When Do Politicians Appeal Broadly?
The Economic Consequences of Electoral Rules in Brazil

Moya Chin*

December 21, 2019

Please click here for the latest version

Abstract

Electoral rules determine how voters' preferences are aggregated and translated into political representation, and their design can lead to the election of representatives who represent broader or narrower constituencies. This project examines how single- and two-round elections in Brazil affect municipal mayoral races using a regression discontinuity design. Two-round elections use two rounds of voting to elect a winner, ensuring that the eventual winner must obtain at least 50% of the vote. Theoretically, this may provide incentives for candidates to secure a broader base of support. Consistent with this, I show that in two-round systems, candidates represent a more geographically diverse group of voters, more resources are allocated to public schools, and there is less variance in resources allocated to public schools across the municipality. I find evidence suggesting that these effects are driven by strategic responses of candidates, rather than differential entry of candidates into races. The findings suggest that two-round systems can lead candidates to secure broader voter bases and subsequently exhibit less political favoritism when implementing policy.

*Harvard University, Department of Economics, mchin@g.harvard.edu. Sincere thanks to my advisors Melissa Dell, Nathan Nunn, Vincent Pons, and Gautam Rao for their extensive advice and guidance. I also thank Alberto Alesina, Juan Pablo Chauvin, Siddharth George, Nir Hak, Ross Mattheis, Tzachi Raz, Jonathan Roth, Juan Francisco Santini, Henrik Sigstad, Niharika Singh, Edoardo Teso, and Clemence Tricaud. Ravi Jagadeesan offered invaluable comments. Roberto Acioly, Renata Dias, and Julia Orrico provided excellent research assistance. This project also benefited from numerous participants at the Development Lunch and Seminar, Political Economy and Culture Lunch, Political Economy Workshop, and Labor and Public Finance Lunch at Harvard and at the NEUDC conference.

1. Introduction

Electoral rules determine how voters' preferences are aggregated and translated into political representation, and their design can lead to the election of representatives who represent broader or narrower constituencies. This is particularly important given the evidence that more inclusive political institutions are beneficial for long-term growth (Acemoglu and Robinson, 2008; Acemoglu et al., 2014; Bardhan and Mookherjee, 2006). This paper studies how one difference in electoral rules – namely, if elections feature a single or two rounds – affects the extent to which elected representatives appeal to a broader constituency and how this, in turn, affects the overall level and distribution of public goods.

I take advantage of a unique policy in Brazil that assigns a municipality's electoral rule based on a threshold of 200,000 registered voters. Municipalities below this threshold elect their mayor in a single-round election, and municipalities above this threshold elect their mayor in a two-round election. Using a regression discontinuity framework that compares municipalities close to the threshold, this paper provides causal empirical evidence that politicians elected under two-round systems secure geographically broader bases of support, provide more resources to public schools, and allocate these resources more equitably.

Two-round systems are the most widely used rule in democratic presidential elections: 64.0% of elections under presidential systems used the two-round system between 2000 and 2010 (Bormann and Golder, 2013 and Figure 1). Together, single- and two-round systems account for 86.7% of elections in this period. In a single-round system, voters vote once and the candidate with the most votes wins. In a two-round system, voters first vote and, if no candidate receives a majority, vote a second time between the top two candidates.[1] This difference generates two important distinctions. First, two-round systems require winners to attain a vote share above 50%. Second, the existence of a second round effectively limits the number of candidates (Lizzeri and Persico, 2005). Even in the first round, the top candidate effectively only needs to be concerned with the runner-up, who poses the threat of either victory in the first round or opposition in the second. Because of these distinctions, it has been argued that two-round systems incentivize candidates to secure a broader base of support and legitimize the winner's position once elected (Bouton, 2013; Bouton and Gratton, 2015).[2] The intuition behind this is that the rules imposed in a two-round system make it more difficult for politicians to win with policies that appeal to a narrow group of voters.

[1] This is the case in Brazil and most countries with two-round systems. Outside of single- and two-round systems, a small number of countries use qualified two-round, electoral college, and alternative vote.

[2] While not a focus of this paper, there is also a large literature arguing that two-round systems allow voters to vote more sincerely in the first round (see Bouton et al., 2019 for a review) and to better communicate their policy preferences to candidates (Piketty, 2000).

To the extent that politicians commit to their campaign promises, the policies they offer in order to win the election may translate into economic consequences. This paper tests the hypothesis that electoral rules affect the level of public goods provided and the manner in which they are allocated *across* the electorate. When electoral rules make it more difficult for politicians to win with a narrow constituency, this reduces their incentive to provide public goods supported by a narrow constituency once in office.

To guide the empirical analysis and explain this intuition, I propose a model of electoral competition to illustrate how two-round systems create incentivizes for politicians to appeal to broader groups of voters, provide more public goods, and allocate public goods differently. I adapt a standard probabilistic voting model where candidates offer policy proposals that (i) specify the overall size of the government budget and (ii) target government resources to specific localities within a municipality, as in Genicot et al. (2018). These policy proposals are announced prior to the election and, in a two-round system, are binding between rounds. In these elections, the top two candidates must contend with a small, third candidate who is non-strategic and commits to allocating all resources to a single locality.

In the model, two-round elections lead to different outcomes, because the second round raises the marginal return for the top two candidates to allocate resources to all localities. First, in a two-round election, to win in the first round, candidates must attain not only the most votes, but must attain at least 50% of the vote. If they do not, candidates must compete in a second round. This conditionality raises the value of each vote. As a result, candidates promise a higher overall budget. Second, candidates in a single-round election face low incentives to allocate resources to the locality dominated by the third candidate. In contrast, in a two-round election, the top two candidates in the first round must look ahead to the possibility of a second round in which the third candidate will not be present. As a result, candidates promise more resources to the locality that the third candidate appeals to, even if they are unlikely to attract those votes in the first round. This increased allocation generates a force to reduce inequality in the allocation of government resources.

I test these predictions empirically across six municipal elections, between 1996 and 2016, in Brazil. I exploit the fact that Brazil's electoral authority assigns the electoral rule for mayor in each election based on the number of registered voters in the municipality and whether this lies above or below an arbitrary threshold. To estimate the causal impact on electoral and economic outcomes after each election, I compare municipalities just above the registered voter threshold with those just below, in a regression discontinuity design. I obtain three main empirical results.

First, candidates in two-round elections receive broader geographical support. Using each candidate's vote count at each polling station, I measure the geographic distribution of voters for specific candidates in the first round in two ways: (i) indices of voter concentration to quantify the *overall* level at which voters are geographically concentrated, and (ii) the standard deviation of candidates' vote shares to quantify a *candidate*-level measure of geographic concentration. In two-round municipalities, voters are less geographically concentrated, corresponding to a reduction that is 27.4–45.6% of the average level of concentration in single-round municipalities. The decrease in concentration only occurs among the top two candidates, indicating that the electoral rule mainly affects the candidates with a chance of winning. I show that this is not driven by the increased number of candidates in two-round elections.[3] These results suggest that two-round elections lead to greater inclusiveness: Elected candidates represent a geographically broader constituency. In turn, I find inclusiveness along another dimension: Voters are more engaged in the political process. While turnout is unaffected, as turning out to vote is mandatory in Brazil, I find that the number of blank and invalid ballots is significantly lower in two-round elections.

Second, once in office, politicians elected under two-round systems provide and allocate public goods differently. I find effects on both the *level* and *distribution* of municipal resources. I look at the provision of a local public good that can be geographically targeted and is the main public good completely under the jurisdiction of the municipality: municipal education. To quantify provision, I measure the level of resources present in public schools in the municipality. I find that (i) public schools in two-round municipalities measure 0.057–0.081 percentiles higher in the national distribution of resources, and (ii) the standard deviation of resources across schools is lower. Schools with the fewest resources in the municipality benefit the most from these additional resources. When politicians secure broader bases of support, they provide more public goods and distribute these resources more evenly across the municipality.

Third, I find evidence that educational outcomes are improved in two-round municipalities. Specifically, drop-out rates are lower and literacy rates higher among cohorts of school age during the electoral term. I find limited improvements along more downstream economic indicators. Two-round municipalities have fewer low-income households, but there are no significant improvements in average income, employment, or night lights. These results suggest that electoral rules impact direct policy outcomes, but have limited effects on more downstream economic outcomes.

Turning to mechanisms, I do not find that two-round elections cause differential selection of candidates. More specifically, I do not find that different types of candidates enter the

[3] The effect of the electoral rule on the number of candidates is known as Duverger's Law, which states that single-round elections will lead to a two-party system, while two-round elections will lead to a multi-party system. This has been formalized, and sometimes challenged, in recent literature (Bouton, 2013; Bouton and Gratton, 2015; Callander, 2005; Cox, 1997; Fujiwara, 2011; Osborne and Slivinski, 1996).

races, nor do different types of candidates win the elections. Candidates in two-round elections are not observably different in terms of demographic characteristics, place of birth, educational attainment, or previous occupation. Winners also do not differ along these observable characteristics. More candidates enter the race in two-round elections – many of these additional candidates are represented by smaller parties and have run in previous elections. However, these candidates are not more likely to win the election.

I find suggestive evidence that the effects I find are driven by the different strategic incentives candidates face in two-round elections. These strategic incentives may affect politicians' behavior once in office through two mechanisms. First, candidates may adopt different strategies during the campaign, which, to the extent that politicians fulfill campaign promises, lead to different behaviors in office. This is in line with the model's predictions. Second, politicians may behave differently in office in anticipation of the forthcoming election and what electoral rule it will follow. If the second mechanism is present, candidates facing different reelection incentives should behave differently. I rule out the second mechanism, as I do not find different treatment effects among mayors facing different reelection incentives. Instead, I find that different outcomes are due to candidates adopting different strategies during the campaign. First, in two-round elections, geographic concentration decreases between the first and second round, suggesting that candidates adopt strategies between rounds to consolidate their voter bases.[4] Second, candidates in two-round elections finance their campaigns differently: They rely less on donations from corporations. To the extent that corporations represent narrower swaths of the electorate, this suggests that candidates in two-round elections adopt strategies that appeal more broadly to individuals rather than corporations. These findings suggest that candidates in two-round elections offer different policy platforms and build more geographically diverse constituencies. When their supporters are less geographically concentrated, mayors allocate public goods in a less geographically concentrated manner.

This study builds on a large, mostly theoretical literature studying the role of electoral rules on political incentives. These studies document how different rules impact electoral accountability (Persson et al., 2003) and personal vote-seeking (Carey and Shugart, 1993).[5] Importantly, these studies highlight how electoral rules incentivize politicians to appeal to broader groups of voters (Myerson, 1993), provide public goods with broader benefits (Lizzeri

[4] There is empirical evidence that candidates qualifying for the second round rally votes from supporters of the candidates eliminated after the first round. Pons and Tricaud (2018) find that in France, the qualification of a third candidate in the second round reduces the top two candidates' vote share, indicating that when the third candidate is not present in the second round (as is always the case in Brazil), the top two candidates capture votes from the third placed candidate's supporters.

[5] See Persson and Tabellini (2003, 2004) for a review of other consequences of electoral rules.

and Persico, 2001, 2005; Persson and Tabellini, 1999), and target public spending to specific groups of voters (Genicot et al., 2018; Milesi-Ferretti et al., 2002). With the exception of Lizzeri and Persico (2005), these studies do not compare single- to two-round systems. In Lizzeri and Persico (2005), politicians in proportional systems provide broad public goods at a higher rate than targetable goods when there is less political competition. They extend their model to argue that two-round elections should lead to higher provision of broad public goods, because the second round effectively limits the number of candidates. This paper provides a theoretical framework that models public goods provision in single- and two-round systems, generates predictions on both the overall level and the exact allocation, and tests these predictions empirically.

This paper adds to a growing empirical literature providing causal evidence on the impacts of local electoral rules. These studies, which compare proportional and single-round systems in addition to single- and two-round systems, have measured the impact on electoral outcomes and fiscal expenditures (Chamon et al., 2018; Cipullo, 2019; Eggers, 2013; Fujiwara, 2011; Pellicer and Wegner, 2013). Chamon et al. (2018) and Fujiwara (2011) also study the Brazilian context. Of particular interest are Bordignon et al. (2016), who compare single- and two-round systems in Italian municipalities and find that municipalities under two-round systems exhibit more policy moderation, as measured by the volatility of a municipal tax rate across elections. In contrast to Bordignon et al. (2016), I study not only an *aggregate* policy outcome – the overall level of public goods provision – but also the *allocation* of this policy across the electorate. This paper's contribution is to provide evidence that electoral rules have economic consequences, both on the level of public goods provision and how these public goods are distributed.

More broadly, this paper connects to a literature examining inequalities in the allocation of state resources. A large literature documents the role of political factors in creating these inequalities – in particular, how politicians politically favor certain subgroups, such as those of the same ethnicity or partisanship. A key insight that emerges is that the extent to which politicians practice political favoritism is reduced when political institutions are stronger, elections are more competitive, and citizens are more broadly engaged in the electoral process (Burgess et al., 2015; Fujiwara and Wantchekon, 2013; Hodler and Raschky, 2014). Notably, Golden and Min (2013) emphasize the importance of policy responsiveness to voter preferences. Electoral rules serve as a key channel through which voter preferences are translated into policy outcomes. The final contribution of this paper is to demonstrate the role of another factor in political favoritism, the electoral rule, and the incentives it creates for politicians to broaden their appeal.

The remainder of this paper is organized as follows. Section 2 presents a theoretical framework for single- and two-round systems. Section 3 describes the context, and Section 4 describes the empirical strategy. Section 5 presents the results on electoral and economic outcomes. Section 6 discusses the results and mechanisms. Section 7 concludes.

2. Theoretical framework

I present a stylized model to illustrate how two-round elections create incentives for politicians to secure a broader base of support and provide public goods differently. My model of electoral competition adapts a standard probabilistic voting model (Burden, 1997; Lindbeck and Weibull, 1987; Persson and Tabellini, 2000) and follows the setup in Genicot et al. (2018) by allowing for targeting of government interventions to specific localities within a municipality. I extend this model by (i) introducing a third non-strategic candidate who appeals to a single locality, (ii) allowing candidates to exert effort to increase the municipal budget, and (iii) adapting it to the context of single- and two-round elections.

The intuition is that two-round elections perform two functions. First, the winner must attain a vote share above 50%. This makes it difficult for politicians who appeal to a minority of the electorate to gain enough votes to win the election. It also raises the marginal value of *every* vote and results in candidates exerting more effort to increase the overall government budget. Second, the existence of a second round effectively limits the number of candidates. Candidates who expect to gather enough first round votes to qualify for the second round but less than the 50% required to win in the first round must behave as though a second round will occur where only one other candidate will stand. This incentivizes candidates to offer policy proposals that appeal to all localities, even those that are likely to vote for another candidate in the first round. While this effect runs counter to a literature documenting the positive impacts of political competition, in a single-round election, higher electoral competition incentivizes candidates to appeal to narrower groups and ignore other voters.[6]

2.1. The environment

Consider an election with three politicians and J localities within a municipality. Politicians are indexed by $c \in \{A, B, C\}$, and localities are indexed by $j \in \{1, 2, \cdots, J\}$ where $J \geq 3$. Each locality has a continuum of voters of mass $1/J$.

[6] A theoretical literature argues that higher electoral competition has the negative effect of incentivizing candidates to focus on narrower groups. Myerson (1993) shows that candidates offer more unequal campaign promises when electoral competition increases in elections using rank-scoring rules. Lizzeri and Persico (2005) extend this model to other electoral rules and introduce a public good, finding a negative effect of political competition on the equality of campaign promises and public good provision.

Prior to election day, politicians simultaneously announce a platform that describes (i) the total government budget, G^c, and (ii) the amount of the government budget to be allocated to each locality, $\mathbf{q}^c = (q_1^c, q_2^c, \ldots, q_J^c)$, where $q_j^c \geq 0$. The politician's budget constraint is:

$$\sum_{j=1}^{J} q_j^c \leq G^c$$

Since each locality has the same number of voters, each voter receives the same fraction of the government budget allocated to their locality. I assume without loss of generality that voters care about the total amount allocated to their locality, q_j^c. In promising a certain budget, politicians face a cost that is quadratic in the size of the budget:

$$\frac{1}{2}\kappa \left(G^c\right)^2$$

for a constant κ. Platforms are binding for politicians between rounds and after the election.[7]

To make solving a three-candidate model more tractable, I assume that the third candidate C is a non-strategic candidate with the following platform:[8]

$$\mathbf{q}^C = \left(0, 0, \ldots, 0, G^C\right)$$

There are several ways to interpret candidate C. In my model, because voters are partitioned into geographic localities, C is a candidate whose supporters are all located in the same geographic area. However, my model easily translates into other interpretations of candidate C. For example, C may be a candidate whose supporters share a common trait and vote for her due to descriptive representation. These traits could include geography, but also other dimensions such as age or race. Another possibility is that C is a single-issue candidate who attracts voters who only care about that issue. In any of these cases, C should be viewed as a small candidate and A and B as front-runners in relation to C. This also empirically matches elections in Brazil. Third placed candidates receive on average 11.9% of the vote, and the vote spread between the second and third placed candidate is on average 23.8%.

Following Genicot et al. (2018), voters in locality j have preferences over government spending. They obtain utility $u_j(q_j)$ from government spending q_j, where $u_j(q_j)$ is strictly

[7]This is not unrealistic, as the time between rounds is often short compared to the length of the campaign. In Brazil, the second round is three weeks after the first. Intuitively, my model translates into contexts where this assumption is relaxed as long as there is some continuity between the two rounds – first, if voters' second round vote depends on a candidate's policy proposal in both rounds; and second, if candidates can change their policy proposals between rounds but are constrained in the extent to which their proposals can change.

[8]I assume that G^C is the highest offer in locality J, ie. that candidate A and B's equilibrium allocations to J are smaller than G^C. See Appendix A.9.

increasing and concave in q_j. In addition to the policy component of voters' preferences, there is an individual shock v_i and a municipality shock δ toward candidate A, which are independently and uniformly distributed:

$$v_i \sim U\left[-\frac{1}{2\psi}, \frac{1}{2\psi}\right] \qquad \delta \sim U\left[-\frac{1}{2\gamma}, \frac{1}{2\gamma}\right]$$

The individual shock, v_i, captures idiosyncratic voter preferences towards candidate A. The municipality shock, δ, captures any political dimensions that swing voters in the municipality as a whole toward candidate A, such as economic shocks, and is independent across rounds.

Voters cast a ballot for the politician who offers them the highest payoff. In localities $j \in \{1, \ldots, J-1\}$, ie. where candidate C has not allocated resources, this amounts to voting for either A or B.[9] In locality J, ie. where candidate C is dominant, voters randomize between voting for C with probability $1-\alpha$ and for either A or B with probability α, depending on whether A or B offers the higher payoff, where $0 < \alpha < 1$.[10]

Thus, in general, voters will vote for A if and only if:

$$u_j(q_j^B) \leq u_j(q_j^A) + v_i + \delta \tag{2.1}$$

In localities $j \in \{1, \ldots, J-1\}$, all voters for whom this is true vote for A.[11] In locality J, a fraction α of voters for whom this is true vote for A. I assume that the marginal utility of locality J relative to that of the other localities, $u'_J(q)/u'_j(q)$, is not too large.[12]

The electoral rules follow those in Brazil. In a single-round system, the candidate with the most votes wins. In a two-round system, if no candidate attains more than 50% of the vote in the first round, a second round occurs with the top two candidates.

2.2. Preliminaries

Let π_t^c be candidate c's total vote share in the municipality in round $t \in \{1, 2\}$. Let $\Delta u_j^{cd} \equiv u_j(q_j^c) - u_j(q_j^d)$ be the difference in utility in locality j between candidate c and d's

[9] Since $u_j(\cdot)$ is strictly increasing, candidates A and B will invest a non-zero amount in these localities, and voters will always vote for either A or B.

[10] Assuming that $\alpha > 0$ performs two functions. First, it guarantees that candidate C gains strictly less than 1/3 in vote share and so never has the most votes. Candidate C will also never have the most votes if $J > 3$. Second, it guarantees a non-zero first order condition for locality J in the single-round election, which allows a direct comparison between single- and two-round elections. This assumption can be relaxed and will yield the same predictions; see Appendix A.10.

[11] For localities $j \in \{1, \ldots, J-1\}$, it is also a requirement that $u_j(q_j^A) + v_i + \delta \geq u_j(q_j^C)$. Because $u_j(q_j^B) \geq u_j(0) = u_j(q_j^C)$, condition (2.1) is sufficient for voters to vote for A in these localities.

[12] In other words, for all $j \in \{1, \ldots, J-1\}$, I assume that $u'_J(q)/u'_j(q) < \dfrac{\psi J + (1-\alpha)\gamma}{\alpha\psi J + (1-\alpha)\gamma}$, for all positive q.

offers.

2.2.a. *Vote shares with three candidates.–* This section applies to single-round elections and the first round of two-round elections.

As in Genicot et al. (2018), I assume that (i) there are voters to be swung in every locality and (ii) all localities are contestable.[13] The probability that candidate A attains a vote share above θ is given by:

$$Pr\left(\pi_1^A \geq \theta\right) = \frac{1}{2} + \frac{\gamma}{\psi}\left[\frac{1}{2} - \left(\frac{J}{J-1+\alpha}\right)\theta + \left(\frac{\psi}{J-1+\alpha}\right)\left(\sum_{j=1}^{J-1}\Delta u_j^{AB} + \alpha\Delta u_J^{AB}\right)\right] \quad (2.2)$$

For a detailed derivation and expressions for other candidates, see Appendix A.2.

With three candidates, I assume that candidate C always receives the lowest vote share: candidate C never wins a single-round election nor makes it to the second round in a two-round election (see Appendix A.5). This simplifies candidate A and B's maximization problem, and shuts down the channel where candidate C poses a different threat to electoral defeat.

2.2.b. *Vote shares with two candidates.–* This section applies to the second round of two-round elections (as mentioned, candidate C never makes it to the second round).

The probability that candidate A attains a vote share above θ is given by:

$$Pr\left(\pi_1^A \geq \theta\right) = \frac{1}{2} + \frac{\gamma}{\psi}\left(\frac{1}{2} - \theta + \frac{\psi}{J}\sum_{j=1}^{J}\Delta u_j^{AB}\right) \quad (2.3)$$

For a detailed derivation and expressions for candidate B, see Appendix A.3.

2.2.c. *Candidates' maximization problem.–* For both candidate A and B, her payoff is 1 if she wins the election and 0 otherwise, minus the effort cost incurred during the campaign. Candidates maximize their expected payoff, so this amounts to maximizing the probability of winning minus the effort cost.

2.3. Equilibrium strategies in a single-round election

In a single-round election, candidate C attains a vote share of $\frac{1-\alpha}{J}$, so the probability of winning is the probability of attaining a vote share above $\frac{1}{2}\left(1-\frac{1-\alpha}{J}\right)$. Using equation (2.2),

[13](i) states that $0 < \pi_{j1}^A < 1$. See Appendix A.1. (ii) states that $0 < Pr\left(\pi_1^A \geq \theta\right) < 1$ and $0 < Pr\left(\pi_1^B \geq \theta\right) < 1$. See Appendix A.4.

for candidate A, this is equivalent to solving the following maximization problem:

$$\max_{G^A, \mathbf{q}^A = \left(q_1^A, \ldots, q_J^A\right)} \frac{1}{2} + \left(\frac{\gamma}{J-1+\alpha}\right)\left(\sum_{j=1}^{J-1} \Delta u_j^{AB} + \alpha \Delta u_J^{AB}\right) - \frac{1}{2}\kappa\left(G^A\right)^2 \qquad \text{s.t.} \sum_j q_j^A \leq G^A$$

In a single-round system, in equilibrium, candidates will promise less resources to locality J in comparison to the other localities (see Appendix A.6):

Prediction 2.1. *In a single-round election, for all $j \in \{1, \ldots, J-1\}$, we have that $q_j^A > q_J^A$.*

Prediction 2.1 results from the fact that, for a given level of government spending, candidates' marginal return to allocating resources to locality J is lower than in other localities, leading to less resources promised to locality J in equilibrium.

2.4. Equilibrium strategies in a two-round election

In a two-round election, the probability of winning is the probability of attaining a vote share above $\frac{1}{2}$ in the first round or, if a second round occurs, of attaining a vote share above $\frac{1}{2}$ in the second round:

$$\begin{aligned} & Pr(A \text{ wins in 1st round}) + Pr(\text{second round occurs}) \cdot Pr(A \text{ wins 2nd round}) \\ & = Pr\left(\pi_1^A \geq \frac{1}{2}\right) + \left(1 - Pr\left(\pi_1^A \geq \frac{1}{2}\right) - Pr\left(\pi_1^B \geq \frac{1}{2}\right)\right) Pr\left(\pi_2^A \geq \frac{1}{2}\right) \end{aligned}$$

Using equations (2.2) and (2.3), this corresponds to the following maximization:

$$\begin{aligned} \max_{G^A, \mathbf{q}^A = \left(q_1^A, \ldots, q_J^A\right)} \quad & \left(\frac{1}{2} + \frac{\gamma}{\psi}\left[\frac{1}{2}\left(\frac{\alpha-1}{J-1+\alpha}\right) + \left(\frac{\psi}{J-1+\alpha}\right)\left(\sum_{j=1}^{J-1} \Delta u_j^{AB} + \alpha \Delta u_J^{AB}\right)\right]\right) \\ & + \frac{\gamma}{\psi}\left(\frac{1-\alpha}{J-1+\alpha}\right)\left[\frac{1}{2} + \frac{\gamma}{J}\sum_{j=1}^{J} \Delta u_j^{AB}\right] - \frac{1}{2}\kappa\left(G^A\right)^2 \\ \text{s.t.} \quad & \sum_j q_j^A \leq G^A \end{aligned}$$

In a two-round system, in equilibrium, candidates will promise less resources to locality J in comparison to the other localities (see Appendix A.7):

Prediction 2.2. *In a two-round election, for all $j \in \{1, \ldots, J-1\}$, we have that $q_j^A > q_J^A$.*

As in the single-round system, for a given level of government spending, candidates' marginal return to allocating resources to locality J is lower than in other localities, leading to less resources promised to locality J in equilibrium.

2.5. Comparing single- to two-round systems

In this section, I compare three outcomes under the single- and the two-round systems: (i) politician's allocations to localities, (ii) politician's choice of the overall budget, and (iii) overall inequality in the allocation of resources. To simplify notation, denote the equilibrium allocations and overall budget for candidate A as q_j^{1R} and G^{1R} (for single-round systems) and q_j^{2R} and G^{2R} (for two-round systems).

