Party Size and Candidate Wealth: Evidence from India

Kartikeya Batra

September 07, 2025*

Abstract

Why are some elected legislators wealthier than others? In this paper, I examine whether the size a of a political party determines the economic background of its candidates. Unlike parties with a national footprint, smaller regional parties in a multi-party system may have limited avenues for resource mobilization. I hypothesize that such parties overcome this disadvantage by fielding wealthier candidates. Using data from state elections across India during 2005-2016, I estimate whether challengers fielded by regional parties in narrowly lost seats are wealthier than those fielded by national parties. RD estimates suggest that challengers fielded by regional parties may, on average, be 44-103% wealthier than those fielded by national parties. These results provide insights into the process of candidate selection in the face of resource constraints.

*I would like to express my sincere gratitude to Sebastian Galiani , Jessica Goldberg and Ethan Kaplan for their invaluable insights and comments.

1 Introduction

What explains spatial differences in the affluence of political representatives? This question is important in light of evidence that a political representative's socio-economic profile impacts outcomes related to governance (Prakash et al., 2019). In this paper, I explore one possible explanation related to the size of political parties. In a multi-party system, political parties of different sizes exist. Whereas some hold political influence across different parts of the country, others may be limited to specific political geographies. I argue that when compared to those in the former category, parties in the latter category have limited avenues to raise funds. I hypothesize that as a result of this difference, regional parties rely more on wealthier candidates. Using state-level election data during the period 2005-2016, I test whether regional parties that barely lose elections to national parties during election cycle t field wealthier challengers during election cycle $t + 1$, when compared to national party challengers contesting against regional party incumbents.

As per the Election Commission of India (2023), there are over 2,500 political parties across India. Of these, six are recognized as "national parties", whereas over 70 are considered as state-level parties, and over 2,500 are unrecognized parties. Among the 6 national parties, only two-the Indian National Congress (INC) and the Bharatiya Janata Party (BJP)- have substantial presence across a large part of India. Since India's independence in 1947, the Indian Prime Minister has belonged to either of these parties for over 90% of the duration. The rest of the parties are mostly limited to a specific state, whereas many have presence within certain regions of a specific state.

Treating the INC and the BJP as national parties, and the rest as regional parties, I test whether the latter field wealthier candidates than the former. There are at least two reasons

to believe that this may be the case. First, unlike most regional parties, the INC and the BJP have been in power at at federal level. It is, therefore, possible that they are better connected with potential donors owing to their past, as well as the possibility of being in power in the future. At the same time, unlike regional parties, these two parties head state governments of different states, which may provide them with additional avenues to mobilize resources. Regional parties, on the other hand, have fewer avenues to raise resources in a centralized manner, and may, therefore, rely on wealthier candidates for their campaigns.

Suggestive evidence points in this direction. Vaishnav (2011) points out, "Because parties are in need of finances to fund activities such as campaigning, voter mobilization and vote buying, they must strategically select candidates who will not be a drain on finite party coffers." Verniers and Jaffrelot (2020) suggest that in the context of national elections, during 2004-2019, while average individual wealth levels for all elected candidates jumped to six times the original level, the same for elected candidates from regional parties rose to thirteen times the baseline value. In this paper, I try to find causal evidence regarding the impact of the politico-geographical reach of a political party on its choice of candidates along the dimension of candidate wealth.

A naive regression of a candidate's wealth on the nature of their political party (national or regional) may face challenges related to endogeneity. I overcome this challenge by looking at seats where national and regional parties are involved in close contests. In such a seat, the victory of a national party candidate against a regional party challenger can be treated as exogenous. Similarly, the victory of a regional party candidate against a national party challenger can also be considered as exogenous. In my sample, I focus on seats that are decided by a narrow margin during election cycle e_t. Following a close election regression discontinuity setup, I compare the wealth of regional and national party challengers during

election cycle e_{t+1}.

In the benchmark specification, I use candidates' self-declared assets as well as net worth (self-declared assets minus self-declared liabilities) as the primary outcome variables. I normalize these variables by dividing them by the number of electors in the constituency. Depending on the outcome variable, estimates suggest that candidates fielded by regional parties in seats lost by narrow margins are 44-103% wealthier than candidates fielded by national parties in similar seats. Results are robust to a battery of empirical checks.

The nature of the results, however, is localized. I focus on those seats that witness contests between regional and nationl parties in both election cycles e_t as well as e_{t+1}. In addition, in the context of this paper, a treated seat is one where a regional party challenges an incumbent national party during election cycle e_{t+1} after being narrowly beaten during election cycle e_t. Control seats, on the other hand, are those where a national party challenges an incumbent regional party during election cycle e_{t+1} after being narrowly beaten during election cycle e_t. Hence, the nature of treatment is bundled, i.e. a regional party that acts as challenger and a national party that is incumbent. I partially address the last concern by controlling for characteristics of the candidates during e_{t+1}. The results remain robust to the inclusion of these covariates.

