

WP/21/227

IMF Working Paper

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

by Moya Chin

***IMF Working Papers* describe research in progress by the author(s) and are published to elicit comments and to encourage debate.** The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

I N T E R N A T I O N A L M O N E T A R Y F U N D

IMF Working Paper

Institute for Capacity Development

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

Prepared by Moya Chin¹

Authorized for distribution by Stephan Danninger

August 2021

IMF Working Papers describe research in progress by the author(s) and are published to elicit comments and to encourage debate. The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

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. Relying on a regression discontinuity design, I contrast single- and two-round elections in Brazilian municipal races. Two-round elections use two rounds of voting to elect a winner, ensuring that the eventual winner obtains at least 50% of the vote. Theoretically, this can provide incentives for candidates to secure a broader base of support. Consistent with this, I show that in two-round elections, candidates represent a more geographically diverse group of voters, public schools have more resources, and there is less variation in resources across public schools. Effects appear to be driven by strategic responses of candidates, rather than differential entry into races. These results suggest that two-round elections can lead candidates to secure broader bases of support and to distribute public goods more broadly.

JEL Classification Numbers: D72, D78, H75

Keywords: Electoral institutions, voting theory, political responsiveness, political favoritism, education resources

Author's E-Mail Address: mchin2@imf.org

¹ Sincere thanks to 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, Ravi Jagadeesan, Ross Mattheis, Tzachi Raz, Jonathan Roth, Juan Francisco Santini, Henrik Sigstad, Niharika Singh, Edoardo Teso, Clemence Tricaud, and seminar participants at Harvard University, NEUDC (2019), and the IMF ICD seminar. Roberto Acioly, Renata Dias, and Julia Orrico provided excellent research assistance. I am grateful to IQSS and the Harvard University Economics Department for financial support. The views expressed here are personal and should not be attributed to the IMF, its Executive Board, or its management.

Table of Contents

Abstract.....	2
I. Introduction.....	6
II. Institutional Context.....	9
III. Empirical Strategy.....	11
A. Econometric framework.....	11
B. Identification	11
C. Data sources	15
D. Measuring geographic concentration of voters.....	16
IV. The Effect of the Two-Round System.....	18
A. The geography of votes.....	18
B. The allocation of municipal resources.....	22
C. Downstream outcomes	25
D. Robustness of main results.....	27
V. Mechanisms.....	29
A. Selection in candidates.....	29
B. Strategic responses by candidates	32
VI. Theoretical Framework.....	34
A. The environment	35
B. Equilibrium strategies.....	37
C. Comparing single- to two-round elections.....	38
VII. Conclusion.....	40
References.....	42
Appendix A. Data Appendix	44
Appendix B. Additional Figures and Tables.....	46
Appendix B-1. Identification.....	46
Appendix B-2. Other results on electoral outcomes.....	48
Appendix B-3. Other results on education resources.....	50
Appendix B-4. Robustness tests on RDD design.....	51
Appendix B-5. Placebo tests	57
Appendix C. Theory Appendix	63
Appendix C-1. Assumption: Swingable voters.....	63
Appendix C-2. Deriving candidate maximization problems.....	64
Appendix C-3. Assumption: Contestability of localities.....	65

Appendix C-4. Assumption: C never makes it to the second round	66
Appendix C-5. First order conditions.....	67
Appendix C-6. Comparing single- to two-round elections: Predictions 3 and 4.....	68
Appendix C-8. Relaxing the assumption that $\alpha > 0$	70

Figures

1: Density of elections around the 200,000 registered voter threshold.....	12
2: Regression discontinuity plots of overall concentration of voters for specific candidates.	19
3: Regression discontinuity plots of the candidate-level concentration in voters.....	20
4: Regression discontinuity plots of the overall level of resources in municipal schools	23
5: Regression discontinuity plots of the distribution of resources in municipal schools.....	24
B-1: Regression discontinuity plots of the probability of falling above/below other policy thresholds	46
B-2: Regression discontinuity plots of pre-treatment population density.....	47
B-3: Regression discontinuity coefficients on pre-treatment population density at different bandwidths	48
B-4: Regression discontinuity coefficients on overall concentration of voters for specific candidates at different bandwidths.....	51
B-5: Regression discontinuity coefficients on the candidate-level concentration in voters at different bandwidths	52
B-6: Regression discontinuity coefficients on overall level of resources in municipal schools at different bandwidths.....	54
B-7: Regression discontinuity coefficients on distribution of resources in municipal schools at different bandwidths.....	55
B-8: Regression discontinuity coefficients on municipal education outcomes at different bandwidths	56
B-9: Regression discontinuity coefficients on overall concentration of voters for specific candidates at different thresholds.....	59
B-10: Regression discontinuity coefficients on the candidate-level concentration in voters at different thresholds	60
B-11: Regression discontinuity coefficients on overall level of resources in municipal schools at different thresholds.....	61
B-12: Regression discontinuity coefficients on distribution of resources in municipal schools at different thresholds.....	62
B-13: Regression discontinuity coefficients on municipal education outcomes at different thresholds.....	63

Tables

1: Regression discontinuity estimates on municipality pre-characteristics	14
2: Regression discontinuity estimates on the geographic concentration of voters	21
3: Regression discontinuity estimates on other electoral outcomes.....	22
4: Regression discontinuity estimates on resources in municipal schools.....	24

5: Regression discontinuity estimates on resources in municipal schools, for schools at different quartiles in the municipal distribution.....	25
6: Regression discontinuity estimates on municipal education and economic outcomes.....	27
7: Regression discontinuity estimates on candidate characteristics.....	30
8: Regression discontinuity estimates on characteristics of winner.....	31
9: Regression discontinuity estimates on political affiliation of candidates.....	32
10: Estimates on the geographic concentration of voters across rounds in two-round elections	33
11: Regression discontinuity estimates on campaign donations.....	34
A-1: Variables used to construct the equipment index, from the <i>Censo Escolar</i>	44
A-2: Variables used to construct the infrastructure index, from the <i>Censo Escolar</i>	45
B-1: Regression discontinuity estimates on the geographic concentration of voters, using vote shares from the top two candidates only.....	48
B-2: Estimates on the geographic concentration of voters, using vote shares from the final round	49
B-3: Regression discontinuity estimates on the geographic concentration of voters, with number of candidates as a control.....	49
B-4: Regression discontinuity estimates on resources in municipal schools, using z-scores. 50	
B-5: Regression discontinuity estimates on the geographic concentration of voters, without controls.....	53
B-6: Regression discontinuity estimates on resources in municipal schools, without controls.....	55
B-7: Regression discontinuity estimates on municipal education outcomes, without controls.....	56
B-8: Placebo regression discontinuity estimates on the geographic concentration of voters, at 285,714 inhabitant threshold.....	57
B-9: Placebo regression discontinuity estimates on the geographic concentration of voters, at 300,000 inhabitant threshold.....	58

I. 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).

Electoral rules define the coalitions that politicians need in order to win, and this can generate incentives for politicians to appeal to broader groups of voters (Lizzeri and Persico, 2005; Myerson, 1993). This broad appeal can lead politicians to provide public goods with broader benefits (Lizzeri and Persico, 2001; Persson and Tabellini, 1999) or engage in less targeting of public spending to specific groups of voters (Genicot et al., 2020; Milesi-Ferretti et al., 2002). On the other hand, factors such as strategic voting can result in the Condorcet loser winning even when using rules that appear to encourage broad representation (Bouton, 2013). Despite this rich theoretical literature, there is limited causal evidence of the political impacts of electoral rules and how these impacts may translate into economic policy.

This article seeks to identify 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 level of public goods provided and the manner in which they are allocated across the electorate. Single- and two-round systems are the most widely used rules in democratic presidential elections (Bormann and Golder, 2013), but endogenous selection of electoral institutions makes studying their causal effects difficult (Cusack et al., 2007). 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.

In a single-round election, voters vote once and the candidate with the most votes wins. In a two-round election, voters first vote and, if no candidate receives a majority, they vote a second time between the top two candidates.² This difference generates two important distinctions. First, two-round elections 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 either threatens victory in the first round or will be the opposition in the second. Because of these distinctions, two-round elections can incentivize candidates to secure a broader base of support (Bouton, 2013; Bouton and Gratton, 2015).³

² This is the case in Brazil and most countries with two-round systems. A small number of countries use qualified two-round systems, where the vote threshold to trigger a second round is something other than 50%, or majority-plurality two-round systems, where the number of candidates in the second round is determined by a vote threshold.

³ While not a focus of this paper, there is also a large literature arguing that two-round elections 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).

The intuition is that the rules imposed in a two-round election make it more difficult for politicians to win with policies that appeal to a narrow group of voters. To the extent that politicians commit to their campaign promises, the policies they offer in order to win the election can have economic consequences. When it is 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 examine this intuition empirically, I exploit Brazil's threshold-based rule to assign the electoral rule for mayor in each election. I employ a regression discontinuity design across six municipal elections, between 1996 and 2016. I estimate the causal effect of the electoral rule by comparing political and economic outcomes in municipalities just above the registered voter threshold with outcomes in those just below. I obtain three main empirical results.

First, candidates in two-round elections receive broader geographical support, suggesting that two-round elections foster greater inclusiveness. Using vote counts at each polling station, I measure the geographic distribution of voters in the first round with (i) indices of voter concentration to quantify the *overall* level at which voters for specific candidates are geographically concentrated, and (ii) the standard deviation of candidates' vote shares across polling stations to quantify a *candidate*-level measure of geographic concentration. In two-round municipalities, voters for specific candidates are less geographically concentrated, corresponding to a 27.4 – 45.6% reduction from the level in single-round municipalities. The main impact of the electoral rule is on the candidates with a chance of winning, as the decrease in concentration only occurs among the top two candidates. These results are not due to the increased number of candidates in two-round elections.⁴ Further, I find that increased inclusiveness does not only occur through representation but also through voter behavior. Voters are more engaged in the political process and cast significantly fewer blank and invalid ballots in two-round elections.

Second, once in office, politicians elected under two-round systems provide public goods differently, both in the *level* and *distribution* of municipal resources. I measure the provision of a local public good that can be geographically targeted and is controlled by the municipal government: public elementary education. In two-round municipalities, (i) the level of resources present in public schools measures 5.7 – 8.1 percentiles higher in the national distribution, and (ii) the standard deviation of these 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, these differences in resources lead to improved education outcomes in two-round municipalities. Specifically, drop-out rates are lower and literacy rates higher among cohorts

⁴ 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).

of school age during the electoral term. Improvements on more downstream economic indicators are limited. Two-round municipalities do not have significantly different levels of income, employment, or night lights. While electoral rules improve direct policy outcomes, they have limited effects on broader economic conditions.

Turning to mechanisms, I argue that the effects of two-round elections are not driven by differential selection of candidates. I do not find that different types of candidates enter the races, nor do they win. Candidates in two-round elections are not observably different from those in single-round elections 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 – these additional candidates often are from smaller parties and have run in previous elections. However, these candidates are not more likely to win.

Instead, I argue that candidates face different strategic incentives in two-round elections, which lead them to adopt different strategies during the campaign. Geographic concentration decreases between the first and second round of two-round elections, suggesting that candidates adjust their strategies between rounds to consolidate their voter bases.⁵ Candidates in two-round elections also rely less on donations from corporations to finance their campaigns. To the extent that corporations represent narrower groups in the electorate, candidates in two-round elections seem to be adopting strategies that appeal more broadly to individuals rather than corporations. Taken together, these results suggest that mayors in two-round municipalities employ different strategies. These strategies result in mayors building less geographically concentrated constituencies and subsequently allocating public goods in a less geographically concentrated manner.

I propose a model to interpret these results in a context where two-round elections impact politician's strategic incentives. I draw on Genicot et al. (2020) to develop a standard probabilistic voting model where candidates offer policy proposals before the first round that (i) specify the overall size of the government budget and (ii) target government resources to specific localities within a municipality. The election features two candidates contending with a small, third candidate who is non-strategic and commits to allocating all resources to a single locality.

My model predicts that candidates in two-round elections appeal more broadly, both by appealing more to all voters and by appealing to minority groups. Candidates appeal more to all voters by offering a higher overall budget. They do so because each vote is more valuable due to the conditionality embedded in the two-round system: candidates must attain both a plurality and 50% of the votes, else a second round is triggered. Candidates appeal to minority groups by offering more resources to the locality where the third candidate promises

⁵ 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.

resources. Should a second round occur, the top two candidates will need to appeal to these voters, due to the third candidate's absence in the second round. This increased allocation generates a force to reduce inequality in the allocation of government resources.

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 in Italy (Cipullo, 2019), France (Eggers, 2013), Morocco (Pellicer and Wegner, 2013), as well as Brazil (Chamon et al., 2019; Fujiwara, 2011). Of particular interest are Bordignon et al. (2016), who compare single- and two-round elections in Italy and find more policy moderation in two-round municipalities, as measured by the volatility of a municipal tax rate across elections. My results complement Bordignon et al. (2016) by studying 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 on 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. This paper demonstrates 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 II describes the context, and Section III describes the empirical strategy. Section IV presents the results. Section V discusses mechanisms. Section VI presents a theoretical framework for single- and two-round elections to interpret the empirical findings. Section VII concludes.

II. INSTITUTIONAL CONTEXT

Municipal governments in Brazil consist of an elected executive (a mayor, *prefeito*) and legislative body (a council of legislators, *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.⁶ Legislators are elected through an

⁶ In the sample, a second round occurs 57% of the time.

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 Eleitorais 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.⁷ 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 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.⁸ 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-government transfers, which comprise the majority of revenues, and local revenues. The bulk of inter-government transfers come from the state or federal level and are either constitutional automatic transfers (*Fundo de Participação do Municípios*) or discretionary transfers (*convênios*). Municipalities have 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 majority of public goods spending is allocated through the annual budgetary process. Funds for specific public works and services can be allocated through bills submitted by the mayor or legislator. While all budgetary 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 both the size of the municipal budget and how municipal funds are allocated. This study focuses on public goods provision in municipal elementary education for several reasons. First, a large fraction of the municipal budget is allocated to education: in 2012, it represented 30.5% of municipal budgets. Second,

⁷ Voters who fail to vote may provide justification to a local electoral judiciary office. Absent this justification, voters 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.