Prediction 2.3 states that (i) candidates promise more to locality J in a two-round system than in a single-round system and (ii) the measure of candidates' promises to the other localities in a two-round system compared to a single-round system is ambiguous.

PREDICTION 2.3. $q_J^{1R} < q_J^{2R}$ *and* $q_j^{1R} \lesseqgtr q_j^{2R}$ *for all* $j \in \{1, \ldots, J-1\}$.

For locality J – where candidate C has promised the entire government budget – in a single-round system, the fact that voters there will strongly favor candidate C leads candidates A and B to ignore those voters, as the marginal return to allocating resources there is low. In contrast, the two-round system, while not producing a completely equitable distribution, incentivizes candidates to allocate more resources to that locality. The presence of a second round where the third candidate is not present raises the marginal return to allocating resources to that locality. As a result, candidates A and B solicit more votes from locality J in a two-round election.

For the other localities – where candidate C has promised 0 – the change in allocations is ambiguous. If the increase in the overall budget G is small (large) and/or the increase in allocation to locality J is large (small), then allocations to the other localities may decrease (increase). The magnitude of these changes will depend on the parameters of the model, such as the cost of effort, and the functional form of $u_j(\cdot)$.

Prediction 2.4 states that the overall government budget promised is higher in two-round systems than in single-round systems:

PREDICTION 2.4. $G^{1R} < G^{2R}$.

In a single-round system, candidates face lower incentives to invest in *all* localities. The marginal return to investing in any locality is higher in two-round elections. These incentives come from the fact that, in two-round elections, there is a conditionality to winning: To win in the first round, candidates must not only attain the most votes, but must attain a majority of votes, otherwise candidates must compete again in a second round. With higher incentives to invest in all localities, candidates in two-round elections exert more effort to increase the government budget.

The proofs for predictions 2.3 and 2.4 are in Appendix A.8.

When comparing inequality in allocations between the two electoral rules, the outcome of interest is how the allocation of resources to locality J – where candidate C has offered the entire government budget – compares to the allocation to the other localities.[14] Prediction 2.5 states that when comparing two-round systems to single-round systems, the level of inequality in government resources promised is ambiguous:

PREDICTION 2.5. *If $q_j^{1R} > q_j^{2R}$ for all $j \in \{1, \ldots, J-1\}$, then the level of inequality is lower in two-round systems compared to single-round systems. If $q_j^{1R} < q_j^{2R}$ for all $j \in \{1, \ldots, J-1\}$, then the change in the level of inequality is ambiguous.*

The increased allocation in two-round elections to locality J is a force to reduce inequality in allocations between J and the other localities. However, if candidates also offer more to the other localities in two-round elections, the change in the level of inequality will depend on whether the increase in resources to the other localities is greater or less than the increase in resources to locality J.

2.5.a. *Discussion.–* While my model partitions the electorate into geographic localities, voters can be partitioned along other dimensions, such as income, race, or ideology. My model easily translates into these other settings, and the predictions yield the same interpretation. First, two-round systems raise the marginal return to allocating resources to groups of voters that are heavily targeted by other candidates. This results in candidates offering campaign promises that also appeal to these voters. Second, candidates face higher incentives to appeal to all voters, leading them to increase the overall government budget.

My model predicts that two-round elections lead to different outcomes because candidates adapt different campaign strategies by offering policy proposals that appeal to a broader group of voters. However, two-round elections may also lead to different outcomes by (i) causing different types of candidates to enter the race or (ii) causing different types of candidates to win the election. While I do not model candidates' entry decisions, I test for these alternative

[14] I focus on the difference in allocations between locality J and the other localities. However, depending on the functional form of $u_j(\cdot)$, it is possible that relative allocations between the other $\{1, \ldots, J-1\}$ localities become more or less unequal in two-round elections. If differences in $u_j(\cdot)$ across the localities are sufficiently small, any changes in inequality between the $\{1, \ldots, J-1\}$ localities will be small relative to changes in inequality caused by changes in locality J. In the case that $u_j(\cdot) = u(\cdot)$, then all $\{1, \ldots, J-1\}$ localities receive the same allocation, and the only disparity that matters is between J and the other localities.

mechanisms in the empirical analysis. I find evidence suggesting that different outcomes in two-round elections are mostly driven by candidates' strategic responses.

3. Institutional context

Municipalities in Brazil are autonomous governmental entities with an elected executive (a mayor, or *prefeito*) and legislative body (a council of legislators, or *camara de vereadores*). Elections for municipal positions are at large and held for all positions simultaneously every four years. Mayors in municipalities with less than 200,000 voters are elected through a single-round system, while in larger municipalities they are elected in a two-round system. In the two-round system in Brazil, if no candidate receives at least 50% of the votes in the first round, then a second round is held 3 weeks later with the top two candidates.[15] Legislators are elected through an open-list proportional system. Voters cast votes for a mayoral candidate, and either for a legislative candidate or a generic vote for the party. Mayors are limited to serving two consecutive terms, while there is no term limit for legislators.

State electoral authorities, the *Tribunais Electorais Regionais*, register citizens and maintain electoral rolls. Several features of Brazilian elections, mandated either in the federal constitution or by law, facilitate voter turnout on election day. First, voter registration is compulsory and must be completed at least 151 days prior to the election. Second, voting is compulsory for all literate Brazilian citizens between 18 and 69 years of age.[16] Third, elections are held on the first Sunday in October, a day when few voters are at work.

The timing of the announcement of the electoral rule for mayor has varied. In earlier elections, the electoral rule was announced 3–4 months prior to the election. In more recent elections, the number of registered voters has been regularly published, allowing the electoral rule to be known much earlier.

Brazilian elections are a multi-party system, with over 30 political parties registered in the 2016 municipal elections. Mayoral candidates are associated with a party and often a coalition of parties, which are formed prior to the election.[17] Party and coalition affiliations serve as important linkages to the state and federal levels of government (Brollo and Nannicini, 2012).

Once elected, mayors have a broad mandate to provide public goods, particularly in education, health, and local infrastructure. Municipal revenue is a combination of inter-

[15] In my sample, a second round occurs 57% of the time.

[16] Voters may provide justification to a local electoral judiciary office. Absent this justification, voters who fail to vote must pay a small fine and those who fail to vote for three consecutive elections are prevented from accessing public services, such as obtaining a passport or government loans.

[17] Seats for the legislative council are allocated based on the number of votes received by candidates or parties in the coalition.

government transfers and local revenues. The majority of the municipal budget is represented by inter-government transfers (the average within my sample is 65.6% of the budget). The bulk of these transfers come from the state or federal level (on average, 78.0%), and are either constitutional automatic transfers (*Fundo de Participação do Municipios*) or discretionary transfers (*convênios*). Municipalities face considerable flexibility in spending these transfers. Among the automatic transfers, 70% of the funds are unrestricted. While 30% are earmarked, municipalities are only restricted to spending this percentage on health and education.

The allocation of spending on public goods occurs through two main channels. First, the majority of public goods are allocated through the annual budgetary process. Second, mayors and legislators can submit bills requesting specific public works or services. While these actions require joint approval by the mayor and legislature, mayors retain veto power and wield significant influence over the process.

As a result, mayors are important for decisions around both the size of the municipal budget and how municipal funds are allocated. This study focuses on public goods provision in municipal education for several reasons. First, a large fraction of the municipal budget is allocated toward education spending: In 2012, it represented 30.5% of municipal budgets. Second, municipal education is a fairly geographically localized public good. Third, unlike other public goods, for which municipalities share joint responsibility with the state or federal government, the provision of elementary education (*Ensino Fundamental*) is one of the few public goods almost entirely under the jurisdiction of the municipality.

4. Empirical strategy

4.1. Econometric framework

The threshold rule for mayoral elections provides a natural candidate for a regression discontinuity design (RDD). I exploit the 200,000 registered voter threshold, which determines a municipality's electoral rule, to estimate the impact of the electoral rule.

Denote $Y_i(0)$ and $Y_i(1)$ as the potential outcome in municipality i if assigned a single-round and two-round system, respectively. Assignment of the electoral rule is determined by the running variable, the number of registered voters X_i. The assignment variable D_i takes on the value of $D_i = 0$ if $X_i < 200,000$ and $D_i = 1$ if $X_i \geq 200,000$.

Following Imbens and Lemieux (2008) and Calonico et al. (2014), to estimate the treatment effect at the discontinuity, I use a local linear regression specification:

$$Y_{it} = \beta_1 D_{it} + \beta_2 X_{it} + \beta_3 X_{it} \cdot D_{it} + \gamma_t + \varepsilon_{it} \tag{4.1}$$

where for municipality i in election year t, X_{it} is the running variable, D_{it} is the assignment variable, γ_t is an election-year fixed effect, and Y_{it} is the outcome of interest. Each observation in equation (4.1) represents a municipality and election year, or municipality-year. Equation (4.1) amounts to fitting two linear regressions using municipality-years close to the left and to the right of the threshold. β_1 represents the estimate of the local average treatment effect. Standard errors are clustered at the municipality level.

Because the treatment effect is identified only at the threshold, equation (4.1) is estimated using municipality-years close to the threshold. The main analysis uses a 50,000 registered voter window, but robustness is provided for other bandwidths as well as bandwidths selected using data-driven methods (Calonico et al., 2014 and Imbens and Kalyanaraman, 2012).[18]

4.2. Identification

In order for β_1 to represent the causal effect of the electoral rule, the conditional expectation of the potential outcomes must be continuous at the threshold. In the following section, I discuss identification and interpretation of the RDD estimates.

4.2.a. *Violations of smoothness.–* The first way the smoothness assumption can be violated is if the threshold choice is motivated by political or economic factors. There appears to be little evidence for this. The choice of 200,000 registered voters as the threshold was somewhat arbitrary and mainly reflected practical concerns regarding the cost of holding a second round (Chamon et al., 2018; Fujiwara, 2011). In addition, the 200,000 registered voter threshold was set in the 1988 Federal Constitution and has not changed since then. It is unlikely that politicians chose the threshold anticipating which municipalities would be above or below this threshold in 1996 and later.

A second possibility is if municipalities sort across the threshold. Practically, it is difficult for municipalities to selectively sort, since voter registration is mandatory and handled by state electoral authorities. Visually, there are no discontinuities in the distribution of registered voters at the threshold (Figure 2). To test this formally, I estimate the size of the discontinuity in the density of the running variable at the threshold (McCrary, 2008). The size of the discontinuity is both small in magnitude (0.169) and insignificant ($p = 0.389$).[19]

A third violation is if other policies also change discretely at the threshold. While a number of policies in Brazil are implemented using thresholds, these are based off population

[18]Because there is a skewed right tail of municipality sizes due to a few extremely large municipalities such as Rio de Janeiro and São Paulo, the data-driven methods would sometimes select bandwidths larger than the support – ie. larger than 200,000. As a result, the optimal bandwidth is calculated on a subset of elections that lies within the support and is symmetrical around the threshold: 0-400,000 voters.

[19]Due to the skewed right tail of municipality sizes, the size of the discontinuity was estimated off a sample of municipality-years excluding those above the 99.9 percentile of registered voters.

counts and not the number of registered voters. Population and the number of registered voters are highly correlated but do not vary one-to-one (Figure C.1). To the best of my knowledge, there are no other policies at the 200,000 registered voter threshold, and most other policy thresholds are far from 200,000 registered voters. Two exceptions are a constitutional amendment in 2000 that places a salary cap for local legislators at 300,000 inhabitants and a constitutional amendment in 2004 that changes the size of the local legislature at 285,714 inhabitants. I address this in two ways. First, I estimate a placebo regression where the electoral rule is assigned at these population thresholds. I show that there are no discontinuities of a similar size in the mayoral electoral outcomes I examine (Tables C.18 and C.19). However, legislator salaries and legislature size can affect economic outcomes, so I do not estimate placebo regressions for my public goods outcomes. Second, I test whether the probability of being above or below these thresholds changes discontinuously at the 200,000 voter threshold. I do not find evidence of a discontinuity (Figure C.2), indicating that any effect of these policies is balanced across the threshold.[20]

The last possibility is that potential confounds change discretely at the threshold. I test this by estimating equation (4.1) on pre-treatment characteristics of municipalities. I discuss this in detail in the following section.

4.2.b. *Balance on pre-treatment characteristics.–* Since the treatment (here, the two-round election) is determined by the number of registered voters, municipalities can move into the treatment group or be treated multiple times.[21] As a result, I define "pre-treatment" in two ways: 1) prior to the introduction of the electoral system in the 1988 Constitution, and 2) prior to the most recent election in which the municipality was untreated or prior to 1996 if the municipality was never untreated.[22] To measure outcomes for (1), I use the 5% population sample of the 1980 census, the most recent census prior to 1988. To measure outcomes for (2), I use outcomes either from the census prior to the most recent year in a single-round system (the 1991, 2000, or 2010 census) or outcomes from the 1991 census, for municipalities that were in a two-round system in 1996.

There is no significant treatment effect on nearly all outcomes measured prior to the 1988 Constitution (Panel A of Table 3) or prior to the most recent election in a single-round

[20]The regression discontinuity estimate on the probability of being above the 300,000 resident threshold is -0.0350 ($p = 0.735$). The regression discontinuity estimate on the probability of being above the 295,714 resident threshold is 0.0516 ($p = 0.667$).

[21]Municipalities can also move out of treatment. However, while municipalities do experience population decline, none moves below the 200,000 voter threshold.

[22]The earliest electoral data available is 1996, so I cannot observe whether municipalities are treated or untreated prior to 1996. Since there is only one unobserved municipal election after the 1988 Constitution (in 1992) and only 45 municipalities have moved across the 200,000 voter threshold between 1996 and 2016, it is unlikely that many municipalities had their electoral rule change between 1992 and 1996.

system (Panel B of Table 3). The treatment effect is estimated on economic characteristics (unemployment rate, literacy, and income), segregation levels[23] (along income and demographics), and income inequality. To rule out the concern that there are factors that change discontinuously at the threshold and affect which municipalities move into treatment and the length of treatment, I test whether population growth is discontinuous at the threshold. If municipalities that grow at slower rates are more likely to be under the threshold and remain there longer, it would pose an issue for causal identification. However, I do not find that population growth is discontinuous at the threshold.

One exception is a large and significant effect on population density. However, there are several reasons to believe that this is a false positive. First, dropping one outlier municipality reduces the coefficient by 38.2%, suggesting that a few municipalities are driving this effect (although the estimate is still significant at the 10% level). My main results are robust to dropping this outlier municipality. In addition, there is no visible discontinuity and the effect disappears at larger bandwidths (Figures C.3 and C.4). Second, the estimate ($p = 0.047$) is not significant after Bonferroni adjusting the significance threshold for the number of hypotheses tested. Third, the regression discontinuity coefficients across the ten outcomes (from the most recent single-round election) are not jointly significant ($p = 0.339$).

While it is not clear how an imbalance in population density impacts the economic and political outcomes of interest, this poses an issue if politicians are manipulating the composition and size of the electorate by, say, moving citizens or manipulating municipality borders. This does not appear to be the case for two reasons. First, I find no differences in the urbanization rate (a difference of 0.065%, where the single-round mean is 95.191%, $p = 0.943$). Second, this effect is seen in 1980, prior to the introduction of the threshold rule (Panel A of Table 3). Third, there is no evidence that changes in municipality area or population growth change discontinuously across the threshold (Panel B of Table 3). Nevertheless, I include population density as a control in all specifications:

$$Y_{it} = \beta_1 D_{it} + \beta_2 X_{it} + \beta_3 X_{it} \cdot D_{it} + \beta_4 Z_{it} + \beta_5 Z_{it} \cdot D_{it} + \gamma_t + \varepsilon_{it}$$

where Z_{it} is the municipality's population density in the most recent census prior to the election.

4.2.c. *Testing smoothness .–* I test whether my outcomes of interest vary smoothly around the discontinuity. To do this, I estimate placebo regressions where the electoral rule is

[23] This is calculated using the entropy index (see Section 4.4 for a detailed calculation), which measures how far each census sector is from equal representation of all groups. This was calculated separately for income and demographics. For income, the groups are defined by bins of income relative to the minimum wage. For demographics, the groups are defined by sex, age, and literacy.

assigned at other registered voter thresholds between 170,000 and 230,000 voters. This tests whether the effects I estimate are concentrated at the actual threshold of 200,000 voters. I show that there is no discontinuity in my outcomes at thresholds where the treatment does not change (Figures C.20, C.21, C.22, C.23, and C.24).

4.2.d. *Registered voters as a running variable.–* Since treatment (moving into a two-round election) is determined by the number of registered voters, there are interesting implications for the interpretation of the regression discontinuity estimate.

As discussed above, municipalities can move into treatment or be treated multiple times. While this does not invalidate causal identification, it affects whether the treatment effects I estimate are the result of single or multiple treatments.

To investigate this, I test whether the probability that a municipality's *previous* election was a two-round system changes discontinuously at the threshold (Figure C.6). Because the regression discontinuity framework identifies the treatment effect *at* the threshold, treatment effects should be interpreted as the result of the change in this probability *at* the threshold. I find that at the 200,000 registered voter threshold, the probability of having previously been treated is zero and there is no discontinuous jump.[24] As a result, the treatment effects identify the effect of moving into a two-round election for the first time.

4.2.e. *Compliance.–* While not an issue for causal identification, imperfect compliance with treatment can affect the interpretation of the causal estimates (Angrist et al., 1996). In this context, compliance was perfect (Figure C.5). All municipalities below the threshold or where the top candidate received at least 50% held one round. All municipalities above the threshold and where the top candidate did not receive at least 50% held two rounds.

4.3. Data sources

4.3.a. *Electoral data.–* Data on municipality elections come from Brazil's electoral authority (*Tribunal Superior Eleitoral*, or TSE). These are available for 6 municipal elections between 1996 and 2016. The electoral data provides information on the candidates running, the party and coalition each candidate belongs to, and the number of votes received. In total, the data encompasses 32,767 municipal elections covering 5,568 municipalities.

Electoral results are available for each polling station (*seçao eleitoral*), allowing me to observe at a very fine level the number of votes each candidate receives. I use this to measure the geographic distribution of voters for specific candidates at both an *overall* and *candidate*

[24]Because no municipalities move under the threshold, all municipalities under the threshold have 0 probability that the previous election was two-round. The regression discontinuity estimate on the probability of being in a two-round system in the previous election is -0.00802 ($p = 0.870$).

level (see Section 4.4 on the measures used). Baseline results use votes from the first round of elections. Results from the first round are used in order to have similar measures between single- and two-round elections, but results are robust to using votes from the final round (the first round in single-round elections and the second round in two-round elections).

4.3.b. *Public goods provision in schools.–* To measure public goods provision in municipal public schools, I use data provided in the 1997-2016 School Census (*Censo Escolar*). This census is conducted annually by the research arm of the Ministry of Education, *Instituto Nacional de Estudos e Pesquisas Educacionais.*

I use the Census to observe the level of resources present in schools offering elementary education. I calculate a measure of resources present in each school across two categories: equipment and infrastructure. Equipment includes *movable* elements, such as the number of computers and availability of air conditioning (see Table B.1 for the full list). Infrastructure includes *immovable* elements, such as the number of classrooms, type of sanitation, and availability of a library (see Table B.2 for the full list). I construct indices of these resources, separately for equipment and infrastructure, by taking the principal component of these variables and computing the school's percentile rank within the country for each year.[25]

4.4. Measuring concentration of voters

In this section, I use the following notation. In municipality m, there are K_m candidates and I_m polling stations. p_{mk} is the fraction of voters for candidate k in municipality m, p_{imk} is the fraction of voters for candidate k in polling station i, n_{im} is the number of voters in polling station i, and N_m is the number of voters in municipality m.

4.4.a. *Overall geographic concentration of voters.–* To measure the *overall* geographic concentration of voters, I use three indices that are commonly used in the racial segregation literature to measure multigroup spatial segregation: the coefficient of variation, the fractionalization index, and the entropy index. These indices and their properties are described in more detail in White (1986) and Reardon and Firebaugh (2002). The indices assume a value of 1 if there is full geographic concentration of voters: Each polling station contains voters for only one candidate. The indices assume a value of 0 if there is full geographic dispersion of voters: Each polling station contains the same composition of voters as the municipality as a whole.

[25] The variables collected in the School Census under each category varied from year to year. This makes it difficult to compare the raw PCA index across years. Calculating a school's percentile rank across all schools for that year allows for valid comparisons across years.

The coefficient of variation, s_m, is defined as:

$$s_m = \frac{1}{K_m - 1} \sum_{k=1}^{K_m} \sum_{i=1}^{I_m} \frac{n_{im}}{N_m} \frac{(p_{imk} - p_{mk})^2}{p_{mk}}$$

This index is interpreted as the square deviation of voter composition in polling stations from voter composition in the municipality. Dividing by $K_m - 1$ keeps the index between 0 and 1. When each polling station has the same composition as the municipality ($p_{imk} = p_{mk}$), the index takes a value of 0.

The fractionalization index, f_m, is calculated in two steps. The average fractionalization within each polling station, $\bar{f}_m$, is defined as:

$$\bar{f}_m = \sum_{k=1}^{K_m} \sum_{i=1}^{I_m} \frac{n_{im}}{N_m} p_{imk} (1 - p_{imk})$$

The overall fractionalization within the municipality, $\hat{f}_m$, is defined as:

$$\hat{f}_m = \sum_{k=1}^{K_m} p_{mk} (1 - p_{mk})$$

The final measure, the fractionalization index, f_m, is defined as:

$$f_m = \frac{\hat{f}_m - \bar{f}_m}{\hat{f}_m}$$

Fractionalization, also known as the interaction index, measures the probability that two members within a population chosen at random are from different groups. There are two ways to interpret f_m. One, f_m is the average concentration within polling stations, normalized by the level in the municipality to keep the index between 0 and 1. Two, f_m is the fraction of concentration in the municipality that is due to differences in voter composition between polling stations. When each polling station has the same concentration as the municipality – or when there are no differences between polling stations and $\hat{f}_m = \bar{f}_m$ – the index takes a value of 0. When each polling station contains only one type of voter – or when there are large differences between polling stations and $\bar{f}_m = 0$ – the index takes a value of 1.

The entropy index, h_m, is also calculated in two steps. The average entropy within each polling station, $\bar{h}_m$, is defined as:

$$\bar{h}_m = -\sum_{k=1}^{K_m} \sum_{i=1}^{I_m} \frac{n_{im}}{N_m} p_{imk} \ln p_{imk}$$

The overall entropy within the municipality, $\hat{h}_m$, is defined as:

$$\hat{h}_m = -\sum_{k=1}^{K_m} p_{mk} \ln p_{mk}$$

The final measure, the entropy index, h_m, is defined as:

$$h_m = \frac{\hat{h}_m - \bar{h}_m}{\hat{h}_m}$$

Entropy measures how far the population is from equal representation of all groups. The interpretation and range of values of the entropy index h_m are the same as that of the fractionalization index f_m.

In my sample, the correlation between these indices is between 0.89 and 0.96.[26] Conceptually, the three indices can be thought of as measures of the average deviation of the composition of polling stations from that of the municipality.

Sensitivity of overall concentration to the number of candidates.– The indices may mechanically change due to the number of candidates. I discuss this briefly below; the majority of this discussion derives from Reardon and Firebaugh (2002).

For the coefficient of variation, before dividing by $K_m - 1$, the index attains a maximum value of $K_m - 1$ and as a result would depend mechanically on the number of candidates. Dividing by $K_m - 1$ nets out this mechanical effect.

For the fractionalization index, $\sum_{i=1}^{I_m} p_{imk}(1 - p_{imk})$ ranges from $1/K_m$ to 1. As a result, the polling station average, $\bar{f}_m$, and municipality measure, $\hat{f}_m$, depend mechanically on the number of candidates. Normalizing $\bar{f}_m$ by $\hat{f}_m$ removes part of this mechanical effect.

For the entropy index, $\sum_{i=1}^{I_m} p_{imk} \ln p_{imk}$ ranges from 0 to $\ln K_m$. As a result, the polling station average, $\bar{h}_m$, and municipality measure, $\hat{h}_m$, depend mechanically on the number of candidates. Normalizing $\bar{h}_m$ by $\hat{h}_m$ removes part of this mechanical effect.

While these indices no longer monotonically depend on the number of candidates, these indices may still be affected by the number of candidates. I address this concern with several robustness exercises in Section 5.4.a.

Candidate-level geographic concentration of voters.– To capture a *candidate*-level measure of the spatial distribution of voters, I use the standard deviation in vote shares across polling stations. This measure describes whether a candidate's supporters are spread across many or concentrated within a few polling stations. For candidates who receive votes from many areas

[26] The correlation coefficient between the coefficient of variation and fractionalization is 0.956; between the coefficient of variation and entropy is 0.892; and between fractionalization and entropy is 0.929.

in the municipality, we should expect to observe the vote share varying less across polling stations. The standard deviation in a candidate's vote share, σ_{mk}, is defined as:

$$\sigma_{mk} = \left(\frac{1}{I_m - 1} \sum_{i=1}^{I_m} \left(p_{imk} - \frac{1}{I_m} \sum_{i=1}^{I_m} p_{imk} \right)^2 \right)^{1/2}$$

Using only the top two candidates.– In some cases, I calculate the indices using only vote shares from the top two candidates. To do this, I assume that only the top two candidates are in the race and use each candidate's vote share out of the top two. These indices should be interpreted as the geographic distribution of voters for the top two candidates. The advantage of doing this is that it allows me to keep keep the number of candidates fixed and ignore the potential dilution of votes from lower-placed candidates. This is done when comparing to the second round of two-round elections or to show robustness of my results to the number of candidates. Note that, by construction, the standard deviation of votes for the 1st place candidate σ_{m1} is the same as that of the 2nd place candidate σ_{m2}.

5. The effect of the two-round system

I present three main results. First, candidates in two-round elections receive broader geographical support. Second, once in office, politicians elected under two-round systems provide more resources to schools and distribute these resources more equitably. These results suggest that politicians in two-round systems are represented by a broader group of voters and that this affects how public goods are provided in the municipality. Third, I find that educational outcomes are improved in two-round municipalities.

5.1. The geography of votes

Do candidates in two-round elections secure broader bases of support? I provide evidence that voters are *overall* less geographically concentrated, and that it is the *top two* candidates who receive support from a geographically broader group of voters.