The rest of the paper is organized as follows. In Section 2, I discuss the existing literature and this paper's contribution to the same. In Section 3, I lay down the context for this study. In Section 4, I explain the empirical strategy, while in Section 5, I provide details regarding the data and its sources. In Section 6, I establish the validity of the main empirical strategy. In Section 7, I show the main results, and in Section 8, I show results from robustness checks. Finally, in Section 9, I discuss the said results, and provide concluding

remarks.

2 Literature Review

The paper contributes to existing literature related to challenger and incumbent spending. The importance of money in elections was noted by Abramowitz (1988). Specifically, that paper focuses on the ability of challengers to raise funds for their campaigns. In the context of the United States of America, Abramowitz (1991) noted the declining ability of House challengers to raise campaign funds. At the same time, there is evidence that challenger spending impacts the challenger's fate significantly. While Jacobson (1990) suggests that challenger spending is more effective than incumbent spending, Green and Krasno (1988) suggest that the role of campaign financing is a complement to the candidate's quality.

In the Indian context, Gowda and Sridharan (2012) cite the National Election Audit (1999) conducted by the Center for the Study of Developing Societies, which suggests that "to be competitive, candidates need to spend at least a certain minimum amount", although above certain thresholds, the correlation between spending and vote shares is lost. However, given data constraints related to election spending in India, the paper will focus on challengers' assets (and net worth) instead of electoral spending. In that regard, Vaishnav (2011) points out that political parties are attracted towards criminal politicians because they possess the ability to self-finance, and they do not act as a drain on their party's resources. Chauchard et al. (2019) suggest that while voters see candidate wealth with suspicion related to corruption and political violence, there are mechanisms that justify victories of "wealth accumulators" despite these negative reactions. However, to my best knowledge, at least in the Indian context, this is among the first few papers that examines the causal impact of a political party's politico-geographic reach on its candidate selection along the dimension of wealth.

3 Context

The Indian legislative system has three tiers. Voters elect representatives to each of these tiers, i.e. Members of Parliament to the Parliament of India (national), Members of Legislative Assembly to their respective state assemblies, and representatives to their areas' local legislative bodies. Each of these elections follows the "first past the post" system. In any given election, a candidate who secures the highest number of votes in a constituency is elected to the legislature from that constituency. In this paper, I will focus on elections to state assemblies. I make this choice (over the the other two types of elections) for multiple reasons.

When compared to national elections, the contest between regional and national parties in state elections is much stiffer. For instance, as of March 2024, almost a dozen Indian states were governed by Chief Ministers from regional parties. At the same time, for almost 90% of India's 76-year old post-independence history, a member belonging to either of the two national parties-Indian National Congress (INC) and Bharatiya Janata Party (BJP)-has been the Prime Minister.[1] In addition, elections to the national parliament are held once every 5 years, whereas there are 31 states (and union territories) where elections are held once every 5 years. These state elections take place in a staggered manner, such that there is at least one state assembly ever year. Hence, these elections provide a much larger number of observations to detect effects. Finally, when compared to local body elections, data availability for state elections is much more robust.

[1]Technically, as per the Election Commission of India, as of 2024, there are 6 national parties, i.e. Bahujan Samaj Party, Bharatiya Janata Party, National People's Party, Communist Party of India (Marxist), Indian National Congress and Aam Aadmi Party. For the purpose of this project, I am considering only the INC and BJP because, unlike the other national parties, they are present across several states, have formed local governments across many of these states and have been the focal point of almost all federal governments in India's history so far.

As explained in the subsequent section, my empirical strategy hinges on the idea that constituencies with stiff competition between a regional and national party at time t will be assessed for outcomes at time *t+1*. It is to be noted that in 2008, electoral boundaries were redrawn across states. This complicates the empirical strategy. For instance, assembly elections were held to the state of Uttar Pradesh in 2002, 2007, 2012 and 2017. While constituencies in the 2002 election are comparable with those in 2007, the same is not true of constituencies in 2007 and 2012. Similarly, in the state of Kerala, elections were held in 2001, 2006, 2011 and 2016. As in the case of Uttar Pradesh, constituencies in 2001 are the same as those in 2006, and those in 2011 are the same as those in 2016. However, I cannot use 2006 as t and 2011 as *t+1*, because the configuration of constituencies changed during the intervening period. In addition, candidate affidavits, which contain information about their finances, criminality and education became compulsory only beginning 2003-04. That data is available only beginning 2004-05 and has been collated until 2017. Finally, there are states where the contest is restricted to the two national parties.

