 election with this method. Elections, after all, have a built-in threshold. In India's first-past-the-post system, a candidate doesn't need a majority of votes to win—only more votes than the nearest rival. The critical line is the **zero margin**, the point where the difference in vote share changes sign. If elections are free and fair, there should be roughly as many constituencies where the ruling party *barely loses* as where it *barely wins*. The curve that plots the number of constituencies by their margin of victory should flow smoothly across that zero line.

When Das applied McCrary's test to the 2019 results, that smoothness broke. The graph showed a bulge just to the right of zero—a sudden concentration of narrow victories and a shortage of narrow losses for the BJP. Running the same check for earlier national elections (2004, 2009, 2014), for the Congress when it was incumbent, and for state elections held after 2019 produced no such bulge. The kink appeared only in the 2019 parliamentary election, and most clearly in states already governed by the BJP.

In statistical terms, it signals that the pattern of close contests departed from what random variation would normally produce. It is the point in the data where a democratic process begins to look like something has nudged it off balance.

3.3. Why This Was the Right First Step

The first move in this paper—testing for a kink in the pattern of close races—was less about proving manipulation than about seeing whether the data themselves looked unusual. The McCrary test offers a clean way to ask that question before venturing into explanations.

It was the right starting point for practical reasons. It uses only publicly available data, so anyone can replicate it; it doesn't require assumptions about why a party might perform better in some places than others; and it can be visualised easily enough for its result to speak for itself. If the shape of the data already looks improbable, the burden of explanation shifts from the statistician to the political process.

Still, beginning with this test was a judgment call. Other ways of checking for irregularities exist, and Das turns to some of them later. But this method let him start with the simplest possible question: *do the results look the way chance would have arranged them if nothing else were at play?* Once that pattern looked odd, the next step was to ask why.

Competing Explanations: Campaigning or Manipulation?

Once the data revealed that the ruling party had won too many knife-edge races, Das faced a fork in the road. Two broad explanations were possible. Either the BJP's campaign machine was so finely tuned that it could sense close contests in advance and push just hard enough to tip them its way, or something outside ordinary campaigning—errors or bias in registration, voting, or counting—had nudged those results.

Das treats these as rival hypotheses.

If the first is true, we should find traces of extra effort where the party narrowly won with door-to-door visits, more rallies, or heavier social-media engagement.

If the second is true, we should instead see traces of administrative or procedural irregularity.

He tests the first before even considering the second.

4.1. Testing the "Better Campaigning" Hypothesis

To test whether exceptional campaigning could explain the imbalance, Das turns to post-election survey data (CSDS 2019) that record how voters experienced the campaign—whether they were contacted by party workers, attended rallies, or encountered messages online. These are rough indicators of what political scientists call *campaign intensity*.

The logic is simple. If the ruling party's machinery was truly decisive, the pattern of campaign contact should be heavier in the constituencies it narrowly won than in those it narrowly lost. But the survey data show no such difference. Voters in close-loss seats report campaign exposure at roughly the same rates as those in close-win seats. In some states the contact is, if anything, slightly higher where the party lost.

Das extends the check to digital outreach (social media and WhatsApp activity) that supposedly gave the BJP its organisational edge. Yet even there, the correlation appears mostly in states it did *not* control, suggesting that online mobilisation substituted for ground work rather than amplifying it. In the states the party already governed—the ones driving the statistical kink—no distinctive campaign surge appears.

The point is not that the BJP didn't campaign effectively; it did, everywhere. But if the distribution of narrow victories were purely the result of superior strategy, we would expect to see sharper traces of that effort where margins were thinnest. The absence of such a pattern weakens the "better-campaigning" explanation and pushes the investigation to a different arena. Whether irregularity might lie not in how parties fought the election, but in how election itself was administered.

The evidence from campaign intensity doesn't explain the anomaly.

If the BJP's superior strategy can't account for the pattern, what else might?

In Part II of this essay, we turn to the harder question: whether small, seemingly administrative biases—in voter registration, turnout, and counting—could together tilt outcomes without breaking the machinery of democracy.

Thanks for reading Third World Econ! Subscribe for free to receive new posts and support my work.