5.1.a. *Geographic concentration of voters.–* The coefficients on the coefficient of variation, fractionalization index, and entropy index are negative, indicating that voters for specific candidates are overall less geographically concentrated in two-round elections (Figure 4 and Panel A of Table 6). The composition of voters in polling stations is closer to the composition of voters in the municipality as a whole, indicated by the coefficient of variation, and the composition of voters in polling stations is on average less concentrated, indicated by the fractionalization and entropy indices. Municipalities in two-round systems experience a reduction of 0.0087, 0.0118, and 0.0082 in the coefficient of variation, fractionalization, and entropy of voters, respectively. These coefficients correspond to 45.6%, 43.9%, and 27.4%, respectively, of the average level in single-round municipalities within the bandwidth.

Next, I test whether all or only some candidates obtain support from geographically broader constituencies. I find that voters for the top two candidates are less concentrated in two-round systems, but that voters for the third and fourth placed candidates in two-round systems are concentrated similarly to voters of the same-placed candidates in single-round systems (Figure 5 and Panel B of Table 6). The estimates for the top two candidates are similar in magnitude: The first placed candidate experiences a 0.0167 reduction in variance (20.9% of the single-round mean), and the second placed candidate experiences a 0.0142 reduction (18.9% of the single-round mean). The estimates for the third and fourth placed candidates are close to zero and insignificant.

The effects on overall concentration are significantly stronger when using vote shares from the top two candidates only (Table 8).[27] This further indicates that the bases of support of the top two candidates drive the reduced concentration of voters in two-round elections.

These effects are not limited to vote shares in the first round. Using only vote shares from the top two candidates, I compare concentration in single-round elections with concentration in the second round of two-round elections. Voters in the second round are also less geographically concentrated (Panel A of Table 7). Additionally, I compare concentration between the first and second round in two-round municipalities. While these estimates are correlational and not causal, concentration in the second round is lower than in the first round, indicating that candidates consolidate their voter bases between rounds (Panel B of Table 7).

5.1.b. *Voter engagement.–* These results suggest that two-round elections lead to greater inclusiveness, as voters from more geographical areas are represented. I find inclusiveness along another dimension: voter behavior. Specifically, I find higher rates of voter engagement in two-round elections. While turnout is unaffected (which is expected, as turnout is mandatory in Brazil), the number of blank and invalid ballots is significantly lower in two-round municipalities (Table 9). The number of blank and invalid ballots plausibly corresponds to the number of dissatisfied or disinterested voters (Gonzales et al., 2019).[28] The reduction

[27] I compare the coefficients in Table 6 with Table 8. The p-values for the difference between the estimates are 0.024 (for the coefficient of variation), 0.052 (for fractionalization), and 0.147 (for entropy).

[28] Ballots can be invalid or blank for a number of reasons. For example, municipalities with higher numbers of illiterate voters will have more blank and invalid ballots (Fujiwara, 2015). Since the illiteracy rate is not discontinuous across the threshold and all municipalities used electronic voting by 2000 (which reduced the number of unintentional errors), I interpret the difference in the number of blank and invalid ballots as voter engagement. Gonzales et al. (2019) provide empirical evidence for this interpretation, as they find that forced

suggests that voters in two-round elections engage in the electoral process at higher rates.

5.2. The allocation of municipal resources

I next investigate the impact of the two-round system on public goods provision. If politicians secure broader bases of support in two-round elections, they may also provide public goods differently once in office. I provide evidence that two-round elections impact both the *level* and *distribution* of municipal resources.

5.2.a. *Municipal schools.–* Municipal schools in two-round municipalities have, on average, more resources than those in single-round municipalities (Figure 10 and Columns 1 and 2 in Table 12). A school in a two-round municipality is, on average, 0.081 percentiles and 0.057 percentiles higher in the national distribution of equipment resources and infrastructure resources, respectively. The coefficient on infrastructure resources is smaller and less significant; this may be because infrastructure is more difficult to manipulate. Allocating new infrastructure, such as gymnasiums, requires significantly more time and capital than allocating equipment, such as computers. This is reflected when looking at the estimates separately by year in the mayoral term: Infrastructure levels are less responsive to the electoral cycle (Column 1 in Table C.7).

In addition to differences in the overall levels, there is less variance in the resources present in schools in two-round municipalities (Figure 11 and Columns 1 and 2 in Table 12). The standard deviation in equipment resources is 0.018 percentiles lower in two-round municipalities (15.9% of the single-round mean). Although the estimate on the standard deviation in infrastructure resources is of a similar magnitude (-0.021 percentiles), the difference is not significant.

It is instructive to estimate effects for schools at different parts of the distribution in the municipality. Intuitively, if the variance in resources is lower in two-round municipalities, we should expect to see that schools with the least (most) resources in the municipality have more (less) resources.

I group schools into quartiles, which are defined by first calculating each school's percentile in the municipality distribution prior to the election, and then assigning the school to one of four quartiles. The increased level of resources is concentrated in schools located at the lower end of the distribution (Table 13).[29] Schools at the bottom 25% of the distribution are 0.082 percentiles higher in equipment resources and 0.116 percentiles higher in infrastructure resources. Schools in the second quartile experience positive, but smaller, gains – 0.066

electoral participation increases the number of blank and invalid ballots cast.

[29]The p-values for the difference in the regression discontinuity estimates between schools at the 1st quartile compared to the 4th quartile is 0.057 (for equipment) and 0.038 (for infrastructure).

percentiles in equipment resources and 0.102 percentiles in infrastructure resources. There is no significant difference in resources at the top of the distribution.

5.3. Downstream outcomes

If mayors provide more public goods and distribute them more equitably, does this translate into downstream economic outcomes? I find, in two-round municipalities, improvements in education outcomes but limited effects on economic outcomes.

5.3.a. *Education outcomes.*– I measure education outcomes in four ways. Using the School Census, I measure the drop-out, failing, and passing rate in schools across the municipality. Using the 2000 and 2010 Demographic Censuses, I measure the literacy rate among cohorts who were of elementary school age during the electoral term. I find that in two-round municipalities, drop-out rates are significantly lower, 1.65 percentage points off a baseline of 3.21 percentage points; literacy rates are also significantly higher among elementary cohorts, 1.20 percentage points off a baseline of 91.45 percentage points (Figure 15 and Panel A of Table 16). While the drop-out rates are lower, this does not necessarily lead to improved passing rates or worsened failing rates. These results suggest that differential public goods provision in two-round municipalities lead to some improvements in education outcomes.

5.3.b. *Economic outcomes.*– If two-round municipalities lead to improved education outcomes, this may result in improved economic outcomes. I find some – though limited– improvements on broader economic outcomes (Panel B of Table 16). Using the 2000 and 2010 Demographic Census, I measure the fraction of low-income households, income per capita, and the unemployment rate. These outcomes are measured between 2 and 10 years after the election.[30] Using the 1997-2013 NOAA night lights series, I measure the mean night lights level in the municipality. I find that the fraction of low-income households is significantly lower. This may be a false positive, as the coefficient is large relative to the mean; controlling for the pre-election level reduces the coefficient to -1.477 percentage points ($p = 0.0872$). Coefficients on the other outcomes suggest that two-round municipalities lead to improved economic outcomes (income per capita is higher, unemployment is lower, and night lights is higher), but that this effect is not significant.

While two-round elections may simply not lead to improved economic outcomes, there may be several reasons that this is a false negative. First, the public goods and policies that Brazilian mayors implement may have limited or no influence on economic outcomes.

[30] For the 1996 elections, outcomes are observed 4 years later in the 2000 Demographic Census. For the 2000, 2004, and 2008 elections, outcomes are observed 10, 6, and 2 years later in the 2010 Demographic Census.

Outside of health and education, mayors are also responsible for local infrastructure, such as public transportation; urban planning; and public health, such as sanitation. It is possible but unlikely that these policies do not influence outcomes such as income and employment. Second, it could be that there are no improvements in economic outcomes in the short term. Outcomes such as income and night lights may take more than 2 to 10 years to improve. Third, improvements in economic outcomes may not be experienced in aggregate, but only among certain populations. For example, I find that increased school resources are concentrated in schools at the bottom of the distribution. This may explain the significant effect on the low-income rate, which reflects improved outcomes for the poorest households, and not on more aggregate economic outcomes.

5.4. Robustness of the main results

5.4.a. *Bias in measures of concentration.–* I address two potential sources of mechanical bias in the concentration indices.

One is the size of the parcels (here, polling stations) used to calculate the indices. The number of voters assigned to each polling station is regulated by the TSE, so in principle all polling stations should be of similar size (on average, there are 272 valid votes at each polling station). Empirically, the difference in the number of valid votes at each polling station is small in magnitude (3.605 votes) and insignificant ($p = 0.764$).

A second concern, although potentially a mechanism, is the increased number of candidates in two-round elections (Table 9). While the indices may be affected by the number of candidates, the direction of bias is not monotonic (see Section 4.4.a). Nevertheless, I perform three robustness checks.

First, while the number of candidates is a bad control, as it is an endogenous outcome, controlling for the number of candidates does not affect the qualitative results (Table C.8). Second, I simulate the effect of adding an additional candidate to all single-round elections.[31] For most measures, the estimated bias is small and, for the coefficient of variation, of the wrong sign (see the row "Potential bias" in Table 6). However, the bias is substantial for the entropy index. Third, I re-calculate all the outcomes using only the vote shares from the top two candidates, in order to maintain the same number of candidates across single- and two-round elections. Doing so does not substantially change the results (Table 8).

[31] I do this by adding in a last-place candidate to each polling station. I assign that candidate the number of votes the average last-place candidate receives (1.5%). I take away a proportionate number of votes from the other candidates, to ensure that the total number of voters remains the same. I then calculate the change between the actual value and the simulated value.

5.4.b. *Calculating the resource index.–* The resource index is constructed by taking the first principal component of each school's resources, then calculating a school's percentile rank in the national distribution. Using a resource index constructed by taking the z-scores of each school's resources, then calculating a school's percentile rank in the national distribution, does not affect the qualitative results (Table C.9).

5.4.c. *RDD design.–* I investigate the robustness of my results to the regression discontinuity design. First, these results are not driven by the choice of bandwidth, whether fixed or chosen by a data-driven method (Figures C.10 and C.11 for voter concentration outcomes; Figures C.12 and C.13 for municipal school outcomes; Figure C.14 for education outcomes). The estimates maintain similar magnitudes and mostly retain significance for bandwidths out to 150,000 voters, although the estimates for the standard deviation in school resources decline and are not significant at larger bandwidths. Second, dropping controls from the regression – namely, population density and election-year fixed effects – does not substantially affect the results, although the results are noisier (Tables C.15, C.16, and C.17).

5.4.d. *Placebo tests.–* As discussed in section 4.2.a, I do not find that policy thresholds at 300,000 inhabitants and at 285,714 inhabitants confound my results. First, in placebo regressions, I do not find similar effects at these thresholds for the mayoral electoral outcomes (Tables C.18 and C.19).[32] Second, I show that any effects of these policies is balanced across both sides of the threshold (Figure C.2).

I also show that there are no discontinuities at placebo thresholds in registered voters (170,000; 180,000; 190,000; 210,000; 220,000; 230,000), indicating that the outcomes are relatively continuous at places where the treatment does not change (Figures C.20 and C.21 for voter concentration outcomes; Figures C.22 and C.23 for municipal school outcomes; Figure C.24 for education outcomes). The treatment effect is isolated to the actual threshold: There are no estimates with the same size and significance as at the actual threshold.

[32] Since the number of inhabitants is not the same as the number of registered voters (nor do they map 1:1), to maintain comparability with the baseline estimates, I use a bandwidth of 125,000 inhabitants. This bandwidth was determined by taking half of the population range of municipalities in my 50,000 voter bandwidth (the smallest municipality is 182,082 inhabitants and the largest 434,474 inhabitants). Since the salary cap was implemented in 2000, I estimate this using elections after 2000. Since the legislature size was implemented in 2004, I estimate this using elections after 2004.

6. Discussion

In my model, two-round elections lead to different outcomes because candidates adjust their strategies by offering policies that appeal to a broader group of voters. However, there may be other explanations. Namely, different types of candidates may *enter* two-round elections, or different types of candidates may *win* two-round elections.[33] In the following section, I explore these explanations and provide suggestive evidence that candidates' strategic behavioral responses explain a larger part of the effect of the two-round system.

6.1. Selection in candidates

Candidates in two-round elections may have a broader group of supporters because different types of candidates *enter* electoral races. For example, two-round elections may incentivize candidates who are more competent or more relatable to enter. I do not find that candidates in two-round elections are significantly different from candidates in single-round elections along observable characteristics: There is no significant difference in age, sex, educational attainment, or state of birth (Panel A of Table 17). Candidates also do not have different occupational backgrounds (Panel B of Table 17).

A second possibility is that different types of candidates *win* in two-round elections. However, I also find no evidence that election winners are observably different along the aforementioned characteristics (Table 18).

I do find differences in political affiliation among candidates who enter the elections. There are more candidates from small parties and who previously ran as mayoral candidates in two-round elections (Panel A of Table 19), but they are not more likely to win (Panel B of Table 19).[34] Incumbent candidates are less likely to win two-round elections, but this result is not robust to other bandwidths. Since smaller parties are more likely to appeal to narrower electorates, it is unlikely that the identity of the candidate's party explains the reduced concentration in vote shares.

Small candidates may enter two-round elections for several reasons. First, they may be building support for subsequent elections, which is one reason more candidates with previous campaign experience enter two-round elections. Second, these candidates may seek to gain

[33] A third possibility is that voters behave differently in two-round elections, either through turnout or strategic voting. As mentioned earlier, in Brazil, turnout is mandatory, and so not a major factor. Regarding strategic voting, Fujiwara (2011) finds that third placed and lower candidates receive higher vote shares in two-round elections and argues that voters behave less strategically. While this paper is not focused on voter behavior, I interpret strategic behavioral responses of candidates as an equilibrium outcome that can arise from the electoral rule directly or indirectly through the electoral rule's impact on voter behavior.

[34] I define a "small party" as any party that is not one of the top 5 parties in Brazil by national membership.

positions in the elected administration. Third, and most interestingly, these candidates may want to influence the top candidates' platforms. In my model, in the two-round election, the third candidate indirectly influences the top candidates' platforms by leading these candidates to offer policies that appeal to her voters. In reality, the third candidate may influence these policies more directly, either by strategically shaping her own policy or direct bargaining.

6.2. Strategic behavioral responses by candidates

I find limited evidence that two-round elections cause different types of individuals to enter or win the elections. Instead, I find evidence that is consistent with candidates adjusting their strategies during the election.

6.2.a. *Reelection incentives.–* I rule out that candidates' different strategic behaviors are caused by changes in reelection incentives. I do not find a significant difference in the treatment effect on municipal school resources between mayors in their first term (eligible for reelection) or their second (Table 14).

However, it may be the case that term limits are not relevant for mayors in my sample. In Brazil, mayors face strong incentives to reward their supporters after the election, even those who are not eligible for reelection. Municipal mayors are viewed as important local operatives who deliver votes for their parties at the state and federal levels, and it is not uncommon for mayors to seek office themselves at the state and federal levels.[35]

This suggests that, in line with the model's predictions, different outcomes in two-round elections come instead from candidates adopting different strategies during the campaign, which I discuss in further detail in the following sections.

6.2.b. *Concentration between the first and second round.–* In Section 5.1, I find that the concentration of voters for the top two candidates decreases between the first and second round of the two-round election. In my model, candidates do not change platforms between rounds. In reality, it is more likely that candidates adjust their strategies but are constrained in the extent to which they can do so. There is empirical evidence that candidates qualifying for the second round rally votes from supporters of the candidates eliminated after the first round (Pons and Tricaud, 2018). The decrease in concentration between rounds suggests that candidates adjust their strategies to rally these voters.

[35]In my sample, 46.2% of mayors stand as candidates in a state or federal election after their term, and 83.7% of mayors have stood at this level either before or after their term. In larger cities, the position of the municipal mayor is often viewed as a stepping stone to higher office.

6.2.c. *Campaign financing.*– For the 2004-2012 elections,[36] I am able to observe one aspect of candidates' strategies: how the campaigns are financed. I find that candidates in two-round elections receive *less* donations, both on average (Panel A of Table 20) and between the top two candidates (Panel B of Table 20). While the outcomes are noisy, the effects are strongest for donations from corporations: Candidates in two-round elections receive *fewer* donations from corporations.[37]

These results suggest that candidates in two-round elections run their campaigns differently. For example, candidates may offer policies that appeal to individuals rather than corporations. To the extent that corporations represent a narrower swath of the electorate, the pattern in donations suggests that candidates in two-round elections adopt strategies to appeal more broadly.

7. Conclusion

A majority of countries use two-round systems to choose their leaders, and an increasing number of countries are adopting this system over time. This paper studies how the electoral rule leads to the election of politicians who represent broader or narrower groups of voters and distribute state resources differently. To identify the effect of the two-round system, I leverage a unique rule in Brazilian municipal elections: Municipalities above a threshold of registered voters hold two rounds, whereas municipalities below this threshold hold a single round. I find that candidates in two-round municipalities are represented by a geographically broader group of voters. Once in office, mayors elected under two-round systems provide more resources to municipal schools and distribute these resources more evenly across schools. I find evidence that downstream educational outcomes are also improved in two-round municipalities.

I present a model to highlight why two-round elections may lead to these empirical results. In my model, the second round raises the marginal return to allocating resources. This creates incentives to (i) increase the government budget and (ii) appeal to voters that candidates in a single-round system would otherwise ignore. The main intuition is that two-round elections perform two functions. First, they require a candidate to attain at least 50% of the vote in order to win. Second, the second round effectively limits the number of candidates to two.

My model proposes that two-round elections lead to different outcomes because candidates adopt different strategies. I find evidence suggesting that two-round systems cause candidates

[36] I exclude the 2016 elections from the sample, as a new campaign finance law was passed in 2016 that banned donations from corporations.

[37] Donors are classified as corporations or individuals, depending on whether a CPF (individual identification number) or CNPJ (corporate identification number) was filed for the donation.

to adjust their behavior rather than cause different types of candidates to enter the races. First, I find evidence indicating that candidates consolidate their bases of support between rounds. Second, candidates in two-round elections adopt more broadly appealing strategies, resulting in fewer corporate donors for their campaigns.

If two-round systems lead to positive outcomes, why is the two-round system not more widely used? The reality is that there are potential trade-offs. First, it may be costly for voters to vote twice in a short span of time. In Brazil, turnout is *lower* in the second round compared to the first round. Uncovering the reasons for this will help better explain the costs of two-round systems. Second, I find that in two-round systems, individuals at lower parts of the distribution benefit. As a result, there may be opposition to implementing two-round systems by richer households or the elite. Identifying barriers to adopting more inclusive institutions is crucial to understanding the process of political reform. Finally, two-round elections may result in better outcomes only when the electorate is composed of many small groups. Brazil is a multi-party system, and the average single-round election has 4.6 candidates running. Incentivizing candidates to incorporate smaller groups in the coalition may lead to better outcomes. This may not translate to contexts where the electorate is composed of two large groups. Providing more empirical evidence of the causal effect of two-round systems in different contexts would greatly advance our overall understanding of electoral systems.

References

Acemoglu, Daron and James A. Robinson (2008). "Persistence of Power, Elites, and Institutions." *American Economic Review* 98.1, pp. 267–293.

Acemoglu, Daron, Tristan Reed, and James A. Robinson (2014). "Chiefs: Economic Development and Elite Control of Civil Society in Sierra Leone." *Journal of Political Economy* 122.2, pp. 319–368.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin (1996). "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91.434, pp. 444–455.

Bardhan, Pranab and Dilip Mookherjee (2006). "Decentralisation and Accountability in Infrastructure Delivery in Developing Countries." *The Economic Journal* 116, pp. 101–127.

Bordignon, Massimo, Tommaso Nannicini, and Guido Tabellini (2016). "Moderating Political Extremism: Single Round versus Runoff Elections under Plurality Rule." *American Economic Review* 106.8, pp. 2349–2370.

Bormann, Nils-christian and Matt Golder (2013). "Democratic Electoral Systems around the world, 1946-2011." *Electoral Studies* 32.2, pp. 360–369.

Bouton, Laurent (2013). "A Theory of Strategic Voting in Runoff Elections." *American Economic Review* 103.4, pp. 1248–1288.

Bouton, Laurent and Gabriele Gratton (2015). "Majority Runoff Elections: Strategic Voting and Duverger's Hypothesis." *Theoretical Economics* 10.2, pp. 283–314.

Bouton, Laurent, Jorge Gallego, Aniol Llorente-Saguer, and Rebecca Morton (2019). "Runoff Elections in the Laboratory." *NBER Working Paper*.

Brollo, Fernanda and Tommaso Nannicini (2012). "Tying Your Enemy's Hands in Close Races: The Politics of Federal Transfers in Brazil." *American Political Science Review* 106.4, pp. 742–761.

Burden, Barry C. (1997). "Deterministic and Probabilistic Voting Models." *American Journal of Political Science* 41.4, pp. 1150–1169.

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró-i Miquel (2015). "The Value of Democracy: Evidence from Road Building in Kenya." *American Economic Review* 105.6, pp. 1817–1851.

Callander, Steven (2005). "Duverger's Hypothesis, the Run-off Rule, and Electoral Competition." *Political Analysis* 13.3, pp. 209–232.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik (2014). "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82.6, pp. 2295–2326.

Carey, John M. and Matthew Soberg Shugart (1993). "Incentives to Cultivate a Personal Vote: A Rank Ordering of Electoral Formulas." *Electoral Studies* 14.4, pp. 417–439.

Chamon, Marcos, Joao De Mello, and Sergio Firpo (2018). "Electoral Rules, Political Competition and Fiscal Expenditures: Regression Discontinuity Evidence from Brazilian Municipalities." *Journal of Development Studies*.

Cipullo, Davide (2019). "Voting Systems, Incentives and Fiscal Policy: Evidence from Runoff and Plurality." *Working Paper.*

Cox, Gary W. (1997). *Making Votes Count: Strategic Coordination in the World's Electoral Systems.* Cambridge University Press.

Eggers, Andrew C. (2013). "Proportionality and Turnout: Evidence from French Municipalities." *Comparative Political Studies* 48.2, pp. 135–167.

Fujiwara, Thomas (2011). "A Regression Discontinuity Test of Strategic Voting and Duverger's Law." *Quarterly Journal of Political Science* 6, pp. 197–233.

— (2015). "Voting Technology, Political Responsiveness and Infant Health: Evidence from Brazil." *Econometrica* 83.2, pp. 423–464.

Fujiwara, Thomas and Leonard Wantchekon (2013). "Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin." *American Economic Journal: Applied Economics* 5.4, pp. 241–255.

Genicot, Garance, Laurent Bouton, and Micael Castanheira (2018). "Electoral Systems and Inequalities in Government Interventions." *NBER Working Paper.*

Golden, Miriam and Brian Min (2013). "Distributive Politics Around the World." *Annual Review of Political Science* 16, pp. 73–99.

Gonzales, Mariella, Gianmarco León-Ciliotta, and Luis R. Martínez (2019). "How Effective Are Monetary Incentives to Vote? Evidence from a Nationwide Policy." *Working Paper.*

Hodler, Roland and Paul A. Raschky (2014). "Regional Favoritism." *Quarterly Journal of Economics* 129.2, pp. 995–1033.

Imbens, Guido and Karthik Kalyanaraman (2012). "Optimal bandwidth choice for the regression discontinuity estimator." *Review of Economic Studies* 79.3, pp. 933–959.

Imbens, Guido W. and Thomas Lemieux (2008). "Regression discontinuity designs: A guide to practice." *Journal of Econometrics* 142.2, pp. 615–635.

Lindbeck, Assar and Jörgen W. Weibull (1987). "Balanced-Budget Redistribution as the Outcome of Political Competition." *Public Choice* 52.3, pp. 273–297.

Lizzeri, Alessandro and Nicola Persico (2001). "The Provision of Public Goods under Alternative Electoral Incentives." *American Economic Review* 91.1, pp. 225–239.

— (2005). "A Drawback of Electoral Competition." *Journal of the European Economic Association* 3.6, pp. 1318–1348.

McCrary, Justin (2008). "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics* 142.2, pp. 698–714.

Milesi-Ferretti, Gian Maria, Roberto Perotti, and Massimo Rostagno (2002). "Electoral systems and public spending." *Quarterly Journal of Economics* 117.2, pp. 609–657.

Myerson, Roger B. (1993). "Incentives to Cultivate Favored Minorities Under Alternative Electoral Systems." *American Political Science Review* 87.4, pp. 856–869.

Osborne, Martin J. and Al Slivinski (1996). "A Model of Political Competition with Citizens-Candidates." *Quarterly Journal of Economics* 111.1, pp. 65–96.

Pellicer, Miquel and Eva Wegner (2013). "Electoral Rules and Clientelistic Parties: A Regression Discontinuity Approach." *Quarterly Journal of Political Science* 8, pp. 1–33.

Persson, Torsten and Guido Tabellini (1999). "The size and scope of government: Comparative politics with rational politicians." *European Economic Review* 43, pp. 699–735.

— (2000). *Political Economics: Explaining Economic Policy*. Cambridge, MA: MIT Press.

— (2003). *The Economic Effects of Constitutions*. Cambridge, MA: MIT Press.

— (2004). "Constitutions and Economic Policy." *Journal of Economic Perspectives* 18.1, pp. 75–98.

Persson, Torsten, Guido Tabellini, and Francesco Trebbi (2003). "Electoral Rules and Corruption." *Journal of the European Economic Association* 1.4, pp. 958–989.

Piketty, Thomas (2000). "Voting as Communicating." *Review of Economic Studies* 67, pp. 169–191.