Considering the said factors in mind, as shown in Appendix A.1, constituencies from 18 states/Union Territories are considered for this study. This amounts to more than 50% of the total states/Union Territories where elections to state assembly elections are held every 5 years. Of these, I use a single set of "t" and "t+1" for 10 states, and two sets of "t" and "t+1"for the remaining 8 states.

4 Empirical Strategy

A naive OLS regression aimed at the identification of the impact of the size of a political party on the wealth of its nominated candidates is shown below (Equation (1)). As per Equation (1), the wealth of candidate i nominated by political party p in constituency c in

year t is regressed on whether the party has a national footprint (or not) in year t, candidate-level controls ($X'_{i,p,c,t}$), party-level controls ($Z'_{p,t}$), constituency-level controls ($J'_{c,t}$), and time fixed effects (δ_t). In this case, β is the coefficient of interest. A positive β will imply that all else equal, candidates belonging to regional parties are, on average, wealthier than those belonging to national parties.

$$CandidateWealth_{i,p,c,t} = \alpha + \beta RegionalParty_{p,t} + X'_{i,p,c,t}\lambda + Z'_{p,t}\eta + J'_{c,t}\kappa + \delta_t + \epsilon_{i,p,t} \quad (1)$$

However, this specification may suffer from omitted variable bias. For instance, candidate wealth may be influenced by unobservables such as a constituency's political culture. There may also be concerns regarding reverse causality. The primary identification strategy in this paper will, therefore, be a sharp regression discontinuity design (Lee, 2008; Imbens and Lemieux, 2008).

Focusing on close contests between national and regional party candidates, I will estimate the impact of the election of a national party candidate in a constituency during an election cycle e_t on the wealth of the challenger candidate fielded by the competing regional party during the subsequent election cycle e_{t+1}. The counterfactual to these treated constituencies will be those where during e_t, a regional party candidate narrowly wins, and is challenged by a national party during e_{t+1}. The methodology is based on existing literature that uses close election RD designs (among others, Lee et al., 2004 and Prakash et al., 2019).

The underlying assumption behind this setup is that at the end of election cycle e_t, constituencies where a national party defeats a regional party by a narrow margin (let's call these type "A" constituencies) are, on average, similar to constituencies where a regional

party defeats a national party by a narrow margin (call these type "B" constituencies) along all dimensions except that incumbents in type "A" constituencies belong to national parties, and those in type "B" belong to regional parties. During election cycle e_{t+1}, a candidate from a regional party will challenge the national party incumbent in type "A" constituencies, and a candidate from a national party will challenge the regional party incumbent in type "B" constituencies. A comparison of the wealth of these challenger candidates belonging to regional (treated) and national (control) political parties will then give us the causal effect of a regional party's challenge against an incumbent national party on the wealth of the former's candidate. The empirical strategy can be formalized as follows (Equation (2)).

$$CandidateWealth_{i,p,c,s,t+1} = \alpha + \beta RegionalPartyChallenger_{c,t+1} + f(Margin_{c,s,t}) + J'_{c,s,t}\kappa + \theta_s + \delta_t + \epsilon_{i,p,c,s,t+1} \quad (2)$$

As mentioned above, I will focus on close contests between national party and regional party candidates. The running variable in this case will be the difference between the national party candidate's vote share and the regional party's vote share during election cycle e_t. Therefore, the cut-off point will be zero. The final sample will contain constituencies that were won with a very narrow margin on either side of the cut-off, i.e. constituencies where national parties just won against regional parties (treated) and those where national parties just lost against regional parties (control) during election cycle e_t. In the benchmark specification, I will use a bandwidth of 5 percentage points. In subsequent sections, I test the robustness of the main results by varying the bandwidth.

The outcome variables ($CandidateWealth_{i,p,c,s,t+1}$) will represent the wealth of the challenger candidate i belonging to party p in constituency c in state s during election cycle e_{t+1}. For the main results, I will use two different outcomes: a challenger's total assets per elector

in the constituency ($\frac{Assets}{Number of Electors}$) and their net wealth per elector ($\frac{Assets-Liabilities}{Number of Electors}$) in the constituency. I use the normalized values of assets and net worth instead of aggregate values because of the vast differences in constituency sizes across the country. For instance, across all state assembly elections held in 2016, as per the SHRUG database, the range of number of electors in a constituency is 10,747 to 60,2472. The same for elections held in 2017 is 12,764 to 8,65,650. Similar variations exist within states as well.