⁸ Seats for the legislative council are allocated based on the number of votes received by candidates or parties in the coalition.

municipal education is a geographically localized public good. This feature allows mayors to geographically target public goods provision. Third, unlike other public goods, for which municipalities share joint responsibility with the state or federal government, elementary education is almost entirely under the jurisdiction of the municipality.

III. EMPIRICAL STRATEGY

A. Econometric framework

The 200,000 registered voter threshold rule for mayoral elections provides a natural candidate for a regression discontinuity design (RDD). 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:

$$Y_{it} = \beta_1 D_{it} + \beta_2 X_{it} + \beta_3 X_{it} \cdot D_{it} + \gamma_t + \varepsilon_{it} \quad (1)$$

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 represents a municipality and election year, or municipality-year. Equation (1) amounts to fitting two linear regressions using municipality-years 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 (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; Imbens and Kalyanaraman, 2012).⁹

B. 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.

Violations of smoothness

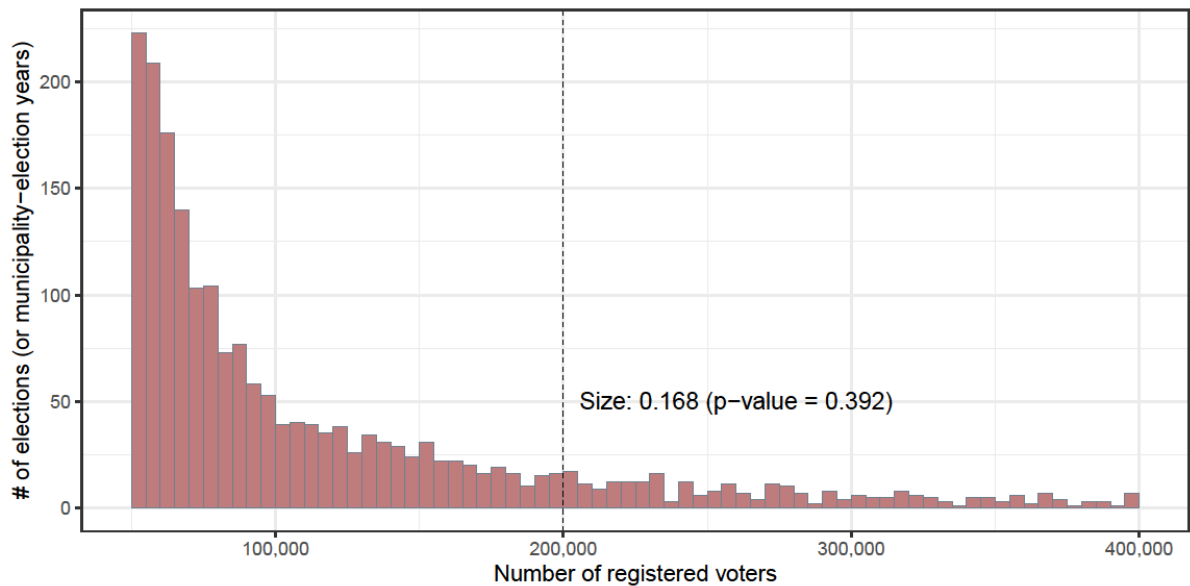
The smoothness assumption can be violated if the threshold choice is motivated by political or economic factors. There appears to be little evidence for this. The choice of 200,000

⁹ 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.

registered voters as the threshold was somewhat arbitrary and mainly reflected practical concerns regarding the cost of holding a second round (Chamon et al., 2019; Fujiwara, 2011). In addition, given that the threshold was set in the federal constitution in 1988, it is unlikely that politicians chose the threshold anticipating which municipalities would be above or below the threshold in 1996 and later.

A second possibility is if municipalities selectively sort across the threshold. Practically, it is difficult for municipalities to do so, since voter registration is mandatory and handled by state electoral authorities. Both visual inspection and the estimated size of the discontinuity in the density of registered voters at the threshold do not provide evidence of sorting (Figure 1).

Figure 1: 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 is estimated based on McCrary (2008). Due to the skewed right tail of municipality sizes, the size of the discontinuity was estimated using a sample excluding those above the 99.9 percentile of registered voters. An “election” is defined as a municipality-election year. Bin sizes are 10,000 voters.

A third violation is if policies other than the electoral rule change discretely at the threshold. While a number of policies in Brazil are implemented using thresholds, these use population counts which, while highly correlated, do not vary one-to-one with the number of registered voters. To the best of my knowledge, there are no other policies at 200,000 registered voters. Two policy thresholds potentially lie close to 200,000 registered voters: 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 show that the electoral outcomes are smooth at these thresholds (Appendix B-5)

and that any effect of these policies are balanced across the electoral rule threshold (Appendix B-1).¹⁰

The last possibility is that potential confounds change discretely at the threshold. I test this by estimating equation (1) on pre-treatment characteristics of municipalities. I discuss this in detail in the following section.

Balance on pre-treatment characteristics

Since treatment (the two-round election) is determined by the number of registered voters, municipalities can move into the treatment group or be treated multiple times.¹¹ As a result, “pre-treatment” can be defined in two ways: 1) prior to the introduction of the threshold rule in the 1988 Constitution, and 2) prior to the most recent election the municipality was untreated or prior to 1996 if the municipality was never untreated.¹² Outcomes for (1) are measured from the 5% population sample of the 1980 census, the census prior to 1988. Outcomes for (2) are measured either from the census prior to the most recent year in a single-round election (the 1991, 2000, or 2010 census) or from the 1991 census, for municipalities that were in a two-round election in 1996.

There is no significant treatment effect on nearly all outcomes measured prior to the 1988 Constitution and prior to the most recent election in a single-round election (Table 1). I test for pre-treatment imbalance on economic characteristics, income and demographic segregation, and income inequality.¹³ Another concern is that there are factors that change discontinuously at the threshold and affect which municipalities move into treatment and the length of treatment, but I do not find that pre-treatment population growth is discontinuous at the threshold.

¹⁰ Appendix B.5 estimates placebo regressions where the electoral rule is assigned at these population thresholds. There are no discontinuities of a similar size in the mayoral electoral outcomes. Because legislator salaries and legislature size can have economic effects, placebo regressions for public goods outcomes are not meaningful. Appendix B.1 shows that the probability of being above or below these thresholds does not change discontinuously at 200,000 voters.

¹¹ Municipalities can also move out of treatment. However, while municipalities do experience population decline, none moves below the 200,000 voter threshold.

¹² The earliest electoral data available is 1996, so I cannot observe whether municipalities are treated or untreated prior to 1996. Since 1992 is the only unobserved municipal election after the 1988 Constitution and only 45 municipalities have moved across the 200,000 voter threshold between 1996 and 2016, it is unlikely that many, if any, municipalities experienced multiple electoral rules between 1992 and 1996.

¹³ Economic characteristics include unemployment, literacy, and low income rate. Segregation is calculated using the entropy index (see Section III.D for the formula), 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. Income inequality is measured using the Gini coefficient.

Table 1: Regression discontinuity estimates on municipality pre-characteristics

<i>Panel A: Characteristics measured prior to the 1988 Constitution</i>					
	% illiterate	% low income	Unempl. rate	Pop. density	
TwoRound	-1.499 (1.057)	-0.751 (1.148)	0.044 (0.145)	-397.678** (184.347)	
Single-round mean	19.029	50.182	1.978	516.067	
Observations	293	293	293	293	
<i>Panel B: Characteristics measured prior to most recent single-round election</i>					
	Muni. area change (%)	Pop. growth (%)	Pop. density	Income seg.	Dem. seg.
TwoRound	-0.601 (2.296)	-0.468 (1.260)	-663.035** (332.555)	-0.003 (0.004)	0.001 (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 (0.590)	-18.964 (33.328)	1.211 (2.290)	-0.652 (1.114)	0.004 (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 km². *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.

One exception is a large and significant effect on population density. However, there are several reasons that this is likely a false positive. First, much of the effect is driven by a single outlier municipality.¹⁴ Visually, there is no discontinuity and the estimate is not robust to other bandwidths (Appendix B-1). 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 would bias 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 municipality borders. Such manipulation is unlikely to be motivated by the electoral rule given that this effect is seen in 1980, prior to the introduction of the threshold rule. In addition, I find no differences in the

¹⁴ Dropping one outlier municipality reduces the coefficient by 38.2%. Although the estimate is significant at the 10% level, this suggests that a few municipalities are driving this effect.

urbanization rate nor do municipality area or population growth change discontinuously across the threshold (Table 1). Nevertheless, I control for population density in all specifications:

$$Y_{it} = \beta_1 D_{it} + \beta_2 X_{it} + \beta_3 X_{it} \cdot D_{it} + \beta_4 Z_{it} + \beta_5 Z_{it} \cdot D_{it} + \gamma_t + \varepsilon_{it} \quad (2)$$

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

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. 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.

C. Data sources

Electoral data

Data on municipality elections come from Brazil's electoral authority (*Tribunal Superior Eleitoral*, or TSE). The electoral data provides information on the candidates running, the party and coalition each candidate belongs to, and the number of votes received. The data encompasses 6 municipal elections between 1996 and 2016, totaling 32,767 elections across 5,568 municipalities.

Electoral results are available for each polling station (*seção 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* level (see Section III.D on the measures used). Baseline results use votes from the first round of elections, to allow comparability 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).

Public goods provision in schools

To measure public goods provision in elementary education, I use the 1997-2016 School Census (*Censo Escolar*), an annual census of schools conducted by the Ministry of Education. I use the Census to calculate the level of resources present in schools across two categories: equipment and infrastructure. Equipment includes *movable* elements, such as the number of computers and availability of air conditioning. Infrastructure includes *immovable* elements, such as the number of classrooms, sanitation, and availability of a library. Appendix A provides a full description of these categories. I construct for each school an index of resources, separately for equipment and infrastructure, by taking the principal

component of the elements and computing the school's percentile rank within the country for each year.¹⁵

D. Measuring geographic concentration of voters

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

Overall geographic concentration of voters

To measure the *overall* geographic concentration of voters, I use three indices from the racial segregation literature that measure multi-group spatial segregation: the coefficient of variation, the fractionalization index, and the entropy index. These indices and their properties are described in White (1986) and Reardon and Firebaugh (2002). The indices assume a value of 1 if there is full geographic concentration of voters, or where each polling station contains voters for only one candidate. The indices assume a value of 0 if there is full geographic dispersion of voters, or where each polling station contains the same composition of voters as the municipality as a whole.

The coefficient of variation, s_m , is defined as:

$$s_m = \frac{1}{K_m - 1} \sum_{k=1}^{K_m} \sum_{i=1}^{I_m} \frac{n_{im}}{N_m} \frac{(p_{imk} - p_{mk})^2}{p_{mk}} \quad (3)$$

s_m 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, then $p_{imk} = p_{mk}$ and $s_m = 0$.

The fractionalization index, f_m , is defined as:

$$f_m = \frac{\hat{f}_m - \bar{f}_m}{\hat{f}_m} \quad \text{where} \quad \begin{aligned} \hat{f}_m &= \sum_{k=1}^{K_m} p_{mk}(1 - p_{mk}) \\ \bar{f}_m &= \sum_{k=1}^{K_m} \sum_{i=1}^{I_m} \frac{n_{im}}{N_m} p_{imk}(1 - p_{imk}) \end{aligned} \quad (4)$$

\hat{f}_m is the fractionalization in the municipality and \bar{f}_m is the average fractionalization across polling stations. Fractionalization, also known as the interaction index, is a measure of

¹⁵ The variables in the School Census varied from year to year, making it difficult to compare the raw PCA index across years. Calculating a school's percentile rank for each year allows for comparison across years.

concentration and 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 across polling stations, normalized by the level in the municipality to keep the index between 0 and 1. Alternatively, 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, then $\hat{f}_m = \bar{f}_m$ and $f_m = 0$. When each polling station contains only one type of voter, or when there are large differences between polling stations, then $\bar{f}_m = 0$ and $f_m = 1$.

The entropy index, h_m , is defined as:

$$h_m = \frac{\hat{h}_m - \bar{h}_m}{\hat{h}_m} \quad \text{where} \quad \begin{aligned} \hat{h}_m &= - \sum_{k=1}^{K_m} p_{mk} \ln p_{mk} \\ \bar{h}_m &= - \sum_{k=1}^{K_m} \sum_{i=1}^{I_m} \frac{n_{im}}{N_m} p_{imk} \ln p_{imk} \end{aligned} \quad (5)$$

\hat{h}_m is the entropy in the municipality and \bar{h}_m is the average entropy across polling stations. Entropy is a measure of concentration and measures how far the population is from equal representation of all groups. The interpretation and range of values of h_m are the same as that of f_m .

In my sample, the correlation between these indices is between 0.89 and 0.96.¹⁶ 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

There is a concern that the indices may be mechanically related to the number of candidates. The majority of the discussion below 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 depends on the number of candidates. Dividing by $K_m - 1$ removes this mechanical effect.

For the fractionalization index, \bar{f}_m attains a maximum of \hat{f}_m and \hat{f}_m attains a maximum of $1 - 1/K_m$. For the entropy index, \bar{h}_m attains a maximum of \hat{h}_m and \hat{h}_m attains a maximum of $\ln K_m$. As a result, before dividing by \hat{f}_m and \hat{h}_m , these indices will depend on the number of candidates. Dividing by \hat{f}_m and \hat{h}_m removes part of the mechanical effect.

¹⁶ 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.

While these indices will not monotonically depend on the number of candidates, these indices may still be affected by the number of candidates. Section IV.D performs several robustness exercises to address this concern.

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 whose voters are spread across many areas in the municipality, we will observe the vote share varying less across polling stations. The standard deviation in a candidate's vote share, σ_{mk} , is defined as:

$$\sigma_{mk} = \left(\frac{1}{I_m - 1} \sum_{i=1}^{I_m} \left(p_{imk} - \frac{1}{I_m} \sum_{i=1}^{I_m} p_{imk} \right)^2 \right)^{1/2} \quad (6)$$

Using only the top two candidates

In some cases, the indices use vote shares from the top two candidates only, in other words assuming that only the top two candidates are in the race. The advantage of doing this is that it fixes the number of candidates and ignores the potential dilution of votes from lower-placed 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} .