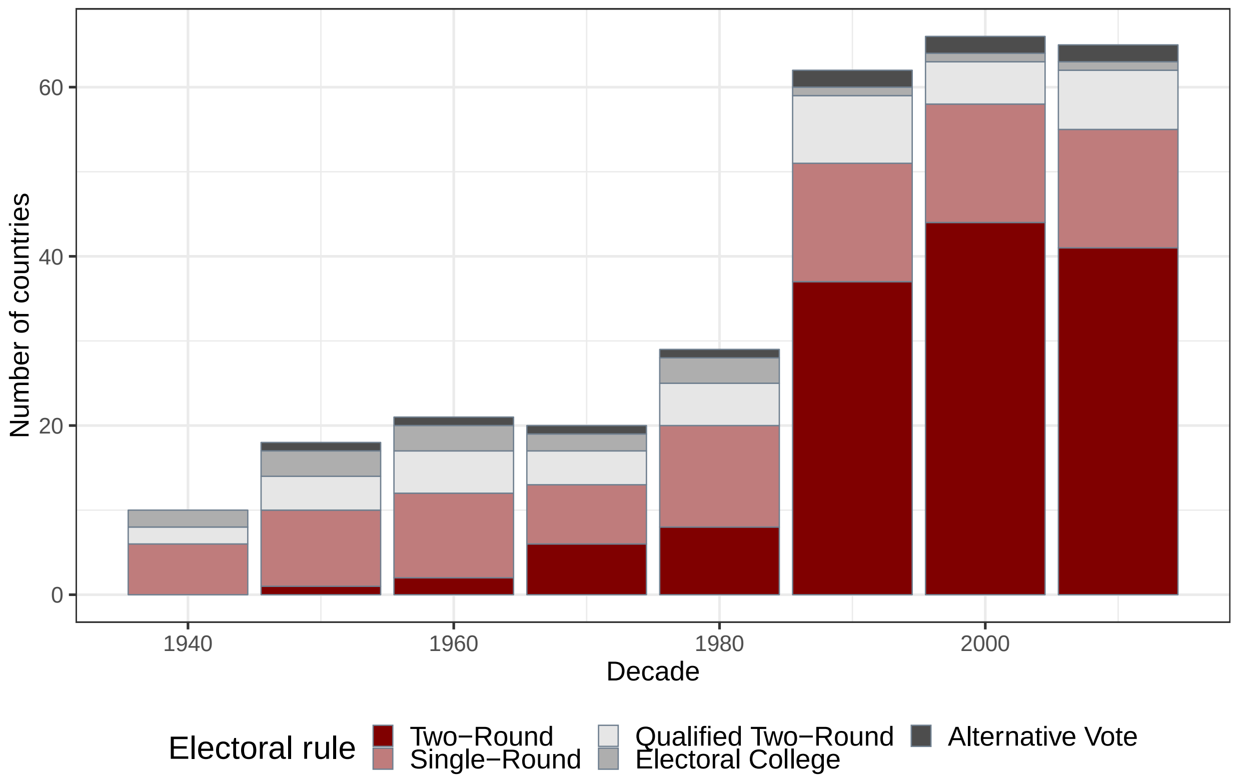
Pons, Vincent and Clemence Tricaud (2018). "Expressive Voting and Its Cost: Evidence from Runoffs with Two or Three Candidates." *Econometrica* 86.5, pp. 1621–1649.

Reardon, Sean F. and Glenn Firebaugh (2002). "Measures of Multigroup Segregation." *Sociological Methodology2* 32, pp. 33–67.

White, Michael J. (1986). "Segregation and Diversity Measures in Population Distribution." *Population Index* 52.2, pp. 198–221.

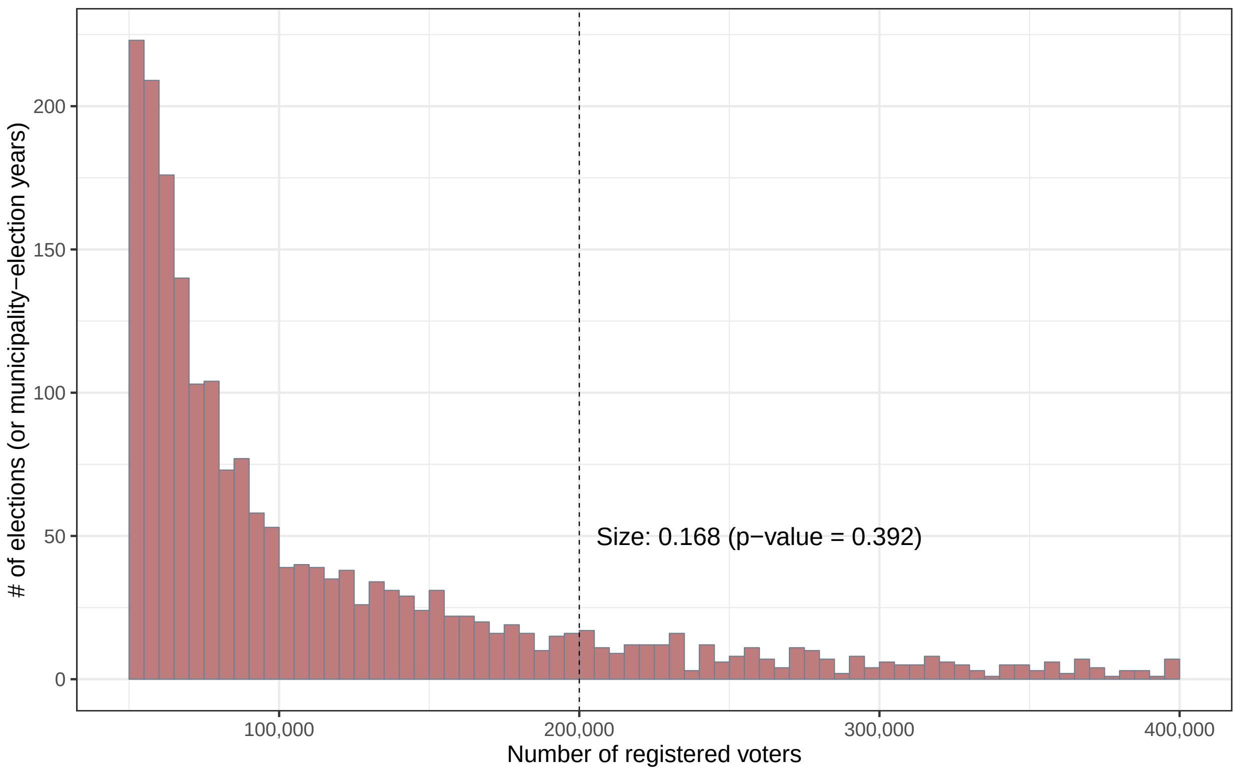
Tables and Figures

Figure 1 Electoral rules in presidential systems around the world



Single-round plurality (light red) and *Two-round majority* (dark red) correspond to the single- and two-round system in Brazil. Vertical axis corresponds to the number of countries which held an election within the specified time period under the specified electoral rule. Data encompasses 497 presidential elections across 71 countries with democratic regimes and presidential systems, between 1946-2016. *Source:* Bormann and Golder (2013).

Figure 2 Density of elections around the 200,000 registered voter threshold



Plot includes only elections with between 50,000 and 400,000 registered voters (6.0% of the universe of elections). Size of the discontinuity in the density of elections uses the McCrary test and is estimated off all elections below the 99.9 percentile in size. An “election” is defined as a municipality-election year. Bin sizes are 10,000 voters.

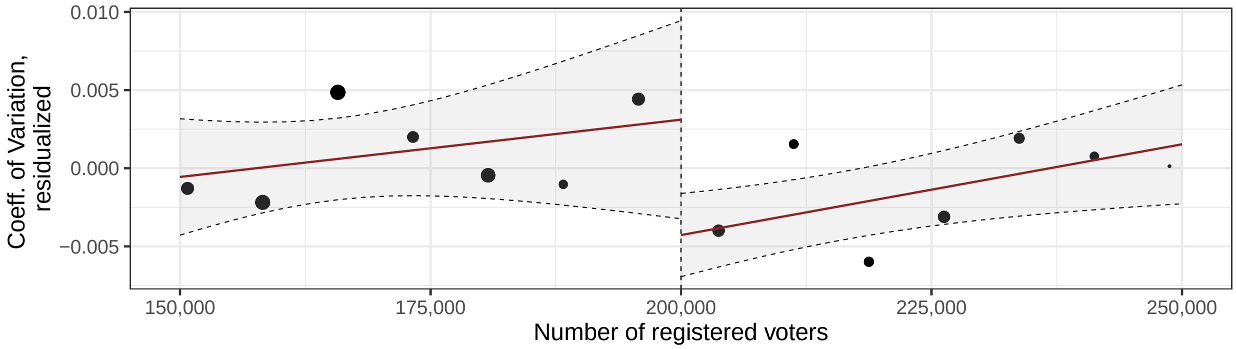
Table 3 Regression discontinuity estimates on municipality pre-characteristics

Panel A: Characteristics measured prior to the 1988 Constitution					
	% illiterate	% low income	Unempl. rate	Pop. density	
TwoRound	-1.499	-0.751	0.044	-397.678^{**}	
	(1.057)	(1.148)	(0.145)	(184.347)	
Single-round mean	19.029	50.182	1.978	516.067	
Observations	293	293	293	293	
Panel B: Characteristics measured prior to most recent single-round election					
	Muni. area change (%)	Pop. growth (%)	Pop. density	Income seg.	Dem. seg.
TwoRound	-0.601	-0.468	-663.035^{**}	-0.003	0.001
	(2.296)	(1.260)	(332.555)	(0.004)	(0.001)
Single-round mean	-3.003	2.863	1,013.396	0.090	0.028
Observations	231	295	295	231	231
	% illiterate	Income per capita	% low income	Unempl. rate	Gini coefficient
TwoRound	0.014	-18.964	1.211	-0.652	0.004
	(0.590)	(33.328)	(2.290)	(1.114)	(0.009)
Single-round mean	7.409	646.917	36.123	11.671	0.540
Observations	295	295	295	295	295

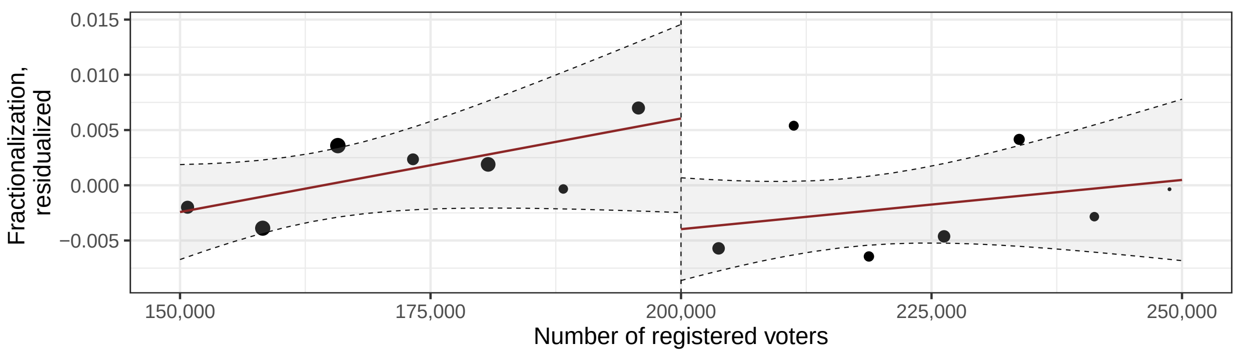
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Panel A: outcomes from the 1980 census. *Panel B:* outcomes either from the census prior to the most recent election in a single-round system or from the 1991 census. F-stat for all treatment effects in Panel B jointly significant: 1.146 ($p = 0.339$). *Muni. area change* is the percentage change in municipality area from the prior census. *Pop. growth* is the percentage change in population from the prior census. *Pop. density* is population density, per km^2. *Income seg.* and *Dem. seg.* refer to income and demographic segregation, respectively, of census tracts (measured using the entropy index). *Income per capita* is average monthly household income per capita, in *reais*. *% low income* is the fraction of households earning between 0 and 50% of the minimum wage. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Standard errors clustered at the municipality level. *Source:* 1980, 1991, 2000, and 2010 Demographic Census.

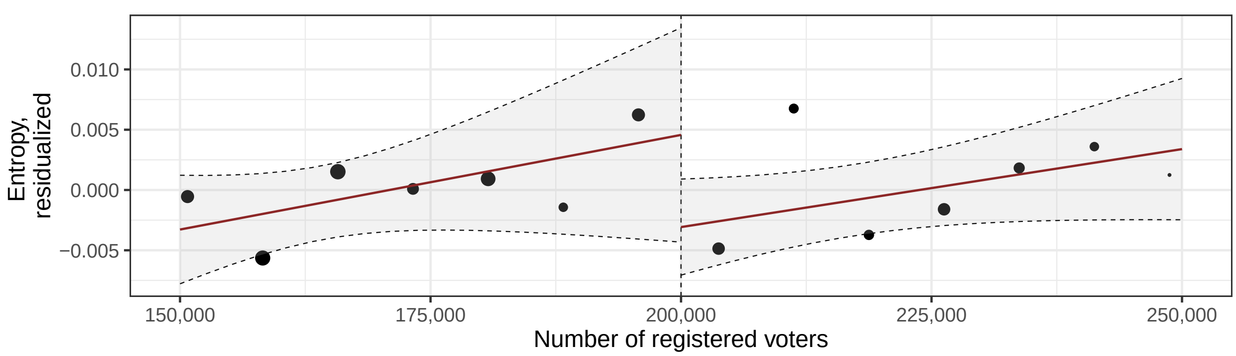
Figure 4 Regression discontinuity plots of overall concentration of voters for specific candidates



(a) Coefficient of variation



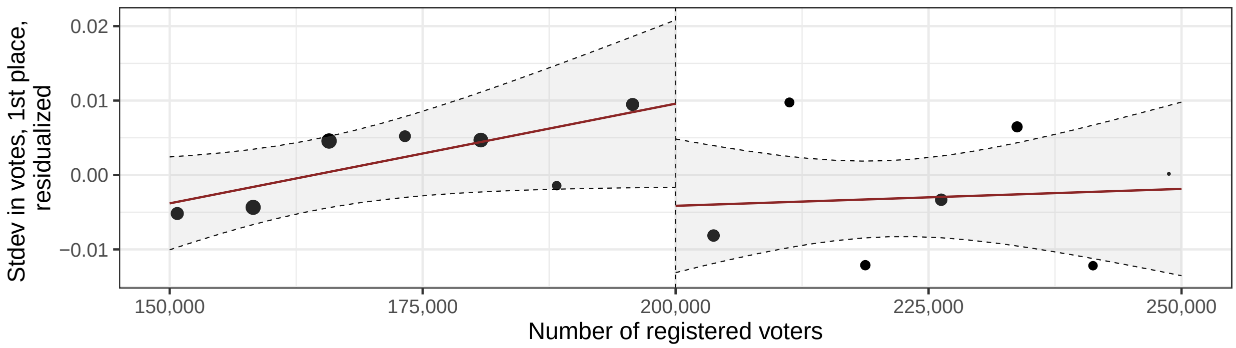
(b) Fractionalization



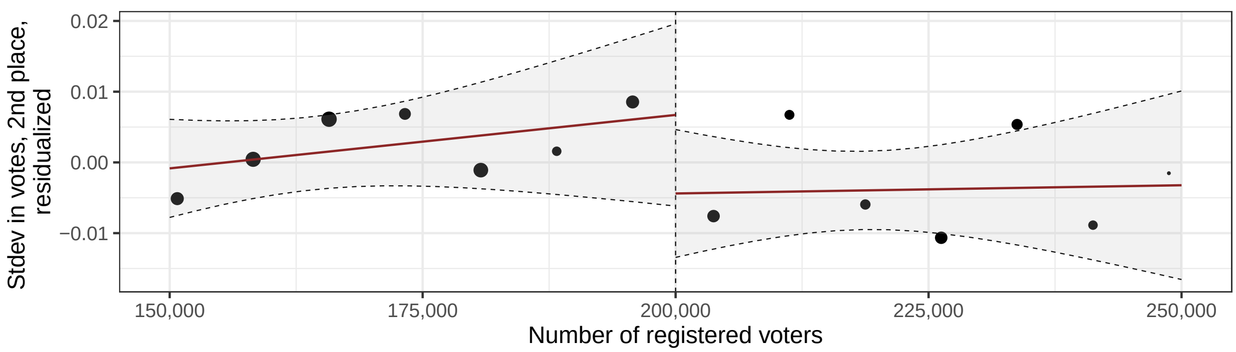
(c) Entropy

Overall concentration of voters for specific candidates, as measured by (a) *Coefficient of variation*, (b) *Fractionalization*, and (c) *Entropy*, using vote counts in polling stations. Vote shares are from the first round. In each panel, each point plots an average value within a 7,500 voter bin. Variables on the vertical axis are residualized by population density and election-year fixed effects. Diameter of the points is proportional to the number of observations. Confidence intervals (dashed lines) represent the 95% confidence intervals of a local linear regression (solid red line) with standard errors clustered at the municipality level.

Figure 5 Regression discontinuity plots of the candidate-level concentration in voters



(a) Standard deviation in votes for the 1st place candidate



(b) Standard deviation in votes for the 2nd place candidate

Standard deviation in a candidate's vote counts across polling stations, for the (a) 1st place and (b) 2nd place candidate. Vote shares are from the first round. In each panel, each point plots an average value within a 7,500 voter bin. Variables on the vertical axis are residualized by population density and election-year fixed effects. Diameter of the points is proportional to the number of observations. Confidence intervals (dashed lines) represent the 95% confidence intervals of a local linear regression (solid red line) with standard errors clustered at the municipality level.

Table 6 Regression discontinuity estimates on the geographic concentration of voters

Panel A: Concentration indices of voters for specific candidates				
	Coefficient of variation	Fractionalization	Entropy	
TwoRound	-0.009^{***}	-0.012^{**}	-0.008^{*}	
	(0.003)	(0.005)	(0.005)	
Potential bias	0.0008	–0.0002	–0.0923	
Single-round mean	0.019	0.027	0.030	
Observations	264	264	264	
Panel B: Standard deviation in vote shares for each candidate				
	1st place candidate	2nd place candidate	3rd place candidate	4th place candidate
TwoRound	-0.017^{**}	-0.014^{*}	-0.005	0.004
	(0.007)	(0.008)	(0.007)	(0.004)
Potential bias	–0.0011	–0.0010	–0.0004	–0.0002
Single-round mean	0.080	0.075	0.042	0.023
Observations	264	264	251	216

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Panel A: overall concentration of voters for specific candidates, as measured by coefficient of variation, fractionalization, and entropy of vote counts in polling stations. *Panel B:* candidate-level concentration of voters, measured by standard deviation in a candidate's vote shares (for the 1st-4th place candidate) across polling stations. *Potential bias* is the simulated effect on the outcome from having an additional candidate in every single-round election. Vote shares are from the first round. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Table 7 Estimates on the geographic concentration of voters across different rounds of elections

	Coefficient of variation	Fractionalization	Entropy	Std Dev of 1st place candidate
Panel A: single- versus two-round elections: vote shares from final round (1st round in single-round compared to 2nd round in two-round)				
TwoRound	-0.020^{**}	-0.023^{***}	-0.017^{***}	-0.019^{**}
	(0.008)	(0.008)	(0.006)	(0.009)
Single-round mean	0.036	0.038	0.029	0.088
Observations	216	216	216	216
Panel B: two-round elections: vote shares in first versus second round (1st round in two-round compared to 2nd round in two-round)				
2ndRound	-0.012^{***}	-0.012^{***}	-0.010^{***}	-0.013^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
First round mean	0.049	0.049	0.037	0.103
Observations	432	432	432	432

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Coefficient of variation, *Fractionalization* and *Entropy* measure the overall concentration of voters for specific candidates, using vote counts in polling stations. *Standard deviation of 1st place candidate* is the standard deviation in the 1st place candidate's vote counts across polling stations. All outcomes use only vote shares from the top two candidates. *Panel A* compares the 1st round results (in single-round elections) with the 2nd round results (in two-round elections) and presents the regression discontinuity estimates. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level. *Panel B* compares the 1st round results (in two-round elections) with the 2nd round results (in two-round elections), using the full sample of elections that held two rounds. *Estimation method:* Standard regression with *2ndround* as the regressor and election fixed effects. Standard errors clustered at the municipality level.

Table 8 Regression discontinuity estimates on the geographic concentration of voters, using vote shares from the top two candidates only

	Coefficient of variation	Fractionalization	Entropy	Std Dev of 1st place candidate
TwoRound	-0.015^{*}	-0.018^{**}	-0.013^{**}	-0.015
	(0.008)	(0.009)	(0.007)	(0.009)
Single-round mean	0.036	0.038	0.029	0.088
Observations	264	264	264	264

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Coefficient of variation, *Fractionalization* and *Entropy* measure the overall concentration of voters for specific candidates, using vote counts in polling stations. *Standard deviation of 1st place candidate* is the standard deviation in the 1st place candidate's vote counts across polling stations. All outcomes use only vote shares from the top two candidates. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

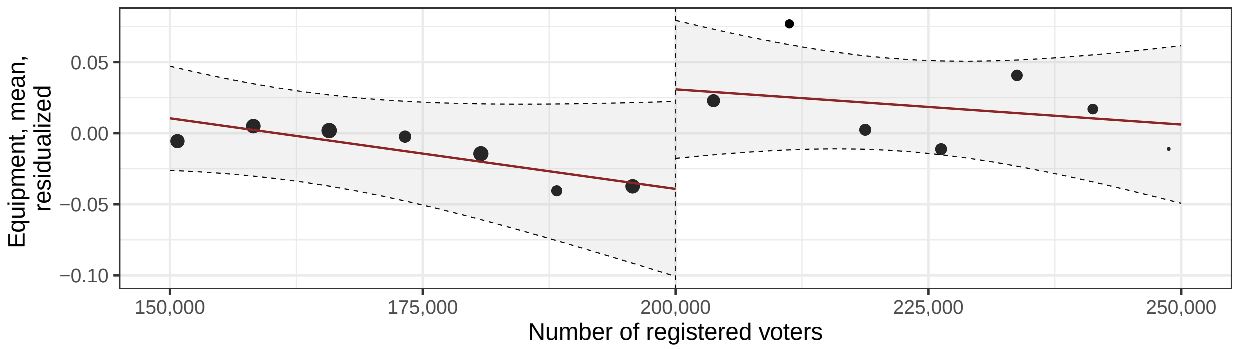
Table 9 Regression discontinuity estimates on other electoral outcomes

	Turnout	Blank/invalid ballots	# candidates
TwoRound	0.006	-3.821^{**}	1.273^{***}
	(0.008)	(1.670)	(0.339)
Single-round mean	0.843	16.524	4.604
Observations	296	296	296

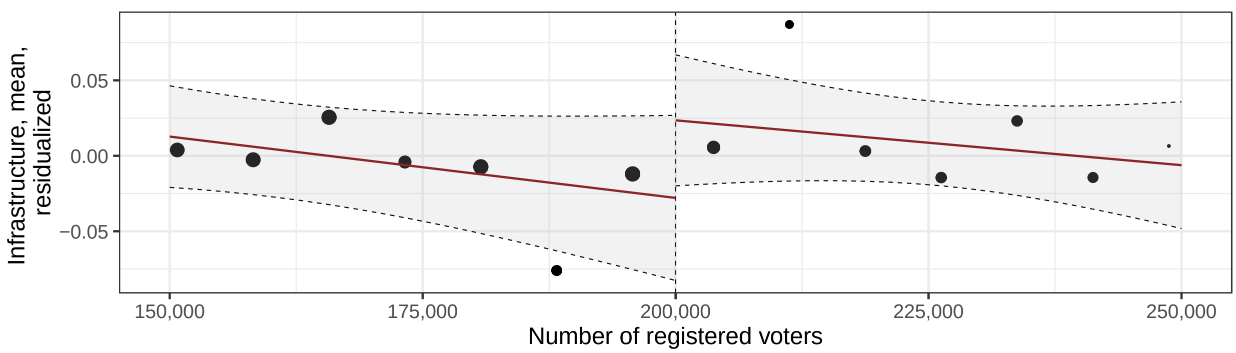
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Turnout is the fraction of eligible voters who cast a ballot in the election. *Blank/invalid ballots* is the sum of ballots (in thousands) that were either blank or voided, and is also equal to turnout minus valid ballots. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Figure 10 Regression discontinuity plots of the overall level of resources in municipal schools



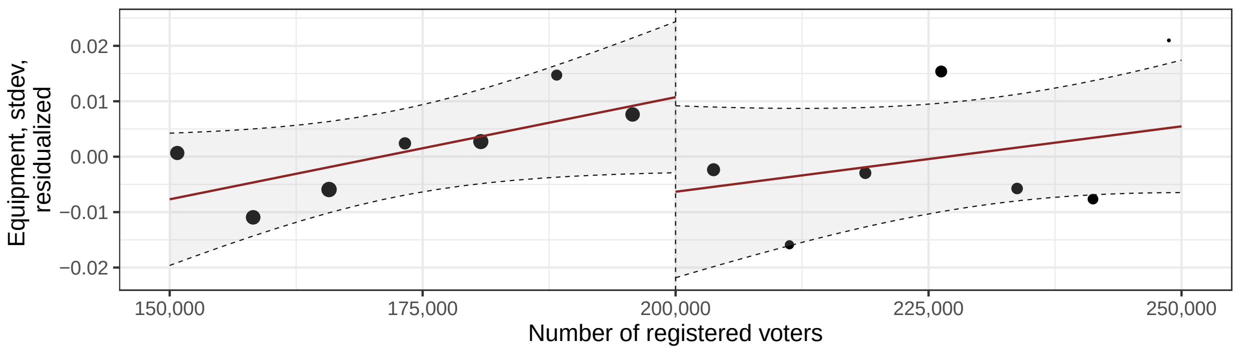
(a) Equipment, mean level of resources



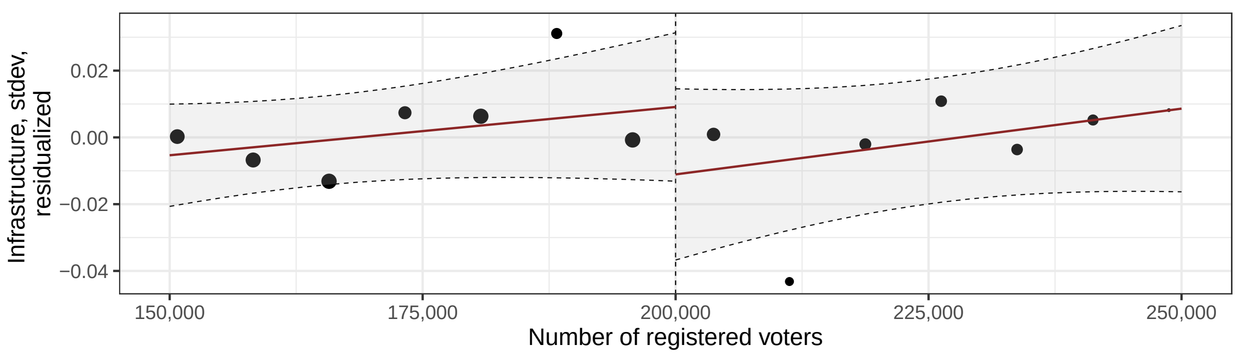
(b) Infrastructure, mean level of resources

Equipment and *Infrastructure* are indices constructed by taking the first principal component of a school's equipment and infrastructure elements, then calculating the school's percentile in the national distribution. *Mean level of resources* is the mean index level across schools in the municipality for (a) equipment and (b) infrastructure. In each panel, each point plots an average value within a 7,500 voter bin. Variables on the vertical axis are residualized by population density and election-year fixed effects. Diameter of the points is proportional to the number of observations. Confidence intervals (dashed lines) represent the 95% confidence intervals of a local linear regression (solid red line) with standard errors clustered at the municipality level.