The variable $RegionalPartyChallenger_{c,t+1}$ is a binary variable (treatment) that equals 1 if during election cycle e_{t+1} in constituency c, a regional party candidate challenges a national party incumbent. In other words, $RegionalPartyChallenger_{c,t+1}$ equals one if during e_t, in constituency c, a national party narrowly won the election against a regional party candidate. Following Cattaneo et al., 2019, the expression $f(Margin_{c,s,t})$ represents a local linear polynomial in distance to the cut-off point, i.e. margin of victory for the national party during election e_t. $J_{c,s,t}$ is a set of constituency-level controls. θ_s and δ_t represent state and time fixed effects respectively. Finally, $\epsilon_{i,p,c,s,t+1}$ is the error term. The coefficient of interest is β. A positive β implies that regional party challengers are wealthier than national party challengers. Given that the winning candidate from time t may not necessarily be fielded again at time *t+1*, this empirical strategy focuses on incumbency at the party level, and not at the individual level.

5 Data

5.1 Outcomes

As per a 2003 Supreme Court of India order, all candidates appearing in elections are required to disclose information regarding their financial, educational, and criminal background by

means of a sworn affidavit submitted to India's Election Commission. This publicly available data has been compiled by the Association for Democratic Reforms, an India-based think-tank, which was further collated by Prakash et al. (2019). I obtain this data from the Socioeconomic High-resolution Rural-Urban Geographic Dataset for India (SHRUG; see Asher et al., 2021 for details). The dataset contains each candidate's self-declared assets and liabilities. I use this data to create two different outcome variables- value of candidate's assets per elector in their constituency, and candidate's net worth (assets minus liabilities) per elector.

5.2 Explanatory Variables

For all election cycles in the sample, I obtain data on each candidate's vote share (running variable) from Jensenius and Verniers (2017). This data is also a part of the SHRUG. The dataset also contains other information regarding the constituency: turnout, number of electors, reservation status (whether it is a general constituency or reserved for historically disadvantaged groups), number of candidates and political parties in contention, etc. I use some of these variables from election cycle e_t as controls (see Appendix A.2). Finally, I obtain constituency-level data on geographical and economic covariates from SHRUG: elevation, ruggedness and employment levels (1998).[2]

6 Validity of the RD Design

A valid sharp RD setup should satisfy two conditions. First, within the bandwidth around the cut-off point, both treated and control constituencies should be balanced on all ex-

[2]The authors of SHRUG obtain raw data on employment from the Economic Census of 1998, and map it to assembly constituencies.

planatory variables except treatment status. Second, we should not see any bunching of observations around the cut-off point. I shall test each of these assumptions.

In order to confirm the validity of the RD design, I first run the McCrary density test to check if there is any manipulation around the cut-off, i.e. where the running variable equals zero. The t-statistic for this test turns out to be -1.44, which makes the coefficient statistically insignificant at conventional levels (see Figure 1).

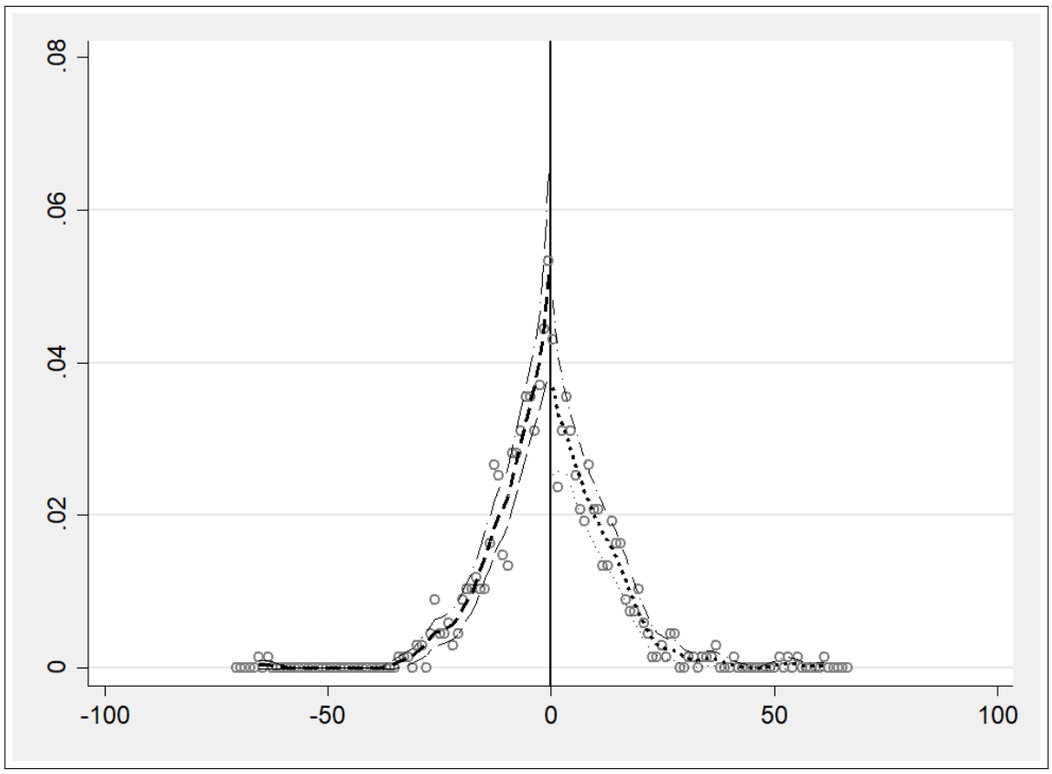

Figure 1: McCrary Test (Bandwidth=5 percentage points)