IV. THE EFFECT OF THE TWO-ROUND SYSTEM

I present three main results. One, candidates in two-round elections receive broader geographical support. Two, once in office, mayors elected under two-round systems provide more resources to schools and distribute these resources more equitably. In other words, these politicians are represented by a broader group of voters and this in turn leads to higher and broader provision of public goods. Three, these differences in public goods provision lead to tangible improvements in downstream education outcomes.

A. The geography of votes

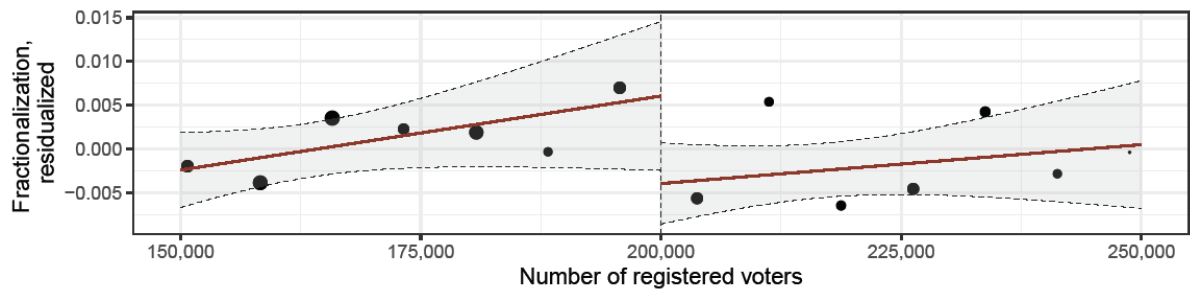
Do candidates in two-round elections secure broader bases of support? I provide evidence that, in two-round elections, voters are *overall* less geographically concentrated, and that it is the *top two* candidates who receive support from a geographically broader group of voters.

Geographic concentration of voters

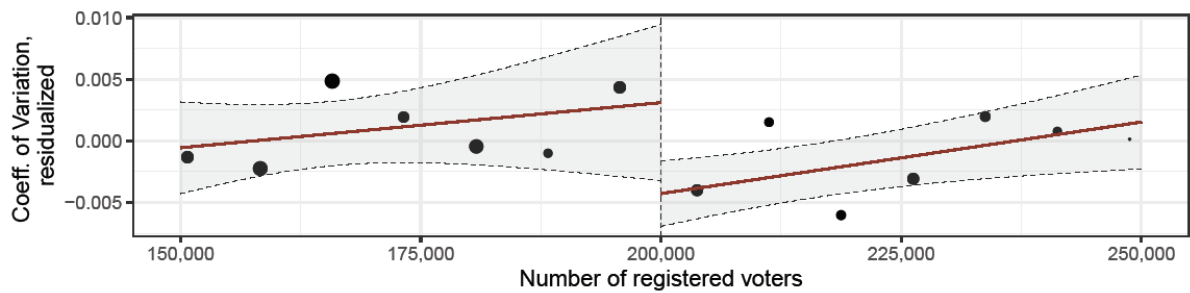
Overall, voters for specific candidates are less geographically concentrated in two-round elections, whether concentration is measured by the coefficient of variation, fractionalization index, or entropy index (Figure 2 and Panel A of Table 2). There is less variation in voter composition between polling stations or, in other words, the composition of voters in polling

stations is closer to the composition of voters in the municipality. Depending on the measure of concentration used, two-round elections experience a reduction of 27.4%–45.6% of the average level of concentration in single-round municipalities, which corresponds to 0.397–0.548 standard deviations of the respective outcomes.

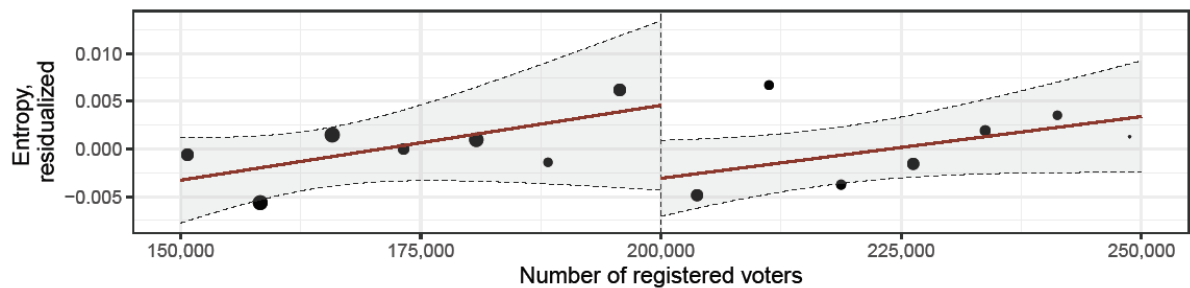
Figure 2: Regression discontinuity plots of overall concentration of voters for specific candidates



(a) Coefficient of variation



(b) Fractionalization



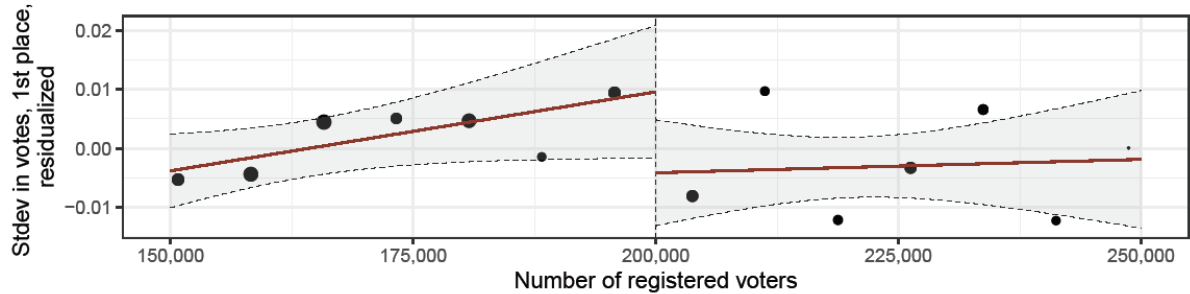
(c) Entropy

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.

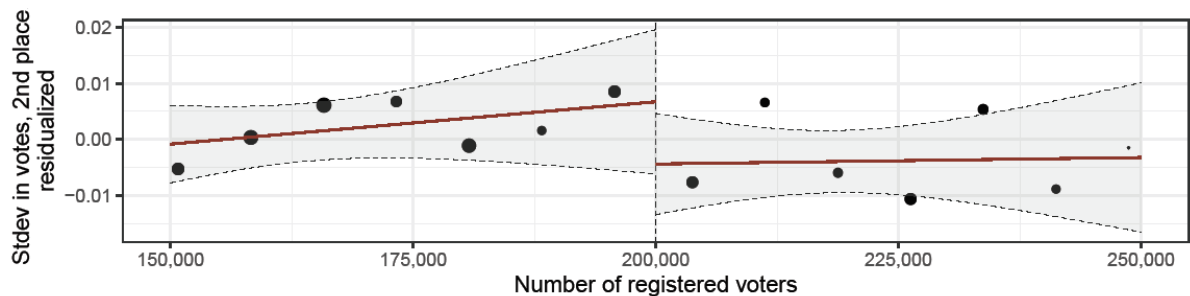
Turning to support for each candidate, not all candidates obtain support from geographically broader constituencies. Voters for the top two candidates are less concentrated in two-round elections, but the concentration of voters for the third and fourth placed candidates is not

significantly different between single- and two-round elections (Figure 3 and Panel B of Table 2). Estimates for the top two candidates are similar in magnitude: there is a 0.0167 reduction in variance of support for the first placed candidate across polling stations (20.9% of the single-round mean and 0.474 standard deviations), and there is a 0.0142 reduction in variance of support for the second placed candidate (18.9% of the single-round mean and 0.430 standard deviations). Estimates for the third and fourth placed candidates are close to zero and insignificant.

Figure 3: 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.

To provide further support that the top two candidates drive the reduced concentration of voters in two-round elections, the effects on overall concentration are significantly stronger when using vote shares from the top two candidates only (Appendix B-2). While it is more intuitive to compare vote shares from the first round, these effects extend to using vote shares from the final round – the first round in single-round elections and the second round in two-round elections (Appendix B-2).

Table 2: Regression discontinuity estimates on the geographic concentration of voters

<i>Panel A: Concentration indices of voters for specific candidates</i>				
	Coefficient of variation	Fractionalization	Entropy	
TwoRound	−0.009*** (0.003)	−0.012** (0.005)	−0.008* (0.005)	
Potential bias	0.0008	−0.0002	−0.0923	
Single-round mean	0.019	0.027	0.030	
Observations	264	264	264	
<i>Panel B: Standard deviation in vote shares for each candidate</i>				
	1st place candidate	2nd place candidate	3rd place candidate	4th place candidate
TwoRound	−0.017** (0.007)	−0.014* (0.008)	−0.005 (0.007)	0.004 (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.

Voter engagement

The pattern of support for specific candidates indicates 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 3). Given evidence that these ballots are often cast by dissatisfied or disinterested voters (Gonzales et al., 2019), the reduction suggests that voters in two-round elections engage in the electoral process at higher rates.¹⁷

¹⁷ 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. Gonzalez et al. (2019) provide empirical evidence for this interpretation, as they find that forced electoral participation increases the number of blank and invalid ballots cast.

Table 3: Regression discontinuity estimates on other electoral outcomes

	Turnout	Blank/invalid ballots	# candidates
TwoRound	0.006 (0.008)	-3.821** (1.670)	1.273*** (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.

B. The allocation of municipal resources

I next investigate the impact 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 resources in municipal schools.

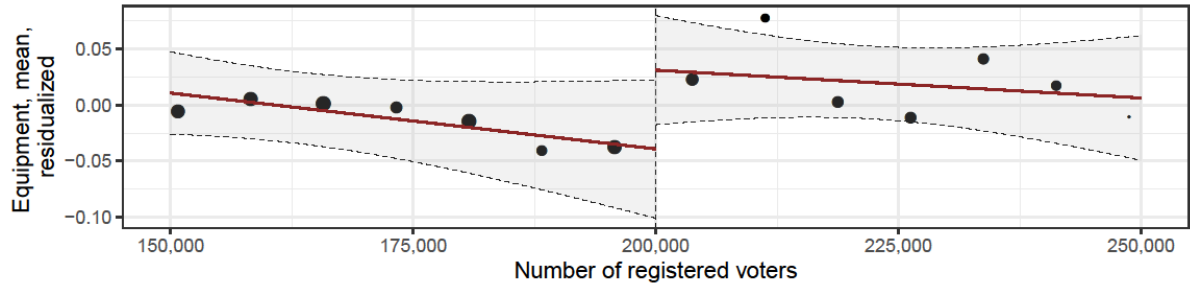
Level of resources in municipal schools

There is a significant increase in resources in municipal schools in two-round municipalities compared to those in single-round municipalities (Figure 4 and Columns 1 and 2 in Table 4). Schools in two-round municipalities are 8.1 and 5.7 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 difficult to manipulate as allocating new infrastructure requires more time and capital than allocating equipment. Empirically, infrastructure resources are less responsive to the electoral cycle (results available on request).

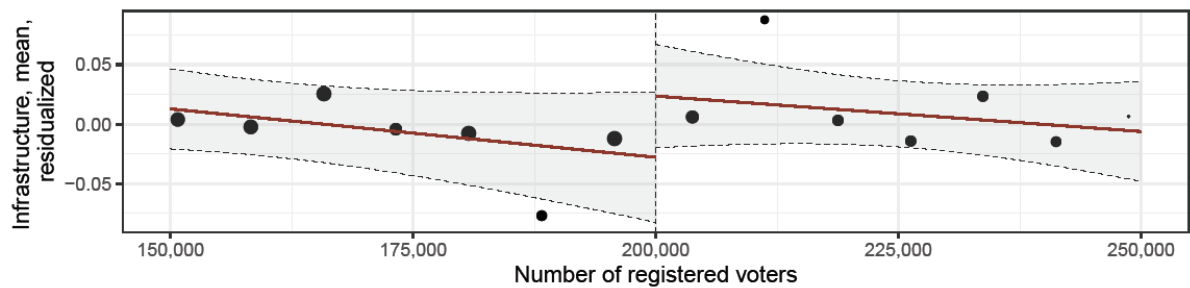
Distribution of resources across municipal schools

In addition to differences in the overall levels, resources are distributed more evenly across schools in two-round municipalities (Figure 5 and Columns 3 and 4 in Table 4). The standard deviation in equipment resources across schools is 1.8 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 (-2.1 percentiles), the difference is not significant.

Figure 4: Regression discontinuity plots of the overall level of resources in municipal schools



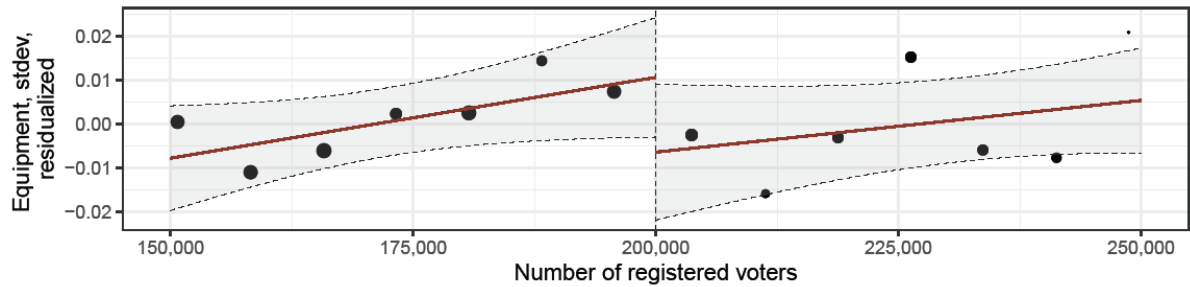
(a) Equipment, mean level of resources



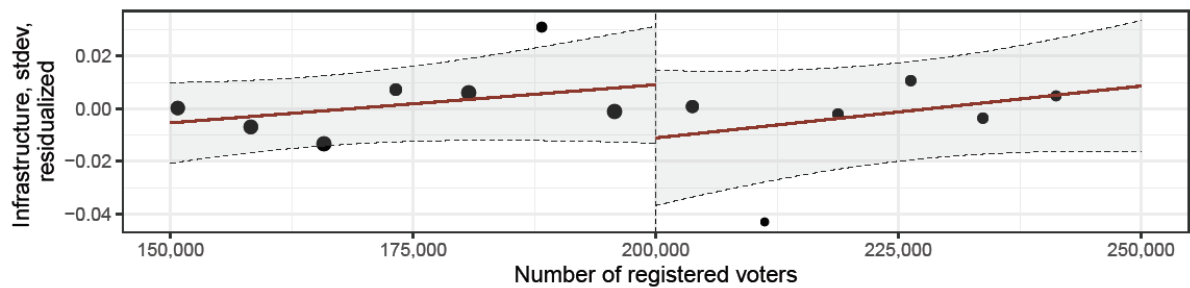
(b) Infrastructure, mean level of resources

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 5: 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 4: Regression discontinuity estimates on resources in municipal schools

	<i>Mean level of resources</i>		<i>Standard deviation in resources</i>	
	Equipment	Infrastructure	Equipment	Infrastructure
TwoRound	0.081** (0.035)	0.057* (0.033)	-0.018* (0.009)	-0.021 (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.