Figure 11 Regression discontinuity plots of the distribution of resources in municipal schools



(a) Equipment, standard deviation in resources



(b) Infrastructure, standard deviation in resources

Equipment and *Infrastructure* are indices constructed by taking the first principal component of a school's equipment and infrastructure elements, then calculating the school's percentile in the national distribution. *Standard deviation in resources* is the standard deviation in the index across schools in the municipality for (a) equipment and (b) infrastructure. In each panel, each point plots an average value within a 7,500 voter bin. Variables on the vertical axis are residualized by population density and election-year fixed effects. Diameter of the points is proportional to the number of observations. Confidence intervals (dashed lines) represent the 95% confidence intervals of a local linear regression (solid red line) with standard errors clustered at the municipality level.

Table 12 Regression discontinuity estimates on resources in municipal schools

	Mean level of resources		*Standard deviation in resources*	
	Equipment	Infrastructure	Equipment	Infrastructure
TwoRound	0.081^{**}	0.057^{*}	-0.018^{*}	-0.021
	(0.035)	(0.033)	(0.009)	(0.016)
Single-round mean	0.738	0.731	0.121	0.157
Observations	820	912	820	912

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Equipment and *Infrastructure* are indices constructed by taking the first principal component of a school's equipment and infrastructure elements, then calculating the school's percentile in the national distribution. The first two columns (*Mean level of resources*) have as the dependent variable the mean index level across schools in the municipality. The last two columns (*Standard deviation in resources*) have as the dependent variable the standard deviation in the index across schools in the municipality. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Table 13 Regression discontinuity estimates on resources in municipal schools, for schools at different quartiles in the municipal distribution

	Mean level of resources in schools at different quartiles			
	1st quartile (Bottom 25%)	2nd quartile	3rd quartile	4th quartile (Top 25%)
Panel A: Equipment				
TwoRound	0.082^{**}	0.066^{*}	0.069^{*}	0.038
	(0.035)	(0.037)	(0.042)	(0.029)
Single-round mean	0.652	0.733	0.781	0.856
Observations	700	728	760	748
Panel B: Infrastructure				
TwoRound	0.116^{**}	0.102^{**}	0.056	0.013
	(0.046)	(0.047)	(0.035)	(0.021)
Single-round mean	0.540	0.689	0.814	0.914
Observations	776	764	784	780

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Equipment and *Infrastructure* are indices constructed by taking the first principal component of a school's equipment and infrastructure elements, then calculating the school's percentile in the national distribution. Dependent variables are the mean index level of equipment (*Panel A*) and infrastructure (*Panel B*) elements, separately by quartiles. Quartiles are defined by the school's percentile in the municipal distribution in the year prior to the election. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

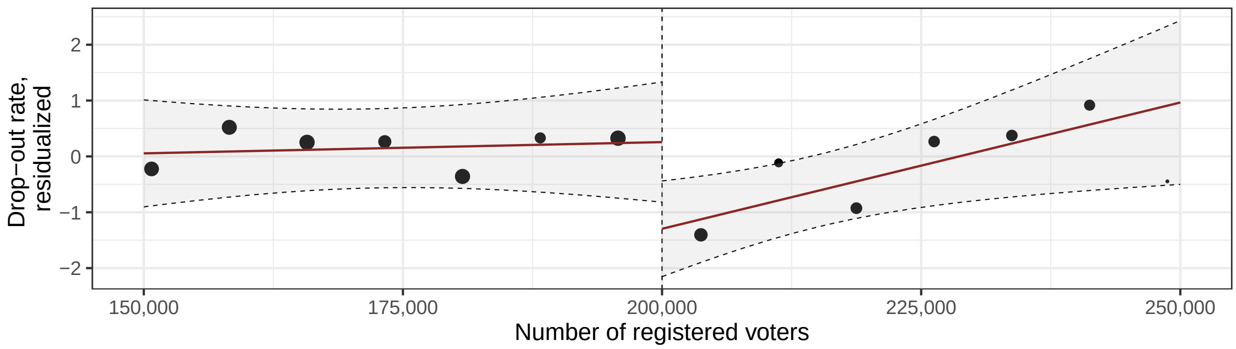
Table 14 Regression discontinuity estimates on resources in municipal schools, by possibility of re-election

	Mean level of resources		*Standard deviation in resources*	
	Equipment	Infrastructure	Equipment	Infrastructure
TwoRound	0.113^{**}	0.088^{**}	-0.019	-0.044^{**}
	(0.047)	(0.045)	(0.013)	(0.018)
TwoRound * FirstTerm	-0.034	-0.018	-0.008	0.006
	(0.033)	(0.029)	(0.011)	(0.015)
Single-round mean	0.738	0.731	0.121	0.157
Bandwidth size	50,000	50,000	50,000	50,000
Observations	789	789	789	789

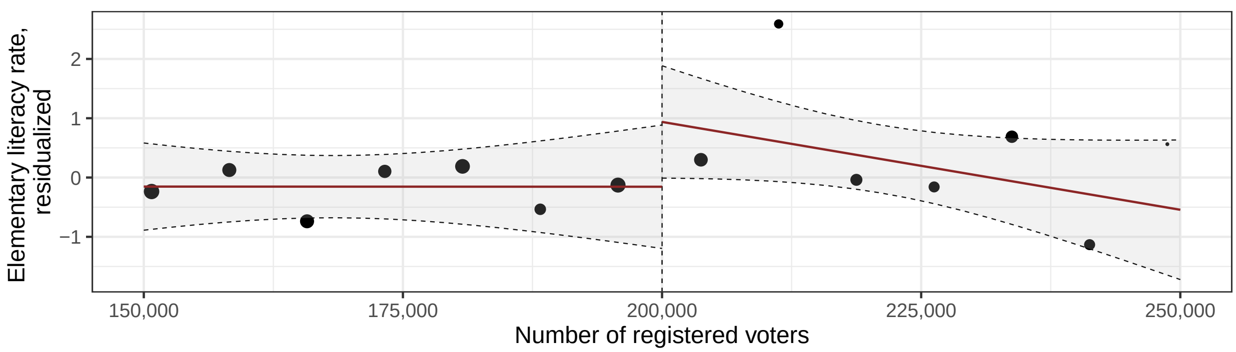
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Heterogeneous treatment effects by re-election incentives. *FirstTerm* is a dummy indicating whether the mayor is a first-term mayor (eligible for reelection). *Equipment* and *Infrastructure* are indices constructed by taking the first principal component of a school's equipment and infrastructure elements, then calculating the school's percentile in the national distribution. The first two columns (*Mean level of resources*) have as the dependent variable the mean index level across schools in the municipality. The last two columns (*Standard deviation in resources*) have as the dependent variable the standard deviation in the index across schools in the municipality. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Figure 15 Regression discontinuity plots of municipal education outcomes



(a) Drop-out rate



(b) Elementary literacy rate

Drop-out rate is from the School Census. It is the mean rate across schools in the municipality. *Elementary literacy rate* is from the 2000 and 2010 Demographic Census. It is the literacy rate of cohorts who are of elementary school age during the mayoral term. In each panel, each point plots an average value within a 7,500 voter bin. Variables on the vertical axis are residualized by population density and election-year fixed effects. Diameter of the points is proportional to the number of observations. Confidence intervals (dashed lines) represent the 95% confidence intervals of a local linear regression (solid red line) with standard errors clustered at the municipality level.

Table 16 Regression discontinuity estimates on municipal education and economic outcomes

Panel A: Education outcomes	Drop-out rate	Failing rate	Passing rate	Elem. literacy rate
TwoRound	-1.649^{**}	-0.747	2.330	1.199^{*}
	(0.667)	(1.115)	(1.459)	(0.710)
Single-round mean	3.211	8.645	88.283	91.445
Observations	909	908	909	177

Panel B: Economic outcomes	Low income rate	Income per capita	Unemployment rate	Night lights
TwoRound	-5.186^{*}	64.667	-0.964	2.715
	(3.079)	(61.782)	(0.635)	(3.306)
Single-round mean	27.929	762.417	9.815	22.527
Observations	177	177	177	763

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Panel A: Municipal education outcomes. *Drop-out rate*, *Failing rate*, and *Passing rate* are from the School Census. They are the mean rate across schools in the municipality and should add up to 1 in each school. *Elem. literacy rate* is from the 2000 and 2010 Demographic Census. It is the literacy rate of cohorts who are of elementary school age during the mayoral term. *Panel B:* Municipal economic outcomes. *Low income rate*, *Income per capita*, and *Unemployment rate* are from the 2000 and 2010 Demographic Census. *Low income rate* is the fraction of households earning between 0 and 50% of the minimum wage. *Income per capita* is the average monthly household income per capita, in *reais*. *Night lights* is from the 1997-2013 NOAA night lights series. It is the mean night lights level in the municipality. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Table 17 Regression discontinuity estimates on candidate characteristics

Panel A: Demographic characteristics of candidates

	Age	Female	Univ. degree	Born same state
TwoRound	0.470	−0.065	−0.045	−0.070
	(1.514)	(0.040)	(0.048)	(0.049)
Single-round mean	49.955	0.129	0.796	0.784
Observations	264	263	263	263

Panel B: Previous occupation of candidates

	Public sector	Technical	Business
TwoRound	−0.008	0.014	0.017
	(0.063)	(0.059)	(0.025)
Single-round mean	0.466	0.388	0.047
Observations	263	263	263

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Outcomes are the average characteristics of candidates in elections. Panel A contains demographic characteristics. *Univ. degree* is the fraction of candidates whose highest educational attainment is university or higher. *Born same state* is the fraction of candidates who were born in the same state as the election. Panel B contains the industry of candidates' stated previous occupation. *Public sector* includes occupations such as elected positions, judiciary, and workers in public administration. *Technical* includes occupations such as scientists, technicians, and artists. *Business* includes occupations such as administrative positions, workers in commerce and services, and business owners. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Table 18 Regression discontinuity estimates on characteristics of the winner

Panel A: Demographic characteristics of winners

	Age	Female	Univ. degree	Born same state
TwoRound	0.160	−0.027	0.023	−0.041
	(2.763)	(0.087)	(0.100)	(0.071)
Single-round mean	51.608	0.112	0.832	0.789
Observations	264	263	263	263

Panel B: Previous occupation of winners

	Public sector	Technical	Business
TwoRound	−0.088	0.069	0.059
	(0.137)	(0.131)	(0.044)
Single-round mean	0.534	0.348	0.043
Observations	263	263	263

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Outcomes are the characteristics of the candidate who won the election. Panel A contains demographic characteristics. *Univ. degree* is an indicator for whether the winner's highest educational attainment is university or higher. *Born same state* is an indicator for whether the winner was born in the same state as the election. Panel B contains the industry of candidates' stated previous occupation. *Public sector* includes occupations such as elected positions, judiciary, and workers in public administration. *Technical* includes occupations such as scientists, technicians, and artists. *Business* includes occupations such as administrative positions, workers in commerce and services, and business owners. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Table 19 Regression discontinuity estimates on political affiliation of candidates

	Previous candidacy	Incumbency	Small party	PT party	Governor's party
Panel A: All candidates					
TwoRound	0.539**	−0.182	0.721**	0.038	−0.084
	(0.259)	(0.155)	(0.334)	(0.100)	(0.136)
Single-round mean	1.652	0.752	2.535	0.636	0.584
Observations	263	263	296	296	263
Panel B: Winner only					
TwoRound	−0.144	−0.201*	−0.008	−0.022	−0.012
	(0.130)	(0.115)	(0.127)	(0.094)	(0.120)
Single-round mean	0.621	0.410	0.369	0.187	0.242
Observations	263	263	296	296	263

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Previous candidacy is whether the candidate ran in a previous mayoral election. *Incumbency* is whether the candidate held the position of mayor in a previous term. *Small party* is any party that is not one of the top 5 parties, by national membership. *PT party* is whether the candidate is from the *Partido dos Trabalhadores*. *Governor's party* is whether the candidate is from the party of the incumbent state governor. Dependent variables are either the number of candidates with that characteristic (*Panel A*) or an indicator for the winner having that characteristic (*Panel B*). *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Table 20 Regression discontinuity estimates on campaign donations

	Donation amounts received by candidates		
	Total	From individuals	From corporations
Panel A: Average donations per candidate			
TwoRound	-0.225	-0.491	-0.673^{*}
	(0.286)	(0.310)	(0.404)
Single-round mean	12.844	10.742	11.782
Observations	154	154	154
Panel B: Total donations among top two candidates			
TwoRound	-0.074	-0.614	-1.023^{*}
	(0.335)	(0.492)	(0.541)
Single-round mean	14.053	11.717	12.846
Observations	154	154	154

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Panel A: Outcomes are log average donation levels, in *reais*, received by candidates (total donations in the election divided by the number of candidates). *Panel B:* Outcomes are log total donations, in *reais*, received by the top two candidates. Donors identified as *Individual* and *Corporation* depending on whether the donor provided a CPF (individual identification number) or CNPJ (corporate identification number). *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

A. Theory Appendix

A.1. Swingable voters

For $0 < \pi_{j1}^{A} < 1$, we need that:

$$u_j(q_j^B) - u_j(q_j^A) - \delta \in \left(-\frac{1}{2\psi}, \frac{1}{2\psi}\right)$$

Let $u(y)$ be the largest possible utility coming from the allocation of government resources. This assumption is satisfied if:

$$\delta \in \left(u(y) - u(0) - \frac{1}{2\psi}, u(y) - u(0) + \frac{1}{2\psi}\right)$$
$$\iff \frac{1}{2\gamma} + u(y) - u(0) < \frac{1}{2\psi}$$

In other words, that swings in municipality vote shares are smaller than the variation in individual preferences. Note that this implies that $\gamma > \psi$ since $u(y) - u(0) > 0$.

A.2. Deriving vote shares with three candidates

Condition (2.1) corresponds to voters for whom $v_i \geq u_j(q_j^B) - u_j(q_j^A) - \delta$. Since there are swingable voters in every locality, π_{jt}^{c} is given by:

$$\pi_{j1}^{A} = \begin{cases} \frac{1}{2} + \psi\left(\Delta u_j^{AB} + \delta\right) & \text{if } j \in \{1, \ldots, J-1\} \\ \alpha\left(\frac{1}{2} + \psi\left(\Delta u_j^{AB} + \delta\right)\right) & \text{if } j = J \end{cases}$$
$$\pi_{j1}^{B} = \begin{cases} \frac{1}{2} + \psi\left(\Delta u_j^{BA} - \delta\right) & \text{if } j \in \{1, \ldots, J-1\} \\ \alpha\left(\frac{1}{2} + \psi\left(\Delta u_j^{BA} - \delta\right)\right) & \text{if } j = J \end{cases}$$
$$\pi_{j1}^{C} = \begin{cases} 0 & \text{if } j \in \{1, \ldots, J-1\} \\ 1-\alpha & \text{if } j = J \end{cases}$$

Candidates' total vote share in the municipality π_t^c is given by:

$$\pi_1^A = \left(\frac{J-1+\alpha}{J}\right)\left(\frac{1}{2} + \psi\delta\right) + \frac{\psi}{J}\left(\sum_{j=1}^{J-1} \Delta u_j^{AB} + \alpha \Delta u_J^{AB}\right)$$
$$\pi_1^B = \left(\frac{J-1+\alpha}{J}\right)\left(\frac{1}{2} - \psi\delta\right) + \frac{\psi}{J}\left(\sum_{j=1}^{J-1} \Delta u_j^{BA} + \alpha \Delta u_J^{BA}\right)$$

$$\pi_1^C = \frac{1-\alpha}{J}$$

The probability that candidates A and B attain a vote share above θ is equivalent to:

$$Pr\left(\pi_1^A \geq \theta\right) \equiv Pr\left[\delta \geq \frac{1}{\psi}\left(\frac{J}{J-1+\alpha}\theta - \frac{\psi}{J-1+\alpha}\left(\sum_{j=1}^{J-1}\Delta u_j^{AB} + \alpha \Delta u_J^{AB}\right) - \frac{1}{2}\right)\right]$$
$$Pr\left(\pi_1^B \geq \theta\right) \equiv Pr\left[\delta \leq \frac{1}{\psi}\left(\frac{1}{2} + \frac{\psi}{J-1+\alpha}\left(\sum_{j=1}^{J-1}\Delta u_j^{BA} + \alpha \Delta u_j^{BA}\right) - \frac{J}{J-1+\alpha}\theta\right)\right]$$

A.3. Deriving vote shares with two candidates

Candidates' vote shares in each locality and the municipality as a whole are given by:

$$\pi_{j2}^A = \frac{1}{2} + \psi\left(\Delta u_j^{AB} + \delta\right) \qquad \pi_{j2}^B = \frac{1}{2} + \psi\left(\Delta u_j^{BA} - \delta\right)$$
$$\pi_2^A = \frac{1}{2} + \psi\delta + \frac{\psi}{J}\sum_{j=1}^{J}\Delta u_j^{AB} \qquad \pi_2^B = \frac{1}{2} - \psi\delta + \frac{\psi}{J}\sum_{j=1}^{J}\Delta u_j^{BA}$$

and the probability of attaining a vote share above θ is equivalent to:

$$Pr\left(\pi_2^A \geq \theta\right) \equiv Pr\left[\delta \geq \frac{1}{\psi}\left(\theta - \frac{1}{2} - \frac{\psi}{J}\sum_{j=1}^{J}\Delta u_j^{AB}\right)\right]$$
$$Pr\left(\pi_2^B \geq \theta\right) \equiv Pr\left[\delta \leq \frac{1}{\psi}\left(\frac{1}{2} - \theta + \frac{\psi}{J}\sum_{j=1}^{J}\Delta u_j^{BA}\right)\right]$$

A.4. Contestability of localities

For $0 < Pr(\pi_1^A \geq \theta) < 1$ and $0 < Pr(\pi_1^B \geq \theta) < 1$, we need that:

$$\frac{1}{\psi}\left[\left(\frac{J}{J-1+\alpha}\right)\theta - \frac{1}{2} - \left(\frac{\psi}{J-1+\alpha}\right)\left(\sum_{j=1}^{J-1}\Delta u_j^{AB} + \alpha\Delta u_J^{AB}\right)\right] \in \left(-\frac{1}{2\gamma}, \frac{1}{2\gamma}\right)$$
$$\frac{1}{\psi}\left[\frac{1}{2} - \left(\frac{J}{J-1+\alpha}\right)\theta + \left(\frac{\psi}{J-1+\alpha}\right)\left(\sum_{j=1}^{J-1}\Delta u_j^{BA} + \alpha\Delta u_J^{BA}\right)\right] \in \left(-\frac{1}{2\gamma}, \frac{1}{2\gamma}\right)$$

which corresponds to the following condition for the first round:

$$\theta \in \left(\left(\frac{J-1+\alpha}{J}\right)\left(-\frac{\psi}{2\gamma} + \frac{1}{2}\right), \left(\frac{J-1+\alpha}{J}\right)\left(\frac{\psi}{2\gamma} + \frac{1}{2}\right)\right) \tag{A.1}$$

For $0 < Pr\left(\pi_2^A \geq \theta\right) < 1$ and $0 < Pr\left(\pi_2^B \geq \theta\right) < 1$, we need that:

$$\frac{1}{\psi}\left[\theta - \frac{1}{2} - \frac{\psi}{J}\sum_{j=1}^{J}\Delta u_j^{AB}\right] \in \left(-\frac{1}{2\gamma}, \frac{1}{2\gamma}\right)$$
$$\frac{1}{\psi}\left[\frac{1}{2} - \theta + \frac{\psi}{J}\sum_{j=1}^{J}\Delta u_j^{BA}\right] \in \left(-\frac{1}{2\gamma}, \frac{1}{2\gamma}\right)$$

which corresponds to the following condition for the second round:

$$\theta \in \left(-\frac{\psi}{2\gamma} + \frac{1}{2}, \frac{\psi}{2\gamma} + \frac{1}{2}\right) \tag{A.2}$$

Claim A.1. *$\theta = \frac{1}{2}\left(1 - \frac{1-\alpha}{J}\right)$ satisfies condition (A.1).*

Both the upper and lower inequalities are satisfied because:

$$-\frac{\psi}{2\gamma} + \frac{1}{2} < \frac{1}{2} < \frac{\psi}{2\gamma} + \frac{1}{2}$$

Claim A.2. *$\theta = \frac{1}{2}$ satisfies condition (A.1).*

The lower inequality is satisfied:

$$\left(\frac{J-1+\alpha}{J}\right)\left(-\frac{\psi}{2\gamma} + \frac{1}{2}\right) < \frac{1}{2}$$

because $\frac{J-1+\alpha}{J} < 1$ and $-\frac{\psi}{2\gamma} + \frac{1}{2} < \frac{1}{2}$.

The upper inequality is equivalent to:

$$\frac{1}{2} < \left(\frac{J-1+\alpha}{J}\right)\left(\frac{\psi}{2\gamma} + \frac{1}{2}\right)$$
$$\iff J > (1-\alpha)\left(\frac{\gamma+\psi}{\psi}\right)$$

which is true so long as J is large enough and γ/ψ is not too large.

Claim A.3. *$\theta = \frac{1}{2}$ satisfies condition (A.2).*

Both the upper and lower inequalities are satisfied because:

$$-\frac{\psi}{2\gamma}+\frac{1}{2}<\frac{1}{2}<\frac{\psi}{2\gamma}+\frac{1}{2}$$

A.5. C never makes it to the second round

For C to never make it to the second round, the probability that candidates A and B attain vote shares above candidate C's must be 1, or π_1^C does not satisfy condition (A.1):

$$\frac{1-\alpha}{J}\leq\left(\frac{J-1+\alpha}{J}\right)\left(-\frac{\psi}{2\gamma}+\frac{1}{2}\right)\qquad\text{or}\qquad\frac{1-\alpha}{J}\geq\left(\frac{J-1+\alpha}{J}\right)\left(\frac{\psi}{2\gamma}+\frac{1}{2}\right)$$

The first inequality (left equation) and second inequality (right equation) are equivalent to:

$$J\geq(1-\alpha)\left(\frac{2\gamma}{\gamma-\psi}+1\right)\qquad\qquad J\leq(1-\alpha)\left(\frac{2\gamma}{\gamma+\psi}+1\right)$$

The first inequality is much more likely to be satisfied, which is true so long as J is large enough and $2\gamma/(\gamma-\psi)$ is not too large.

A.6. Prediction (2.1)

The first order conditions of the single-round maximization are:

$$\left(\frac{\gamma}{J-1+\alpha}\right)u_j'(q_j^A)=\lambda_{1R}\qquad\text{for } j\in\{1,\dots,J-1\}$$

$$\left(\frac{\gamma}{J-1+\alpha}\right)\alpha u_j'(q_j^A)=\lambda_{1R}\qquad\text{for } j=J$$

$$\kappa G^A=\lambda_{1R}$$

where λ_{1R} is the Lagrange multiplier of the budget constraint in a single-round system.

The ratio in marginal utilities between localities is:

$$\text{between } j \text{ and } j'\text{:}\quad\frac{u_j'(q_j^A)}{u_{j'}'(q_{j'}^A)}=1\qquad\forall\quad j,j'\in\{1,\dots,J-1\}$$

$$\text{between } j \text{ and } J\text{:}\quad\frac{u_j'(q_j^A)}{u_J'(q_J^A)}=\alpha\qquad\forall\quad j\in\{1,\dots,J-1\}\tag{A.3}$$

Equation (A.3) implies that $u_j'(q_j^A)<u_J'(q_J^A)$. Since $u_j(\cdot)$ is strictly increasing and strictly concave and $u_J'(q)/u_j'(q)$ is not too large, this implies that $q_j^A>q_J^A$.

A.7. Prediction (2.2)

The first order conditions of the two-round maximization are:

$$\begin{aligned}
&\left(\frac{\gamma}{J-1+\alpha}\right)\left(1+\frac{(1-\alpha)\gamma}{\psi J}\right)u'(q_j^A)=\lambda_{2R} && \text{for } j\in\{1,\dots,J-1\}\\
&\left(\frac{\gamma}{J-1+\alpha}\right)\left(\alpha+\frac{(1-\alpha)\gamma}{\psi J}\right)u'(q_j^A)=\lambda_{2R} && \text{for } j=J\\
&\kappa G^A=\lambda_{2R}
\end{aligned}$$

where λ_{2R} is the Lagrange multiplier of the budget constraint in a two-round system.

The ratio in marginal utilities between localities is:

$$\begin{aligned}
&\text{between } j \text{ and } j'\text{:} \quad \frac{u_j'(q_j^A)}{u_{j'}'(q_{j'}^A)}=1 && \forall \quad j,j'\in\{1,\dots,J-1\}\\
&\text{between } j \text{ and } J\text{:} \quad \frac{u_j'(q_j^A)}{u_J'(q_J^A)}=\frac{\alpha+\frac{(1-\alpha)\gamma}{\psi J}}{1+\frac{(1-\alpha)\gamma}{\psi J}} && \forall \quad j\in\{1,\dots,J-1\}
\end{aligned} \tag{A.4}$$

Equation (A.4) implies that $u_j'(q_j^A)<u_J'(q_J^A)$. Since $u_j(\cdot)$ is strictly increasing and strictly concave and $u_J'(q)/u_j'(q)$ is not too large, this implies that $q_j^A>q_J^A$.

A.8. Comparing single- to two-round elections

I first establish three lemmas.

Lemma A.4. $\dfrac{u_j'(q_j^{1R})}{u_j'(q_j^{2R})}\dfrac{G^{2R}}{G^{1R}}>1$ *for all* $j\in\{1,\dots,J\}$.