In Table 1, I show results from balance checks within the bandwidth. To arrive at these estimates, I use the empirical strategy outlined in Equation (2), but without covariates. Of the 15 covariates, I find that the estimate on the treatment variable is statistically significant

for two variables: at the 5% level for whether the winner at the baseline was a member of socially disadvantaged groups (scheduled caste/SC or scheduled tribe/ST), and at the 10% level for whether the runner-up at baseline was a member of socially disadvantaged groups (scheduled caste/SC or scheduled tribe/ST). These results are not very worrying, for even if the coefficients are null, one expects one to two estimates to be statistically significant by chance. Furthermore, I include all these covariates in the benchmark specification.

Table 1: Balance Checks (Baseline Covariates)

	(1) RD Estimate
National Party=INC?	-0.246
	(0.132)
Winner SC/ST (t)?	-0.388**
	(0.035)
Runner-up SC/ST (t)?	-0.325*
	(0.067)
Winner female (t)?	0.037
	(0.894)
Runner-up female (t)?	-0.251
	(0.184)
Turnout percentage (t)	-0.047
	(0.735)
Effective number of parties (t)	0.010
	(0.969)
No. of candidates (t)	-0.095
	(0.615)
SC Reserved constituency?	-0.235
	(0.354)
Elevation	-0.105
	(0.463)
Ruggedness	-0.205
	(0.155)
Percentage employed (1998)	0.024
	(0.832)
Winner's age (t)	0.185
	(0.499)
Runner-up's age (t)	-0.178
	(0.601)
Number of Electors (t)	-0.080
	(0.367)
N	243

p-values in parentheses

*p<0.10, **p<0.05, ***p<0.01

RD estimate using local linear polynomial. All outcome variables are standardized. The specification controls for state and year fixed effects. Robust standard errors in parentheses. Number of observations for age-related variables is 142.

7 Main Results

I report the main results in Table 2. While all estimates are positive, not all are statistically significant at the 5% level. Given that the number of observations in the specifications without any controls (columns 1 and 4), and those with all controls except age-related variables (columns 2 and 5) is 243, statistical insignificance may be a result of lack of power. On the other hand though, when I include variables related to candidates' age at baseline, 101 observations drop out due to lack of data. Estimates obtained from this sample are statistically significant (columns 3 and 6). The coefficients are consistently high across all specifications. As per these estimates, the impact of a regional party narrowly losing an election to a national party during e_t on its candidate's wealth during e_{t+1} is anywhere in the range of 44-103% of the control mean.[3] The control mean for challenger's net worth per elector, i.e. challenger's net worth per elector in tightly contested constituencies where national party candidate challenges an incumbent regional party during e_{t+1} is INR 148. The same for the national party challenger's total assets per elector is INR 165 per elector.

Table 2: Main Results

	Log of Challenger's Total Assets per Elector			Log of Challenger's Total Net Worth per Elector		
	(1)	(2)	(3)	(4)	(5)	(6)
Regional Party Challenger=1	0.535*	0.368	0.709**	0.577**	0.445*	0.643**
	(0.289)	(0.265)	(0.304)	(0.282)	(0.243)	(0.314)
Bandwidth (pp.)	5	5	5	5	5	5
Baseline Covariates	None	All ex age	All	None	All ex age	All
N	243	243	142	241	241	142

Robust standard errors in parentheses

*p<0.10, **p<0.05, ***p<0.01

Local linear regression coefficient estimated using triangular kernels. All specifications control for state fixed effects. Robust standard errors in parentheses

[3]This range comes from estimates with minimum and maximum magnitudes, i.e. 0.368 and 0.709.

8 Robustness Checks

8.1 Choice of Bandwidth

In the main specification, I use an arbitrarily chosen narrow bandwidth (margin of victory), i.e. 5 percentage points. In Figure 2, I show that results are robust to the choice of different bandwidths, both narrower and wider.