To better understand how resources are distributed, I estimate effects for schools at different parts of the distribution in the municipality. If the variance in resources is lower, schools with the least (most) resources in the municipality should have more (less) resources. I group schools into quartiles, which are defined by each school's percentile in the municipal distribution prior to the election. The increased level of resources is concentrated in schools located at the lower end of the distribution (Table 5). Schools in the bottom quartile are 8.2 percentiles higher in equipment resources and 11.6 percentiles higher in infrastructure resources. Schools in the second quartile also experience gains, though smaller – 6.6 percentiles in equipment resources and 10.2 percentiles in infrastructure resources. There is no significant difference in resources in schools at the top of the distribution.

Table 5: Regression discontinuity estimates on resources in municipal schools, for schools at different quartiles in the municipal distribution

	<i>Mean level of resources in schools at different quartiles</i>			
	1st quartile (Bottom 25%)	2nd quartile	3rd quartile	4th quartile (Top 25%)
<i>Panel A: Equipment</i>				
TwoRound	0.082** (0.035)	0.066* (0.037)	0.069* (0.042)	0.038 (0.029)
Single-round mean	0.652	0.733	0.781	0.856
Observations	700	728	760	748
<i>Panel B: Infrastructure</i>				
TwoRound	0.116** (0.046)	0.102** (0.047)	0.056 (0.035)	0.013 (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.

C. Downstream outcomes

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

Education outcomes

I measure four education outcomes. Using the School Census, I measure the drop-out, failing, and passing rate in public municipal schools. Using the 2000 and 2010 Demographic Censuses, I measure the literacy rate among cohorts who were of elementary school age during the electoral term. In two-round municipalities, drop-out rates are significantly lower, 1.65 percentage points off a baseline of 3.21 percentage points, and literacy rates are significantly higher among elementary cohorts, 1.20 percentage points off a baseline of 91.45 percentage points (Panel A of Table 6). While differences in the failing and passing rates are not significant, the direction of the estimates suggests improvements in two-round municipalities: failing rates are lower and passing rates higher.

Economic outcomes

While improvements in education outcomes may lead to improvements in broader economic outcomes, I do not find that this is the case in two-round municipalities (Panel B of Table 6). Using the 2000 and 2010 Demographic Census, I measure the fraction of low-income households, income per capita, and the unemployment rate. The caveat is that these outcomes are measured between 2 and 10 years after the corresponding election.¹⁸ Using the 1997-2013 NOAA night lights series, I measure the mean night lights level in the municipality. The fraction of low-income households is significantly lower in two-round municipalities, but this may be a false positive, as the coefficient is unrealistically large relative to the mean. The direction of the estimates on the other outcomes suggests that two-round elections lead to improved economic outcomes (income per capita is higher, unemployment is lower, and night lights is higher), but not significantly so.

There may be several reasons why I do not observe improved economic outcomes. One, two-round elections may have no effect on the broader economy in Brazil. While mayors are responsible for a range of public goods, such as health, education, local infrastructure, and urban planning, their level of influence on outcomes such as income and employment may not be high. Two, improvements in economic outcomes may not occur in the short term. Outcomes such as income and night lights may take more than 2 to 10 years to improve. Three, 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.

¹⁸ 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.

Table 6: Regression discontinuity estimates on municipal education and economic outcomes

<i>Panel A: Education outcomes</i>				
	Drop-out rate	Failing rate	Passing rate	Elem. literacy rate
TwoRound	-1.649** (0.667)	-0.747 (1.115)	2.330 (1.459)	1.199* (0.710)
Single-round mean	3.211	8.645	88.283	91.445
Observations	909	908	909	177
<i>Panel B: Economic outcomes</i>				
	Low income rate	Income per capita	Unemployment rate	Night lights
TwoRound	-5.186* (3.079)	64.667 (61.782)	-0.964 (0.635)	2.715 (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.

D. Robustness of main results

Bias in measures of concentration

I address two potential sources of mechanical bias in the concentration indices.

One concern 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 number of valid votes at each polling station varies smoothly across the threshold (results available on request).

A second concern, although potentially a mechanism, is the increased number of candidates in two-round elections (Table 3). As discussed in Section III.D, while the indices may be affected by the number of candidates, the direction of bias is not monotonic. Nevertheless, I perform three robustness checks, which can be found in Appendix B-2.

One, while the number of candidates is a bad control, as it is an endogenous outcome, including it as a control does not affect the qualitative results. Two, I simulate the effect of adding an additional candidate to all single-round elections.¹⁹ 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 2). However, the bias is substantial for the entropy index. Third, to maintain the same number of candidates across single- and two-round elections, using only the vote shares from the top two candidates does not substantially change the results.

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 z-scores of each school's resources, rather than the first principal component, does not affect the qualitative results (Appendix B-3).

RDD design

I investigate the robustness of my results to the regression discontinuity design (Appendix B-4). The results are not driven by the choice of bandwidth, whether fixed or chosen by a data-driven method. 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. In addition, dropping controls from the regression – namely, population density and election-year fixed effects – does not substantially affect the results, although the results are noisier.

Placebo tests

As discussed in section III.B, I do not find that policy thresholds at 285,714 and 300,000 inhabitants confound my results. In placebo regressions, there are no similar effects on the electoral outcomes at these thresholds (Appendix B-5).²⁰

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 (Appendix B-5). The

¹⁹ The simulation adds to each polling station a last placed candidate who receives a vote share equal to the average vote share of last placed candidates (1.5%). A proportionate number of votes is taken from the other candidates, to ensure that the total number of voters remains the same. The estimated bias is the change between the actual outcome and the simulated outcome.

²⁰ 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.

treatment effect is isolated to the actual threshold: there are no estimates with the same size and significance as at the actual threshold.

V. MECHANISMS

Broadly, there are two reasons why two-round elections can lead to different outcomes. Selection is one mechanism: different types of candidates may *enter* two-round elections, or different types of candidates may *win* two-round elections. Strategic incentives are another mechanism: candidates adopt different behaviors during the campaign and in office.²¹ In the following section, I explore these explanations and provide suggestive evidence that candidates' strategic responses explain a larger part of the effect of the two-round election.

A. Selection in candidates

Candidates in two-round elections may have a broader group of supporters because different types of candidates *enter* electoral races. I do not find significant differences in the observable characteristics of candidates: age, sex, educational attainment, state of birth, or occupational background (Table 7). I also do not find that different types of candidates *win* in two-round elections, as winners are not observably different along these characteristics (Table 8). While candidates may differ along other characteristics I am not able to measure, such as charisma or competence, I find no evidence that observably different types of candidates enter or win two-round elections.

There are differences in political affiliation among candidates who enter the races. In two-round elections, there are more candidates from small parties and who previously ran as mayoral candidates, but they are not more likely to win (Table 9).²² 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, I argue that this cannot explain the reduced concentration in vote shares.

Why do small candidates enter two-round elections, particularly when they are not more likely to win? One motivation is to build support for subsequent elections, which explains why more candidates with previous campaign experience enter two-round elections. A second motivation is to gain positions in the elected administration. Third, and most interestingly, these candidates may seek to influence the top candidates' platforms, either through the political competition they pose or through direct bargaining.

²¹ 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 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.

²² I define a "small party" as any party that is not one of the top 5 parties in Brazil by national membership.

Table 7: Regression discontinuity estimates on candidate characteristics

<i>Panel A: Demographic characteristics of candidates</i>				
	Age	Female	Univ. degree	Born same state
TwoRound	0.470 (1.514)	-0.065 (0.040)	-0.045 (0.048)	-0.070 (0.049)
Single-round mean	49.955	0.129	0.796	0.784
Observations	264	263	263	263

<i>Panel B: Previous occupation of candidates</i>			
	Public sector	Technical	Business
TwoRound	-0.008 (0.063)	0.014 (0.059)	0.017 (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.

Table 8: Regression discontinuity estimates on characteristics of winner

<i>Panel A: Demographic characteristics of winners</i>				
	Age	Female	Univ. degree	Born same state
TwoRound	0.160 (2.763)	-0.027 (0.087)	0.023 (0.100)	-0.041 (0.071)
Single-round mean	51.608	0.112	0.832	0.789
Observations	264	263	263	263

<i>Panel B: Previous occupation of winners</i>			
	Public sector	Technical	Business
TwoRound	-0.088 (0.137)	0.069 (0.131)	0.059 (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.

Table 9: Regression discontinuity estimates on political affiliation of candidates

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

B. Strategic responses by candidates

I find limited evidence for selection as a mechanism. Instead, I argue that candidates face different strategic incentives in two-round elections, and that these incentives lead candidates to adjust their strategies during the campaign. I provide two pieces of suggestive evidence.

One, I find that the concentration of voters for the top two candidates decreases between the first and second round of two-round elections (Table 10).²³ The decrease in concentration between rounds suggests that candidates use the period between rounds (about three weeks) to adjust their strategies to rally voters who supported eliminated candidates, a strategy that has been documented in other work (Pons and Tricaud, 2017).

²³ This is not a causal effect, as I estimate a regression that compares concentration in the first round with concentration in the second round, using all elections holding two rounds.

Table 10: Estimates on the geographic concentration of voters across rounds in two-round elections

	Coefficient of variation	Fractionalization	Entropy	Std Dev of 1st place candidate
2ndRound	-0.012*** (0.001)	-0.012*** (0.001)	-0.010*** (0.001)	-0.013*** (0.001)
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. Regression 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.

Two, using data from the 2004-2012 elections,²⁴ I document that candidates in two-round elections finance their campaigns differently (Table 11). Campaign donations reflect the strategies candidates adopt to appeal to voters, and thus provide one indication that candidates face different incentives during the election. Candidates in two-round elections receive *fewer* donations, both on average and between the top two candidates. While the outcomes are noisy, these effects are strongest for donations from corporations: candidates in two-round elections receive *fewer* donations from corporations.²⁵ To the extent that corporations represent a narrower swath of the electorate, this pattern suggests that candidates in two-round elections are appealing to broader groups in the electorate.

²⁴ Data on campaign finances is not available before the 2004 elections. I exclude the 2016 elections from the sample, as a new campaign finance law was passed in 2016 banning donations from corporations.

²⁵ 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.

Table 11: Regression discontinuity estimates on campaign donations

	<i>Donation amounts received by candidates</i>		
	Total	From individuals	From corporations
<i>Panel A: Average donations per candidate</i>			
TwoRound	-0.225 (0.286)	-0.491 (0.310)	-0.673* (0.404)
Single-round mean	12.844	10.742	11.782
Observations	154	154	154
<i>Panel B: Total donations among top two candidates</i>			
TwoRound	-0.074 (0.335)	-0.614 (0.492)	-1.023* (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).

VI. THEORETICAL FRAMEWORK

I present a stylized model to interpret the empirical results in a context where two-round elections create incentives for politicians to appeal to a broader group of voters 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, 2008) and follows the setup in Genicot et al. (2020) by allowing for targeting of government interventions to specific localities within a municipality. I extend their 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.

My model predicts that candidates in two-round elections appeal more broadly, both by appealing more to all voters and by appealing to minority groups. This occurs through two channels. First, two-round elections embed a conditionality requiring winners to attain a vote share above 50%, else a second round is triggered. This raises the marginal value of *every* vote and results in candidates appealing more to all voters. Second, the existence of a second round effectively limits the number of candidates. Candidates who expect to qualify for the second round but not attain 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 appeal to minority groups, who, in the model, vote for the third 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.²⁶

A. 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 $j \in \{1, 2, \dots, J\}$ where $J \geq 3$. Each locality has a continuum of voters of mass $1/J$.

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, $\mathbf{q}^c = (q_1^c, q_2^c, \dots, q_J^c)$, 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:

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

for a constant κ . Platforms are binding for politicians between rounds and after the election.²⁷

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:²⁸

$$\mathbf{q}^C = (0, 0, \dots, 0, G^C)$$

²⁶ 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.

²⁷ 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.

²⁸ 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 C.7.

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. This easily translates into other interpretations. For example, C may be a candidate whose supporters share a common trait and vote for her due to descriptive representation. These traits can 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 interpretation 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%.

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 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:

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

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 $j \in \{1, \dots, J-1\}$, ie. where candidate C has not allocated resources, this amounts to voting for either A or B .²⁹ 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 $0 < \alpha < 1$.³⁰

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

$$u_j(q_j^B) \leq u_j(q_j^A) + v_i + \delta \quad (7)$$

²⁹ 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 .

³⁰ 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 C.8.

In localities $j \in \{1, \dots, J-1\}$, all voters for whom this is true vote for A .³¹ 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.³²

B. Equilibrium strategies

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 offers; and π_t^c be the total vote share in the municipality for candidate c in round $t \in \{1, 2\}$. Candidates' payoff is 1 if they win 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. In equilibrium, candidate A and candidate B will have the same strategy, so I solve the model for candidate A . See Appendices C.2 and C.5 for the derivations and proofs in this section.

With three candidates (in single-round elections and in the first-round of two-round elections), 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 C.4). This simplifies candidate A and B 's maximization problem, and shuts down the channel where candidate C poses a threat to electoral defeat.