Proof. For $j\in\{1,\dots,J-1\}$, combining the first round first order conditions in Appendix A.6:

$$\frac{u_j'(q_j^{1R})}{u_j'(q_j^{2R})}\frac{G^{2R}}{G^{1R}}=1+\frac{(1-\alpha)\gamma}{\psi J}>1$$

For $j=J$, combining the second round first order conditions in Appendix A.7:

$$\frac{u_J'(q_J^{1R})}{u_J'(q_J^{2R})}\frac{G^{2R}}{G^{1R}}=1+\frac{(1-\alpha)\gamma}{\alpha\psi J}>1$$

□

Lemma A.5. $\frac{u_j'(q_j^{1R})}{u_j'(q_j^{2R})} < \frac{u_J'(q_J^{1R})}{u_J'(q_J^{2R})}$

Proof. Comparing the ratio of marginal utilities in equations (A.3) and (A.4), the ratio is smaller in the single-round system compared to the two-round system:

$$\frac{u_j'(q_j^{1R})}{u_J'(q_J^{1R})} < \frac{u_j'(q_j^{2R})}{u_J'(q_J^{2R})} \iff \alpha < \frac{\alpha + \frac{(1-\alpha)\gamma}{\psi J}}{1 + \frac{(1-\alpha)\gamma}{\psi J}}$$

which is true because $\alpha < 1$. □

Lemma A.6. *If $q_j^{1R} > q_j^{2R}$ for one $j \neq J$ then $q_{j'}^{1R} > q_{j'}^{2R}$ for all other $j' \in \{1, \ldots, J-1\}$.*

Proof. If $q_j^{1R} > q_j^{2R}$, then $u_j'(q_j^{1R}) < u_j'(q_j^{2R})$ because $u_j(\cdot)$ is strictly concave. The first order conditions in Appendices A.6 and A.7 establish that the marginal utilities between all $j, j' \in \{1, \ldots, J-1\}$ are equal. Then we must have that $u_{j'}(q_{j'}^{1R}) < u_{j'}(q_{j'}^{2R})$ and that $q_{j'}^{1R} > q_{j'}^{2R}$. □

A.8.a. *Proof of allocations in locality J.–* I prove that $q_J^{1R} < q_J^{2R}$.

Proof. I prove by contradiction. Assume that $q_J^{1R} \geq q_J^{2R}$. Then $u_J'(q_J^{1R}) \leq u_J'(q_J^{2R})$. By lemma A.4, we must have that $G^{2R} > G^{1R}$. To satisfy the budget constraint, we must have that $q_j^{1R} < q_j^{2R}$ for some $j \neq J$ and, by lemma A.6, for all $j \neq J$. Then $u_j'(q_j^{1R}) > u_j'(q_j^{2R})$. However, this violates lemma A.5, and so we must have $q_J^{1R} < q_J^{2R}$. □

A.8.b. *Proof of overall budget.–* I prove that $G^{1R} < G^{2R}$.

Proof. I prove by contradiction. Assume that $G^{1R} \geq G^{2R}$. Since $q_J^{1R} < q_J^{2R}$, to satisfy the budget constraint, we must have that $q_j^{1R} > q_j^{2R}$ for some $j \neq J$ and, by lemma A.6, for all $j \neq J$. Then $u_j'(q_j^{1R}) < u_j'(q_j^{2R})$. However, this violates lemma A.4, and so we must have $G^{1R} < G^{2R}$. □

A.8.c. *Proof of allocations in other localities..–* I show that $q_j^{1R} \lesseqgtr q_j^{2R}$.

Proof. From the first order conditions in Appendices A.6 and A.7 and since $G^{1R} < G^{2R}$, we have that $\lambda_{1R} < \lambda_{2R}$. Then, for $j \in \{1, \ldots, J-1\}$:

$$\frac{u_j'(q_j^{1R})}{u_j'(q_j^{2R})} < 1 + \frac{(1-\alpha)\gamma}{\psi J} \implies \frac{u_j'(q_j^{1R})}{u_j'(q_j^{2R})} \gtrless 1$$

which implies that $q_j^{1R} \lesseqgtr q_j^{2R}$. □

A.9. Candidate C's budget

In general, for every set of utility functions $(u_1(\cdot), \cdots, u_J(\cdot))$, there exists a G^C such that G^C is the highest offer in locality J. I show this for the case where $u_j(\cdot) = \beta_j \ln(\cdot)$ and for the two-round election (since q_J is higher in a two-round election).

The first order conditions with respect to q_J and G are:

$$\begin{aligned}
&\frac{\gamma}{J-1+\alpha}\left(\alpha + \frac{(1-\alpha)\gamma}{\psi J}\right)\frac{\beta_J}{q_J} = \lambda_{2R} \\
&\kappa G = \lambda_{2R}
\end{aligned}$$

We can write G as a function of q_J:

$$G = \frac{1}{\kappa}\frac{\gamma}{J-1+\alpha}\left(\alpha + \frac{(1-\alpha)\gamma}{\psi J}\right)\frac{\beta_J}{q_J}$$

Since $q_J < q_j$ for all $j \neq J$ (prediction (2.2)), then $q_J < G/J$, implying:

$$q_J < \left(\frac{1}{\kappa J}\frac{\gamma}{J-1+\alpha}\left(\alpha + \frac{(1-\alpha)\gamma}{\psi J}\right)\beta_J\right)^{1/2} \equiv \Gamma$$

So long as $G^C > \Gamma$, then candidate C's allocation to locality J will be the highest offer there. This will be true so long as κ or J is large enough and γ/ψ is not too large.

A.10. Relaxing the assumption that $\alpha > 0$

Assume $\alpha = 0$, so that candidate C receives all the votes in locality J. Then vote shares in the first round, or when all 3 candidates are present, are given by:

$$\begin{aligned}
\pi_1^A &= \left(\frac{J-1}{J}\right)\left(\frac{1}{2} + \psi\delta\right) + \frac{\psi}{J}\sum_{j=1}^{J-1}\Delta u_j^{AB} \\
\pi_1^B &= \left(\frac{J-1}{J}\right)\left(\frac{1}{2} - \psi\delta\right) + \frac{\psi}{J}\sum_{j=1}^{J-1}\Delta u_j^{BA} \\
\pi_1^C &= \frac{1}{J}
\end{aligned}$$

and the probability of attaining a vote share above θ is given by:

$$Pr\left(\pi_1^A \geq \theta\right) = \frac{1}{2} + \frac{\gamma}{\psi}\left[\frac{1}{2} - \left(\frac{J}{J-1}\right)\theta + \left(\frac{\psi}{J-1}\right)\sum_{j=1}^{J-1}\Delta u_j^{AB}\right]$$

$$Pr\left(\pi_1^B \geq \theta\right) = \frac{1}{2} + \frac{\gamma}{\psi}\left[\frac{1}{2} - \left(\frac{J}{J-1}\right)\theta + \left(\frac{\psi}{J-1}\right)\sum_{j=1}^{J-1}\Delta u_j^{BA}\right]$$

Vote shares for the second round with candidates A and B are the same as when $\alpha > 0$.

A.10.a. *Equilibrium strategies in a single-round election.–* The maximization is:

$$\max_{G^A, \mathbf{q}^A = \left(q_1^A, \ldots, q_J^A\right)} \frac{1}{2} + \frac{\gamma}{J-1}\sum_{j=1}^{J-1}\Delta u_j^{AB} - \frac{1}{2}\kappa\left(G^A\right)^2 \qquad \text{s.t.} \quad \sum_j q_j^A \leq G^A$$

The first order conditions are:

$$\begin{aligned} &\frac{\gamma}{J-1}u'(q_j^A) = \lambda'_{1R} && \forall\ j \in \{1, \ldots, J-1\} \\ &\kappa G^A = \lambda'_{1R} \end{aligned}$$

where λ'_{1R} is the Lagrange multiplier of the budget constraint in a single-round system when $\alpha = 0$. I show that in equilibrium, the optimal strategy is to allocate $q_J^A = 0$.

Say $q_J^A > 0$. Consider the following deviation: $\left(q_J^A\right)' = q_J^A - \epsilon$ and $\left(q_k^A\right)' = q_k^A + \epsilon$ for some $k \neq J$ and $\epsilon > 0$. Candidate C's vote share is unchanged, so the threshold for winning remains $\frac{1}{2}\left(1 - \frac{1}{J}\right)$. The net change in the probability of winning is given by:

$$\begin{aligned} &\left[\frac{1}{2} + \frac{\gamma}{J-1}\left(\Delta\left(u_k^{AB}\right)' + \sum_{j\neq k,J}\Delta u_j^{AB}\right)\right] - \left[\frac{1}{2} + \frac{\gamma}{J-1}\left(\Delta u_k^{AB} + \sum_{j\neq k,J}\Delta u_j^{AB}\right)\right] \\ &= \frac{\gamma}{J-1}\left[u_k\left(\left(q_k^A\right)'\right) - u_k\left(q_k^A\right)\right] > 0 \end{aligned}$$

where the last line follows because $\epsilon > 0$ and $u_k(\cdot)$ is strictly increasing. There is a deviation that strictly increases the probability of winning. As a result, any $q_j^A > 0$ cannot be optimal so the optimal strategy is to allocate $q_J^A = 0$.

A.10.b. *Equilibrium strategies in a two-round election.–* The maximization is:

$$\begin{aligned} \max_{G^A, \mathbf{q}^A = \left(q_1^A, \ldots, q_J^A\right)} &\left(\frac{1}{2} + \frac{\gamma}{\psi}\left[-\left(\frac{1}{J-1}\right)\frac{1}{2} + \left(\frac{\psi}{J-1}\right)\sum_{j=1}^{J-1}\Delta u_j^{AB}\right]\right) \\ &+ \frac{\gamma}{\psi}\left(\frac{1}{J-1}\right)\left[\frac{1}{2} + \frac{\gamma}{J}\sum_{j=1}^{J}\Delta u_j^{AB}\right] - \frac{1}{2}\kappa\left(G^A\right)^2 \quad \text{s.t.} \quad \sum_j q_j^A \leq G^A \end{aligned}$$

The first order conditions are:

$$\left(\frac{\gamma}{J-1}\right)\left(1+\frac{\gamma}{\psi J}\right)u'(q_j^A) = \lambda'_{2R} \qquad \text{for } j \in \{1,\ldots,J-1\}$$

$$\left(\frac{\gamma}{J-1}\right)\left(\frac{\gamma}{\psi J}\right)u'(q_j^A) = \lambda'_{2R} \qquad \text{for } j = J$$

$$\kappa G^A = \lambda'_{2R}$$

where λ'_{2R} is the Lagrange multiplier of the budget constraint in a two-round system when $\alpha = 0$.

From this, we can see that $q_J^A > 0$. Thus, $q_J^{1R} < q_J^{2R}$, and prediction (2.3) holds.

Similarly, $G^{1R} < G^{2R}$. Assume not. To satisfy the budget constraint, we need that $q_j^{1R} > q_j^{2R}$ for some $j \neq J$ and by extension all $j \neq J$. Then $u_j'(q_j^{1R}) < u_j'(q_j^{2R})$. However, then $\frac{u_j'(q_j^{1R})}{u_j'(q_j^{2R})}\frac{G^{2R}}{G^{1R}} < 1$, which violates the first order conditions. Thus, $G^{1R} < G^{2R}$ and prediction (2.4) holds.

We also have that $q_j^{1R} \lesseqgtr q_j^2$. From the first order conditions, we have that $\frac{u_j'(q_j^{1R})}{u_j'(q_j^{2R})} < 1 + \frac{\gamma}{\psi J}$ and so $q_j^{1R} \lesseqgtr q_j^2$. Predictions (2.3) and (2.5) follow.

B. Data Appendix

B.1. Variables from the *Censo Escolar*

Table B.1 Variables used to construct the equipment index

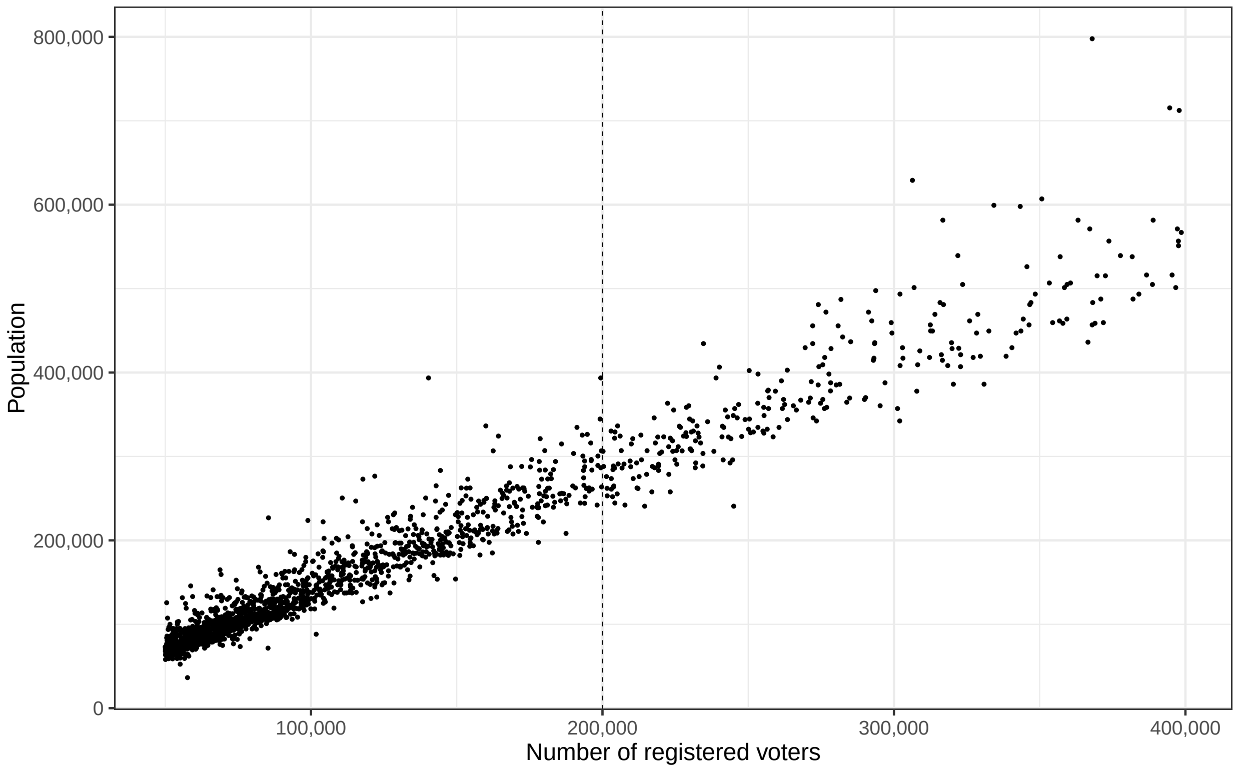
Variable	2000	2001-2003	2004	2005-2006	2007	2008	2009-2011	2012-2016
VCR/DVD	x	x	x	x	x	x	x	x
TV	x	x	x	x	x	x	x	x
satellite dish	x	x	x	x	x	x	x	x
satellite dish with internet			x	x				
overhead projector	x	x	x	x	x	x	x	x
projector			x	x				x
fax		x	x	x				x
copier		x	x	x	x	x	x	x
stereo/sound system	x	x	x	x				x
camera/camcorder			x	x				x
drinking fountain			x	x				
special needs accom.	x	x	x	x	x	x		
fan		x	x	x				
air conditioning		x	x	x				
computers	x	x	x	x	x	x	x	x
printer	x	x	x	x	x	x	x	x
local network	x	x	x					
internet	x	x	x		x	x	x	x
broadband						x	x	x

Table B.2 Variables used to construct the infrastructure index

Variable	1997	1998	1999-2000	2001-2003	2004-2006	2007-2008	2009-2011	2012-2016
principal office	x	x	x	x	x	x	x	x
secretary office	x	x	x	x	x			x
teacher lounge		x	x	x	x	x	x	x
teacher housing								x
library	x	x	x	x	x	x	x	x
reading room				x	x		x	x
video library / room			x	x	x			
toy library					x			
auditorium					x			x
solarium					x			
science lab	x	x	x	x	x	x	x	x
computer lab	x	x	x	x	x	x	x	x
other lab		x	x	x	x			
kitchen	x	x	x	x	x	x	x	x
food pantry	x	x	x	x	x			x
cafeteria	x	x	x	x	x			x
warehouse				x	x			x
schoolyard	x	x	x	x	x			x
green area								x
sports field	x	x	x	x	x	x	x	x
pool				x	x			
gymnasium					x			
playground	x	x	x	x	x	x	x	x
laundry					x			x
sanitation	x	x	x	x	x	x	x	x
special needs accomm.	x	x	x	x	x	x	x	x
classrooms	x	x	x	x	x	x	x	x

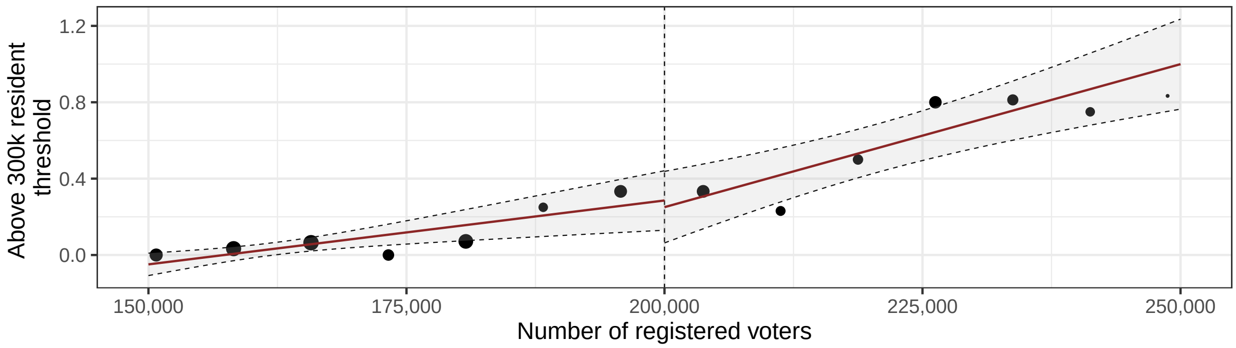
C. Additional Figures and Tables

Figure C.1 Relationship between municipality population and number of registered voters

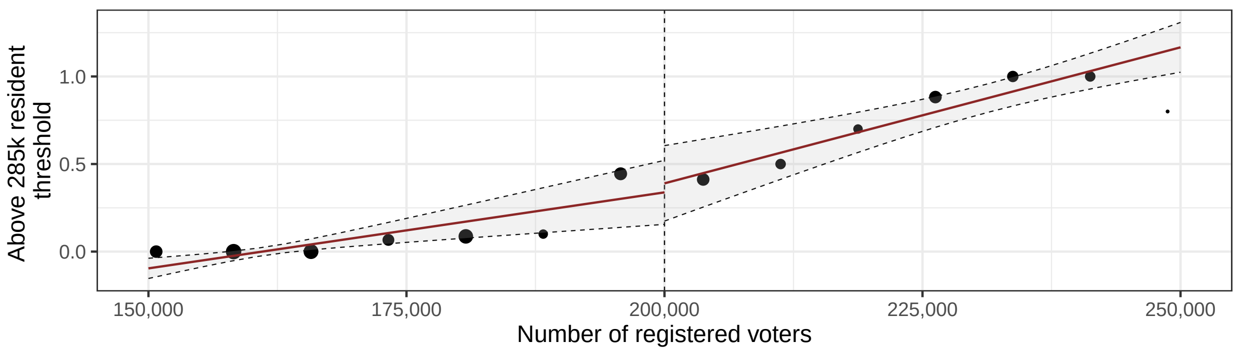


The vertical axis is municipality population in the most recent census prior to the election. Plot includes only elections with between 50,000 and 400,000 registered voters (6.0% of the universe of elections).

Figure C.2 Regression discontinuity plots of the probability of falling above/below other policy thresholds



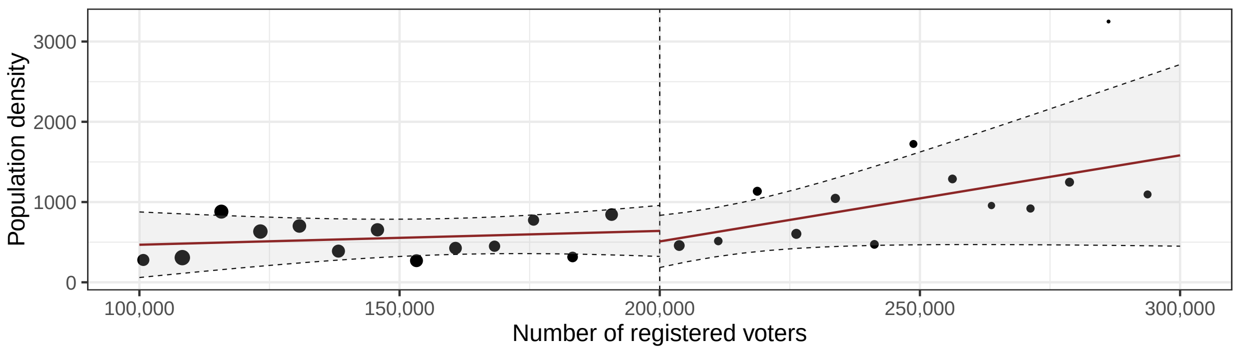
(a) Threshold: 300,000 inhabitants



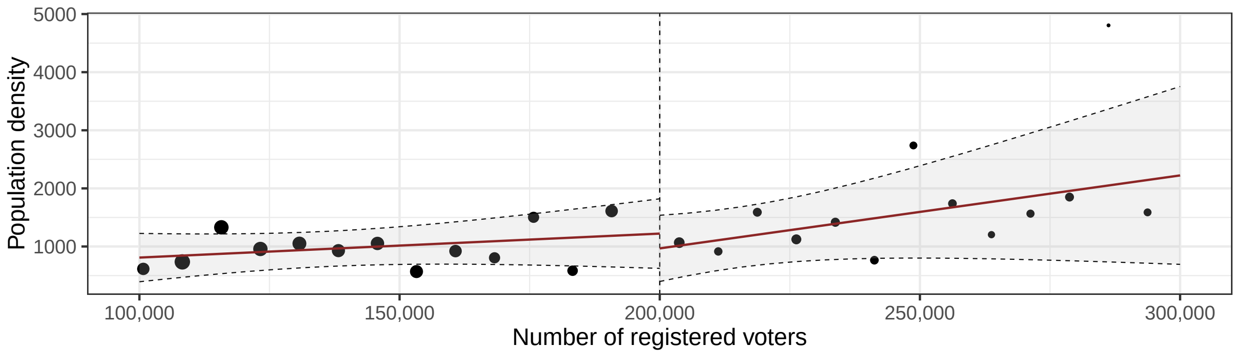
(b) Threshold: 285,714 inhabitants

The vertical axis is the fraction of elections above (a) the 300,000 resident threshold and (b) the 285,714 resident threshold. At 300,000 residents, a salary cap for municipal legislators comes into effect. At 285,714 residents, the size of the legislature changes. In each panel, each point plots an average value within a 7,500 voter bin. Diameter of the points is proportional to the number of observations. Confidence intervals (dashed lines) represent the 95% confidence intervals of a local linear regression (solid red line) with standard errors clustered at the municipality level.

Figure C.3 Regression discontinuity plots of pre-treatment population density



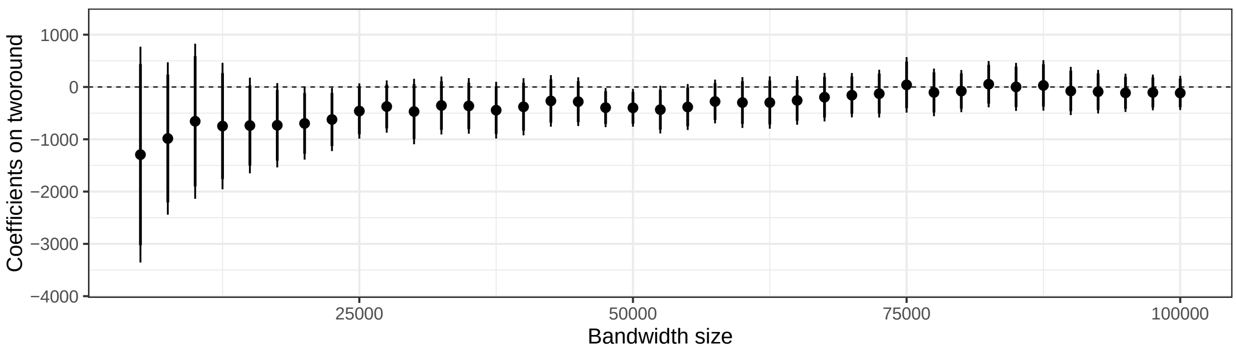
(a) Measured prior to the 1988 Constitution



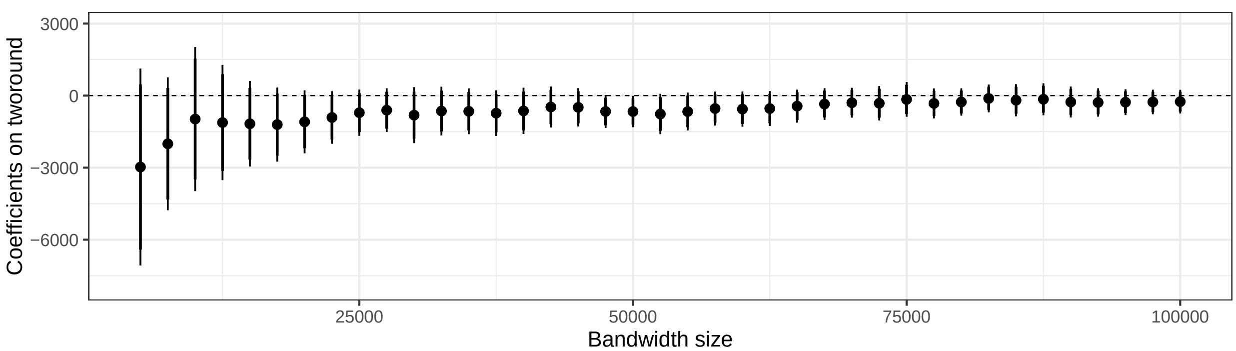
(b) Measured prior to the most recent single-round

Population density measured (a) in the 1980 census or (b) in the census prior to the most recent year in a single-round system or in the 1991 census (b). In each panel, each point plots an average value within a 7,500 voter bin. Diameter of the points is proportional to the number of observations. Confidence intervals (dashed lines) represent the 95% confidence intervals of a local linear regression (solid red line) with standard errors clustered at the municipality level.