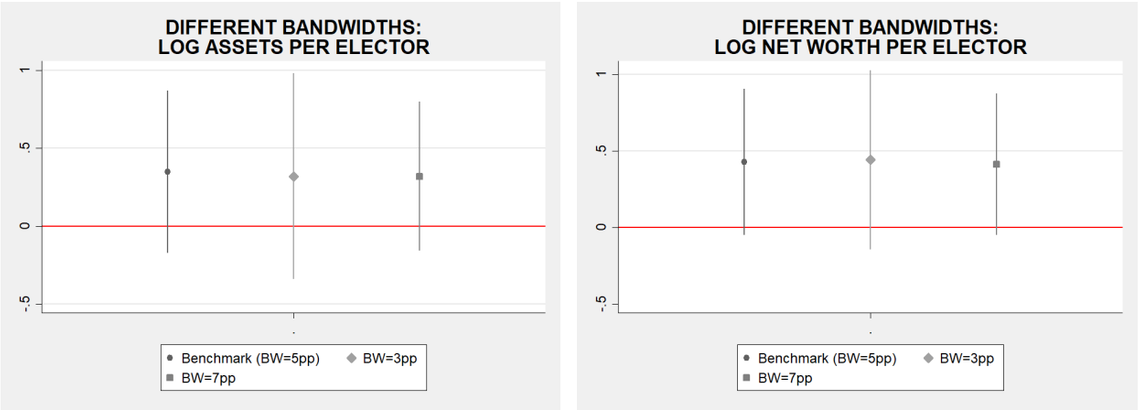

Figure 2: Robustness Checks: Different Bandwidths

8.2 Different Kernel Functions

In Figure 3, I show that results are robust to the usage of different types of kernels. In the benchmark specification, I use a triangular kernel function to assign weights to observations. In Figure 3, I show that results do not change if I use rectangular or Epanechnikov kernel functions.

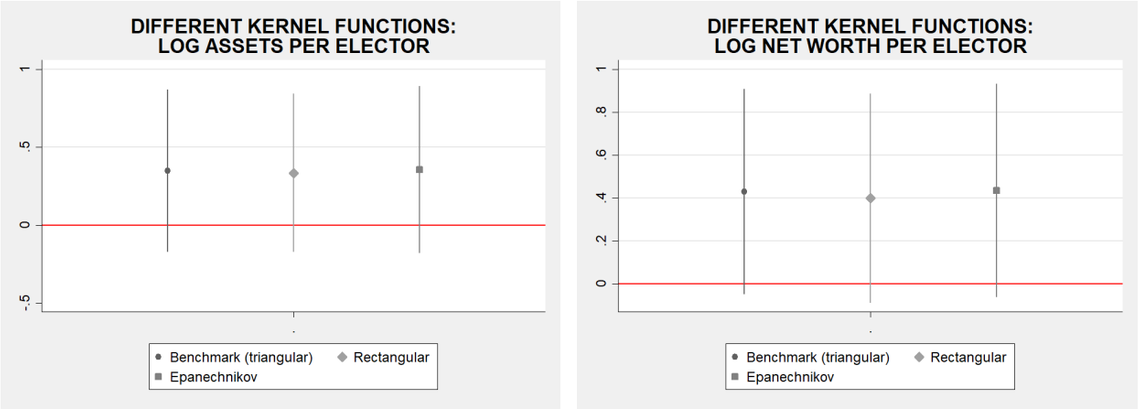

Figure 3: Robustness Checks: Different Kernel Functions

8.3 Inclusion of e_{t+1} Controls

It is plausible that the treatment, i.e. election of a national party winner against a regional party during e_{t+1} may impact other characteristics of candidates during e_{t+1}. These include factors such as candidates' age, education, criminal history, etc. This may also include the incumbent party's candidate's net worth. If these factors are correlated with challenger's wealth during e_{t+1}, then the coefficient on the treatment variable may be picking up the impact of treatment on these variables, and not challenger net worth. I address this concern by controlling for these factors in the benchmark specification. As shown in Figure 4, after the inclusion of these controls, the estimates remain stable, and the confidence intervals shrink.

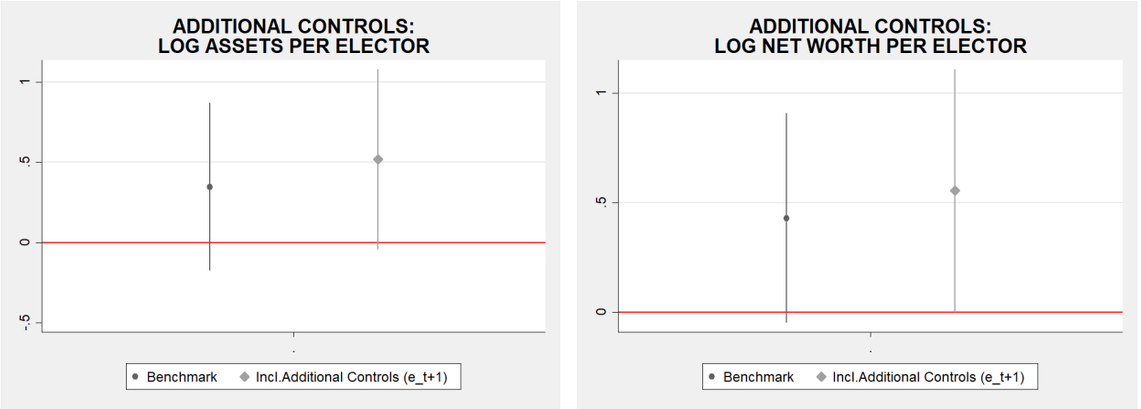

Figure 4: Robustness Checks: Inclusion of Additional Controls (e_{t+1})