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 over half of the remaining votes:

$$\Pr\left(\pi_1^A \geq \frac{1}{2}\left(1 - \frac{1-\alpha}{J}\right)\right)$$

This corresponds to the following maximization problem:

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

³¹ For localities $j \in \{1, \dots, J-1\}$, 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 (7) is sufficient for voters to vote for A in these localities.

³² In other words, for all $j \in \{1, \dots, J-1\}$, I assume that $\frac{u'_j(q)}{u'_j(q)} < \frac{\psi_{J+(1-\alpha)\gamma}}{\alpha\psi_J + (1-\alpha)\gamma}$ for all positive q . This ensures that, in equilibrium, candidates allocate strictly less resources to J than to other localities.

Prediction 1 states that, in a single-round election, in equilibrium, candidates will promise less resources to locality J in comparison to the other localities.

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

This 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.

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:

$$\Pr\left(\pi_1^A \geq \frac{1}{2}\right) + \left(1 - \Pr\left(\pi_1^A \geq \frac{1}{2}\right) - \Pr\left(\pi_1^B \geq \frac{1}{2}\right)\right) \Pr\left(\pi_2^A \geq \frac{1}{2}\right)$$

This corresponds to the following maximization:

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

Prediction 2 states that, in a two-round election, in equilibrium, candidates will promise less resources to locality J in comparison to the other localities.

Prediction 2. *In a two-round election, for all $j \in \{1, \dots, J-1\}$, we have that $q_j^A > q_j^A$.*

As in the single-round election, 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.

C. Comparing single- to two-round elections

In this section, I compare three outcomes under the single- and the two-round elections: (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 elections) and q_j^{2R} and G^{2R} (for two-round elections). See Appendix C.6 for the proofs in this section.

Prediction 3 states that (i) candidates promise more to locality J in a two-round election than in a single-round election and (ii) candidates' promises to the other localities in a two-round election compared to a single-round election is ambiguous.

Prediction 3. $q_j^{1R} < q_j^{2R}$ and $q_j^{1R} \leq q_j^{2R}$ for all $j \in \{1, \dots, J - 1\}$.

In a single-round election, because voters in locality J – where candidate C has promised the entire government budget – strongly favor candidate C , the marginal return to allocating resources there is low and candidates A and B ignore these voters. In contrast, in the two-round election, there is the possibility of a second round where C is not present and thus the marginal return to allocating resources to that locality is higher. As a result, while not offering a completely equitable distribution, candidates A and B solicit more votes from locality J in a two-round election, even in the first round.

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 4 states that the overall government budget promised is higher in two-round elections than in single-round elections.

Prediction 4. $G^{1R} < G^{2R}$.

In a single-round election, candidates face lower incentives to invest in *all* localities. These incentives result 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.

When comparing inequality in allocations between the two electoral rules, the outcome of interest is how the allocation of resources to locality J compares to the allocation to the other localities.³³ Prediction 5 states that when comparing two-round elections to single-round elections, the level of inequality in government resources promised is ambiguous.

³³ 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 $\{1, \dots, J - 1\}$ 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 $\{1, \dots, J - 1\}$ 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 $\{1, \dots, J - 1\}$ localities receive the same allocation, and the only disparity that matters is between J and the other localities.

Prediction 5. *If $q_j^{1R} > q_j^{2R}$ for all $j \in \{1, \dots, J - 1\}$, then the level of inequality is lower in two-round elections compared to single-round elections. If $q_j^{1R} < q_j^{2R}$ for all $j \in \{1, \dots, J - 1\}$, 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 .

Discussion

My model aligns with an interpretation of the empirical results where two-round elections lead candidates to adopt different campaign strategies by offering policy proposals that appeal to a broader group of voters. Candidates in two-round elections appeal more broadly, both by appealing more to all voters and by appealing to minority groups. Candidates in two-round elections face higher incentives to appeal to all voters, leading them to increase the overall government budget. Candidates in two-round elections also appeal more to minority groups, because two-round elections raise the marginal return to allocating resources to groups of voters that are heavily targeted by other candidates. 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.

VII. CONCLUSION

A majority of countries use two-round elections to choose their leaders, and an increasing number of countries are adopting this electoral rule 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 election, 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 find evidence suggesting that two-round elections cause candidates to adjust their strategies 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.

I present a model to highlight how two-round elections can lead to these empirical results, through a channel where candidates adjust their strategies. 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 election 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.

If two-round elections lead to positive outcomes, why is the two-round election 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 elections. Second, I find that, in two-round elections, individuals at lower parts of the distribution benefit. As a result, there may be opposition to implementing two-round elections 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. In this context, 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 elections in different contexts would greatly advance our overall understanding of electoral systems.

REFERENCES

- Acemoglu, D., & Robinson, J. A. (2008). Persistence of Power, Elites, and Institutions. *American Economic Review*, 98(1), 267–293.
- Acemoglu, D., Reed, T., & Robinson, J. A. (2014). Chiefs: Economic Development and Elite Control of Civil Society in Sierra Leone. *Journal of Political Economy*, 122(2), 319–368.
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, 91(434), 444–455.
- Bardhan, P., & Mookherjee, D. (2006). Decentralisation and Accountability in Infrastructure Delivery in Developing Countries. *The Economic Journal*, 116, 101–127.
- Bordignon, M., Nannicini, T., & Tabellini, G. (2016). Moderating Political Extremism: Single Round versus Runoff Elections under Plurality Rule. *American Economic Review*, 106(8), 2349–2370.
- Bormann, N., & Golder, M. (2013). Democratic Electoral Systems around the world, 1946–2011. *Electoral Studies*, 32(2), 360–369.
- Bouton, L. (2013). A Theory of Strategic Voting in Runoff Elections. *American Economic Review*, 103(4), 1248–1288.
- Bouton, L., & Gratton, G. (2015). Majority Runoff Elections: Strategic Voting and Duverger’s Hypothesis. *Theoretical Economics*, 10(2), 283–314.
- Bouton, L., Gallego, J., Llorente-Saguer, A., & Morton, R. (2019). Runoff Elections in the Laboratory. *NBER Working Paper*.
- Brollo, F., & Nannicini, T. (2012). Tying Your Enemy’s Hands in Close Races: The Politics of Federal Transfers in Brazil. *American Political Science Review*, 106(4), 742–761.
- Burden, B. C. (1997). Deterministic and Probabilistic Voting Models. *American Journal of Political Science*, 41(4), 1150–1169.
- Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., & Padró-i-Miquel, G. (2015). The Value of Democracy: Evidence from Road Building in Kenya. *American Economic Review*, 105(6), 1817–1851.
- Callander, S. (2005). Duverger’s Hypothesis, the Run-off Rule, and Electoral Competition. *Political Analysis*, 13(3), 209–232.
- Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6), 2295–2326.
- Chamon, M., De Mello, J., & Firpo, S. (2019). Electoral Rules, Political Competition and Fiscal Expenditures: Regression Discontinuity Evidence from Brazilian Municipalities. *Journal of Development Studies*, 55(1), 19–38.
- Cipullo, D. (2019). Voting Systems, Incentives and Fiscal Policy: Evidence from Runoff and Plurality. *Working Paper*.
- Cox, G. W. (1997). *Making Votes Count: Strategic Coordination in the World’s Electoral Systems*. Cambridge, UK: Cambridge University Press.
- Cusack, T. R., Iversen, T., & Soskice, D. (2007). Economic Interests and the Origins of Electoral Systems. *American Political Science Review*, 101(3), 373–391.
- Eggers, A. C. (2013). Proportionality and Turnout: Evidence from French Municipalities. *Comparative Political Studies*, 48(2), 135–167.
- Fujiwara, T. (2011). A Regression Discontinuity Test of Strategic Voting and Duverger’s Law. *Quarterly Journal of Political Science*, 6, 197–233.

- Fujiwara, T. (2015). Voting Technology, Political Responsiveness and Infant Health: Evidence from Brazil. *Econometrica*, 83(2), 423–464.
- Fujiwara, T., & Wantchekon, L. (2013). Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin. *American Economic Journal: Applied Economics*, 5(4), 241–255.
- Genicot, G., Bouton, L., & Castanheira, M. (2020). Electoral Systems and Inequalities in Government Interventions. *NBER Working Paper*.
- Golden, M., & Min, B. (2013). Distributive Politics Around the World. *Annual Review of Political Science*, 16, 73–99.
- Gonzales, M., León-Ciliotta, G., & Martínez, L. R. (2019). How Effective Are Monetary Incentives to Vote? Evidence from a Nationwide Policy. *Working Paper*.
- Hodler, R., & Raschky, P. A. (2014). Regional Favoritism. *Quarterly Journal of Economics*, 129(2), 995–1033.
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615–635.
- Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, 79(3), 933–959.
- Lindbeck, A., & Weibull, J. W. (1987). Balanced-Budget Redistribution as the Outcome of Political Competition. *Public Choice*, 52(3), 273–297.
- Lizzeri, A., & Persico, N. (2001). The Provision of Public Goods under Alternative Electoral Incentives. *American Economic Review*, 91(1), 225–239.
- Lizzeri, A., & Persico, N. (2005). A Drawback of Electoral Competition. *Journal of the European Economic Association*, 3(6), 1318–1348.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), 698–714.
- Milesi-Ferretti, G. M., Perotti, R., & Rostagno, M. (2002). Electoral systems and public spending. *Quarterly Journal of Economics*, 117(2), 609–657.
- Myerson, R. B. (1993). Incentives to Cultivate Favored Minorities Under Alternative Electoral Systems. *American Political Science Review*, 87(4), 856–869.
- Osborne, M. J., & Slivinski, A. (1996). A Model of Political Competition with Citizens-Candidates. *Quarterly Journal of Economics*, 111(1), 65–96.
- Pellicer, M., & Wegner, E. (2013). Electoral Rules and Clientelistic Parties: A Regression Discontinuity Approach. *Quarterly Journal of Political Science*, 8(4), 1–33.
- Persson, T., & Tabellini, G. (1999). The size and scope of government: Comparative politics with rational politicians. *European Economic Review*, 43(4–6), 699–735.
- Persson, T., & Tabellini, G. (2008). Electoral Systems and Economic Policy. In D. A. Wittman & B. R. Weingast (Eds.), *The Oxford Handbook of Political Economy*.
- Piketty, T. (2000). Voting as Communicating. *Review of Economic Studies*, 67(1), 169–191.
- Pons, V., & Tricaud, C. (2018). Expressive Voting and Its Cost: Evidence from Runoffs with Two or Three Candidates. *Econometrica*, 86(5), 1621–1649.
- Reardon, S. F., & Firebaugh, G. (2002). Measures of Multigroup Segregation. *Sociological Methodology*, 32, 33–67.
- White, M. J. (1986). Segregation and Diversity Measures in Population Distribution. *Population Index*, 52(2), 198–221.

Appendix A. DATA APPENDIX

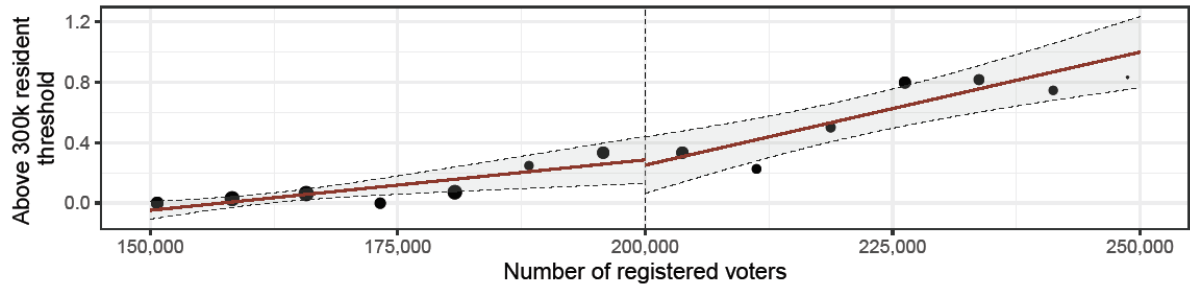
Table A-1: Variables used to construct the equipment index, from the *Censo Escolar*

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

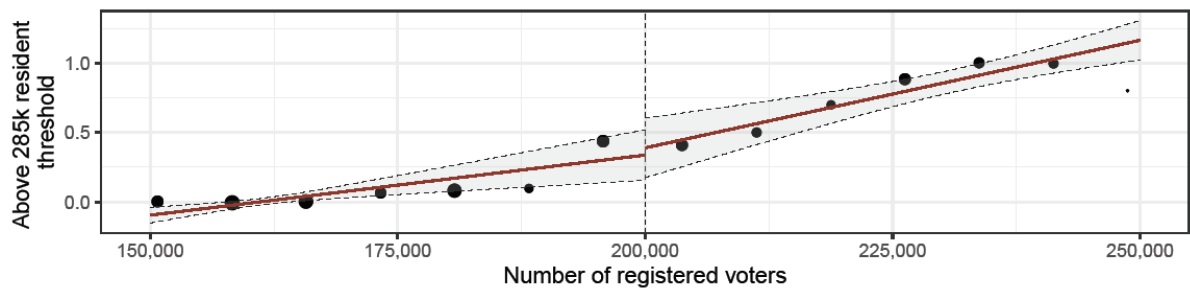
Appendix B. ADDITIONAL FIGURES AND TABLES

Appendix B-1. Identification

Figure B-1: Regression discontinuity plots of the probability of falling above/below other policy thresholds