Figure C.4 Regression discontinuity coefficients on pre-treatment population density at different bandwidths



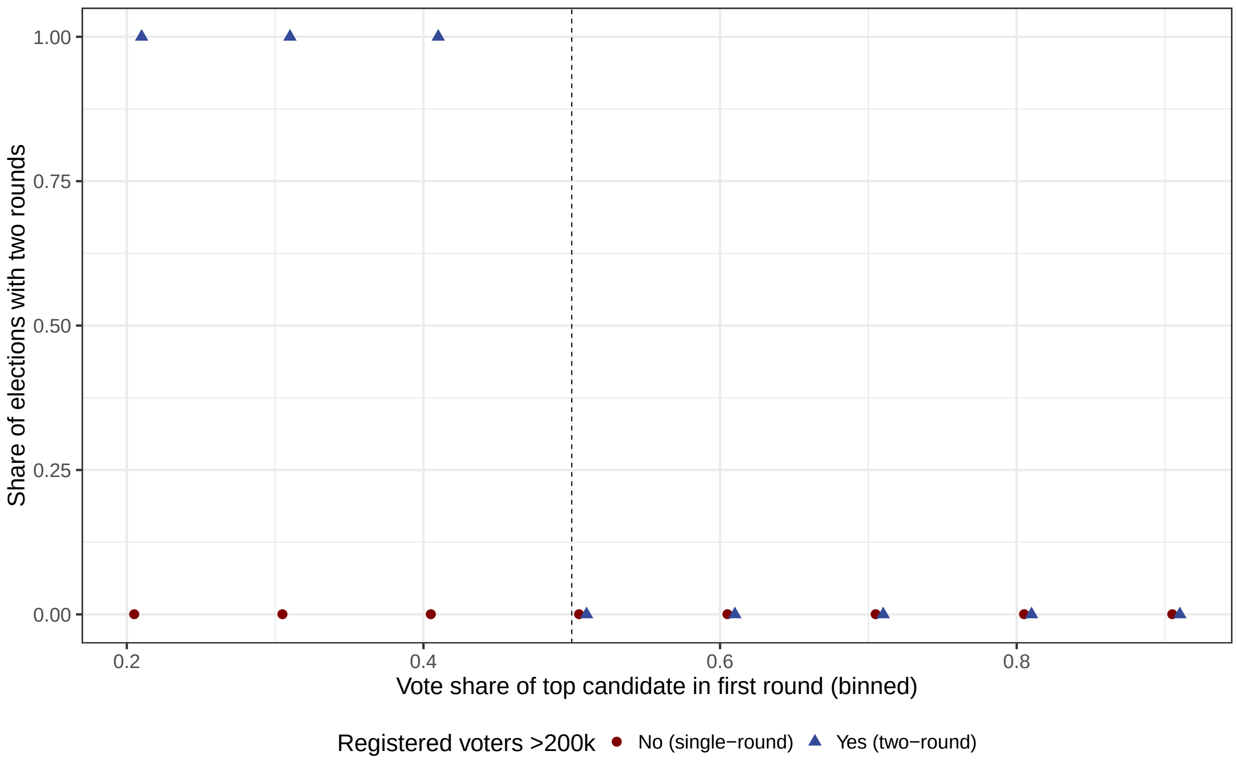
(a) Measured prior to the 1988 Constitution



(b) Measured prior to the most recent single-round

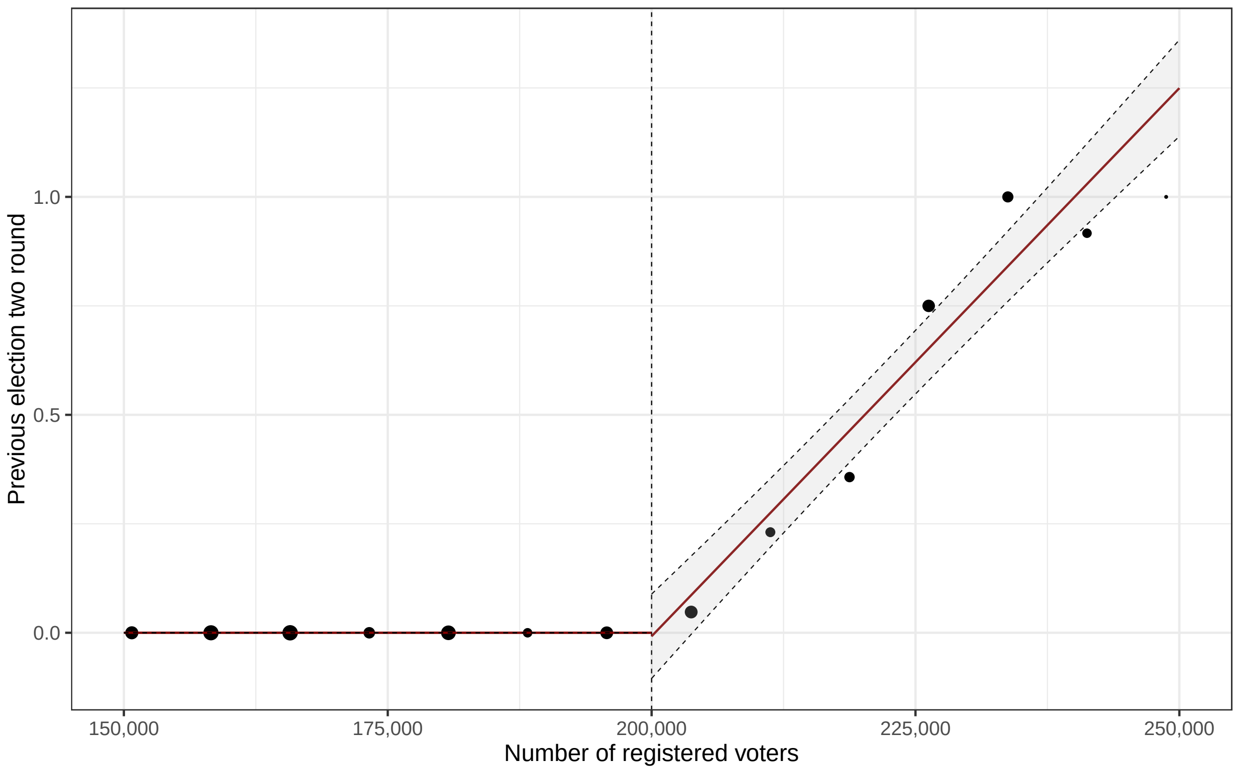
Population density measured (a) in the 1980 census or (b) in the census prior to the most recent year in a single-round system or in the 1991 census. The thicker vertical lines represent the 90% confidence interval and the thinner vertical lines represent the 95% confidence interval. *Estimation method:* Local linear regression with the specified voter bandwidth and election-year fixed effects. Standard errors clustered at the municipality level. *Source:* 1980, 1991, 2000, and 2010 Demographic Census.

Figure C.5 Compliance with treatment assignment



The vertical axis is the fraction of elections that held two rounds of elections. The horizontal axis is the vote share of the top candidate in the first round. Municipalities below the 200,000 registered voter threshold (and thus should always hold one round) are denoted by red circles. Municipalities above the 200,000 registered voter threshold (and thus should hold two rounds if no candidate receives 50% in the first round) are denoted by blue triangles. Bin sizes are 10% vote share.

Figure C.6 Regression discontinuity plot of probability of treatment in the previous election



The vertical axis is the probability that the *previous* election was a two-round election. The horizontal axis is the number of registered voters in the *current* election. In each panel, each point plots an average value within a 7,500 voter bin. Diameter of the points is proportional to the number of observations. Confidence intervals (dashed lines) represent the 95% confidence intervals of a local linear regression (solid red line) with standard errors clustered at the municipality level. Note that because all observations to the left of the threshold are 0, there are no standard errors.

Table C.7 Regression discontinuity estimates on resources in municipal schools, by year in the term

	Mean level of resources		*Standard deviation in resources*	
	Equipment	Infrastructure	Equipment	Infrastructure
Panel A: First year in the electoral term				
TwoRound	0.077^{**}	0.052	-0.013	-0.021
	(0.039)	(0.035)	(0.011)	(0.018)
Observations	197	227	197	227
Panel B: Second year in the electoral term				
TwoRound	0.091^{**}	0.059^{*}	-0.018^{*}	-0.022
	(0.039)	(0.033)	(0.011)	(0.017)
Observations	197	228	197	228
Panel C: Third year in the electoral term				
TwoRound	0.087^{**}	0.059^{*}	-0.026^{**}	-0.022
	(0.037)	(0.034)	(0.011)	(0.016)
Observations	197	228	197	228
Panel D: Fourth year in the electoral term				
TwoRound	0.070^{**}	0.059^{*}	-0.015	-0.020
	(0.033)	(0.034)	(0.010)	(0.017)
Bandwidth size	50,000	50,000	50,000	50,000
Observations	229	229	229	229

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Mayoral terms are for four years. Each panel displays the estimate separately for the 1st year (*Panel A*), 2nd year (*Panel B*), 3rd year (*Panel C*), and 4th year (*Panel D*) of the term. *Equipment* and *Infrastructure* are indices constructed by taking the first principal component of a school's equipment and infrastructure elements, then calculating the school's percentile in the national distribution. The first two columns (*Mean level of resources*) have as the dependent variable the mean index level across schools in the municipality. The last two columns (*Standard deviation in resources*) have as the dependent variable the standard deviation in the index across schools in the municipality. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Table C.8 Regression discontinuity estimates on the geographic concentration of voters, with number of candidates as a control

Panel A: Concentration indices of voters for specific candidates				
	Coefficient of variation	Fractionalization	Entropy	
TwoRound	-0.005	-0.010^{**}	-0.009^{*}	
	(0.003)	(0.005)	(0.005)	
Single-round mean	0.019	0.027	0.030	
Observations	264	264	264	
Panel B: Standard deviation in vote shares for each candidate				
	1st place candidate	2nd place candidate	3rd place candidate	4th place candidate
TwoRound	-0.016^{**}	-0.012	-0.010	-0.002
	(0.007)	(0.008)	(0.007)	(0.004)
Single-round mean	0.080	0.075	0.042	0.023
Observations	264	264	251	216

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Panel A: overall concentration of voters for specific candidates, as measured by concentration indices (coefficient of variation, fractionalization, and entropy) of vote counts in polling stations. *Panel B:* candidate-level concentration of voters, measured by standard deviation in a candidate's vote shares (for the 1st-4th place candidate) across polling stations. Vote shares are from the first round. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Number of candidates included as a control. Standard errors clustered at the municipality level.

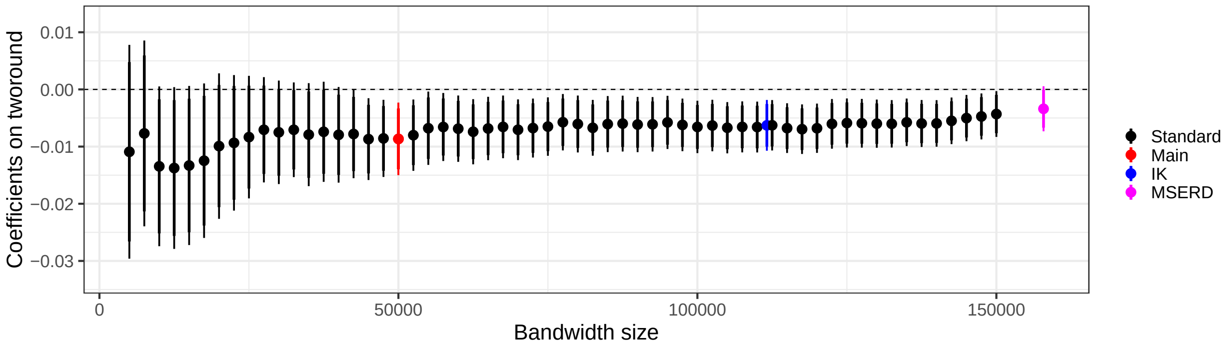
Table C.9 Regression discontinuity estimates on resources in municipal schools, using z-scores

	Mean level of resources		*Standard deviation in resources*	
	Equipment	Infrastructure	Equipment	Infrastructure
TwoRound	0.079^{**}	0.069^{*}	-0.014	-0.007
	(0.033)	(0.037)	(0.009)	(0.017)
Single-round mean	0.724	0.739	0.120	0.146
Observations	820	912	820	912

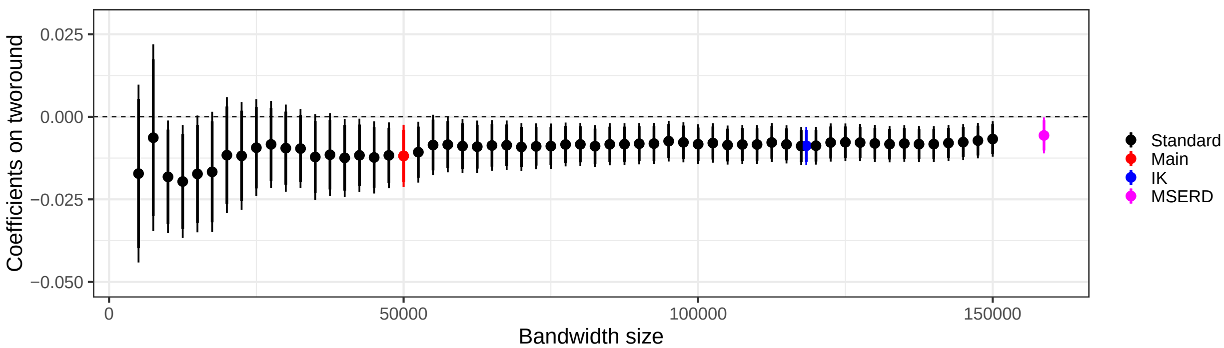
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Equipment and *Infrastructure* are indices constructed by taking the z-score of a school's equipment and infrastructure elements, then calculating the school's percentile in the national distribution. The first two columns (*Mean level of resources*) have as the dependent variable the mean index level across schools in the municipality. The last two columns (*Standard deviation in resources*) have as the dependent variable the standard deviation in the index across schools in the municipality. *Estimation method:* Local linear regression with a 50,000 voter bandwidth. Standard errors clustered at the municipality level.

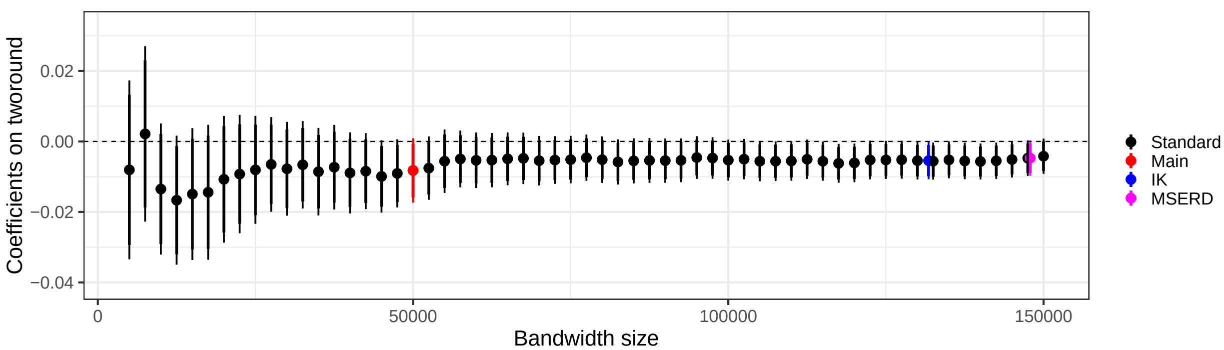
Figure C.10 Regression discontinuity coefficients on overall concentration of voters for specific candidates at different bandwidths



(a) Coefficient of variation



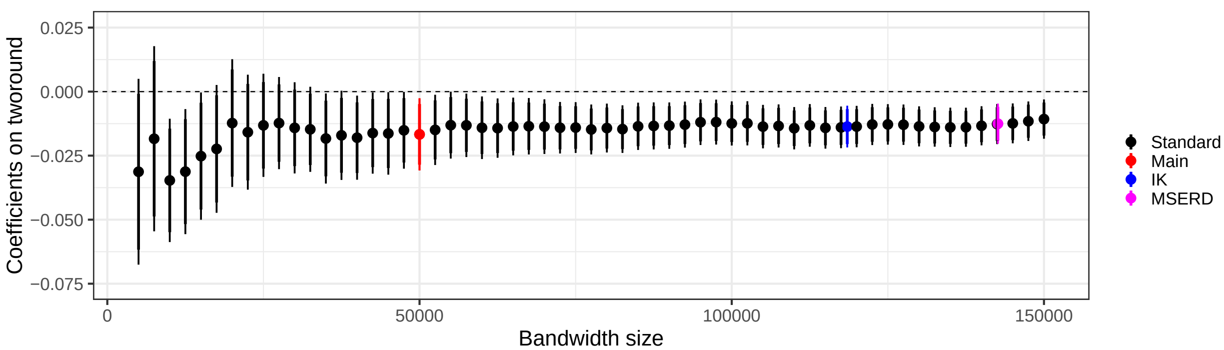
(b) Fractionalization



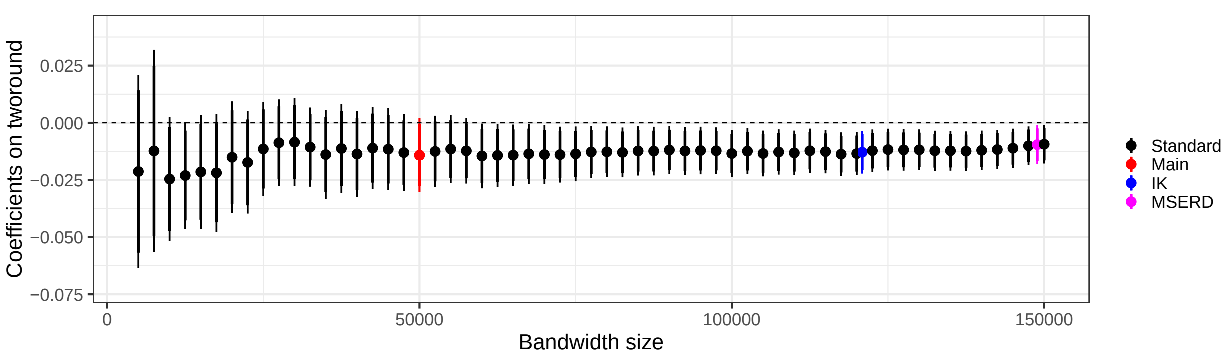
(c) Entropy

Overall concentration of voters for specific candidates, as measured by (a) *Coefficient of variation*, (b) *Fractionalization*, and (c) *Entropy*, using vote counts in polling stations. Vote shares are from the first round. The thicker vertical lines represent the 90% confidence interval and the thinner vertical lines represent the 95% confidence interval. *Estimation method:* Local linear regression with election-year fixed effects and with the specified voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Figure C.11 Regression discontinuity coefficients on the candidate-level concentration in voters at different bandwidths



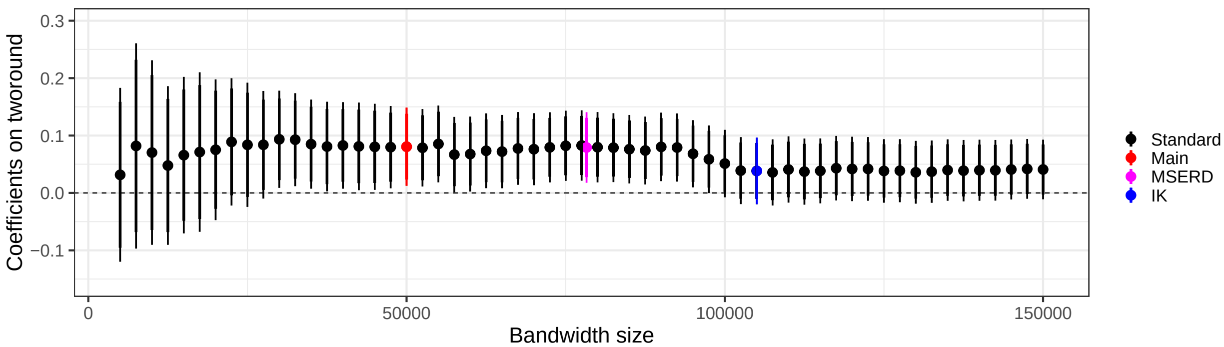
(a) Standard deviation in votes for 1st place candidate



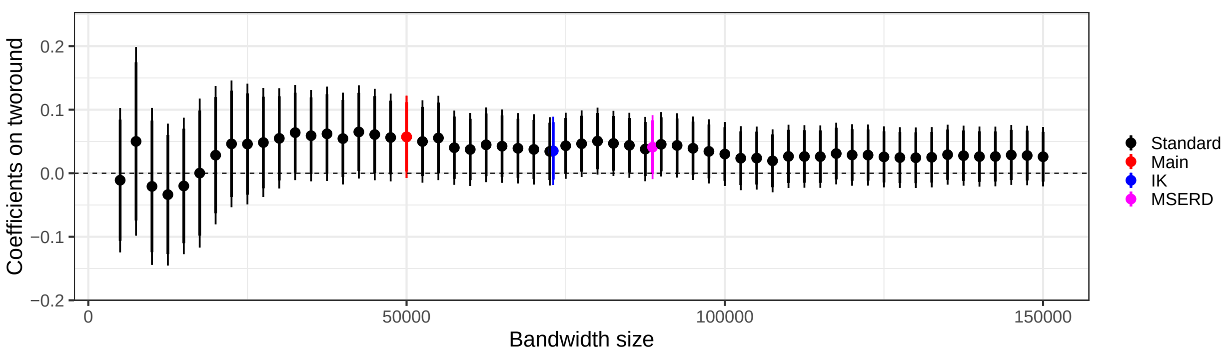
(b) Standard deviation in votes for 2nd place candidate

Standard deviation in a candidate's vote counts across polling stations, for the (a) 1st place and (b) 2nd place candidate. Vote shares are from the first round. The thicker vertical lines represent the 90% confidence interval and the thinner vertical lines represent the 95% confidence interval. *Estimation method:* Local linear regression with election-year fixed effects and with the specified voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Figure C.12 Regression discontinuity coefficients on overall level of resources in municipal schools at different bandwidths



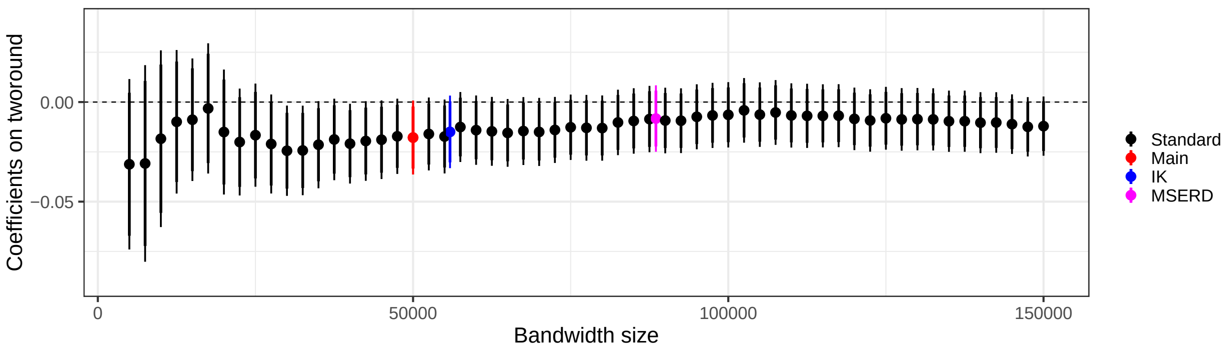
(a) Equipment, mean level of resources



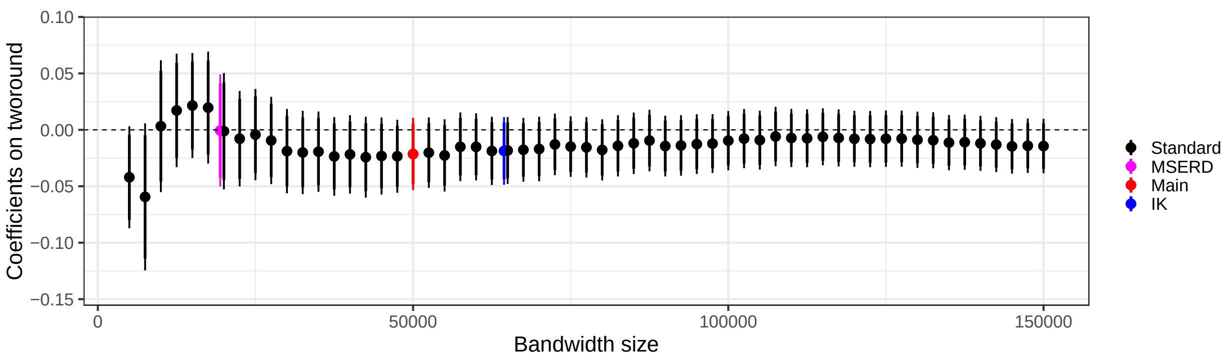
(b) Infrastructure, mean level of resources

Equipment and *Infrastructure* are indices constructed by taking the first principal component of a school's equipment and infrastructure elements, then calculating the school's percentile in the national distribution. *Mean level of resources* is the mean index level across schools in the municipality for (a) equipment and (b) infrastructure. The thicker vertical lines represent the 90% confidence interval and the thinner vertical lines represent the 95% confidence interval. *Estimation method:* Local linear regression with election-year fixed effects and with the specified voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Figure C.13 Regression discontinuity coefficients on distribution of resources in municipal schools at different bandwidths



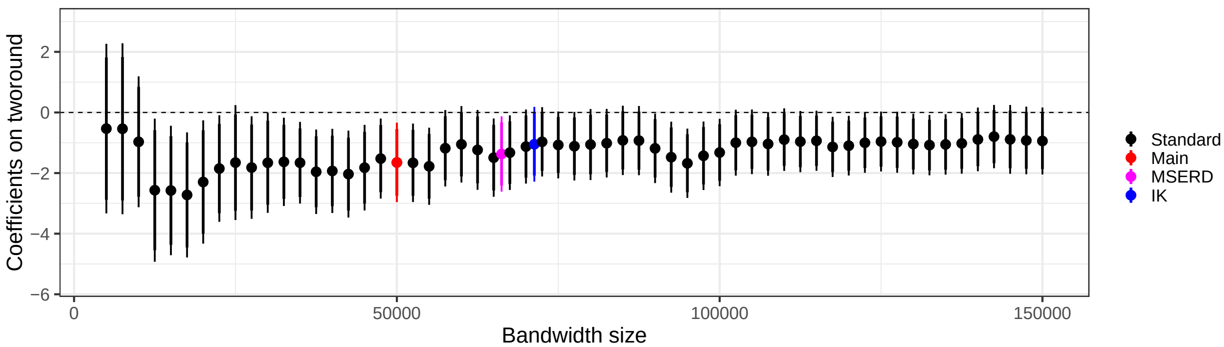
(a) Equipment, standard deviation in resources