9 Discussion and Conclusion

Electoral success depends on several important factors, including electoral funding. In the Indian context, the distinction between regional and national parties is important. While national parties such as the Indian National Congress and the Bharatiya Janata Party are spread across several states and are usually at the center of all federal governments, regional parties are restricted to only a handful of states (and in many cases one state). As a result, it is reasonable to assume that the former have access to larger pools of financial resources than the latter.

As a result, I hypothesize that regional parties are more dependent on wealthy individuals, who can finance their own campaigns as well as contribute to the party's resource pool. Focusing on state-level assembly elections, I estimate whether regional party candidates who challenge incumbent national parties are wealthier than national party candidates who challenge incumbent regional parties. Using a sharp RD design, I focus on seats where regional parties narrowly win/lose against national parties during election cycle e_t. Seats where regional parties narrowly lose against national parties (treated) are those where they field challengers during election cycle e_{t+1}. whereas in seats where national parties narrowly lose

against regional parties (control) are those where the former field challengers during e_{t+1}. The treatment effect is the difference between wealth of the regional party challenger and the national party challenger during election cycle e_{t+1}.

My estimates suggest that on average, regional parties field wealthier challengers than national parties during election cycle e_{t+1}. The magnitudes of these estimates suggest that these effects are in the range of 44-103% of the control mean. While results are robust to standard checks, I also add candidate-level controls from e_{t+1} to ensure that effects are not driven by candidate-level characteristics. The estimates are stable even after the inclusion of these controls.

Given that these estimates are localized-both in terms of focusing on constituencies with close contests as well as focusing on constituencies with contests between regional and national parties-they should be interpreted with caution. There may also be reasons to believe that these estimates are a lower bound of the true treatment effect. For instance, across several states (such as Bihar, Punjab and Maharashtra), there are instances where a regional party, in alliance with a national party, is competing against another national party or another alliance of a regional party and national party. For example, for a very long time in the state of Maharashtra, the Indian National Congress (a national party) and the Nationalist Congress Party (NCP, a regional party) have allied against the Bharatiya Janata Party (a national party) and the Shiv Sena (a regional party). In such cases, in a regional-national party contest, the regional party may not be too dependent on wealthy candidates because their alliance partner (a national party) may also be contributing to the former's resource pool. Similarly, candidates may not have all their wealth/assets listed in their own name. Instead, a substantial share of their true wealth may be amassed in the name of their family members. I cannot estimate these proportions separately, which makes it a limitation of the

analysis. Overall, however, these results open up avenues for future research in the relatively underexplored area of electoral finance in developing countries like India.

A Appendix

A.1 States in Sample

- * For Delhi, I consider 2013 and 2015 because of a peculiar scenario where owing to an unclear mandate, Delhi, after the assembly election in late 2013, witnessed President's rule for almost a year and a fresh round of polls was conducted in early 2015.
- ** Because of Jammu and Kashmir's erstwhile special status as a state, elections used to be held every 6 years there.
- *** The state of Telangana was created in 2014. However, pooling the districts that now comprise Telangana, I used it as a state even in 2009.

	Year "t"	Year "t+1"
Andhra Pradesh	2009	2014
Assam	2001 2011	2006 2016
Bihar	2010	2015
Delhi*	2013	2015
Goa	2002 2012	2007 2017
Haryana	2000 2009	2005 2014
Jammu & Kashmir**	2008	2014
Jharkhand	2009	2014
Karnataka	2008	2013
Kerala	2001 2011	2006 2016
Maharashtra	2009	2014
Odisha	2009	2014
Puducherry	2001	2006
Punjab	2002 2012	2007 2017
Tamil Nadu	2001 2011	2006 2016
Telangana***	2009	2014
Uttar Pradesh	2002 2012	2007 2017
West Bengal	2001 2011	2006 2016

Table 3: Description of Sample

A.2 Control Variables

- **National Party=INC?:** Whether the national party in question is the Indian National Congress (INC) or not. If it is the BJP (the second of the 2 national parties), this variable equals zero.

- **Winner SC/ST (t)?:** Whether the winning candidate at baseline (t) belonged to socially disadvantaged groups (Scheduled caste or scheduled tribe).