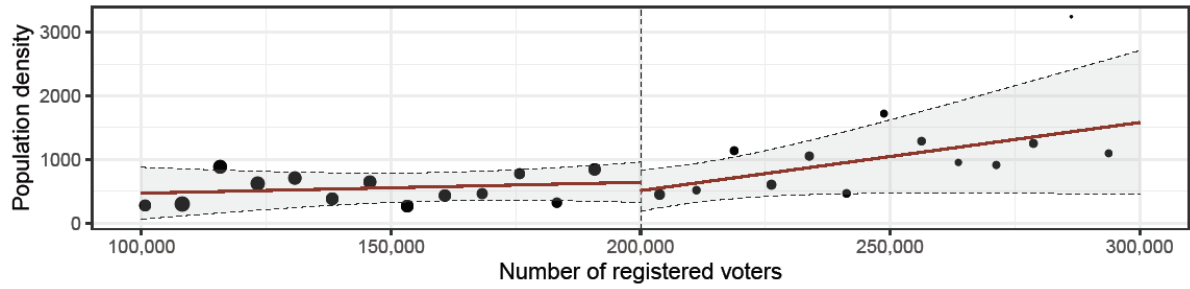


(a) Threshold: 300,000 inhabitants

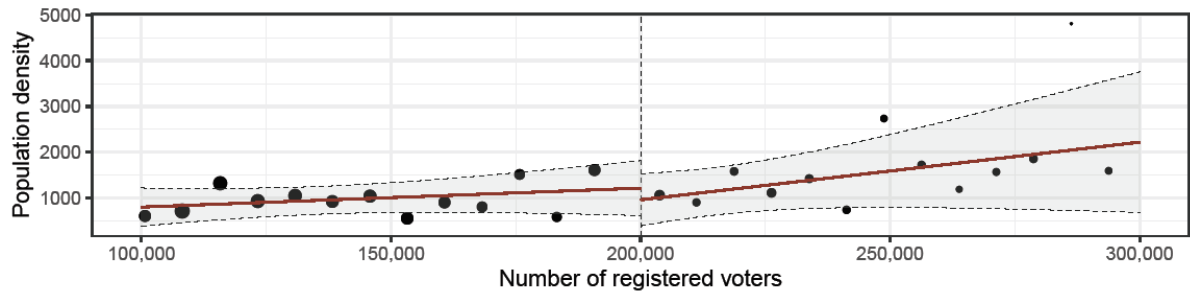


(b) Threshold: 285,714 inhabitants

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 B-2: Regression discontinuity plots of pre-treatment population density

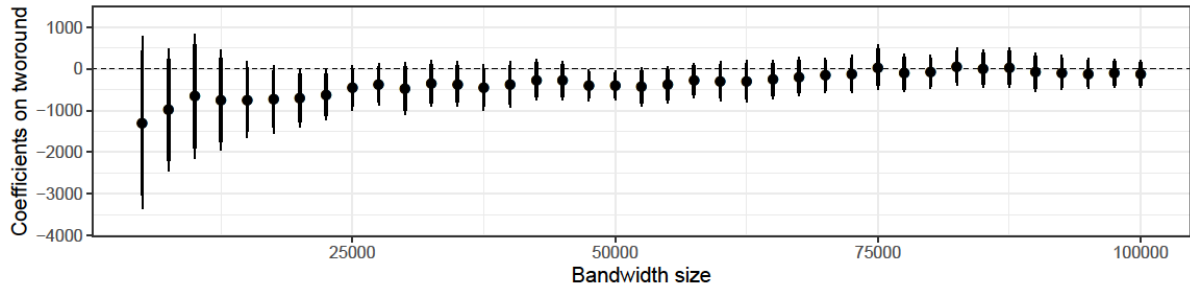
(a) Measured prior to the 1988 Constitution



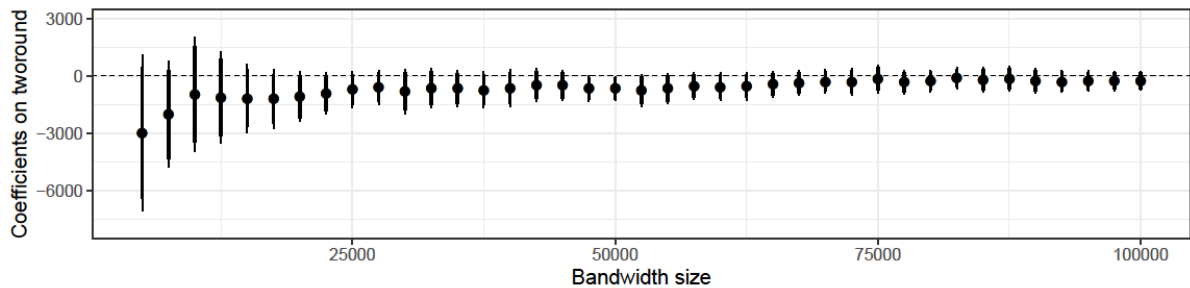
(b) Measured prior to the most recent single-round

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 B-3: 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.

Appendix B-2. Other results on electoral outcomes

Table B-1: 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* (0.008)	-0.018** (0.009)	-0.013** (0.007)	-0.015 (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.

Table B-2: Estimates on the geographic concentration of voters, using vote shares from the final round

	Coefficient of variation	Fractionalization	Entropy	Std Dev of 1st place candidate
TwoRound	-0.020** (0.008)	-0.023*** (0.008)	-0.017*** (0.006)	-0.019** (0.009)
Single-round mean	0.036	0.038	0.029	0.088
Observations	216	216	216	216

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

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. Regression compares the 1st round results (in single-round elections) with the 2nd round results (in two-round elections).

Table B-3: Regression discontinuity estimates on the geographic concentration of voters, with number of candidates as a control

<i>Panel A: Concentration indices of voters for specific candidates</i>				
	Coefficient of variation	Fractionalization	Entropy	
TwoRound	-0.005 (0.003)	-0.010** (0.005)	-0.009* (0.005)	
Single-round mean	0.019	0.027	0.030	
Observations	264	264	264	
<i>Panel B: Standard deviation in vote shares for each candidate</i>				
	1st place candidate	2nd place candidate	3rd place candidate	4th place candidate
TwoRound	-0.016** (0.007)	-0.012 (0.008)	-0.010 (0.007)	-0.002 (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.

Appendix B-3. Other results on education resources

Table B-4: Regression discontinuity estimates on resources in municipal schools, using z-scores

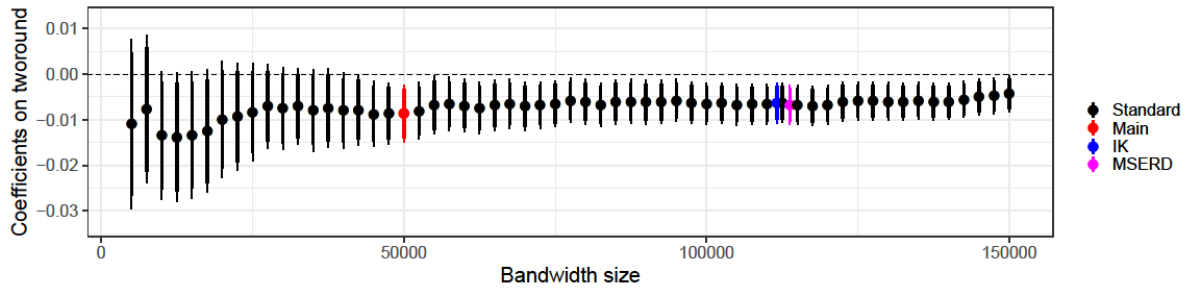
	<i>Mean level of resources</i>		<i>Standard deviation in resources</i>	
	Equipment	Infrastructure	Equipment	Infrastructure
TwoRound	0.079** (0.033)	0.069* (0.037)	-0.014 (0.009)	-0.007 (0.017)
Single-round mean	0.724	0.739	0.120	0.146
Observations	820	912	820	912

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

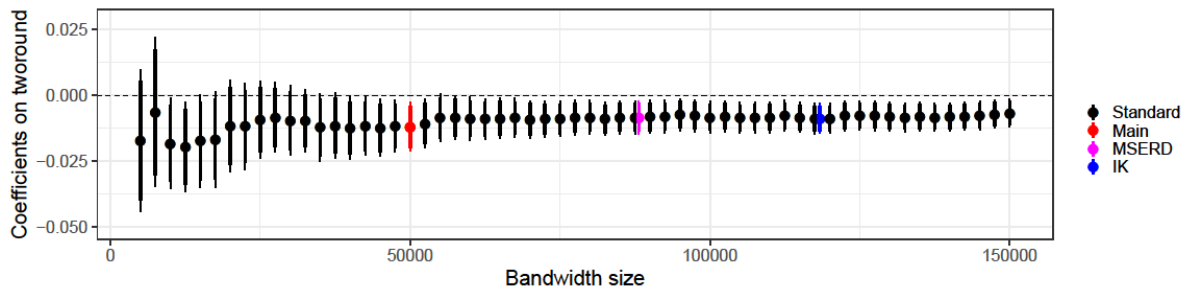
Equipment and *Infrastructure* are indices constructed by taking the z-score of a school's equipment and infrastructure elements, then calculating the school's percentile in the national distribution. The first two columns (*Mean level of resources*) have as the dependent variable the mean index level 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.

Appendix B-4. Robustness tests on RDD design

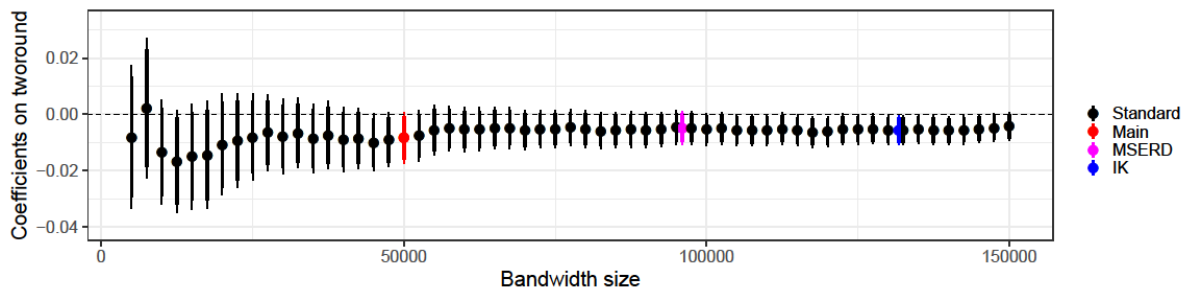
Figure B-4: Regression discontinuity coefficients on overall concentration of voters for specific candidates at different bandwidths



(a) Coefficient of variation



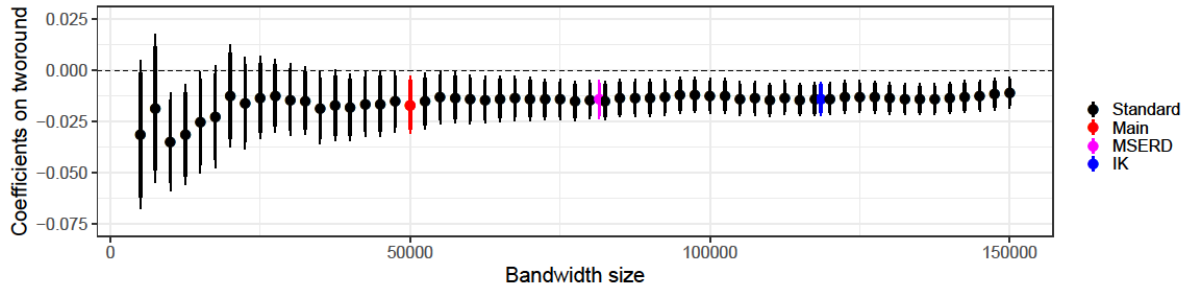
(b) Fractionalization



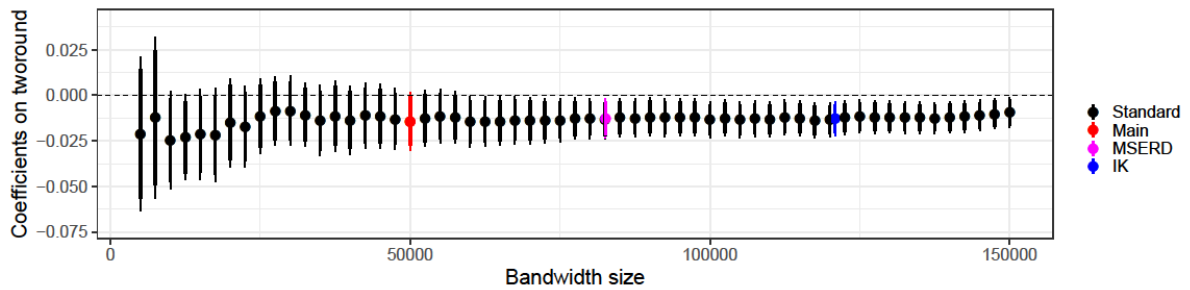
(c) Entropy

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.

Figure B-5: 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.

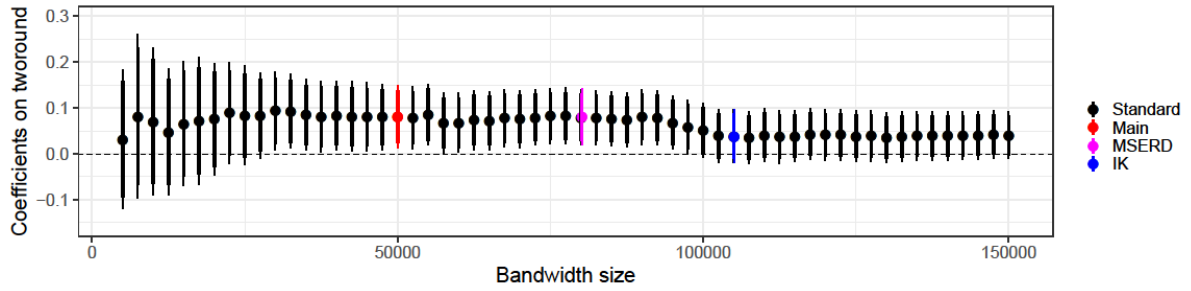
Table B-5: Regression discontinuity estimates on the geographic concentration of voters, without controls

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

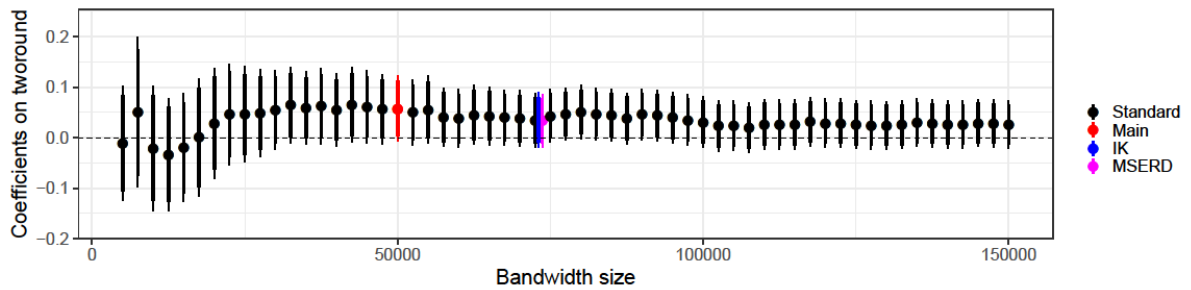
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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.