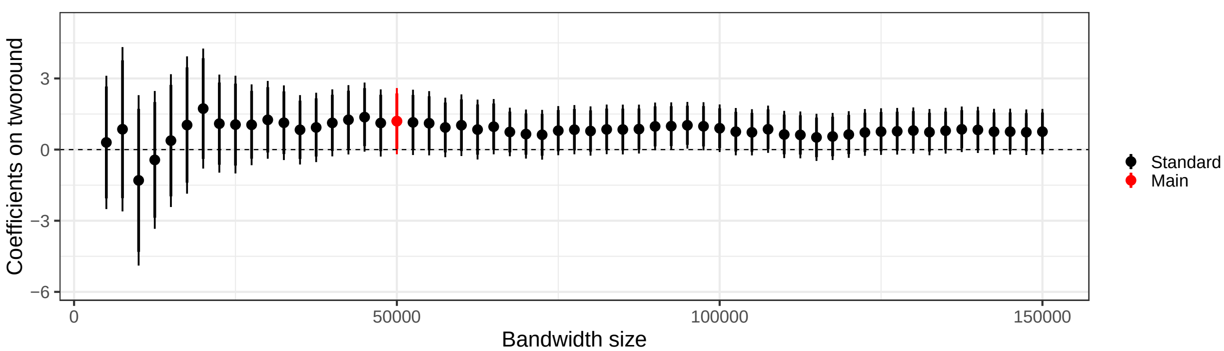
(b) Infrastructure, standard deviation in resources

Equipment and *Infrastructure* are indices constructed by taking the first principal component of a school's equipment and infrastructure elements, then calculating the school's percentile in the national distribution. *Standard deviation in resources* is the standard deviation in the index across schools in the municipality for (a) equipment and (b) infrastructure. The thicker vertical lines represent the 90% confidence interval and the thinner vertical lines represent the 95% confidence interval. *Estimation method:* Local linear regression with election-year fixed effects and with the specified voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Figure C.14 Regression discontinuity coefficients on municipal education outcomes at different bandwidths



(a) Drop-out rate



(b) Elementary literacy rate

Drop-out rate is from the School Census. It is the mean rate across schools in the municipality. *Elementary literacy rate* is from the 2000 and 2010 Demographic Census. It is the literacy rate of cohorts who are of elementary school age during the mayoral term. The thicker vertical lines represent the 90% confidence interval and the thinner vertical lines represent the 95% confidence interval. IK and MSERD bandwidths not shown for *Elementary literacy rate*, as the bandwidth chosen was larger than the support. *Estimation method:* Local linear regression with election-year fixed effects and with the specified voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Table C.15 Regression discontinuity estimates on the geographic concentration of voters, without controls

Panel A: Concentration indices of voters for specific candidates				
	Coefficient of variation	Fractionalization	Entropy	
TwoRound	-0.008^{**}	-0.010^{**}	-0.007	
	(0.003)	(0.004)	(0.005)	
Single-round mean	0.019	0.027	0.030	
Observations	264	264	264	
Panel B: Standard deviation in vote shares for each candidate				
	1st place candidate	2nd place candidate	3rd place candidate	4th place candidate
TwoRound	-0.012^{*}	-0.011	-0.003	0.005
	(0.006)	(0.008)	(0.006)	(0.005)
Single-round mean	0.080	0.075	0.042	0.023
Observations	264	264	251	216

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Panel A: overall concentration of voters for specific candidates, as measured by concentration indices (coefficient of variation, fractionalization, and entropy) of vote counts in polling stations. *Panel B:* candidate-level concentration of voters, measured by standard deviation in a candidate's vote shares (for the 1st-4th place candidate) across polling stations. Vote shares are from the first round. *Estimation method:* Local linear regression with a 50,000 voter bandwidth. Standard errors clustered at the municipality level.

Table C.16 Regression discontinuity estimates on resources in municipal schools, without controls

	Mean level of resources		*Standard deviation in resources*	
	Equipment	Infrastructure	Equipment	Infrastructure
TwoRound	0.068^{*}	0.036	-0.019^{*}	-0.014
	(0.035)	(0.029)	(0.011)	(0.015)
Single-round mean	0.738	0.731	0.121	0.157
Observations	821	916	821	916

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Equipment and *Infrastructure* are indices constructed by taking the first principal component of a school's equipment and infrastructure elements, then calculating the school's percentile in the national distribution. The first two columns (*Mean level of resources*) have as the dependent variable the mean index level across schools in the municipality. The last two columns (*Standard deviation in resources*) have as the dependent variable the standard deviation in the index across schools in the municipality. *Estimation method:* Local linear regression with a 50,000 voter bandwidth. Standard errors clustered at the municipality level.

Table C.17 Regression discontinuity estimates on municipal education outcomes, without controls

	Drop-out rate	Failing rate	Passing rate	Elem. literacy rate
TwoRound	-1.340^{*}	-0.051	1.291	2.918
	(0.755)	(1.167)	(1.686)	(2.030)
Single-round mean	3.211	8.645	88.283	91.445
Observations	913	912	913	178

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Drop-out rate, *Failing rate*, and *Passing rate* are from the School Census. They are the mean rate across schools in the municipality and should add up to 1 in each school. *Elem. literacy rate* is from the 2000 and 2010 Demographic Census. It is the literacy rate of cohorts who are of elementary school age during the mayoral term. *Estimation method:* Local linear regression with a 50,000 voter bandwidth. Standard errors clustered at the municipality level.

Table C.18 Placebo regression discontinuity estimates on the geographic concentration of voters, at 300,000 inhabitant threshold

Panel A: Concentration indices of voters for specific candidates				
	Coefficient of variation	Fractionalization	Entropy	
TwoRound	0.005	0.005	0.005	
	(0.004)	(0.005)	(0.005)	
Single-round mean	0.019	0.024	0.027	
Observations	471	471	471	
Panel B: Standard deviation in vote shares for each candidate				
	1st place candidate	2nd place candidate	3rd place candidate	4th place candidate
TwoRound	0.001	0.0001	0.001	−0.003
	(0.008)	(0.007)	(0.006)	(0.005)
Single-round mean	0.075	0.072	0.040	0.021
Observations	471	471	444	373

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

At 300,000 inhabitants, a 2000 constitutional amendment placing a cap on local legislator salaries comes into effect. *Panel A:* overall concentration of voters for specific candidates, as measured by concentration indices (coefficient of variation, fractionalization, and entropy) of vote counts in polling stations. *Panel B:* candidate-level concentration of voters, measured by standard deviation in a candidate's vote shares (for the 1st-4th place candidate) across polling stations. Vote shares are from the first round. Includes only elections after 2000. *Estimation method:* Local linear regression with election-year fixed effects and a 125,000 inhabitant bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

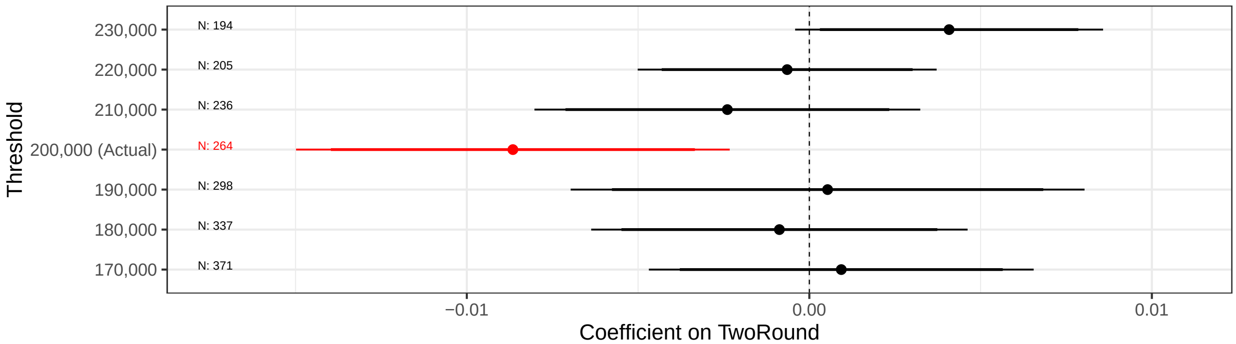
Table C.19 Placebo regression discontinuity estimates on the geographic concentration of voters, at 285,714 inhabitant threshold

Panel A: Concentration indices of voters for specific candidates				
	Coefficient of variation	Fractionalization	Entropy	
TwoRound	0.004	0.005	0.007*	
	(0.003)	(0.004)	(0.004)	
Single-round mean	0.019	0.024	0.027	
Observations	423	423	423	
Panel B: Standard deviation in vote shares for each candidate				
	1st place candidate	2nd place candidate	3rd place candidate	4th place candidate
TwoRound	0.005	0.003	0.009*	0.006
	(0.005)	(0.006)	(0.005)	(0.004)
Single-round mean	0.075	0.071	0.040	0.022
Observations	424	423	400	331

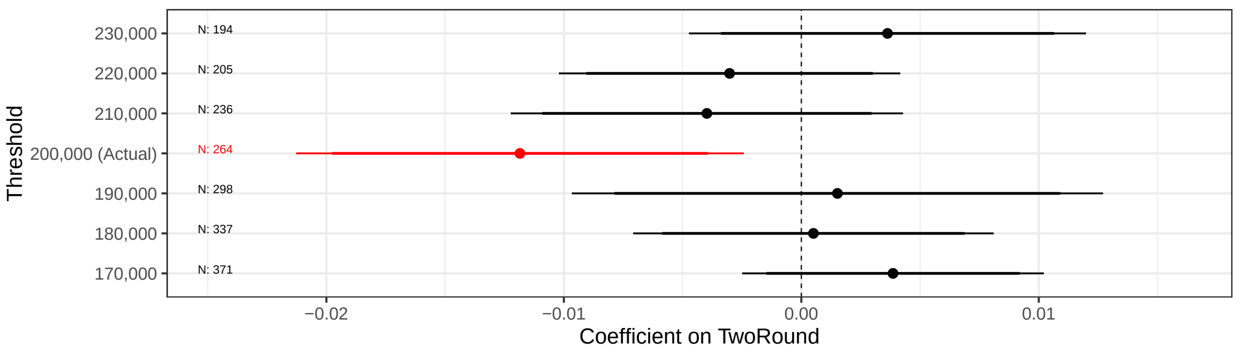
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

At 285,714 inhabitants, a 2004 constitutional amendment changing the size of the local legislature comes into effect. *Panel A:* overall concentration of voters for specific candidates, as measured by concentration indices (coefficient of variation, fractionalization, and entropy) of vote counts in polling stations. *Panel B:* candidate-level concentration of voters, measured by standard deviation in a candidate's vote shares (for the 1st-4th place candidate) across polling stations. Vote shares are from the first round. Includes only elections after 2004. *Estimation method:* Local linear regression with election-year fixed effects and a 125,000 inhabitant bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

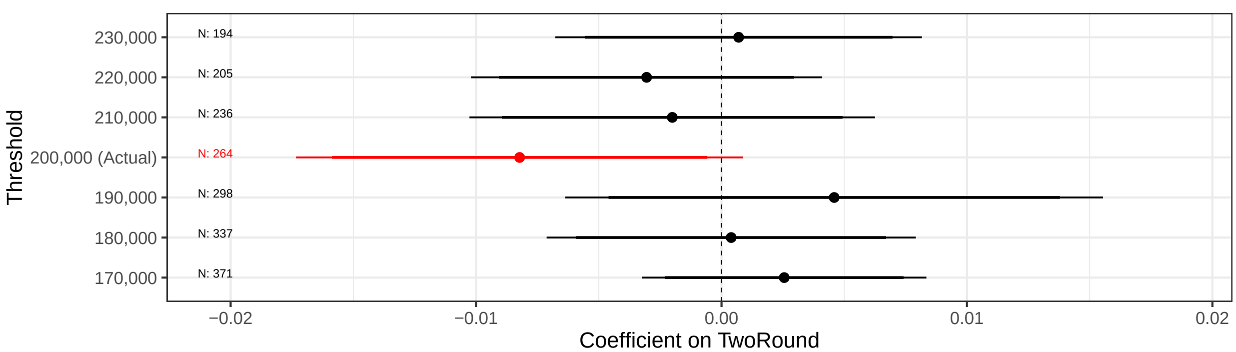
Figure C.20 Regression discontinuity coefficients on overall concentration of voters for specific candidates at different thresholds



(a) Coefficient of variation



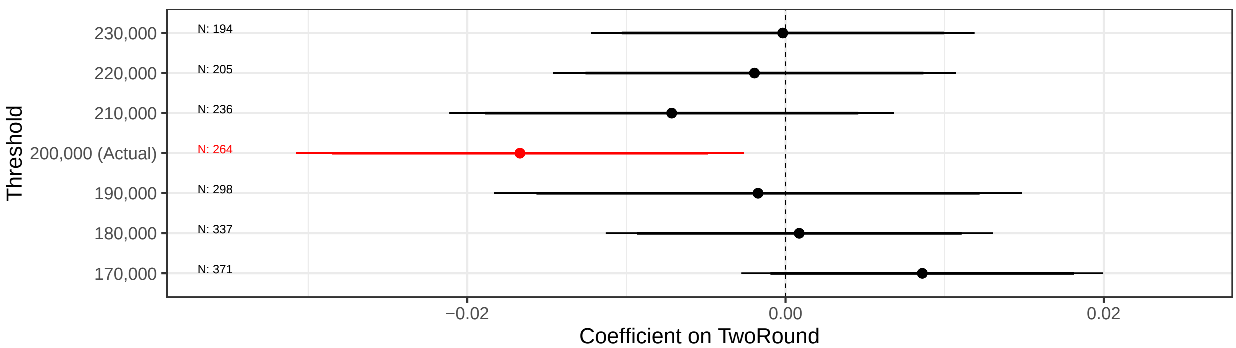
(b) Fractionalization



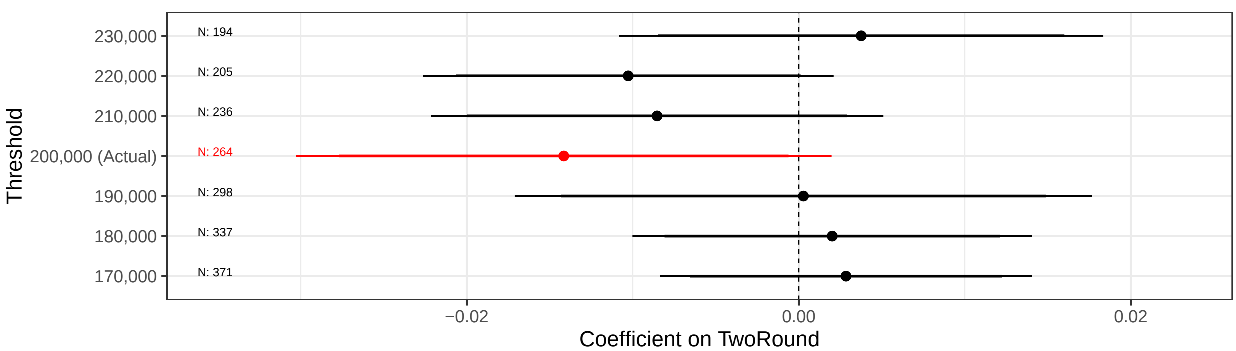
(c) Entropy

Overall concentration of voters for specific candidates, as measured by (a) *Coefficient of variation*, (b) *Fractionalization*, and (c) *Entropy*, using vote counts in polling stations. Vote shares are from the first round. The thicker horizontal lines represent the 90% confidence interval and the thinner horizontal lines represent the 95% confidence interval. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Figure C.21 Regression discontinuity coefficients on the candidate-level concentration in voters at different thresholds



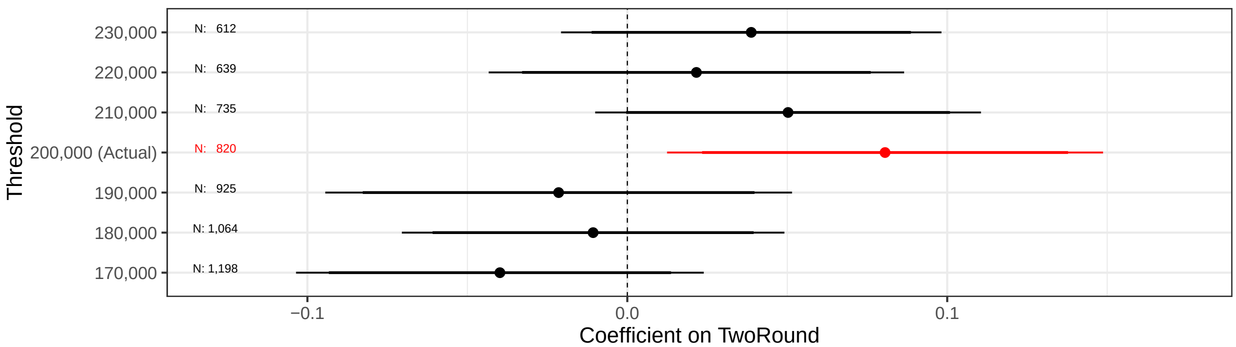
(a) Standard deviation in votes for 1st place candidate



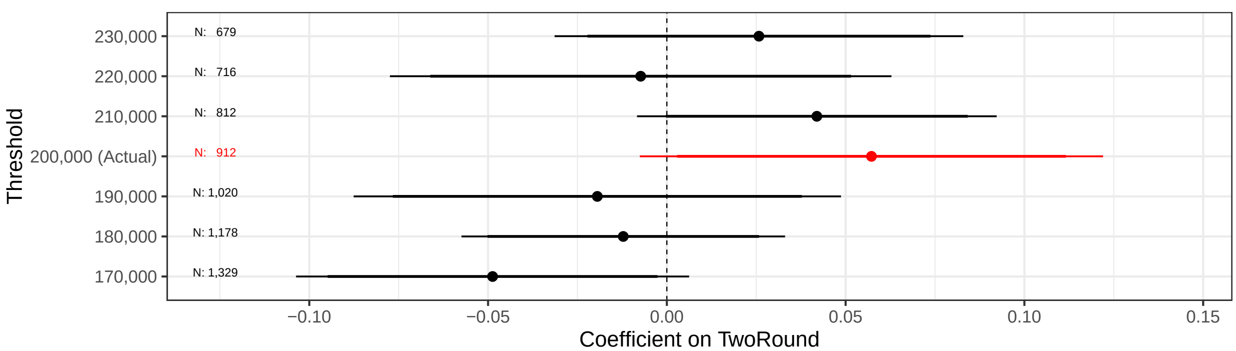
(b) Standard deviation in votes for 2nd place candidate

Standard deviation in a candidate's vote counts across polling stations, for the (a) 1st place and (b) 2nd place candidate. Vote shares are from the first round. The thicker horizontal lines represent the 90% confidence interval and the thinner horizontal lines represent the 95% confidence interval. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Figure C.22 Regression discontinuity coefficients on overall level of resources in municipal schools at different thresholds



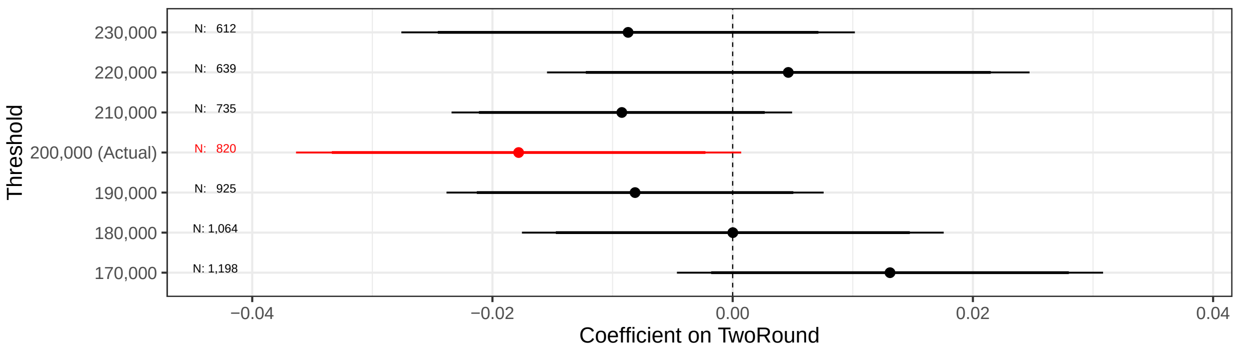
(a) Equipment, mean level of resources



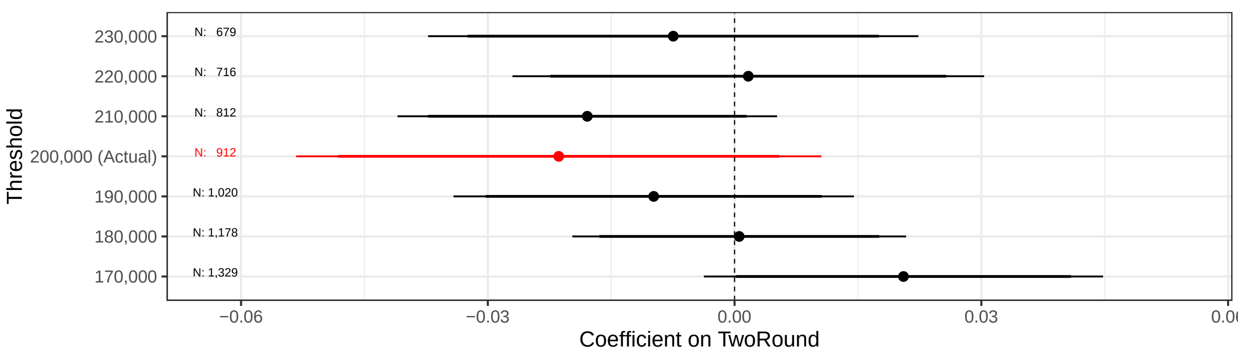
(b) Infrastructure, mean level of resources

Equipment and *Infrastructure* are indices constructed by taking the first principal component of a school's equipment and infrastructure elements, then calculating the school's percentile in the national distribution. *Mean level of resources* is the mean index level across schools in the municipality for (a) equipment and (b) infrastructure. The thicker horizontal lines represent the 90% confidence interval and the thinner horizontal lines represent the 95% confidence interval. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Figure C.23 Regression discontinuity coefficients on distribution of resources in municipal schools at different thresholds



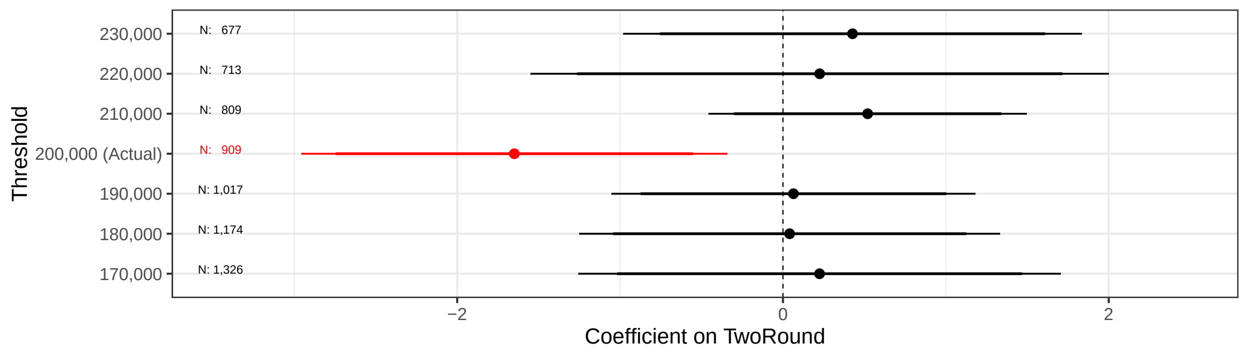
(a) Equipment, standard deviation in resources



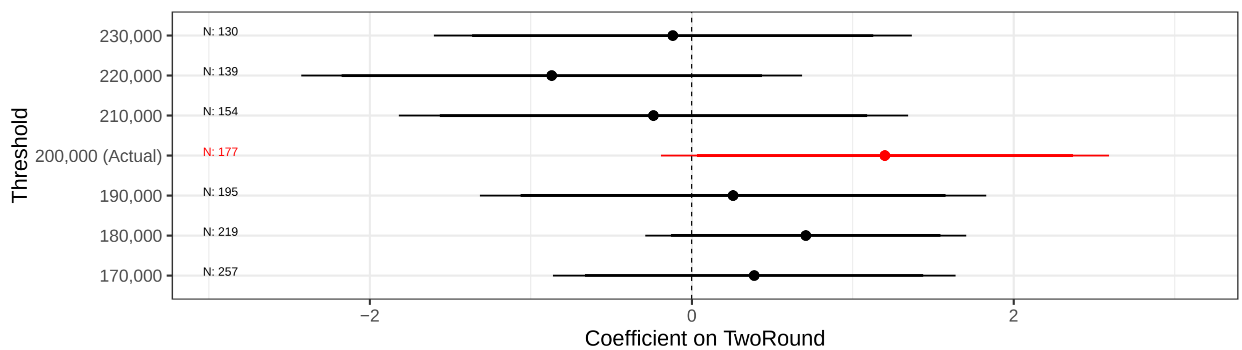
(b) Infrastructure, standard deviation in resources

Equipment and *Infrastructure* are indices constructed by taking the first principal component of a school's equipment and infrastructure elements, then calculating the school's percentile in the national distribution. *Standard deviation in resources* is the standard deviation in the index across schools in the municipality for (a) equipment and (b) infrastructure. The thicker horizontal lines represent the 90% confidence interval and the thinner horizontal lines represent the 95% confidence interval. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.

Figure C.24 Regression discontinuity coefficients on municipal education outcomes at different thresholds



(a) Drop-out rate



(b) Elementary literacy rate

Drop-out rate is from the School Census. It is the mean rate across schools in the municipality. *Elementary literacy rate* is from the 2000 and 2010 Demographic Census. It is the literacy rate of cohorts who are of elementary school age during the mayoral term. The thicker horizontal lines represent the 90% confidence interval and the thinner horizontal lines represent the 95% confidence interval. *Estimation method:* Local linear regression with election-year fixed effects and a 50,000 voter bandwidth. Population density included as a control separately across the cutoff. Standard errors clustered at the municipality level.