- **Runner-up SC/ST (t)?:** Whether the runner-up at baseline (t) belonged to socially disadvantaged groups (Scheduled caste or scheduled tribe).

- **Winner female (t)?:** Whether the winning candidate at baseline (t) was a female or not.

- **Runner-up female (t)?:** Whether the runner-up candidate at baseline (t) was a female or not.

- **Turnout percentage (t):** Percentage of total voters who came out to vote during election cycle e_t.

- **Effective number of parties (t):** Effective number of parties derived from vote shares of each candidate in a constituency.

- **Number of candidates (t):** Number of candidates contesting the election during e_t.

- **SC Reserved constituency?:** Whether the constituency is reserved for representatives belonging to scheduled castes or not.

- **Elevation:** Average elevation of the constituency. Elevation data is from SRTM (Shuttle Radar Topography Mission) at 1-arc second resolution. Its Digital Object Identifier (DOI) number is /10.5066/F7PR7TFT.

- **Ruggedness:** Average ruggedness of the constituency. The ruggedness measure TRI is constructed from the elevation data using the method from Riley et al (1999), and used by Nunn and Puga (2012) 'Ruggedness: The Blessing of Bad Geography in Africa'. The TRI expresses the elevation difference between adjacent pixels. It is the square root of the mean squared difference in elevation values between a pixel and the eight cells immediately surrounding it.

- **Percentage employed (1998):** Percentage of electors employed in economic establishments (except government administration, national defense and agriculture) in a constituency.

- **Winner's age (t):** Age of the winner during election cycle e_t.

- **Runner-up's age(t):** Age of the runner-up during election cycle e_t.

- **Number of electors (t):** Total number of electors in a constituency (those who voted plus those who did not) during election cycle e_t.

References

Abramowitz, A., 1988. Explaining Senate Election Outcomes. American Political Science Review , 385–403.

Abramowitz, A., 1991. Incumbency, Campaign Spending, and the Decline of Competition in the U.S. House. Journal of Politics , 34–57.

Asher, S., Lunt, T., Matsuura, R., Novosad, P., 2021. Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the shrug open data platform. The World Bank Economic Review 35, 845–871.

Asher, S., Novosad, P., 2019. Socioeconomic High-resolution Rural-Urban Geographic Dataset for India (SHRUG). Harvard Dataverse doi:https://doi.org/10.7910/DVN/DPESAK.

Asher, S., Novosad, P., 2020. Rural Roads and Local Economic Development. American Economic Review , 797–823.

Cattaneo, M.D., Idrobo, N., Titiunik, R., 2019. A Practical Introduction to Regression Discontinuity Designs: Foundations. Cambridge Elements: Quantitative and Computational Methods for Social Science .

Chauchard, S., Klašnja, M., Harish, S., 2019. Getting Rich Too Fast? Voters' Reactions to Politicians' Wealth Accumulation. The Journal of Politics 81.

Gowda, R., Sridharan, E., 2012. Reforming India's Party Financing and Election Expenditure Laws. Election Law Journal 11.

Green, D., Krasno, J., 1988. Salvation for the Spendthrift Incumbent: Reestimating the Effects of Campaign Spending in House Elections. American Journal of Political Science , 884–907.

Imbens, G.W., Lemieux, T., 2008. Regression discontinuity designs: a guide to practice. Journal of Econometrics , 615–635.

Jacobson, G., 1990. The Effects of Campaign Spending in House Elections: New Evidence for Old Arguments. American Journal of Political Science 9, 321–350.

Jensenius, F.R., Verniers, G., 2017. Studying Indian politics with large-scale data: Indian election data 1961–today. Studies in Indian Politics .

Lee, D., 2008. Randomized experiments from non-random selection in US House elections. Journal of Econometrics , 675–697.

Lee, D., Moretti, E., Butler, M.J., 2004. Do Voters Affect or Elect Policies? Evidence from the U.S. House. Quarterly Journal of Economics 119, 807–59.

Prakash, N., Rockmore, M., Uppal, Y., 2019. Do criminally accused politicians affect economic outcomes? Evidence from India. Journal of Development Economics 141.

Vaishnav, M., 2011. The Market for Criminality: Money, Muscle and Elections in India. SSRN Electronic Journal doi:10.2139/ssrn.1928011.

Verniers, G., Jaffrelot, C., 2020. The reconfiguration of India's political elite: profiling the 17th Lok Sabha. Contemporary South Asia , 242–254.

Vishnoi, A., 2020. At Rs 1,450 cr, BJP got 61% funding via Electoral Bonds before LS polls. The Economic Times. URL: https://economictimes.indiatimes.com/news/politics-and-nation/at-rs-1450-cr-bjp-got-61-funding-via-electoral-bonds-before-ls-polls/articleshow/73181670.cms?from=mdr.