Figure B-6: Regression discontinuity coefficients on overall level of resources in municipal schools at different bandwidths



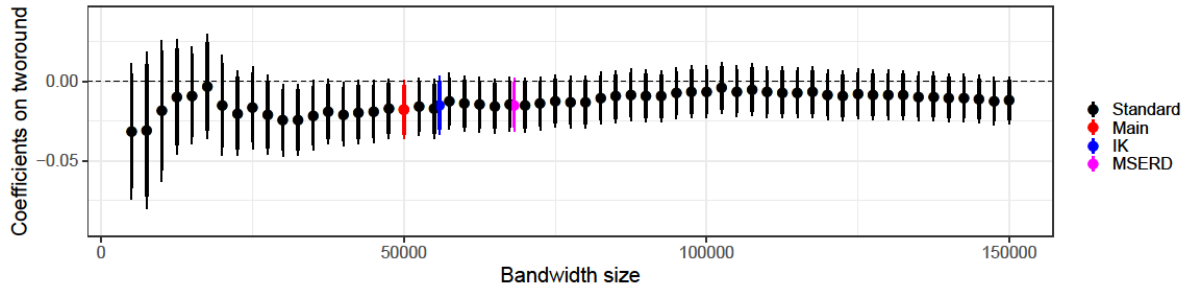
(a) Equipment, mean level of resources



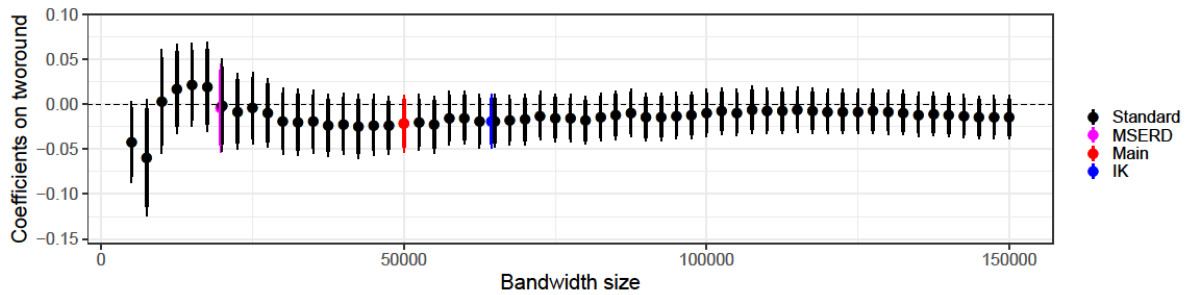
(b) Infrastructure, mean level of resources

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.

Figure B-7: 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.

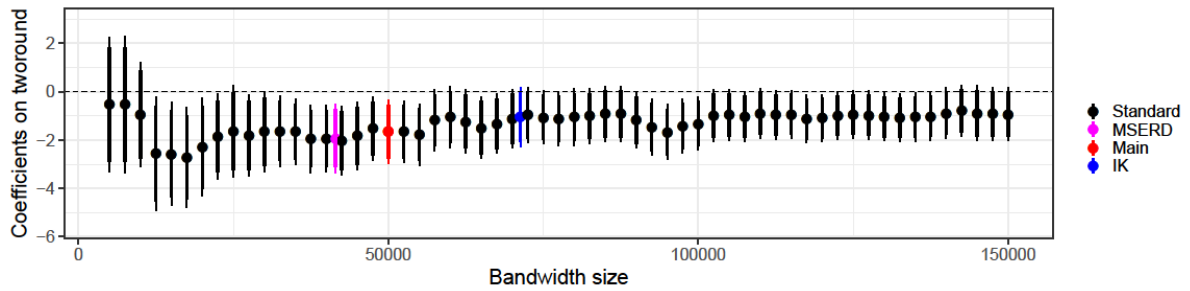
Table B-6: Regression discontinuity estimates on resources in municipal schools, without controls

	<i>Mean level of resources</i>		<i>Standard deviation in resources</i>	
	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

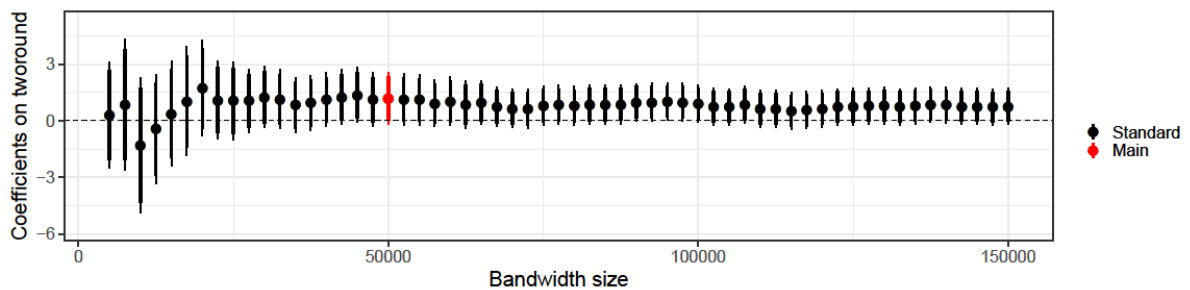
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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.

Figure B-8: 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.

Table B-7: Regression discontinuity estimates on municipal education outcomes, without controls

	Drop-out rate	Failing rate	Passing rate	Elem. literacy rate
TwoRound	-1.340* (0.755)	-0.051 (1.167)	1.291 (1.686)	2.918 (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.

Appendix B-5. Placebo tests

Table B-8: Placebo regression discontinuity estimates on the geographic concentration of voters, at 285,714 inhabitant threshold

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

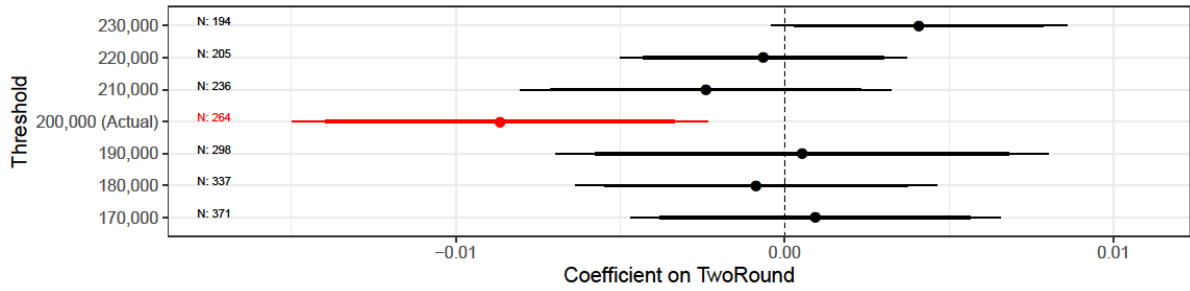
Table B-9: Placebo regression discontinuity estimates on the geographic concentration of voters, at 300,000 inhabitant threshold

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

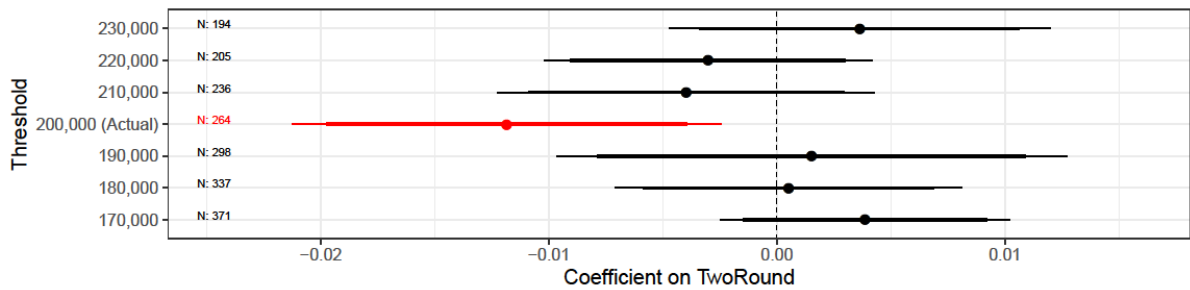
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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.

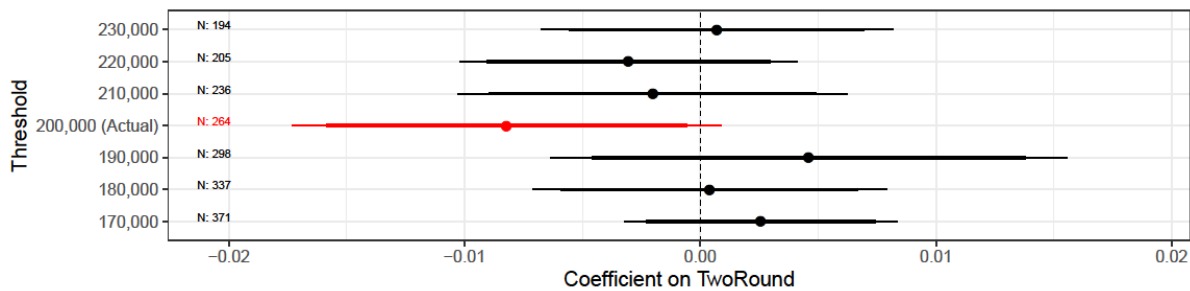
Figure B-9: Regression discontinuity coefficients on overall concentration of voters for specific candidates at different thresholds



(a) Coefficient of variation



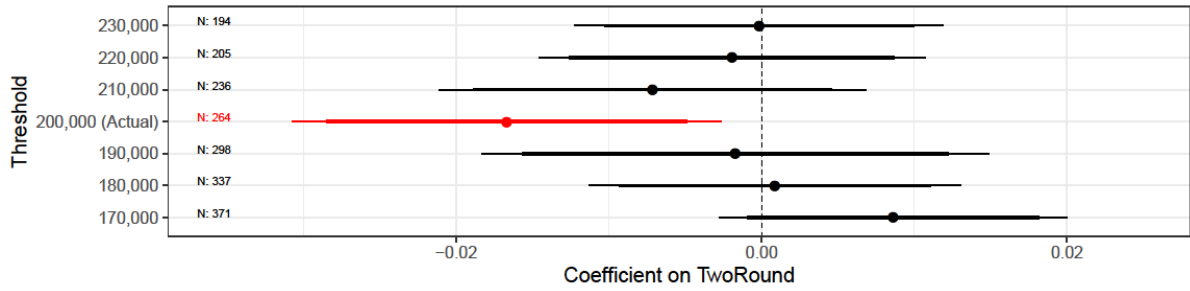
(b) Fractionalization



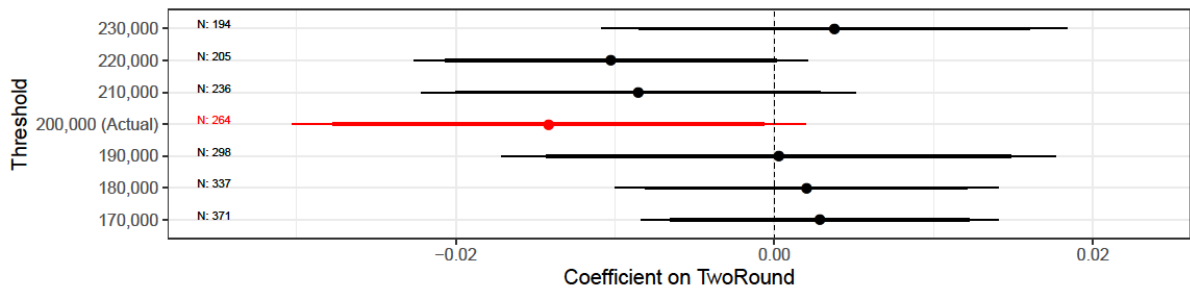
(c) Entropy

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.

Figure B-10: Regression discontinuity coefficients on the candidate-level concentration in voters at different thresholds



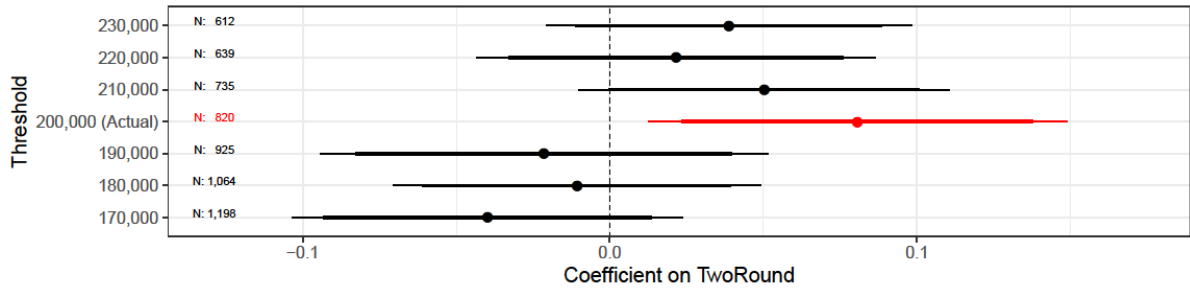
(a) Standard deviation in votes for 1st place candidate



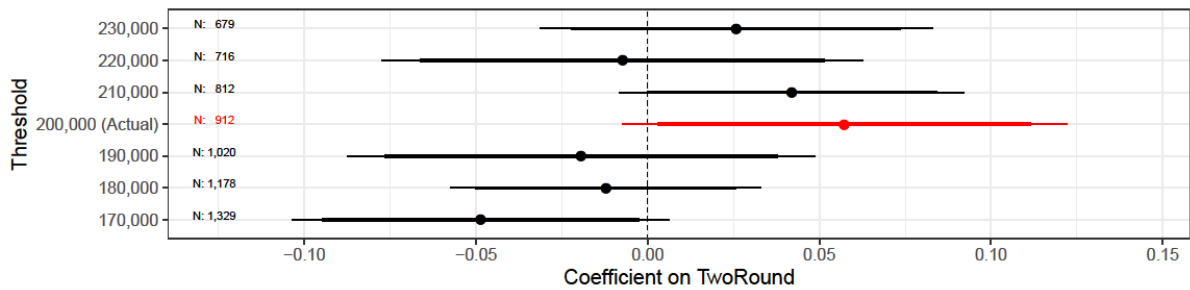
(b) Standard deviation in votes for 2nd place candidate

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.

Figure B-11: Regression discontinuity coefficients on overall level of resources in municipal schools at different thresholds



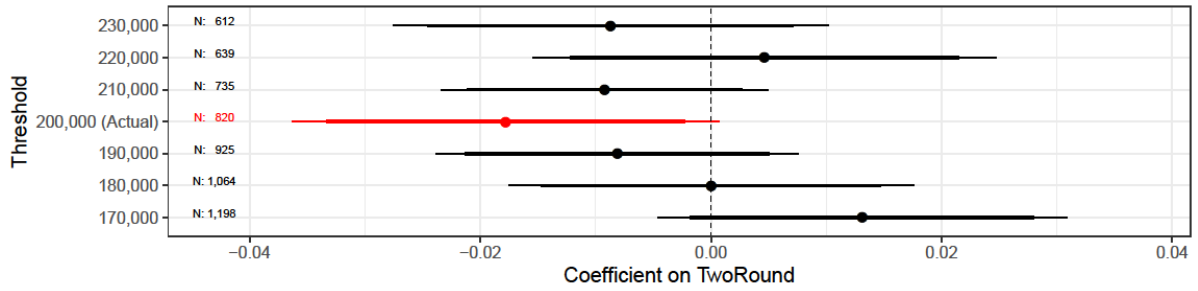
(a) Equipment, mean level of resources



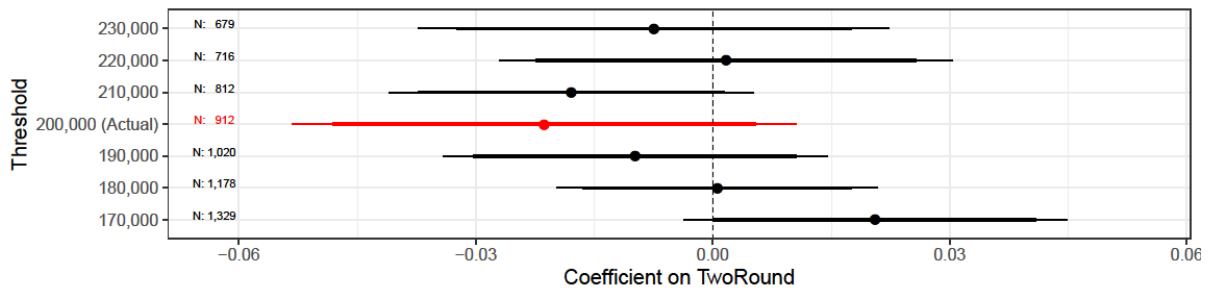
(b) Infrastructure, mean level of resources

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.

Figure B-12: Regression discontinuity coefficients on distribution of resources in municipal schools at different thresholds



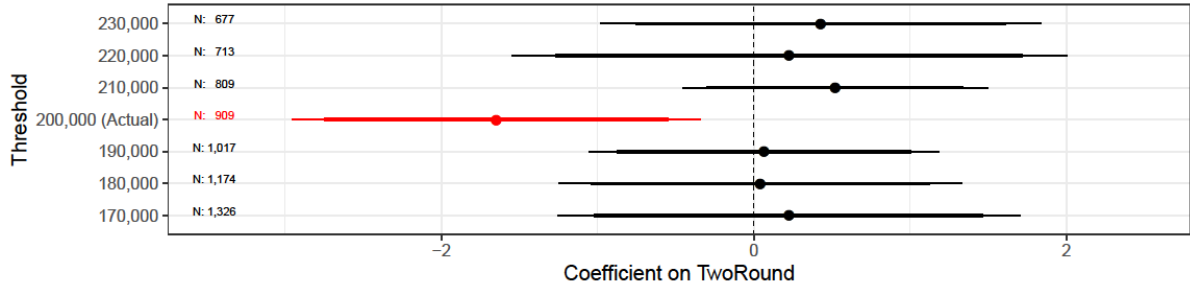
(a) Equipment, standard deviation in resources



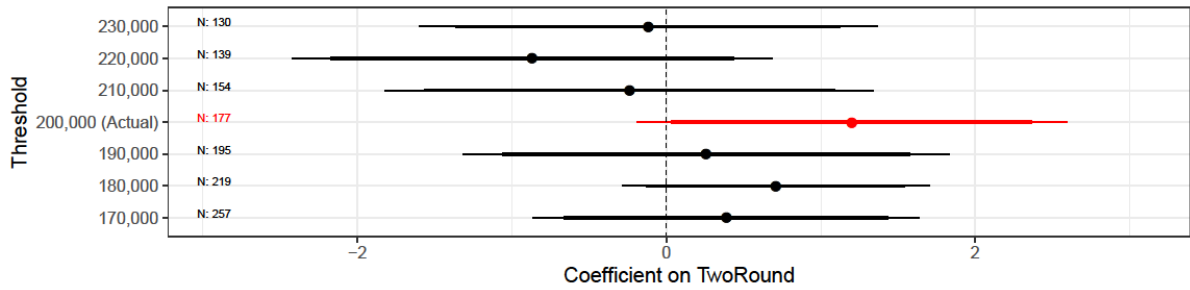
(b) Infrastructure, standard deviation in resources

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.

Figure B-13: 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.

Appendix C. THEORY APPENDIX

Let π_{jt}^c be candidate c 's vote share in locality j in round $t \in \{1, 2\}$.

Appendix C-1. Assumption: 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:

$$\begin{aligned} \delta &\in \left(u(y) - u(0) - \frac{1}{2\psi}, u(y) - u(0) + \frac{1}{2\psi}\right) \\ &\Leftrightarrow \frac{1}{2\psi} + u(y) - u(0) < \frac{1}{2\psi} \end{aligned}$$

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 $u(y) - u(0) > 0$.

Appendix C-2. Deriving candidate maximization problems

This section assumes that (i) there are voters to be swung in every locality and (ii) all localities are contestable. In other words, (i) $0 < \pi_{j1}^A < 1$ (see Appendix C-1), and (ii) $0 < Pr(\pi_1^A \geq \theta) < 1$ and $0 < Pr(\pi_1^B \geq \theta) < 1$ (see Appendix C-3).

With three candidates

Condition 7 corresponds to voters for whom $v_i \geq u_j(q_j^B) - u_j(q_j^A) - \delta$. π_{jt}^c is given by:

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

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

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

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

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

With two candidates

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

$$\begin{aligned}\pi_{j2}^A &= \frac{1}{2} + \psi(\Delta u_j^{AB} + \delta) & \pi_{j2}^B &= \frac{1}{2} + \psi(\Delta u_j^{BA} - \delta) \\ \pi_2^A &= \frac{1}{2} + \psi\delta + \frac{\psi}{J} \sum_{j=1}^J \Delta u_j^{AB} & \pi_2^B &= \frac{1}{2} - \psi\delta + \frac{\psi}{J} \sum_{j=1}^J \Delta u_j^{BA}\end{aligned}$$

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

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

Appendix C-3. Assumption: Contestability of localities

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

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

which corresponds to the following condition for the first round:

$$\theta \in \left(\left(\frac{J-1+\alpha}{J} \right) \left(-\frac{\psi}{2\gamma} + \frac{1}{2} \right), \left(\frac{J-1+\alpha}{J} \right) \left(\frac{\psi}{2\gamma} + \frac{1}{2} \right) \right) \quad (C1)$$

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

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

which corresponds to the following condition for the second round:

$$\theta \in \left(-\frac{\psi}{2\gamma} + \frac{1}{2}, \frac{\psi}{2\gamma} + \frac{1}{2} \right) \quad (C2)$$

Claim 1. $\theta = \frac{1}{2} \left(1 - \frac{1-\alpha}{J}\right)$ satisfies condition C1.

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 2. $\theta = \frac{1}{2}$ satisfies condition C1.

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:

$$\begin{aligned} \frac{1}{2} &< \left(\frac{J-1+\alpha}{J}\right) \left(\frac{\psi}{2\gamma} + \frac{1}{2}\right) \\ \Leftrightarrow J &> (1-\alpha) \left(\frac{\gamma+\psi}{\psi}\right) \end{aligned}$$

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

Claim 3. $\theta = \frac{1}{2}$ satisfies condition C2.

Both the upper and lower inequalities are satisfied because:

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

Appendix C-4. Assumption: 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 C1:

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

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

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

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.

Appendix C-5. First order conditions

Prediction 1

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

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

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

The ratio in marginal utilities between localities is:

$$\begin{aligned} \text{between } j \text{ and } j': \quad \frac{u'_j(q_j^A)}{u'_{j'}(q_{j'}^A)} &= 1 && \forall j, j' \in \{1, \dots, J-1\} \\ \text{between } j \text{ and } J: \quad \frac{u'_j(q_j^A)}{u'_J(q_J^A)} &= \alpha && \forall j \in \{1, \dots, J-1\} \quad (C3) \end{aligned}$$

Equation C3 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 $q_j^A > q_j^A$.

Prediction 2

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

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

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

The ratio in marginal utilities between localities is:

$$\begin{aligned}
\text{between } j \text{ and } j': & \quad \frac{u'_j(q_j^A)}{u'_{j'}(q_{j'}^A)} = 1 & \quad \forall j, j' \in \{1, \dots, J-1\} \\
\text{between } j \text{ and } J: & \quad \frac{u'_j(q_j^A)}{u'_J(q_J^A)} = \frac{\alpha + \frac{(1-\alpha)\gamma}{\psi J}}{1 + \frac{(1-\alpha)\gamma}{\psi J}} & \quad \forall j \in \{1, \dots, J-1\} \quad (C4)
\end{aligned}$$

Equation C4 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 $q_j^A > q_J^A$.

Appendix C-6. Comparing single- to two-round elections: Predictions 3 and 4

I first establish three lemmas.

Lemma 1. $\frac{u'_j(q_j^{1R})G^{2R}}{u'_j(q_j^{2R})G^{1R}} > 1$ for all $j \in \{1, \dots, J\}$.

Proof. For $j \in \{1, \dots, J-1\}$, combining the first round first order conditions in Appendix C-5:

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

For $j = J$, combining the second round first order conditions in Appendix C-5:

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

■

Lemma 2. $\frac{u'_j(q_j^{1R})}{u'_j(q_j^{2R})} < \frac{u'_J(q_J^{1R})}{u'_J(q_J^{2R})}$.

Proof. Comparing the ratio of marginal utilities in equations C3 and C4, the ratio is smaller in the single-round system compared to the two-round system:

$$\frac{u'_j(q_j^{1R})}{u'_j(q_j^{2R})} < \frac{u'_J(q_J^{1R})}{u'_J(q_J^{2R})} \Leftrightarrow \alpha < \frac{\alpha + \frac{(1-\alpha)\gamma}{\psi J}}{1 + \frac{(1-\alpha)\gamma}{\psi J}}$$

which is true because $\alpha < 1$.

■

Lemma 3. If $q_j^{1R} > q_j^{2R}$ for one $j \neq J$ then $q_{j'}^{1R} > q_{j'}^{2R}$ for all other $j' \in \{1, \dots, J-1\}$.

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 Appendix C-5 establish that the marginal utilities between all $j, j' \in \{1, \dots, J-1\}$ are equal. Then we must have that $u_{j'}(q_j^{1R}) < u_{j'}(q_j^{2R})$ and that $q_j^{1R} > q_j^{2R}$. ■

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 1, 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 3, for all $j \neq J$. Then $u'_j(q_j^{1R}) > u'_j(q_j^{2R})$. However, this violates lemma 2, and so we must have $q_j^{1R} < q_j^{2R}$. ■

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 3, for all $j \neq J$. Then $u'_j(q_j^{1R}) < u'_j(q_j^{2R})$. However, this violates lemma 1, and so we must have $G^{1R} < G^{2R}$. ■

Proof of allocations in other localities

I show that $q_j^{1R} \leq q_j^{2R}$.

Proof. From the first order conditions in Appendix C-5 and since $G^{1R} < G^{2R}$, we have that $\lambda_{1R} < \lambda_{2R}$. Then, for $j \in \{1, \dots, J-1\}$:

$$\frac{u'_j(q_j^{1R})}{u'_j(q_j^{2R})} < 1 + \frac{(1-\alpha)\gamma}{\psi J} \Rightarrow \frac{u'_j(q_j^{1R})}{u'_j(q_j^{2R})} \leq 1$$

which implies that $q_j^{1R} \leq q_j^{2R}$. ■

Appendix C-7. Assumption: Candidate C's budget

In general, for every set of utility functions $(u_1(\cdot), \dots, u_J(\cdot))$, 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:

$$\left(\frac{\gamma}{J-1+\alpha}\right)\left(\alpha + \frac{(1-\alpha)\gamma}{\psi J}\right)\frac{\beta_J}{q_J} = \lambda_{2R}$$

$$\kappa G = \lambda_{2R}$$

We can write G as a function of q_J :

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

Since $q_J < q_j$ for all $j \neq J$ (prediction 2), then $q_J < \frac{G}{j}$, implying:

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

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.

Appendix C-8. 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:

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

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

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

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

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

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

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

Equilibrium strategies in a single-round election

The maximization is:

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

The first order conditions are:

$$\begin{aligned} \left(\frac{\gamma}{J-1}\right) u_j'(q_j^A) &= \lambda'_{1R} & \text{for } j \in \{1, \dots, J-1\} \\ \kappa G^A &= \lambda'_{1R} \end{aligned}$$

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: $(q_j^A)' = q_j^A - \epsilon$ and $(q_k^A)' = q_k^A + \epsilon$ 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:

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

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 $q_j^A = 0$.

Equilibrium strategies in a two-round election

The maximization is:

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

The first order conditions are:

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

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

From this, we can see that $q_j^A > 0$. Thus, $q_j^{1R} < q_j^{2R}$, and prediction 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 $\frac{u'_j(q_j^{1R}) G^{2R}}{u'_j(q_j^{2R}) G^{1R}} < 1$, which violates the first order conditions. Thus, $G^{1R} < G^{2R}$ and prediction 4 holds.

We also have that $q_j^{1R} \leq q_j^2$. 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 $q_j^{1R} \leq q_j^2$. Predictions 3 and 5 follow.